Three Essays in Auctions, Contests, and Leaderboards

by

John C. Lin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2011

Doctoral Committee:
Professor Yan Chen, Chair
Professor Thomas A. Finholt
Professor Jeffrey K. MacKie-Mason
Assistant Professor Yusuf Can Masatlioglu

To my parents, whose love gave me life.

Acknowledgments

First, I like to thank my advisor and committee chair, Yan Chen, whose generous gift of herself has helped me to be the scholar that I am today. Without her guidance, wisdom and encouragement, I could not have completed the long and sometimes arduous journey of completing my dissertation. I thank my committee members, Jeffrey MacKie-Mason, Thomas Finholt, and Yusufcan Masatlioglu for their advice and guidance throughout the dissertation writing process. Chapter 2 is based on my joint work with Yan Chen, Thomas Finholt, and Kan Takeuchi. This chapter has been published in the 2010 issue of Experimental Economics. Chapters 3 and 4 were completed under the supervision of Yan Chen.

I thank the participants of the Incentive-Centered Design group and the Behavioral Experimental Economics Laboratory at the School of Information, especially Lian Jian, Tracy Liu, Grace Young Joo Jeon, Ming Jiang, Rick Wash, and Anna Osepayshvili. I thank the support and the constant encouragements of my colleagues Archer Batcheller and Kevin Nam from the School of Information.

I thank the financial support of the STIET Fellowship funded by the National Science Foundation, the School of Information, as well as the Rackham Graduate School.

I thank my parents, Victor Lin and Ming-Chu Chang for their love and for the many sacrifices that they made for me to pursue my interests and dreams. I thank my sisters Jenny Lin and Lulu Lin for believing in me, for the support and encouragement of my brother-in-laws Hoover Tai and Ryan Chung, and for the wonderful smiles of my nephews Timothy and Joshua and my niece Isabella. I give thanks and praise to the Trinitarian God and for the spiritual family that Jesus Christ has blessed me with.

Table of Contents

List of Tables

List of Figures

Figure

List of Appendices

Abstract

In this dissertation, I study three resource allocation problems. In the first study, I present joint work with Yan Chen, Thomas Finholt, and Kan Takeuchi on a laboratory study evaluating three different package auction mechanisms for allocating access to large scientific equipments. I found that mechanisms that allow the bidders to express their intensity of preferences do equally well in maximizing the efficiency of the allocation. This is true regardless if the mechanism allows for multiple rounds of bidding before the final allocation.

In the second study, we evaluate the predictions of a theory of contest in the laboratory. The research question is given a fixed amount of prize money, how can we best partition it to encourage the highest level of total group effort. A theory of contests developed by Moldovanu et al. (2007) predicts that the optimal structure of prizes depend on the distribution of the ability factors of the contestants. Roughly speaking, if the contestants' ability factors are distributed uniformly, then it is better to combine the prize money into one grand prize. However, if a few contestants have high ability factors but the rest have low ability factors, then it is better to have as many prizes as there are contestants. In our laboratory study, we found that one of the predictions holds but the other fails because when people are unsure of the optimal strategy, they tend to exert more effort.

In the third study, we present an empirical study evaluating the effectiveness of social information in encouraging people to contribute product reviews on Amazon.com. We found empirical evidence that shows that people respond to social information when their action can more readily improve their ranking on a leaderboard.

Chapter 1

Introduction

Resources are scarce. We want to design institutions to allocate resources well. In this dissertation, I address three types of resource allocation problems. First, how do we allocate access to large, shared-use scientific equipment to research scientists when we care about allocating the resources to the researchers that have the highest values from using the equipment while insuring some level of equity to insure that everyone has a chance to use the equipment? Second, given a fixed amount of prize money, how can we allocate the monetary resources into prizes to encourage the highest level of total effort by those participating in the contests? Third, how can we allocate non-monetary incentives, such as recognition of people's effort through the user of leaderboards, to encourage product reviewers on Amazon to contribute product reviews.

In the first study, I present joint work with Yan Chen, Thomas Finholt, and Kan Takeuchi on a laboratory study evaluating three different package auction mechanisms for allocating access to large scientific equipments. The three mechanisms are Knapsack, generalized Vickrey-Clark-Groves (VCG), and Resource Allocation Design (RAD). Knapsack and VCG are one-shot mechanisms where bidders can only submit their bids once, whereas RAD is an iterative mechanism where bids are submitted over several rounds before a final allocation is made. Knapsack allows only ordered preferences to be expressed, but not the intensity of the preferences, whereas VCG and RAD both allow bidders to express the intensity of their preferences. If the goal of the mechanism designer is to maximize efficiency, both the VCG and the RAD mechanisms perform equally well. If the goal of the mechanism designer is to maximize the number of users of the equipment, Knapsack is the preferred mechanism.

In the second study, we evaluate the predictions of a theory of contest in the laboratory. The research question is given a fixed amount of prize money, how can we best partition it to encourage the highest level of total group effort. A theory of contests developed by Moldovanu et al. (2007) predicts that the optimal structure of prizes depends on the distribution of the ability factors of the contestants. Roughly speaking, if the contestants' ability factors are distributed uniformly, then it is better to combine the prize money into one grand prize. However, if a few contestants have high ability factors but the rest have low ability factors, then it is better to have as many prizes as there are contestants. In our laboratory study, we found that the first prediction holds but the second fails. The second prediction fails because when people are unsure of the optimal strategy, they tend to exert more effort.

In the third study, we present an empirical study evaluating the effectiveness of social information in encouraging people to contribute product reviews. Amazon does not provide any direct monetary incentives for the reviewers to write reviews, but they do provide two publicly viewable leaderboards to recognize the contributions of the product reviewers. One leaderboard, named Classic, weights the lifetime number of reviews more heavily than the New leaderboard, which was introduced on August 23, 2008. We found that reviewers ranked under the New leaderboard are more likely to respond to the review-writing behavior of their neighbors in the New ranking. This suggests that the social information presented by the New ranking mechanism is more motivating than the Classic ranking mechanism.

Chapter 2

Package Auctions: A Laboratory Study

2.1 Introduction

As the Internet increases possibilities for collaborative research and shared resources across geographically dispersed groups, a key question arises in how best to allocate these resources. This study focuses on one type of shared resource system called the "collaboratory". Researchers in a collaboratory interact with colleagues, access instrumentation, and share data and computational resources without regard to physical location (Wulf, 1993). Collaboratories exist within multiple scientific communities, including space physics, medicine, software engineering, and neuroscience (Finholt, 2002).

One collaboratory is the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) which focuses on accelerating and improving earthquake engineering research. The NEES collaboratory connects earthquake engineering researchers to 16 state-of-the-art laboratory facilities in the United States designed to combine their capabilities through the Internet. Through NEES, remote participants can access collaboratory studies, observe experiments, and under highly controlled circumstances, control experiments in real-time.

A critical element of the NEES vision, and of collaboratories generally, is that Internet-mediated methods for conducting and observing research should increase resource access. In the past, proximity to unique instruments has created differential access to instruments and to the community of scientists who use the instruments (Hagstrom, 1965; Traweek, 1992). In NEES, a critical goal of making the 16 new sites accessible over the Internet is to broaden use of the facilities, particularly among researchers at institutions that lack earthquake engineering research equipment.

At an abstract level, the problem confronting NEES is a mechanism design problem. That is, NEES controls instrument time designated for shared use at the 16 laboratories (i.e., equipment used by researchers in the broader earthquake engineering community. Economically, this instrument time has value for researchers, as demand for instrument time exceeds supply. A second element of the problem is that contiguous time slots are more valuable than the sum of separate slots, that is, user valuation for multiple slots exhibits synergy. Therefore, package auctions might be an appropriate mechanism to determine an efficient allocation of equipment time.

A variety of problems has used package auction mechanisms for resource allocation, such as the Federal Communications Commission (FCC) spectrum auctions (see, e.g., Brunner et al. (2006)) the Business-to-Business (B2B) auction (Katok and Roth, 2004), and transportation procurement in the private (Ledyard et al., 2002) and public sectors (Cantillon and Pesendorfer, 2006). We refer to Milgrom (2007) for an overview of recent developments in the use of package auctions and exchanges to allocate resources.

In this paper, we construct a laboratory environment, within which we evaluate the performance of three allocation mechanisms. Although the allocation of equipment time for NEES guides our choice of environment and parameters, our results provide new insights relevant to other scheduling and allocation problems. Compared with past research on scheduling, our paper differs in both the allocation mechanisms used and the characteristics of the environment. First, we allow multiple slots and package bidding. By contrast, Olson and Porter (1994) compare four mechanisms in a laboratory setting but assign each agent at most one slot.

When agents can demand multiple slots with complementarities, the scheduling problem is considerably more complex. In our study, we allow multiple slots similar to Ledyard et al. (1996), where they compare variants of the Adaptive User Selection Mechanism (AUSM) with a committee process. Like Ledyard et al. (1996), we also allow multiple slots and package bidding. Second, we test a new ascending package auction mechanism, the Resource Allocation Design (RAD), which is created by merging features of the AUSM and the FCC Simultaneous Multiple Round auction designs. This hybrid mechanism is shown to perform better than either parent (Kwasnica et al., 2005). We compare the performance of RAD, VCG and a Knapsack mechanism, to determine which provides the most efficient allocation of resources in a laboratory setting.

The rest of the paper is divided into five sections. In Section 2.2, we identify the three allocation mechanisms and describe their respective features. In Section 2.3, we describe our experimental design. Section 2.4 presents our results, and we discuss the results in the final section.

2.2 Allocation Mechanisms

In this section, we outline our three allocation mechanisms. For each, we assume that a critical feature of the instrument time allocation problem is that contiguous time slots are more valuable than the sum of separate slots. In the case of the NEES collaboratory, the difficulty of experiment set-up and teardown dictates a preference for consecutive time intervals to minimize installation effort. We consider two package auction mechanisms, VCG and RAD, compared with an ordinal ranking mechanism, Knapsack, selected as a best-case representation of current allocation methods. The VCG auction (Vickrey, 1961; Clarke, 1971; Groves, 1973) is an important standard for nearly all mechanism design work, and for auctions in particular. Therefore, we use it as a benchmark to assess the performance of our ascending bid auction and the Knapsack mechanism. Among ascending bid auctions that allow package bidding, we choose RAD, which to our best knowledge, no other ascending auctions has passed it in terms of efficiency in experimental investigations (Kwasnica et al., 2005; Brunner et al., 2006). To introduce our allocation mechanisms, we first set up a simple framework that allows us to explain the auctions clearly.

In this framework, let $N = \{1, \ldots, n\}$ be a finite set of bidders. Let $K = \{1, \ldots, k\}$ represent the set of objects to be sold, and $X = \{0, 1\}^k$ represent the set of combinations of objects. Let $\mathscr{B}_i$ be a set of *bids* of bidder i. In this design, each bid is a pair, $b = (\mathbf{x}, p)$, where $\mathbf{x} \in X$ corresponds to the packages desired and $p \in \mathbb{R}_+$ is the bid price. Let $v_i : X \to \mathbb{R}_+$ be bidder i's *valuation function* that assigns a value to a package. For iterative auctions, let $t = 1, 2, 3, \ldots$ represent the iterations or rounds. Finally, let $\delta = (\delta_1, \ldots, \delta_b)$ be an indicator vector, where $\delta_j \in \{0, 1\}$ indicates whether bid (x_j, p_j) is *winning* or *losing*, and where b is the total number of bids.

In the allocation context, the auctioneer wants to maximize efficiency. If the bidder bids truthfully, she can achieve this goal by solving the winner determination problem. We now define a truthful bid: a bid $(\mathbf{x}, p) \in \mathscr{B}_i$ is a *truthful bid* if $v_i(\mathbf{x}) = p$. As the winner determination problem is part of every package auction design, we formally define this problem for our experiment.

The package auction winner determination problem is to maximize the sum of all bid prices, indicating each bid as *winning* or *losing*, under the constraint that each item can be sold to, at most, one bidder:

$$\max_{\delta} \sum_{j=1}^{b} \delta_j p_j \quad \text{subject to} \quad \sum_{j \in \{j: \delta_j = 1\}} x_j \leq (1, 1, \ldots, 1). \tag{2.1}$$

The allocation associated with the solution to the winner determination problem maximizes the aggregate surplus when every bidder bids for all packages and the bids are all truthful.

2.2.1 The VCG Auction

The VCG auction with package bidding is dominant strategy incentive compatible, that is, bidding one's true valuation is always optimal. Furthermore, it implements the efficient outcome. The VCG auction is comprised of the following steps:

1. At the beginning of each auction, bidders select the packages they would like to bid on and the amount they would like to bid for each package.
2. Once all bidders have submitted their bids, the auctioneer solves the winner determination problem (Equation (2.1)), by choosing the combination of submitted bids that yields the highest sum of bids.
3. To compute prices, the auctioneer then chooses each winning bidder as a pivotal bidder. The auctioneer solves the winner determination problem again, ignoring the bids of the pivotal bidder. The auctioneer then compares the sum of bids generated by this allocation with those generated when no bids are excluded. The difference is the rebate for the pivotal bidder.
4. At the end of the auction, a winning bidder's price is the difference between her bid and her rebate.

Despite its attractive theoretical properties, the VCG auction has some practical disadvantages. In the package auction context, there are three main concerns.[1] First, the VCG auction might be vulnerable to collusion. For example, bidders have the incentive to use shill bidders to manipulate the allocation and prices in their favor. Second, it might suffer from a monotonicity problem,[2] that is adding bidders may reduce equilibrium revenues. Third, previous laboratory experiments show that in the single object case, some experimental subjects do not learn the dominant strategy, and consistently overbid in second-price auctions (Kagel, 1995).

[1]Note the first two problems do not appear when all goods are substitutes for all bidders.
[2]See Milgrom (2004) Chapter 8 for examples.

When the VCG auction is extended to allow package bidding, it is more efficient. Laboratory experiments indicate that the VCG auction can achieve high allocation efficiency in small multi-object auctions with package bidding (Isaac and James, 2000), and that it outperforms a complex ascending bid auction in terms of efficiency and revenue (Chen and Takeuchi, 2005). It is an open question whether it can retain its performance in more complex environments.[3] This study addresses this issue by expanding the number of packages.

2.2.2 Resource Allocation Design (RAD)

The RAD mechanism is an iterative ascending bid package auction proposed by Kwasnica et al. (2005). The RAD mechanism uses the package bidding of the AUSM mechanism (Banks et al., 1989) and the eligibility and price improvement rules of the Milgrom FCC Design (Milgrom, 2004).

The *eligibility rule* places an upper limit on the number of items a bidder can bid on in a round, based on one's past bidding. A bidder's eligibility is the maximum number of items she is allowed to bid on in round t, which is exactly the number of items on which she had active bids in round $t-1$. In the first round, she can bid on any and all packages. In each subsequent round, she is allowed to bid on as many packages that jointly contain as many items as she has placed bids on in the previous round. The effect of the eligibility rule can be briefly summarized as a "use it or lose it" rule.

The *price improvement rule* specifies a minimum price for each package based on the bids submitted in the previous round and a price improvement factor. The price improvement rule, in conjunction with the eligibility rule, helps to drive an auction to a close. We now describe the algorithm.

1. RAD is a simultaneous, multi-round auction in which bidders may submit bids on packages in each round. Each bidder can bid on packages that are feasible given her eligibility constraint.
2. In the first round, an acceptable bid must equal or exceed the minimum opening bid for each package. After each subsequent round, prices are calculated for each package on the basis of bids received in this round. Bids must equal or exceed the prices for each package.
3. The auctioneer first solves the winner determination problem, (Equation (2.1)), by maximizing revenue, subject to feasibility. Bids selected this way are called

[3] For a comprehensive analysis of the VCG auction, including its practical limitations, see Ausubel and Milgrom (2006).

provisional winning bids.

4. The auctioneer then calculates a market clearing price for each object, Π^t. "The pricing rule calculates prices that reflect (as closely as possible) the marginal sales revenue of each package based on bids received" (Brunner et al., 2006). Prices for packages are given by the sum of the prices of all items in the package. More specifically, let W^t be the set of provisional winning bids at round t, and $L^t = B^t \setminus W^t$ be the set of losing bids at t. To compute prices Π^{t+1}, we solve the following problem:

$$\min_{\Pi^t, Z, g} Z$$

Subject to:

$$\sum_{k \in K} \Pi^t_k x_{jk} = p_j, \text{ for all } b_j = (p_j, x_j) \in W^t$$

$$\sum_{k \in K} \Pi^t_k x_{jk} + g_j \geq p_j, \text{ for all } b_j = (p_j, x_j) \in L^t$$

$$0 \leq g_j \leq Z, \text{ for all } b_j \in L^t$$

$$\Pi^t \geq 0,$$

where Z is the bound on the gap between the bid price for a package and the sum of prices for individual objects in a package.

5. Provisional winning bids are automatically resubmitted.
6. At the end of each round, bidders receive information on all provisionally winning bids.
7. The auction closes if the provisional winning bids remain the same for two consecutive rounds. In this case, provisionally winning bids become winning bids and used to determine the final allocation.
If the closing rule is not met, the auction proceeds to another round, and the minimum bid prices increase by a factor above the market clearing prices. In our experiment, this factor is set to 10%.

2.2.3 Knapsack with Ordinal Ranking

Many scientific communities allocate instrument time via an individual, a committee (Olson and Porter, 1994) or some formal or informal optimization procedure that uses ordinal ranking information from potential instrument users. The Knapsack mechanism is an idealized representation of the latter type of decision. In this mechanism, everyone submits ordinal rankings of packages, from the top choice (# 1) to the last choice (# k for k packages). The top choice is awarded k points, second choice $k-1$ points, ..., and so on. A computer then allocates goods to maximize the total number of points subject to the constraint that a bidder can win, at most, one package. If several allocations yield the same total points, the computer program randomly chooses one of the allocations.

However, the Knapsack mechanism with ordinal ranking poses two problems: first, truth-telling is not a dominant strategy. [4] and second, bidders cannot indicate the intensity of their preferences. Therefore, we expect that the Knapsack mechanism will generate lower efficiency than either VCG and RAD will. Because of the prevalence of knapsack-like mechanisms in scheduling, we use it as a benchmark of typical allocation approaches.

2.3 Experimental Design

Our experimental design reflects both theoretical and technical considerations. Specifically, we are interested in two important questions. First, how do the VCG, RAD and Knapsack mechanisms compare in performance? Second, how do subjects respond to the incentives in each mechanism? In this section we describe our economic environment and experimental procedures.

2.3.1 Economic Environment

The environment is designed to capture the essential aspects of a scientific collaboratory, such as NEES. As the number of approved projects for NEES range from 9 to 20 each year, we include 9 researchers in each experimental session. There are two dimensions of heterogeneity in our model: the size of the project and the time preference. We operationalize these two dimensions as researcher (or bidder) types. Large project researchers have larger grants and projects, which take longer to complete than small project researchers. In the time preference dimension, we have three types: those who prefer early slots, those who prefer late slots, and those who are indifferent.

Based on actual NEES operation horizons, we assume the facility is available for 24 months. We set 1 month as a minimum unit. In this model, let $\underline{m}$ be the

[4] For example, suppose we have 3 bidders, {A, B, C}, and 3 months, {1, 2, 3}. Suppose that each bidder can get, at most, 1 month. If $1 \succ_i 2 \succ_i 3$ for $i = A, B$, and $3 \succ_i 2 \succ_i 1$ for $i = C$, and $v_i(m) > 0$ for any i and m. Truthful preference revelation by all bidders will results in the allocations {(A 1), (B 2), (C, 3)} or {(A 2), (B 1), (C, 3)}, with equal probability. However, suppose bidder A misrepresents her preference by ranking only her top choice, month 1, while bidders B and C truthfully rank all months. The unique allocation is {(A 1), (B 2), (C, 3)}, which makes A strictly better off.

required minimum number of months to run a project, with $\underline{\mathrm{m}} = 3$ for large projects and $\underline{\mathrm{m}} = 2$ for small projects. Let $\underline{v}_i$ denote the value to bidder i from the minimum number of months. Each additional month to $\underline{\mathrm{m}}$ gives a researcher value β_i, but she has no value for any additional time beyond $\bar{m}$. The discount factor, δ_i, denotes a researcher's time preference, while *lag* denotes the months between the ideal beginning month and the actual beginning month, m_{begin}. A researcher with access to the facility from m_{begin} to m_{end} has the following utility function:

$$v_i(m_{\text{begin}}, m_{\text{end}}) = \begin{cases} 0 & \text{if } m_{\text{end}} - m_{\text{begin}} + 1 < \underline{\mathrm{m}}, \\ \delta_i^{lag}\left(\underline{v}_i + \beta_i(m_{\text{end}} - m_{\text{begin}} - \underline{\mathrm{m}} + 1)\right), & \text{if } \underline{\mathrm{m}} \leq m_{\text{end}} - m_{\text{begin}} + 1 \leq \bar{m}, \\ \delta_i^{lag}\left(\underline{v}_i + \beta_i(\bar{m} - \underline{\mathrm{m}})\right), & \text{if } m_{\text{end}} - m_{\text{begin}} + 1 > \bar{m} \end{cases} \quad (2.2)$$

where

$$lag = \begin{cases} m_{\text{begin}} - 1 & \text{for researchers who prefer early months,} \\ m_{\text{begin}} - 23 & \text{for researchers who prefer late months,} \\ 0 & \text{for researchers who are indifferent.} \end{cases} \quad (2.3)$$

The discount factor, δ_i, is set to 0.9, 1.0 and 1.2 for those who prefer early, are indifferent and prefer late months respectively; parameters $\underline{v}_i$ and β_i are i.i.d. draws from the following uniform distributions at the beginning of each auction:

$$\beta_i \sim U[10, 20], \quad (2.4)$$

$$\underline{v}_i \sim U[20, 100] \quad \text{for small projects,} \quad (2.5)$$

$$\underline{v}_i \sim U[20, 150] \quad \text{for large projects.} \quad (2.6)$$

A small project bidder has values for packages that last for 2 to 4 months. A large project bidder has values for packages that last for 3 to 5 months.[5] As a result, a small project bidder can bid on packages that start on month 1 to month 23, while large project bidders can bid on packages starting from month 1 to month 22. This means that a large project bidder bids on 63 packages, while a small project bidder bids on 66 packages.

[5] Again, these parameters are chosen based on the NEES environment, where allocations of 6 months or longer are politically infeasible.

In sum, the values of the packages are determined by the value of the minimal package, the bidder discount factor, and the value of an additional month, β. The value of the minimal package in turn is determined by the bidder's type. Given the starting month, a larger package is (weakly) preferred to a smaller one by any type of bidder.

Bidder ID	Project Type	Time Preference	(δ_i)	Package Size
1	Large	Prefer late	(1.2)	3 to 5 months
2	Large	Indifferent	(1.0)	3 to 5 months
3	Large	Prefer early	(0.9)	3 to 5 months
4	Small	Prefer late	(1.2)	2 to 4 months
5	Small	Indifferent	(1.0)	2 to 4 months
6	Small	Indifferent	(1.0)	2 to 4 months
7	Small	Prefer early	(0.9)	2 to 4 months
8	Small	Prefer early	(0.9)	2 to 4 months
9	Small	Prefer early	(0.9)	2 to 4 months
Total Demand				21 to 39 months

Table 2.1 Design Parameters: Bidder Preferences

Table 2.1 summarizes the main features of bidder preferences. Each experimental session has 9 bidders, where 3 have large projects and 6 small projects. The 3 large project bidders each have different discount factors, while among the 6 small project bidders, one has a discount factor of 1.2, two have 1.0, and three have 0.9. Early months are more competitive than late ones, as earlier potential scientific discoveries are more valuable. The total demand for months is between 21 and 39 months, while the total supply is 24 months.

2.3.2 Experimental Procedures

In our experiment, each session required exactly 9 participants. The participants were recruited from the University of Michigan, including both the graduate and undergraduate population, with majors in science, math, or engineering. At the beginning of each session participants drew from a deck of index cards to determine their project type (large or small) and their discount factor (1.2, 1, or 0.9). After instructions, participants took a quiz to test their understanding of the mechanism. The experimenters then reviewed the answers with the participants.

There were no practice auctions in any of the experimental conditions. In the VCG and Knapsack treatments, participants had 7 minutes per auction to input their bids, while in the RAD treatment, they had 4 minutes for the first round and 2 minutes for all subsequent rounds in each auction. There were a total of 8 auctions in total in each of the VCG and Knapsack sessions, and 3 to 5 auctions in each of the RAD sessions.[6]

At the end of each VCG or RAD auction or Knapsack allocation, each participant was given only the following feedback: (1) the package that they won, if any and (2) the profit they made. Furthermore, at the end of each RAD round, the participants could see: (1) the provisional winning package, if any; (2) the minimum market prices of all the packages; and (3) the set of bids available given the eligibility constraint. As the participants in RAD could see the minimum prices of the packages, calculated to reflect all the bids received, they might infer the bids of others.

At the conclusion of each auctions, participants tallied their cumulative earnings, filled out a short demographic survey, and wrote down their strategies. Participants were paid based on their experimental profits at the end of each session.

Mechanism	# sessions	# participants	Exchange Rate
VCG	5	45	12
RAD	5	45	4
Knapsack	5	45	20

Table 2.2 Features of Experimental Sessions

Table 2.2 presents the features of the experimental sessions, including mechanisms, number of experimental sessions, number of participants in each condition and exchange rates. Overall, 15 independent computerized sessions were conducted in the RCGD laboratory at the University of Michigan from July 2005 to April 2006. No participant was used in more than one session, yielding a total of 135 participants across all treatments. Each session lasted approximately 2.5 hours. In addition to their auction earnings, participants could earn money based on their quiz answers. A participant with fully correct answers could earn up to $5. The average total earnings were $34.44, and the standard deviation was $16.21. Further data are available from the authors upon request. Experimental instructions are included in the supplemental material.

[6]Recall that RAD is an iterative auction; therefore, the number of rounds in each auction is endogenous and varies from auction to auction. For the same reason, the exchange rate for RAD is lower than that for the VCG or Knapsack treatments.

2.4 Results

In this section, we first examine individual bidder behavior in each allocation mechanism. We then compare the aggregate performance of the mechanisms in terms of efficiency and equity.

2.4.1 Individual Behavior in VCG

Our VCG auction experiment consists of 5 independent sessions, each of which has 9 subjects. Therefore, we have 45 subjects in total. In each experimental session, a subject participates in 8 auctions. In each auction, a subject can bid on any of the 63 (or 66) packages.

A bidder's strategy in a VCG package auction has two dimensions: whether to bid on a package, and how much to bid. A bidder's strategy on either dimension affects her profit. Recall it is a weakly dominant strategy for each bidder to bid on all packages, and to bid her true value.

	Active Bids%	Bid/Value
Auction 1	0.291	0.551
Auction 2	0.434	0.574
Auction 3	0.441	0.582
Auction 4	0.480	0.708
Auction 5	0.499	0.726
Auction 6	0.537	0.734
Auction 7	0.565	0.742
Auction 8	0.587	1.038

Table 2.3 Bidding Under VCG

Table 2.3 presents the proportion of active bids and the bid/value ratio under the VCG auction, averaged across all sessions. Bidders in a VCG auction, on average, bid on 47.94% of the packages, and bid 70.69% of their true value. However, both the proportion of active bids and the Bid/Value ratio are increasing over time, indicating that bidders are learning the weakly dominant strategy.

To investigate which factors induce a higher proportion of active bids, we use a probit model with robust clustering at the session level. Table 2.4 presents results from four probit specifications. The dependent variable is PlaceBid, a dummy variable, which equals one if a bidder places a bid on a package and zero otherwise. Independent variables include Value of a package; AuctionNumber, which captures the effect of learning; PreferEarly, a dummy variable which equals one if a bidder prefers early months, and zero otherwise; PreferLate, a dummy variable which equals one if a bidder prefers late months and zero otherwise; and LargeProject, a dummy variable which equals one if the bidder represents a large project, and zero otherwise. While specification (1) is our basic model, (2) adds the effects of learning, (3) adds the effect of time preferences, and (4) investigates the effect of project size (large vs. small).

Result 1 (Decision to Bid) *While the proportion of active bids is significantly less than one, bidders are significantly more likely to bid on packages with higher values. The proportion of active bids increases significantly over time. Controlling for the value of packages, bidders who prefer early months are significantly more likely to bid on packages.*

Support 1 *As bidders interact, we analyze the bidding probability based on the mean proportions of active bids for each of the 5 sessions, which are* $(0.676, 0.287, 0.413, 0.526, 0.495)$ *respectively. The sign test rejects the null hypothesis that the proportion of active bids is equal to one (*$p = 0.043$*). Table 2.4 presents results from probit specifications. The coefficients are probability derivatives. Increasing Value by 1 increases the likelihood of bidding on a package by 0.4% (*$p < 0.01$*). The likelihood of bidding also increases in the AuctionNumber by 3.9% (*$p < 0.01$*), indicating a significant learning effect. Controlling for Value and AuctionNumber, PreferEarly increases the likelihood of bidding on a package by 20.9% (*$p < 0.05$*).* ∎

In our laboratory setting, 4 bidders prefer early months while only 2 bidders prefer late months. Thus, the significant coefficient on PreferEarly indicates the effect of the competition among bidders who prefer early months.

We now explore the second dimension of the bidding strategy, how much to bid on a package. We classify VCG bidders into three categories: underbidder, truthful bidder and overbidder. Specifically, we run the following simple OLS regression on active bids for each bidder with robust clustering at the auction level:

$$\text{Bid}_{pt} = \beta \text{Value}_{pt} + \varepsilon_{pt}, \tag{2.7}$$

where p denotes package and t denotes the auction number. We then test the null hypothesis: $\hat{\beta} = 1$. Based on the results, we classify each bidder into one of the three categories. A bidder is classified as an *underbidder* if we can reject the hypothesis of truthful bidding at the 5% level and the coefficient is below 1. She is classified as a *truthful bidder* if we cannot reject the hypothesis of truthful bidding at the 5%

	Dependent Variable: PlaceBid			
	(1)	(2)	(3)	(4)
Value	0.004	0.004	0.005	0.006
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
AuctionNumber		0.039	0.039	0.039
		(0.003)***	(0.003)***	(0.003)***
PreferEarly			0.197	0.209
			(0.107)*	(0.100)**
PreferLate			0.149	0.19
			(0.173)	(0.153)
LargeProject				-0.101
				(0.072)
Observations	23400	23400	23400	23400

Notes:
1. Robust standard errors in parentheses are adjusted for clustering at the session level.
2. Significant at: * 10% level; ** 5% level; *** 1% level.

Table 2.4 The Likelihood of Bidding on a Package in VCG Auctions

level. Lastly, she is classified as an *overbidder* if we can reject the hypothesis at the 5% level and the coefficient is above 1. We summarize the bidding behavior in the VCG auction in the following result.

Result 2 (Bid Price in VCG) *Bidders in a VCG auction, on average, bid 76.1% of their true value. Of our participants, 73.3% can be classified as underbidders, 20.0% as truthful bidders and 6.7% as overbidders.*

Support 2 *The OLS estimate of $\hat{\beta}$ in Equation (2.7) is 0.646, with the robust standard errors clustered at the session level equal to 0.071. A two-sided Wald test rejects the null hypothesis of bids being equal to values ($p = 0.008$). We classify bidders using individual regressions at the individual level. The average R^2 of individual regressions is* 0.895*, with a standard deviation of* 0.116*.* ∎

Our finding that most bidders either underbid or bid their true value in VCG auctions is consistent with those in Isaac and James (2000) and Chen and Takeuchi (2005). However, most previous laboratory studies of single-unit VCG auctions find that bidders tend to overbid in such environments (Kagel, 1995). As the VCG package auction is more complex than the single-unit version is, subjects in this case might start cautiously and thus bid below value.

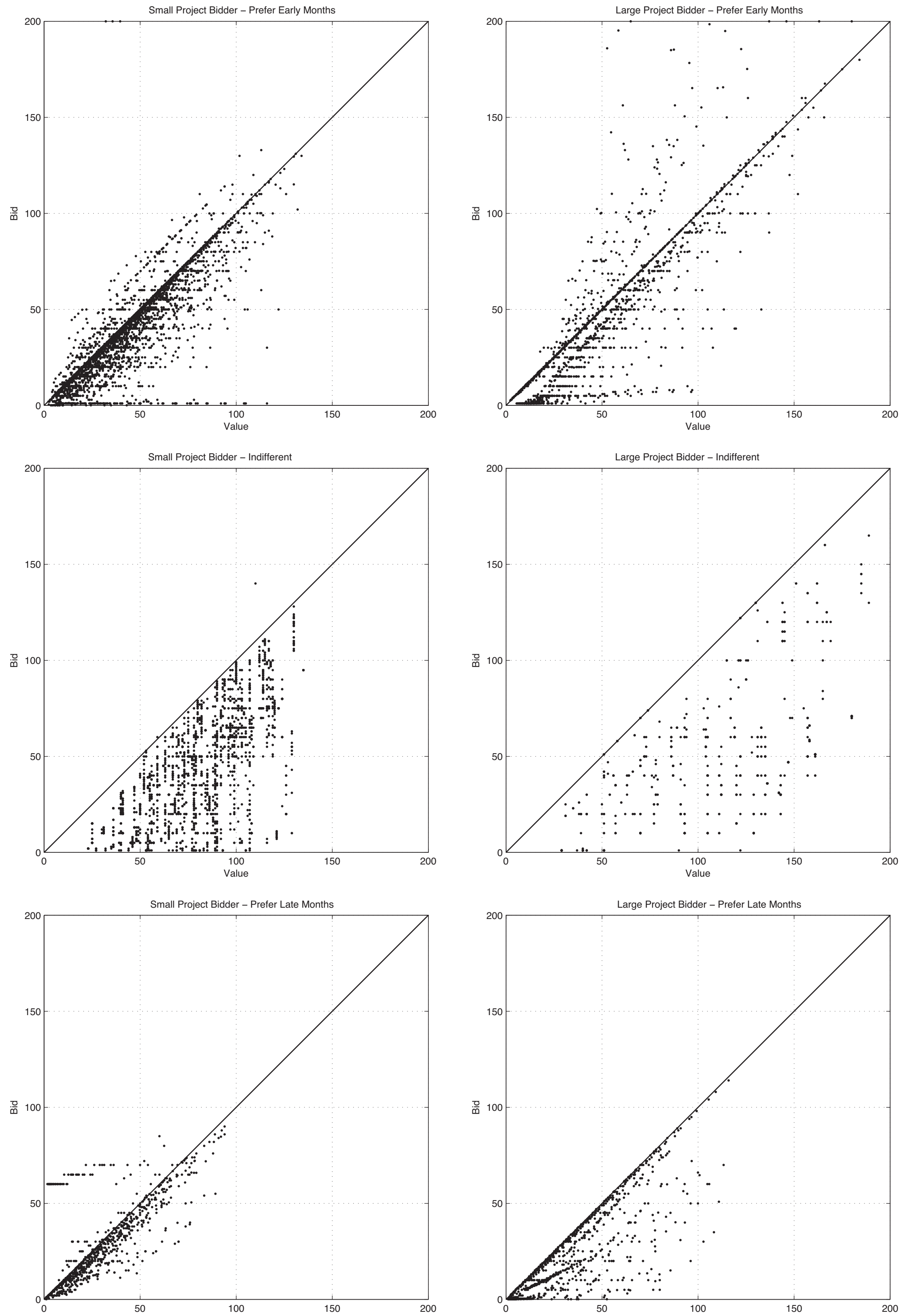


Figure 2.1 Bidding Behavior in VCG Auctions

Figure 2.1 presents the scatter plot of raw bids under the VCG auction for small (left column) and large (right column) project bidders. These bids show that early months are more competitive; most of the bids over value (above the 45 degree line) are from researchers who prefer early months.

	Dependent Variable: OverBid	
	(1)	(2)
Value	0.0001	-0.0002
	(0.0002)	(0.000)
PreferEarly	0.240	0.227
	(0.099)***	(0.095)***
PreferLate	0.321	0.151
	(0.156)***	(0.126)*
PreferEarly*Startmonth		-0.003
		(0.002)***
PreferLate*Startmonth		0.000
		(0.001)
Indifferent*Startmonth		-0.001
		(0.000)***
Observations	11217	11217

Robust standard errors in parentheses are adjusted for clustering
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.5 The Likelihood of Overbidding in VCG Auctions

We use a probit model to examine the likelihood of overbidding. Table 2.5 presents the results from two specifications. The dependent variable is OverBid, a dummy variable, which equals one if a bid on a package is greater than the value and zero otherwise. Independent variables include Value of a package, PreferEarly, PreferLate, Indifferent, and Startmonth, the beginning month of a package. Specification (1) indicates that bidders who prefer early or late months are more likely to overbid compared with indifferent bidders. Specification (2) reveals the characteristics of packages bidders are likely to overbid on. While the interaction between PreferEarly and Startmonth is negative and significant, the interaction between PreferLate and Startmonth is not significant, with the estimated coefficient being zero. The likelihood of overbidding is decreasing in Startmonth for bidders who prefer early months and for those who are indifferent, but not for bidders who prefer late months. This result also highlights the effect of competition among bidders.

2.4.2 Individual Behavior in the Knapsack Mechanism

Our Knapsack experiment consists of 5 independent sessions, each of which has 9 subjects. Therefore, we have a total of 45 subjects. In each experimental session, a subject participates in 8 allocations. In each allocation, a subject can rank any of the 63 (or 66) packages between 1 and 63 (or 66).

As in a VCG auction, a participant's strategy in a Knapsack allocation has two dimensions: whether to rank a package and how to rank it. A participant's strategy on either dimension affects her profit. Recall that truth-telling is not a dominant strategy in our Knapsack mechanism. However, truthful revelation might be a focal point.

	Active Bids%	Bid/Value
Auction 1	0.510	1.304
Auction 2	0.591	1.241
Auction 3	0.580	1.207
Auction 4	0.597	1.111
Auction 5	0.599	1.106
Auction 6	0.585	1.110
Auction 7	0.579	1.104
Auction 8	0.585	1.105

Table 2.6 Ranking Under Knapsack

We define an analogous Bid/Value ratio for the Knapsack mechanism as follows. For the submitted ranking r for a package x, we find the corresponding package, x^r, which yields the same ranking r if it is submitted truthfully. We then calculate the ratio of the value of x^r to the value of x, $v_i(x^r)/v_i(x)$. We use this ratio as the Bid/Value ratio under the Knapsack mechanism. For submitted ranks, we regress the value of x^r on the value of the package x with no constant, and then classify the bidders with the same criteria as in the VCG auction.

Table 2.6 presents the proportion of active rankings and the Bid/Value ratio under the Knapsack mechanism, averaged across all sessions. Bidders in a Knapsack allocation, on average, rank 57.8% of the packages, with a Bid/Value ratio of 115.8%.

To investigate which factors induce a higher proportion of active bids, we use a probit model with robust clustering at the session level. Table 2.7 presents the results from four specifications. The dependent variable is PlaceRank, a dummy variable, which equals one if a participant ranks a package and zero otherwise. Independent

	Dependent Variable: PlaceRank			
	(1)	(2)	(3)	(4)
Value	0.004	0.005	0.006	0.006
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
AllocationNumber		0.007	0.008	0.008
		(0.007)	(0.008)	(0.008)
PreferEarly			0.159	0.155
			(0.099)	(0.100)
PreferLate			0.154	0.143
			(0.108)	(0.146)
LargeProject				0.030
				(0.160)
Observations	23400	23400	23400	23400

Notes:
1. Robust standard errors in parentheses are adjusted for clustering at the session level.
2. Significant at: * 10% level; ** 5% level; *** 1% level.

Table 2.7 The Likelihood of Ranking Packages in Knapsack Allocations

variables include Value of a package; AllocationNumber, which captures the effect of learning; PreferEarly, a dummy variable which equals one if a participant prefers early months, and zero otherwise; PreferLate, a dummy variable which equals one if a participant prefers late months, and zero otherwise; and LargeProject, a dummy variable which equals one if the participant represents a large project, and zero otherwise. Again, while specification (1) is our basic model, specification (2) adds the effects of learning, (3) adds the effects of time preferences, and (4) investigates the effects of player type (large vs. small).

Result 3 (Whether to Rank a Package) *While the proportion of active rankings is significantly less than one, bidders are significantly more likely to rank packages with higher values. The proportion of active rankings does not increase over time. Controlling for the value of packages, bidder time preferences have no significant effect on the likelihood of ranking packages.*

Support 3 *As bidders interact with each other within a session, we analyze the ranking probability based on the mean proportion of active bids for each of the five sessions, which are* $0.491, 0.608, 0.762, 0.437, and 0.594$ *respectively. The sign test rejects the null of the proportion of active rankings equal to one (*$p = 0.041$*, one-sided). Table 2.7 presents the results from probit specifications. The coefficients are probability derivatives. An increase* *in the Value of a package by 1 increases the likelihood of bidding on a package by 0.45% ($p < 0.001$). The likelihood of bidding does not significantly increase in the Allocation-Number ($p = 0.294$), indicating no learning. Neither time preference nor project size has a significant effect on the likelihood of ranking a package ($p > 0.10$).* ∎

We now explore the second dimension of ranking strategy, how to rank a package. We first look at a scatter plot presentation of raw bids in Knapsack allocations.

Figure 2.2 presents the scatter plot of the raw bids under the Knapsack mechanism for small and large project bidders. The horizontal axis represents the true value of a ranked package, $v_i(x)$, while the vertical axis represents the value of the package for the reported rank, $v_i(x^r)$. Note that the panels for bidders who are indifferent have only three values for all packages per auction.

In contrast to the VCG auction, time preference does not seem to be a significant factor in over-ranking (above the 45 degree line).

Using a similar approach to classify the participants as in the VCG auction, we find the following result.

Result 4 (Ranking in Knapsack) *Of the participants in the Knapsack allocation, 20.0% can be classified as underbidders, 64.4% as truthful bidders and 15.6% as overbidders.*

Support 4 *The OLS estimate of $\hat{\beta}$ in Equation (2.7) is 1.005, with the robust standard errors clustered at the session level equal to 0.004. A two-sided Wald test rejects the null hypothesis of bids being equal to values ($p < 0.001$). We classify bidders using individual regressions. The average R^2 of individual regressions is* 0.945, *with a standard deviation of* 0.081. ∎

Since the Knapsack mechanism lacks a theoretical benchmark, we take the empirical distribution of values as given and use simulations to generate our benchmarks for comparison purposes. In particular, we look at two categories of simulations. In the first category, which we refer to as "benchmark" simulations, we rank the packages on behalf of the bidders to evaluate the surplus generated under different ranking strategies. There are four sets of simulations under this category, determined by two dimensions. The first dimension is the number of packages. The second dimension is whether we rank the packages truthfully or randomly. The four cases then are: (a) rank all the packages truthfully, (b) rank a random number of packages truthfully, (c) rank all the packages randomly and finally, (d) rank a random number of packages randomly.

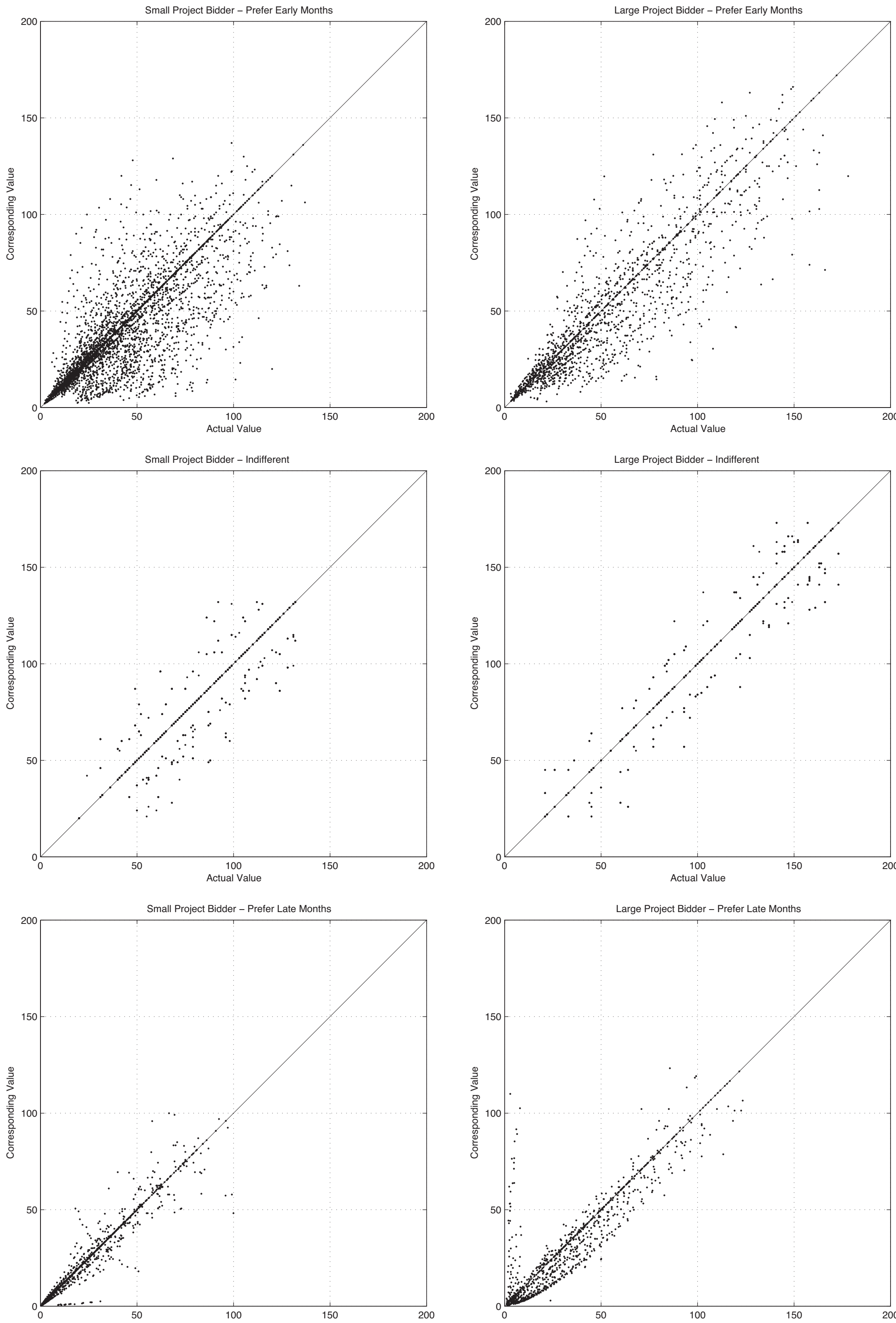


Figure 2.2 Ranking Behavior in the Knapsack Mechanism

We now describe the algorithm for (d), as the methods for the other three categories can be inferred accordingly. Recall that we have 5 independent sessions, each with 8 allocations, yielding a total of 40 allocations. In running the simulations, we submit rankings on behalf of the bidders. For each of the 40 allocations, and for each of the 9 bidders in a given allocation, we randomly pick the number of packages the bidder would rank from a discrete uniform distribution on the integer values with a support of 0 to 63 for a large project bidder and 0 to 66 for a small project bidder. While there is only one way to rank all the packages truthfully, for each of the other three cases, we run 25 iterations and compute the average of the total surplus.

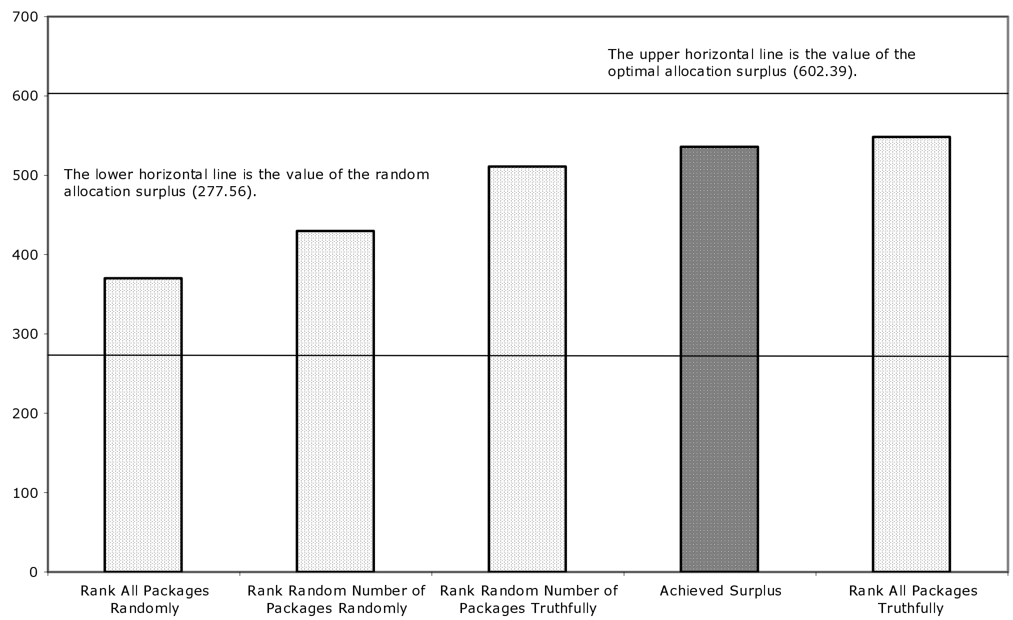


Figure 2.3 Average Total Surplus from Benchmark Simulations

Figure 2.3 presents the average total surplus for each of the four benchmark simulations, as well as the actual achieved total surplus in the experiments. The upper horizontal line is the total surplus of the optimal allocation, while the lower horizontal line is the total surplus from a random allocation process. Note that the actual achieved surplus in the experiment is close to what would result from ranking all packages truthfully.

Result 5 (Truthful Ranking) *With the empirical distribution of values for the 40 allocations, ranking all bids truthfully yields the highest surplus, followed by ranking a random number of bids truthfully, ranking a random number of bids randomly, and ranking all bids* *randomly.*

Support 5 *Sign tests reject the null of equal surplus for each pairwise comparison of aggregate surplus of the four simulations and the achieved surplus ($p < 0.001$).* ■

To investigate whether participant strategies are the best responses to the empirical distributions, we manipulate the rankings in two ways. First, we examine whether we can increase bidder profit by ranking more packages. For each allocation, holding the rankings of the other bidders constant, we choose one bidder, and complete her rankings truthfully, with the following result.

Result 6 (Strategic Truncation) *When rankings of packages are completed truthfully for each bidder, one at a time, the average bidder profit remains the same as the actual observed profit.*

Support 6 *For each session, we compute the average simulated profits by completing the ranking truthfully for a given bidder. We then compare the average simulated profits with actual session-level average profits. We find that out of 360 ranking vectors, a simulated completion of 8 vectors (2.22%) increases bidder profit; in 16 cases (4.44%), it decreases; and in 336 out of 360 cases (93.33%), bidder profit remains the same. A two-sided permutation test fails to reject the null of equal profits ($p = 0.4174$).* ■

To find out whether bidder profit might increase if each were to rank packages truthfully, we choose a bidder, and re-rank all her bids truthfully, while keeping others' rankings constant, which leads to the following result.

Result 7 (Strategic Ranking) *When bidder rankings of packages are corrected to truthful rankings, one bidder at a time, average bidder profit remains the same.*

Support 7 *For each session, we compute the average simulated profits by ranking all packages truthfully for each bidder, one at a time. We then compare the average simulated profits with actual session-level average profits. We find that in 60 out of 351 cases (17.09%), bidder profit increases; in 38 out of 351 cases (10.83%), it decreases; and in 253 out of 351 cases (72.08%), bidder profit remains the same.*[7] *A two-sided permutation test fails to reject the null of equal profits ($p = 0.3824$).* ■

Together, Results 6 and 7 indicate that, on average, bidder profits would not increase significantly, had they been more truthful. Another interpretation is that actual bidder ranking strategies are close to the best responses to the empirical distribution of strategies.

[7] The simulation has only 351 observations, as we could not complete one of the simulated sessions within 48 hours.

2.4.3 Bidding Behavior in RAD

In the RAD mechanism, since bidding above the package value exposes the bidder to the risk of a negative profit, we expect bids to be below or equal to value. We now define the bid/value ratio under RAD as follows:

$$\text{Bid/Value Ratio}(x_j) = \frac{\text{Submit Price}(x_j) - \text{Market Price}(x_j)}{\text{Value}(x_j) - \text{Market Price}(x_j)} = \frac{p_i(x_j) - \Pi_j x_j}{v_i(x_j) - \Pi_j x_j} \quad (2.8)$$

Result 8 (Bids in RAD) *The mean Bid/Value ratio is significantly less than 1.*

Support 8 *Among 8651 bids placed by all bidders, only 53 (0.6%) are above their values. The mean Bid/Value ratio is 0.541 and the 95% confidence interval (clustered at the session level) is [0.435, 0.647]. A one-sided Wald test rejects the null hypothesis of the bid/value ratio being one at the 0.03% level.* ∎

We now explore the response of bidding behavior to the eligibility constraint. For a given set of submitted bids from a bidder, there are packages that the bidder does not bid on. We conjecture that those remaining packages must satisfy one or both of the following: i) their temporary profits are negative (i.e., unprofitable) and ii) any additional bid will violate the eligibility constraint (i.e., ineligible). However, we find that a significant number of packages are profitable and eligible, implying that bidders fail to bid on profitable packages.

Result 9 (Eligibility) *Bidders submit significantly fewer bids than their eligibility allows, leaving profitable packages on the table.*

Support 9 *In 5 sessions, there are 19 auctions and 124 rounds. Thus, the total number of bidding sets (number of rounds times the number of bidders) is 1116. In 884 rounds, we observe that bidders left profitable packages on the table. For each of those 884 rounds, we compare the used eligibility and the given eligibility. The mean values are 10.46 and 13.54, respectively. The t-test shows the difference is significant at the 0.01% level.*

Together, Results 8 and 9 indicate that while most bidders place bids below their value, they do not effectively use the eligibility rule.

2.4.4 Aggregate Performance

In this section, we report the aggregate performance of the three allocation mechanisms in terms of efficiency and equity.

Two measures have been used to compute the efficiency for an auction outcome. In the first, we let x_i be the package that is actually allocated to participant i, and x_i^* be the package allocated in the optimal allocation, whereby the sum of bidder values is maximized. A simple efficiency measure takes the ratio of total surplus from the actual and the optimal allocations. In contrast, a normalized efficiency measure uses the difference between the actual (optimal) surplus and the average surplus resulting from 10,000 random allocations,[8] x_i^{rand}. Specifically,

$$\text{Efficiency} = \frac{\sum_i v_i(x_i) - \sum_i v_i(x_i^{rand})}{\sum_i v_i(x_i^*) - \sum_i v_i(x_i^{rand})}. \tag{2.9}$$

As both measures have been used in the literature, we conduct our efficiency analysis using both and find similar results. In what follows, we report the normalized efficiency measure.

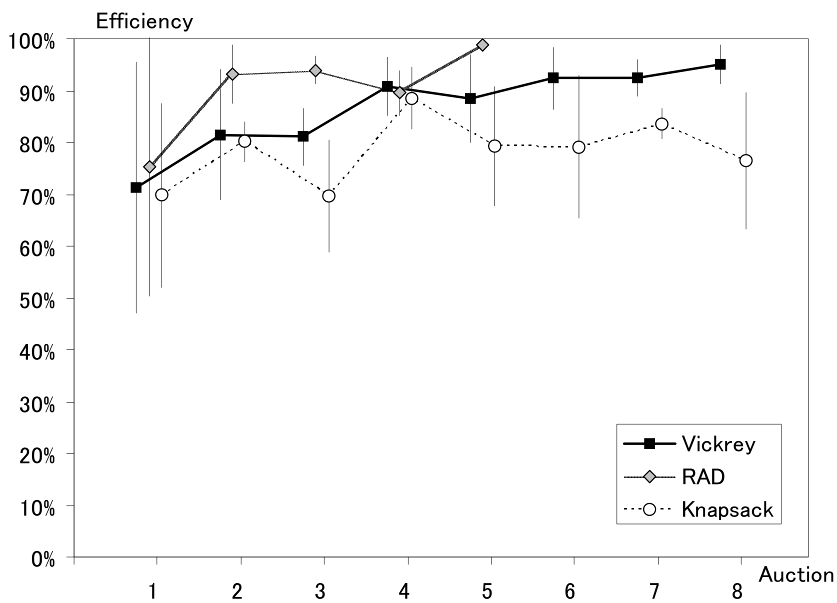


Figure 2.4 Efficiency Comparison Across Mechanisms

Figure 2.4 presents the average time series efficiency and standard deviation (error bars) of the three mechanisms. Consistent with individual learning under the VCG auction, the average efficiency under VCG increases over time and is higher

[8]In implementing the random allocation, we randomly generate 10,000 queues using a uniform distribution of participants in queue positions. For each queue, the first participant is given a package randomly chosen from all the packages. The second participant is given a package randomly chosen from the remaining packages, and so on.

than Knapsack. RAD also generates higher efficiency than Knapsack. Note that RAD has only 3 to 5 auctions in each session, while both VCG and Knapsack have 8 auctions.[9] While RAD generates higher efficiency in the early auctions, VCG catches up by auction the fourth auction.

Result 10 (Efficiency) *VCG and RAD each generate significantly higher efficiency than Knapsack. The efficiency comparison between RAD and VCG is insignificant.*

Support 10 *We compute the session average for the last 2 auctions in each session. For session average efficiency comparisons, we obtain the following results, where the p-value for each one-sided permutation test is written under the inequality sign.*

$$\begin{aligned} &Knapsack \underset{\mathbf{0.03}}{<} VCG \underset{0.38}{<} RAD, \\ &Knapsack \underset{\mathbf{0.03}}{<} RAD. \end{aligned}$$

Taking into account the effects of learning, we next look at the average efficiency of the last two auctions in each session, and find the following result:

$$\begin{aligned} &Knapsack \underset{\mathbf{0.01}}{<} VCG \underset{0.72}{<} RAD, \\ &Knapsack \underset{\mathbf{0.01}}{<} RAD. \end{aligned}$$

■

While the efficiency index measures whether time slots are allocated to the bidders who value them the most, we use the Gini coefficient to measure the distributional equity among participants. Given the set of allocations, $(x_1, x_2, \ldots, x_n)$, the Gini coefficient based on the profits of allocated packages is defined as follows:

$$\text{Gini Coefficient} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \|\pi_i(x_i) - \pi_j(x_j)\|}{2n \sum_{k=1}^{n} \pi_k(x_k)}.$$

When we calculate the Gini coefficient based on profits, we normalize all profits by subtracting the largest negative profit, so all profits are non-negative. Since the Knapsack mechanism does not involve payments, the Gini coefficients based on

[9]To avoid laying the error bars on top of each other, points are shown slightly offset.

value are more comparable. We define the Gini coefficient based on the values of allocated packages as follows:

$$\text{Gini Coefficient} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\|v_i(x_i) - v_j(x_j)\|}{2n\sum_{k=1}^{n} v_k(x_k)}.$$

Note that a higher Gini Coefficient corresponds to greater inequality, with a value of zero in case of perfect equality, and one in the case of perfect inequality.

In the auction literature, it is unusual to use equity as an outcome measure. However, in equipment time scheduling problems, equity is an important measure. One of the policy goals of the National Science Foundation (NSF) is to increase the utilization of its facilities by broadening participation. Therefore, the NSF cares about the equitable use of facilities.

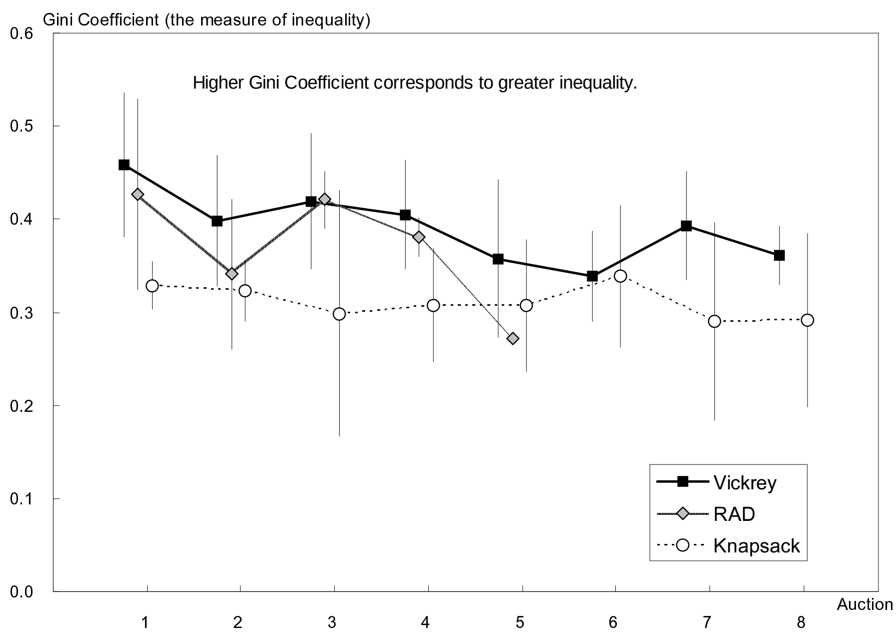


Figure 2.5 Equity Comparison Across Mechanisms

Figure 2.5 presents the average time series Gini coefficients (based on value) and the standard deviations (error bars) for each of the three mechanisms. One striking feature is that Knapsack is roughly the lower envelope of the observations, indicating a more equitable distribution of time slots than the other two mechanisms.

Result 11 (Equity) *The Knapsack mechanism is significantly more equitable than either the VCG or RAD mechanism, based on profit or value. While the equity comparison based* *on the value of VCG and RAD is insignificant, profits are significantly more unequal under RAD.*

Support 11 *Permutation tests of the session averages yield the following results:*

$$\text{(1) Profit:}\quad Knapsack \underset{\mathbf{0.01}}{<} VCG \underset{\mathbf{0.01}}{<} RAD, \quad Knapsack \underset{\mathbf{0.01}}{<} RAD$$

$$\text{(2) Value:}\quad Knapsack \underset{\mathbf{0.01}}{<} VCG \underset{0.55}{<} RAD, \quad Knapsack \underset{\mathbf{0.01}}{<} RAD.$$

■

Therefore, there is an efficiency and equity tradeoff among the three mechanisms. While the two auction mechanisms, VCG and RAD, are significantly more efficient than the ordinal ranking Knapsack mechanism is, the latter is more equitable than either of the auction mechanisms. This is because bidders cannot express the intensity of their preferences under the Knapsack mechanism, which in turn leads to a worse match of needs and allocations (efficiency), but a more equitable allocation.

2.5 Discussion

This paper reports an experimental investigation of three scheduling mechanisms for a relatively complex environment designed to mimic key aspects of actual collaboratory, resource allocation problems such as those of the NEES collaboratory. In our study, we capture variation in both resources (i.e., large vs. small projects) and in time preferences (i.e., early, middle, or late in a 24-month period). Experimental participants are assigned randomly to project types and asked to bid for instrument time using one of three mechanisms – two package auctions (VCG and RAD) and one ordinal ranking (Knapsack).

The results show that Knapsack is more equitable than either VCG or RAD. However, both VCG and RAD are more efficient. Remember the Knapsack mechanism does not allow bidders to express the intensity of their preferences. In the hypothetical collaboratory, this favors small project bidders and makes it easier for them to obtain slots. However, both of the auction mechanisms are better at giving the right people the right slots, albeit at the expense of equity. Therefore, a choice among these mechanisms comes down to a tradeoff between efficiency and equity.

This study has several limitations. First is the question of whether laboratory experiment can simulate real world phenomena. However, there is extensive evidence that experimental participants accurately represent economic behavior within the domain of auctions and other mechanisms. Similarly, while the hypothetical collaboratory cannot model all the nuances of an actual collaboratory, it is possible to operationize relevant aspects as parameters within an experimental design. Perhaps most important, it is not our purpose to use experimental methods to capture the full fidelity of collaboratory use. Rather, the experimental approach gives us a sufficiently realistic arena within which to examine particular aspects of collaboratory use with an eye toward future work that on allocation.

We see four critical next steps in building on the insights in this paper. First, we believe the results will have greater validity when participants are members of authentic scientific and engineering communities. That is, we would like to use practicing scientists and engineers – in the expectation that these participants will identify more strongly with the instrument allocation task and therefore have greater investment in bidding outcomes (i.e., valuations will be more truthful and accurate). Second, we would like to conduct surveys of scientific and engineering communities that depend on scarce resources (e.g., astronomy) to identify better the current mechanisms used to allocate instrument time. In this paper, for instance, we adopted Knapsack as a proxy for typical allocation mechanisms based on evidence that some variant of Knapsack is used to allocate time at the Chandra X-ray observatory. Third, assuming we find other mechanisms in use, we would like to include these as comparisons in future experiments. Finally, our larger ambition is to use the body of experimental results to inform adoption of specific allocation mechanisms within scientific and engineering communities. Ideally we would like to find multiple communities to conduct field experiments contrasting mechanisms. Finally, we would like to observe how real-life participants might use bidding points or a bidding budget. The significant efficiency gain from an ordinal ranking mechanism to an auction mechanism warrants attention to the feasibility of using money or bidding points in real facility or instrument scheduling problems. As NEES sites are used for research outside of NSF private interests, such as oil exploration companies or construction firms, these might be asked to use currency to bid. In this situation, our research points to package auctions, such as VCG or RAD, as an efficient equipment time allocation mechanism.

Chapter 3

Contest for Status: A Laboratory Study

3.1 Introduction

Contests are everywhere. In the class room, we compete for grades with our classmates. At work, we compete with our colleagues for promotion. In online commerce sites such as Amazon.com, product reviewers compete for ranking with each other based on the quality and the quantity of reviews they have written. Donors to non-profit organizations compete against each other for status. We can think of these contests as multi-prize all-pay auctions. They are multi-prize because there are rewards not just for the top-performer, but for people at different points at the ranking ladder. They are all-pay because the effort or money used in achieving a particular ranking is not refunded.

One of the goals of contest designers is to maximize the total group effort produced when people are engaged in contests. The teacher may want to maximize the total effort that students put toward learning a subject matter. The manager may want to maximize the total effort of a sales force. Amazon.com may want to maximize the effort that product reviewers put toward the reviewing of products. A fundraiser may want to maximize the total contribution made by their donors.

What then is the optimal design of contests to achieve the highest total group effort? The primary design dimension we will focus on in this paper is the choice of the number of status/reward categories. For example, can we motivate students to study more by grading the students based on letter grades such as A, B, C, D, F (fewer reward categories) or based on percentages (more rewards categories)? Can we motivate a sales force by rewarding only the top performer with a big bonus (fewer reward categories), or with several smaller prizes (more reward categories)? Would Amazon.com's product reviewers and donors to non-profit organizations be motivated by a complete listing based on their contribution level (more reward categories), or would they be motivated by being grouped into fewer categories?

According to Moldovanu et al. (2007), the answer depends on the distribution of ability factors of the population competing for the prizes. If the distribution of ability factors satisfies the increasing failure rate property (for example, when the ability types are uniformly distributed), then the total expected effort of the group will be higher when we have more status categories. If the distribution of ability factors is sufficiently concave (e.g. when there are more low types than high types), then it is better to have just two status categories with one person in the top category and everyone else in the bottom category.[1] We design and run a series of experiments to test these two predictions.

We now highlight our results. First, the data from the laboratory supports the first prediction but not the second. That is, we found that by having more reward categories by assigning each person their own rank, the subjects put forth more effort in both the IFR and the Concave treatments. Second, on the whole, subjects exert less effort in the laboratory than predicted by the theory, in contrast with the findings in other experiments of single or multi-prize all-pay auctions. Third, when subjects are unable to calculate the equilibrium, they tend to under-estimate their probability of winning and exert more effort than predicated.

We review the relevant theories and experiments in contest design in Section 3.2 before we focus on the model described in Moldovanu et al. (2007) that we use in our experiment. In Section 3.3 we outline the experimental design. In Section 3.4, we report the aggregate results. In Section 3.5, we look at three models of individual decision making to see which model fits the data the best, and we conclude in Section 3.6.

[1]The precise definition of IFR and Concavity are as follows. For a continuous random variable X, we define probability density function to be $f(x)$ and the cumulative distribution function to be $F(x)$. We can define the hazard function as $f(X)/(1-F(X))$. A random variable exhibits IFR if the hazard function is nondecreasing in x from below. A random variable exhibits decreasing failure rate (DFR) if the hazard function is nonincreasing in x from below. DFR implies concavity but not vice versa.

3.2 Literature Review

3.2.1 Theories and Experiments

The model in Moldovanu et al. (2007) follows and refines the models on tournaments by Moldovanu and Sela (2001, 2006). In Moldovanu and Sela (2001), the authors study how to partition a fixed amount of award money to elicit the maximum total group effort. They show that if contestants have linear or concave cost functions, it is optimal to put all the money into one grand prize and reward only the top performer. However, if contestants have convex cost functions, it is optimal to partition the reward money into smaller (nonidentical) prizes and reward them to the top performers. Dubey and Geanakoplos (2004) and Kalra and Shi (2001) derive similar results.

In Moldovanu and Sela (2006), the authors study how to partition the participants into different group sizes. If the cost functions are linear, a single grand static contest yields the highest expected total effort. If the designer wants to maximize the expected highest effort, and if there is a sufficient number of competitors, it is best to split competitors into two groups and to have a final round of contest among the top performer from each group. If the cost functions are convex, the total expected effort is maximized either by splitting contestants into several sub-contests or by awarding prizes to all finalists. Fu and Lu (2009) derive similar results.

In summary, Moldovanu and Sela (2001) look at the optimal partition of prizes, Moldovanu and Sela (2006) study the optimal partition of people, and Moldovanu et al. (2007) examine the optimal partition of the combination of prizes and people. The models in Moldovanu and Sela (2001, 2006) assume everyone has the same type of cost function, where as Moldovanu et al. (2007) provide a more general model by allowing for heterogeneity of costs.

The experimental work closely follows the development of the theoretical literature. Müller and Schotter (2007) test the model in Moldovanu and Sela (2001) and found that the aggregate predictions held true. Under linear cost functions, one grand prize generates higher total expected effort, and under convex (quadratic) cost, two prizes generate higher total expected effort. On the individual level, however, subjects' behavior deviated substantially from the theoretical prediction. Instead of bidding as an increasing function of ability, the bids bifurcate. Low ability workers drop out (exert zero or very low effort) while high ability workers work too hard (exert more effort than predicted by theory). The authors suggest that this behavior can be explained by subjects having a distorted perception of the probability of winning: when the probability is small, the subjects approximate that probability to zero, and when the probability is close to one, the subjects approximate the probability to one. They also found that subjects overbid when the cost function is linear, as consistent with the literature in rent-seeking contests (Davis and Reilly, 1998; Potters et al., 1998), single-unit all-pay auctions (Gneezy and Smorodinsky, 2006; Amann and Leininger, 1998), and multiple-unit all-pay auctions (Barut et al., 2002). However, Müller and Schotter (2007) found that under convex cost, subjects underbid, which is a surprising result.

Sheremeta (2008) tested the model in Moldovanu and Sela (2006). Four lottery contests were implemented: a contest with one grand prize, a contest with two prizes (equal in one case and unequal in the other), and a contest with two subcontests.[2] The aggregate results are consistent with that of the theoretical prediction. He found that the contest with one grand prize generated the highest effort levels and that equal prizes produced lower efforts than unequal prizes. He also found in general, subjects overbid in all four contests and that joint contests generate higher efforts than the equivalent sub-contests. In his experiment, everyone had the same cost.

Compared with the previous experimental studies of contests, we study contests where there are multiple prizes and where the ability factors are drawn from a distribution.

3.2.2 A Theory of Contest by Moldovanu et al. (2007)

We first lay out the key features of the model of contests by Moldovanu et al. (2007).[3] Let i index the N people in the game ($i = 1, 2, ..., N$). Let j index the K status categories ($j = 1, 2, ..., K$), where $j = 1$ is the highest status category and $j = K$ is the lowest status category. Let S_j the number of people in each status category. The number of people in the top status category is S_1, and the number of people in the bottom status category is S_K.

At the beginning of the game, the distribution of ability factors is announced

[2]The difference between the contest with two prizes and a contest with two subcontests is that in the two subcontests case, the participants are split into two separate subgroups and are only competing with those in their subgroup.

[3]In what follows, we sketch out the model to the extent that is necessary to understand the two predictions from the model that we wish to test. For the mathematical details of the complete model, please consult the original paper.

to everyone in the game. An ability factor is then drawn for each person. Each person knows his own ability factor but not the realized values of the other players. Once an individual receives his ability factor a_i, he decides on how much effort e_i he wants to bid. His ranking is based on the effort that he exerts in comparison with the efforts exerted by others in his group. The person with the highest effort is ranked the highest, the person with the second highest effort is ranked the second, and so forth.

The value for achieving status category k is:

$$V(k) = \underbrace{\sum_{j>k} S_j}_{\text{\# of people below i's status category}} - \underbrace{\sum_{j<k} S_j}_{\text{\# of people above i's status category}} \tag{3.1}$$

where e_i is the effort that an individual chooses, and a_i is the ability factor of that individual drawn from a probability distribution function.

The value depends linearly on the number of people below and above one's own status category k, but not the number of people in one's own status category do not affect it. The designer of the status game has two design dimensions that he can control: the number of status categories (K) and the number of positions within each status category($S_{j \in K}$). The number of status categories can be as few as one (everyone in the same status category) or as many as there are people (each person occupies his or her own status category).

The utility function is then:

$$U_i = \underbrace{\sum_{j \in K} \text{Prob}(S_j) * V(S_j)}_{\text{Expected Benefit}} - \underbrace{\frac{e(a_i)}{a_i}}_{\text{Cost}} \tag{3.2}$$

We can then proceed to find the risk neutral symmetric equilibrium where the bid function is monotonically increasing in ability factor and incentive-compatible (people will reveal their true type). In solving the utility maximization problem using the first and second order condition, we need to solve a first order linear differential equation. For the detailed derivation, please consult Moldovanu et al. (2007).

3.3 Experimental Design

We design the experiment to test the following two predictions:

Prediction 1: When the distribution of ability factor satisfies the increasing failure rate[4] (IFR) property, the contest designer would yield the maximum expected total effort by employing complete ranking, where everyone occupies their own status category. The uniform distribution is an example of a distribution that satisfies the IFR property.

Prediction 2: When the distribution of ability factor is sufficiently concave,[5] the contest designer would yield the maximum expected total effort by employing two status categories with one person in the top category and everyone else in the bottom category.

3.3.1 Economics Environment

We employed a 2x2 factorial design to test the two hypotheses. In one dimension, there are two ranking schemes: two status ranking with one person in the top category (TwoStatus), and complete ranking (Complete). In the other dimension, there are two different types of distribution functions: increasing failure rate (IFR) and sufficiently concave (Concave). The four treatments are designated as: IFR-TwoStatus, IFR-Complete, Concave-TwoStatus, and Concave-Complete. We conduct three independent sessions per treatment.

Subjects compete in groups of $N = 4$. In the TwoStatus treatment, $K = 2$ and in the Complete treatment, $K = 4$.[6] The decision problem for the TwoStatus and the Complete ranking cases are summarized in Table 3. The reason that the total prize money is the same under the two reward structures is not accidental. In the Moldovanu et al. (2007) paper, the structure of an individual's utility function is a function of the number of people above him in status categories and below him in status categories. This way of constructing the utility function will always lead to the same total status, or prize money, available in a contest.[7]

[4] See footnote 1 for the mathematical definition.

[5] See footnote 1 for the mathematical definition.

[6] Note that for the TwoStatus ranking case, we also have the freedom to decide how many people will be in the top category and how many people will be in the bottom category. For simplicity, when we refer to the two status category case, we will always assume the case that there is one person in the top category, and everyone else in the bottom category.

[7] We have added a constant of 5 to all the values. Adding a constant to the value function does not change the results of the model.

Ranking Based on Effort	Two Status Categories	Complete Ranking
1	8 francs	8 francs
2	4 francs	6 francs
3	4 francs	4 francs
4	4 francs	2 francs

Table 3.1 Prize Structure Table

Many distribution functions can satisfy the IFR or the Concave properties. We chose to use two different parameterizations of the Weibull distribution for the sake of consistency. The Weibull distribution is a two parameter distribution with a shape parameter α and a scale parameter β. For the IFR distribution, the α was set to 1.5. For the Concave distribution, the α was set to 0.5. The scale parameter β was set to the value of 10 for both the IFR and the Concave Weibull distributions. Table 3 summarizes the functional forms of the two distributions. And we plot the functions in Figures 3.1 and 3.2.

	IFR Property	Concave Property
Distribution Function	Weibull[$\alpha = 1.5, \beta = 10$]	Weibull[$\alpha = 0.5, \beta = 10$]
CDF	$1 - e^{-0.03x^{1.5}}$	$1 - e^{-0.32x^{0.5}}$
PDF	$0.05e^{-0.03x^{1.5}}x^{0.5}$	$\frac{0.16e^{-0.32x^{0.5}}}{x^{0.5}}$

Table 3.2 Summary of Distribution Functions of the Ability Factor

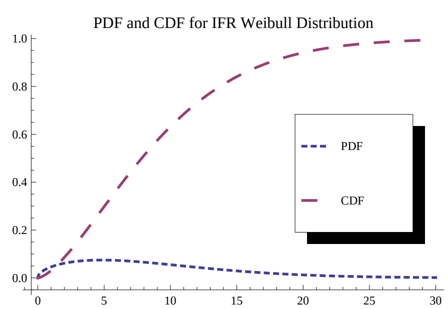


Figure 3.1 The probability distribution of the IFR distribution.

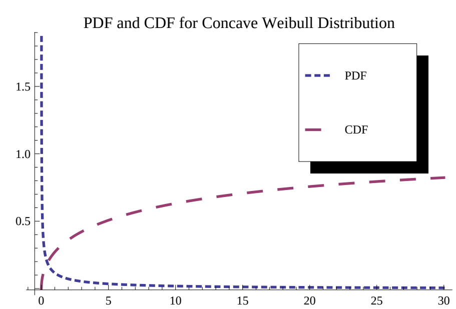


Figure 3.2 The probability distribution of the Concave distribution.

If the simplicity of the functional form of the distribution is more important than the consistency of the functional form, simpler distributional forms satisfy the IFR property or the Concavity property. For example, uniform distribution satisfies the IFR property, and the power distribution satisfies the Concave property. This is a trade off in experimental design. We chose to have consistency of form and to eschew the simpler distributions (uniform or power) so that the results are not driven by the familiarity subjects may have with the more common distributions.

3.3.2 Experimental Procedures

The sessions were conducted from May of 2009 to July of 2009. We ran 3 independent sessions per treatment, leading to a total of 12 completed sessions. In our experiment, each session required exactly 12 participants. No participant was used in more than one session, yielding 144 subjects across all sessions. The participants were recruited from the University of Michigan, including undergraduate and graduate students as well as staff members. Table 3 summarizes the features of the experimental sessions, including the number of sessions and the number of subjects.

	# sessions	# subjects	# periods	group size	exchange rate
IFR-TwoStatus	3	36	20	4	3.33 francs/$
IFR-Complete	3	36	20	4	3.33 francs/$
Concave-TwoStatus	3	36	20	4	3.33 francs/$
Concave-Complete	3	36	20	4	3.33 francs/$

Table 3.3 Summary of Treatments

Subjects were given instructions at the beginning of the session. The subjects were given 10 minutes to review the instructions on their own, and the experimenter read the instructions aloud and answered subjects' questions. After the instructions, participants were given 5 minutes to answer five review questions to test their understanding of the experiment. The experimenter reviewed the answers with the participants before proceeding to the experiment. There were no practice periods and each session ran for 20 periods. In each period, subjects had 2 minutes to make a decision. To help subjects make more informed decisions, we provided a calculator for subjects to calculate their hypothetical payoffs for different effort levels. At the end of each period, they had twenty seconds to look at their result. In each period, subjects could also look at their results from the previous periods. After 20 periods, the subjects completed a short questionnaire and a survey, and paid the amount they made. On average each session lasted about 2 hours.

The subjects were only paid for the decisions that they made in the 20 periods. The subjects in the experiment received a $5 show up fee. If a subject's cumulative earning, including the show up fee, becomes negative, the subject is declared bankrupt and we end the session.[8] The exchange rate was 3.33 francs per dollar (or 0.3 dollar per franc). The average total earnings were \$29.21, including the \$5 show up fee, and the standard deviation was \$4.06. Further data are available from the author upon request. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). Experimental instructions are included in Appendix B.

3.4 Aggregate Results

We first look at the aggregate results. We do this by testing the two comparative statics predictions regarding the total group effort from the Moldovanu et al. (2007) paper. We show that the experimental data supports one comparative statics prediction but not the other. Next, we compare the observed profit versus the equilibrium profit. We find that people made less money than the equilibrium prediction. We then look at how well people were able to sort themselves into the correct rank according to their ability factor, and found that people were able to do so 71% of the time across all treatments.

Before we proceed with the data analysis, we want to make sure that the realized draw of ability factors follow the same distribution in the two IFR treatments and in the two Concave treatments. We did this by running the Kolmogorov-Smirnov (K-S) test comparing all the realized ability factors draw from the three sessions of the IFR-TwoStatus treatment with the draw from the three sessions of the IFR-Concave treatment. The null hypothesis is that the two sets of realized values are from the same distribution, and the alternative hypothesis is that the two sets of realized values are from two different distributions. The test yields a p-value of 0.560, which does not reject the null at the 1%, 5%, and 10% levels. We perform a similar test for the two Concave treatments and that yields a p-value of 0.082, which does not reject the null at the 1% or 5% levels. It is interesting to note that the p-value is substantially smaller for the Concave treatments than for the IFR treatments. This suggests that there is more variance in the Concave distribution.

$$\text{Effort Ratio} = \frac{\text{Avg. Observed Effort Per Person}}{\text{Avg. Equilibrium Effort Per Person}}$$

[8]This happened once in the Concave-TwoStatus experiment in period 9. We have excluded this session from the data analysis and this is not part of the three independent sessions used for the Concave-TwoStatus treatment.

Treatment	Avg. Effort		Effort	Avg. Profit		Profit
	Obs. (std.)	Equ. (std.)	Ratio (std.)	Obs. (std.)	Equ. (std.)	Ratio
IFR-TwoStatus	9.50 (14.85)	9.92 (13.82)	0.96	4.19 (1.57)	4.34 (1.55)	0.97
IFR-Complete	16.17 (14.48)	17.16 (14.82)	0.94	3.35 (1.85)	3.50 (1.89)	0.96
Concave-TwoStatus	13.57 (31.21)	16.64 (31.09)	0.82	4.52 (1.63)	4.59 (1.60)	0.98
Concave-Complete	18.30 (53.64)	14.27 (22.77)	1.28	4.00 (2.24)	4.34 (2.07)	0.92

Table 3.4 Aggregate Statistics

$$\text{Profit Ratio} = \frac{\text{Avg. Observed Profit Per Person}}{\text{Avg. Equilibrium Profit Per Person}}$$

In Table 3, we summarize the key statistics for the four treatments. We report the average effort and profit per person, observed and in equilibrium. We also compute the corresponding simple average by dividing the observed effort by the equilibrium effort and by dividing the observed profit by the equilibrium profit. It is of interest to note that the exerted effort level in the Concave-Complete case has double the variance as the predicted effort level, where as in the other three treatments, the standard deviation between the observed and the predicted effort level are about the same.

We now turn to testing the two main comparative statics predictions from the theory. We state the alternative hypotheses below. The general null hypothesis is that there is no difference between treatments.

Hypothesis 1 *The effort exerted is higher in the IFR-Complete treatment than in the IFR-TwoStatus treatment.*

Result 1 *Under the IFR distribution of ability factors, total group effort is significantly higher in the Complete ranking treatment than in the Two-Status ranking treatment, as predicted by the theory.*

Support 1 *We first calculate the average observed effort per person*[9] *for each of the 6 sessions under the IFR distribution. We ran a one-sided permutation test on the 3 sessions of the IFR-Complete treatment and 3 sessions of the IFR-TwoStatus treatments.*

$$\textit{IFR-TwoStatus} < \textit{IFR-Complete}(p = 0.05).$$

Thus we reject the null hypothesis in favor of the alternative hypothesis that the IFR-Complete treatments yielded more group effort than the IFR-TwoStatus treatments at the 5% level.

[9]The result also holds when we calculate the average observed effort per group.

Hypothesis 2 *The effort exerted is higher under the Concave-TwoStatus treatment than under the Concave-Complete treatment.*

Result 2 *Under the Concave distribution of ability factors, there is no statistically significant difference between the total effort under Complete ranking and the TwoStatus ranking treatments.*

Support 2 *We first calculate the average observed effort per person*[10] *for each of the 6 sessions under the IFR distribution. We ran a one-sided permutation test on the 3 sessions of the Concave-Complete treatment and 3 sessions of the Concave-TwoStatus treatments.*

$$\textit{Concave-TwoStatus} > \textit{Concave-Complete}(p = 1.00).$$

Therefore we cannot reject the null hypothesis at the conventional significance level.

Given the result, we were interested to know if Concave-Complete yield more profit than Concave-TwoStatus. We found that is the case at p=0.05.

$$\textit{Concave-TwoStatus} < \textit{Concave-Complete}(p = 0.05).$$

The result from the permutation tests show that the theory does not consistently predict subjects' choice of effort in the laboratory. In particular, having complete ranking always lead to higher group effort regardless of the distribution of ability factors. Next, we investigate the relationship between the observed effort and profit with respect to the predicted equilibrium effort and profit.

In experiments on tournaments and all-pay auctions, a consistent finding has been that subjects tend to overbid (Amann and Leininger, 1998; Davis and Reilly, 1998; Potters et al., 1998; Barut et al., 2002; Gneezy and Smorodinsky, 2006; Sheremeta, 2008). The only exception that we are aware of is when subjects face a convex cost function, they tend to underbid (Müller and Schotter, 2007). However, in our experiments, we find evidence that shows subjects underbid in our expeirments. We discuss this in Result 3.

Hypothesis 3 *Null : In experiments, subjects overbid relative to the equilibrium prediction.*

Result 3 *Subjects underbid in three of the four treatments when we look at the average bid by running an OLS regression.*

[10]The result also holds when we calculate the average observed effort per group.

Support 3 *The regression model is specified in Equation (3.3). We run an OLS regression with no constant term, clustering the errors at the session level. Table 3 reports the results. We report the estimated coefficient and the p value testing null hypothesis that the coefficient is equal to 1. If the coefficient equals to 1, it means subjects bid exactly according to theory in that treatment.*

We see that under the OLS regression, the subjects in IFR-TwoStatus, IFR-Complete, and Concave-TwoStatus have coefficients less than one and the p-values less than 0.10.

$$\begin{aligned} \text{ObservedEffort}_{it} = \; & \beta_1 * \text{EqulibriumEffort}(a_{it}) + \\ & \beta_2 * \text{EqulibriumEffort}(a_{it}) * \text{IFRComplete} + \\ & \beta_3 * \text{EqulibriumEffort}(a_{it}) * \text{ConcaveTwoStatus} + \\ & \beta_4 * \text{EqulibriumEffort}(a_{it}) * \text{ConcaveComplete} + \\ & \varepsilon_{it}, \end{aligned} \tag{3.3}$$

$$\begin{aligned} \text{ObservedProfit}_{it} = \; & \beta_1 * \text{EqulibriumProfit}(a_{gt}) + \\ & \beta_2 * \text{EqulibriumProfit}(a_{it}) * \text{IFRComplete} + \\ & \beta_3 * \text{EqulibriumProfit}(a_{it}) * \text{ConcaveTwoStatus} + \\ & \beta_4 * \text{EqulibriumProfit}(a_{it}) * \text{ConcaveComplete} + \\ & \varepsilon_{it}, \end{aligned} \tag{3.4}$$

where

$$\text{IFRComplete} = \begin{cases} 1 & \text{if IFR-Complete treatment,} \\ 0 & \text{other treatments} \end{cases}$$

$$\text{ConcaveTwoStatus} = \begin{cases} 1 & \text{if Concave-TwoStatus treatment,} \\ 0 & \text{other treatments} \end{cases}$$

$$\text{ConcaveComplete} = \begin{cases} 1 & \text{if Concave-Complete treatment,} \\ 0 & \text{other treatments} \end{cases}$$

Hypothesis 4 *Null: Observed Profit = Equilibrium Profit for each of the four treatments.*

The regression is specified in Equation 3.4.

Result 4 *Subjects earned less money than the equilibrium prediction in all four treatments.*

Support 4 *For each treatment, we regress the observed individual profit against the equilibrium individual profit with no constant term and cluster the errors at the session level.*[11]

	Bidding Coefficients Equation 3.3	Profit Coefficients Equation 3.4
IFR-TwoStatus (β_1)	0.77 (p=0.09)	0.93 (p=0.01)
IFR-Complete ($\beta_1 + \beta_2$)	0.84 (p=0.04)	0.92 (p=0.03)
Concave-TwoStatus ($\beta_1 + \beta_3$)	0.76 (p<0.01)	0.97 (p=0.06)
Concave-Complete ($\beta_1 + \beta_4$)	1.40 (p<0.01)	0.92 (p<0.01)

Table 3.5 The null hypothesis tested is if the coefficient equals to 1.

Next, we examine how well the subjects sorted themselves according to their ability factors. In the development of the model, we solve for the bid function that is monotonically increasing in ability factors. This means that the efficient allocation would be to assign the person with the highest ability factor to the highest rank, the person with the second highest ability factor to the second rank, and so forth. In the following analysis, we look at the number of people in a group that was correctly assigned to the correct rank. If all four people in a group achieved the rank predicted by their ability factor, then that group would be counted under column with the heading "4" in Table 3. Note that it is not logically possible to assign three people to the correct rank.

	% of Groups With the Following # of Correctly Assigned Rank			
Treatments	0	1	2	4
IFR-TwoStatus	16%	25%	44%	15%
IFR-Complete	10%	17%	44%	29%
Concave-TwoStatus	20%	26%	31%	23%
Concave-Complete	7%	11%	41%	41%

Table 3.6 Sorting Results

In 27% of the groups, all four subjects achieved the ranking predicted by their ability factor.

[11]The reason we run the regression with no constant is because we want to impose the restriction that if the theory predicts that subjects will not make a profit, than subjects in the experiment will not make money.

The results are summarized in Table 3. Note that if people were to sort themselves correctly, we would achieve 100% sorting. If subjects were randomly assigned a rank, then about 4% ($\frac{1}{4!} = \frac{1}{24}$) of the time, we would achieve perfect sorting. [12]

3.5 Individual Level Behavior

The main findings from the aggregate result section are that the data from the experiment supports the comparative static prediction in the case of IFR distribution of ability factors (Result 1) but not in the case of Concave distribution (Result 2). Now, we are interested in identifying the best individual decision making model amongst a set of four models that fits the data by looking at the model with the lowest mean-squared deviation (MSD). But before we look at the different individual decision model, it is important to check that the realized ability factors do in fact come from the same distributions. That is, the ability factors from the two IFR treatments are statistically the same, and the ability factors from the two Concave treatments are statistically the same.

The individual decision model used in Moldovanu et al. (2007) makes the following three assumptions:

Assumption 1 *Individuals are risk-neutral.*

Assumption 2 *Individuals weight probabilities linearly.*

The models we will look at can be divided into two categories: equilibrium based models and heuristic models. The equilibrium based models relax Assumption 1 or Assumption 2 but maintain that players are solving for the Bayes-Nash equilibrium. The heuristic model relaxes the assumption that subjects are able to solve for the Bayes-Nash equilibrium.

3.5.1 Model Descriptions

In this section we look at three models. The first two models use the Bayes-Nash solution concept but relax the risk-neutral assumption and the linear weighting of probability assumption. The third model is a heuristic model where we assume that subjects' bids are simply a linear function of their own ability factor.

[12] There is no column for three correctly assigned rank because that is not logically possible.

Model 1 *Relax Risk Neutral Assumption*

The first model relaxes the risk-neutrality assumption by using the power form of the CRRA utility function. Andreoni and Harbaugh (2010) shows that this functional form of incorporating risk performs well in predicting actual behavior in the lab. In addition, we also included a wealth term (w) to capture the wealth effect.

$$U = X^{1-\delta} \tag{3.5}$$

$$\begin{cases} -1 < \delta < 0 & \text{if risk-loving,} \\ \delta = 0 & \text{if risk-neutral,} \\ 0 < \delta < 1 & \text{if risk-averse.} \end{cases}$$

The utility function from (3.2) becomes:

$$U_i = \sum_{j \in K} \underbrace{\text{Prob}(S_j) * [V(S_j) - \left(\frac{e(a_i)}{a_i}\right) + w]^{1-\delta}}_{\text{Expected Payoff in Status Category j}} \tag{3.6}$$

Once we incorporate the risk parameter, the maximization problem involves solving for a nonlinear differential equation. In general such problem does not have closed-form analytical solution, but we can solve for the solution numerically. [13] The wealth effect is added to test how sensitive the model specification is to the base line wealth level from which subjects compute their earnings from the experiment to be.

Model 2 *Relax Linear Probability Weighting Assumption*

This model incorporates nonlinear weighting of probability by subjects. Nonlinear probability weighting is a feature of Prospect Theory (Kahneman and Tversky, 1979). We examine the effect of this on the equilibrium prediction. Abdellaoui (2000) summarizes different functional forms for modeling nonlinear probability weighting. We use the formulation of nonlinear probability weighting in Tversky and Kahneman (1992) because it has the simplest form, which reduces the computational time in obtaining the numeric solution. We assume that bidders are risk-neutral so that it becomes a one parameter model in γ:

$$\text{f}(p) = \frac{p^\gamma}{[p^\gamma + (1-p)^\gamma]^{\frac{1}{\gamma}}} \tag{3.7}$$

[13] The numerical solutions were obtained using Mathematica.

$$0 \leq p \leq 1, 0 < \gamma \leq 1$$

The utility function from (3.2) becomes:

$$U_i = \sum_{j \in K} \underbrace{\mathrm{f}(\mathrm{Prob}(S_j)) * [V(S_j) - \left(\frac{e(a_i)}{a_i}\right)]}_{\text{Expected Payoff in Status Category j}} \tag{3.8}$$

The difference between the original maximization problem Equation (3.2) and the maximization problem after we incorporate nonlinear probability weighting Equation (3.8) is simply that the probability is weighted. And similar to the effect of incorporating a risk parameter, the maximization problem with nonlinear probability weighting becomes a nonlinear differential equation. This type of problem generally does not have a closed-form analytical solution, but we can solve for the solution numerically.

Model 3 *Linear In Ability Factor (LIAF) Heuristic*

In the fixed-cost bidder model, the subjects use a decision rule that is linear in ability factor. This corresponds to a decision rule where a subject first decides how much cost he is willing to bear. He then chooses the effort level by multiplying the ability factor with the cost factor. This heuristic does not take into account the probability distribution of the ability factor, nor does it take into account the strategic aspects of having different reward structures.

This heuristic corresponds to the regression stated in Equation 3.9. We examined the effects of ability factors on the observed effort level. The parameter of interest is β_{LIAF}, the average fixed-cost. We did not include a constant term in the specification of the regression model. In doing so, we are imposing a rationality constraint. That is, a subject would exert 0 effort if he receives an ability factor of 0. If a subject were to exert a positive amount of effort when the ability factor is zero, the cost he incurs would be infinite.

$$\mathrm{ObservedEffort}_{it} = \beta_{LIAF} * a_{it} + \varepsilon_{it} \tag{3.9}$$

3.5.2 Model Selection

Of the three models we have described, we will select the model that best fit the data using the mean-squared error (MSE) criteria. The process of model selection is done in two stages.

Recall that there are three independent sessions are conducted for each of the four treatments. In the first stage, we take the data from two of the three sessions for each treatment and do in-sample estimation for each of the three models. In the first stage, we find the parameter that would minimize the MSE. In the two equilibrium models (Models 1 and 2), we perform a numeric grid search over the parameter space in intervals of 0.1. [14] For the Risk Model (Model 1), we search over the space $-1 < \delta < 1$. For the Nonlinear Probability Model, we search over the space $0 \leq \gamma \leq 1$. For the heuristic model 3, we run the regression specified for each model would minimizes the MSE.

In the second stage, we use the estimated parameter from the first stage to do out-of-sample prediction on the third session in each treatment, and calculate the MSE. The model with the lowest MSE summed across the four treatments is the best model. This is similar to the approach used in model selection in Chen and Khoroshilov (2003) and Camerer et al. (2004).

There are two ways we can do in-sample estimation. One approach is to assume that there is a representative agent for each of the four treatments. In this case, we would estimate the parameters separately for each of the four treatments, using two of the three sessions to do the in-sample estimation. We then use the estimated parameter to calculate the MSE for the third session in each treatment for the out-of-sample forecast. In estimating the parameters for the heuristic model, we are estimating the parameter for each treatment separately. For example, the specification of Equation 3.9 means we are estimating the coefficient of interest (β_{LIAF}) separately for each treatment. We could have estimated the coefficients for the treatments simultaneously for all treatments similar to the approach specified in Equation 3.3 using treatment level dummy variables. The reason we decided to estimate the coefficient one treatment at a time is to make the estimation process as similar as possible across all three models. In estimating the parameters for the equilibrium models, we were only able to estimate the parameter of interest treatment by treatment, so we decided to do the same for the heuristic models as well.

[14] We do a numeric grid search because we cannot solve the problem algebraically.

Another approach is to assume that there is a representative agent for all four treatments, thus yielding only one parameter estimate for all four treatments under a model. In this approach, we use the data from the same eight sessions that were used for the in-sample estimation described above, but now we estimate only one parameter for each of the five models. We then use this one parameter to calculate the MSE for the remaining four sessions (one from each treatment) for the out-of-sample forecast.

We have reported the in-sample parameter estimate and the corresponding out-of-sample MSE for the representative agent for all treatment case in Table 3 and for the representative agent for all treatments case in Table 3. We estimated the risk model with two different initial wealth levels of 0 and 400 to see if there is a wealth effect.

In the representative agent per treatment case, the rankings are: risk model with wealth level of 400; risk model with wealth level of 0; nonlinear probability model; and the ad-hoc fixed-cost model. In the representative agent for all treatments case, the rankings are: risk model with wealth level of 0, nonlinear probability weighting model, risk model with wealth level of 400, and the ad-hoc fixed-cost model. The fact that the equilibrium based models fit the data better than the heuristic models suggests that on the whole, players are sophisticated in the sense that they act as if they were able to approximate the Bayes-Nash equilibrium. This is of interest especially because the game in question is complex. Note that the MSE for the Nonlinear-Prob model is higher when there is only one representative agent per treatment compared with a representative agent for all treatments. This may seem surprising at first, since we would expect the MSE to be higher when there is more constraint as is the case in one representative agent for all treatments. However, this is not necessarily true because the MSE was calculated out-of-sample. Wealth level does not substantially change the results.

Treatments	Equilibrium Models			Ad-Hoc Model
	Risk MSE ($-1 < \delta < 1$)		Nonlinear-Prob	Fixed-Cost
	w=0	w = 400	MSE ($0 \leq \gamma \leq 1$)	MSE (β_{LIAF})
IFR-TwoStatus	168.42 ($\delta = 0.99$)	154.96 ($\delta = 0.99$)	168.68 ($\gamma = 0.7$)	204.17 ($\beta_{LIAF} = 1.02$)
IFR-Complete	89.37 ($\delta = 0.5$)	94.57 ($\delta = -0.99$)	91.66 ($\gamma = 0.7$)	94.26 ($\beta_{LIAF} = 1.72$)
Concave-TwoStatus	221.57 ($\delta = -0.99$)	262.65 ($\delta = -0.99$)	261.93 ($\gamma = 0.9$)	241.35 ($\beta_{LIAF} = 0.59$)
Concave-Complete	1828.10 ($\delta = 0.99$)	15.18 ($\delta = 0.99$)	1911.95 ($\gamma = 1.0$)	1930.79 ($\beta_{LIAF} = 0.83$)
Total MSE	2307.46	527.36	2434.22	2470.57
Model Ranking	2	1	3	4

Table 3.7 Model Selection - Representative Agent Per Treatment

Treatments	Equilibrium Models			Ad-Hoc Model
	Risk MSE ($-1 < \delta < 1$)		Nonlinear-Prob	Fixed-Cost
	w=0	w = 400	($0 \leq \gamma \leq 1$)	MSE (β_{LIAF})
IFR-TwoStatus	168.42 ($\delta = 0.99$)	154.70 ($\delta = -0.99$)	154.83 ($\gamma = 1.0$)	252.72 ($\beta_{LIAF} = 0.74$)
IFR-Complete	94.32	94.57	94.67	258.22
Concave-TwoStatus	316.15	262.65	263.56	260.60
Concave-Complete	1828.10	1913.31	1911.95	1872.22
Total MSE	2406.99	2425.23	2425.01	2643.76
Model Ranking	1	3	2	4

Table 3.8 Model Selection - Representative Agent for All Treatments

3.6 Conclusion

Is it better to have fewer or more status categories in designing a ranking scheme that would motivate people to contribute? Moldovanu et al. (2007) propose a model that ties the choice of the number of status categories to the distribution of ability factors. The theory concludes that the choice of the number of status categories corresponds to the distribution of ability factors. We investigate two predictions from their model. First, when the distribution of ability factors exhibits increasing failure rate (IFR), it is optimal to have as many status categories as there are participants. Second, when the distribution of ability factors is sufficiently concave, it is optimal to have only two status categories with only one person in the top status category. The data from our experiment supports the first prediction but does not support the second.

To understand better the aggregate level results, we look at three models of individual decision making to see which model fits the data the best. The top individual models that performed the best based on MSE criteria are when bidders are risk-averse. This implies that the Bayes-Nash equilibrium with risk best fit the laboratory data.

The reason the data does not support the second prediction is subjects exerted a higher level in the Concave-Complete case compared with the three other cases. The exerted effort level in the Concave-Complete case has double the variance as the predicted effort level. These two facts combine suggests the possibility that subjects have a harder time of finding and playing the equilibrium strategy, however, we do not know why the Concave-Complete case is harder than the other three cases. When the subjects are uncertain of the actual probabilities, they tend to under-estimate their chances of winning a higher reward and thus increase their effort level.

The laboratory data show that the model does a good job of predicting the behavior of human subjects in the lab overall. However, if subjects are unable to find the equilibrium, they tend to be under-confident and exert more effort than the model predicted. The implication of this result for contest designers is that they need to assess how well the participants will be able to calculate their probability of winning.

Chapter 4

Using Leaderboards to Motivate Contribution to Public Goods: A Field Study on Amazon

4.1 Introduction

Amazon is an online merchant that sells everything from books to consumer electronics to jewelry. In 2009, Amazon sold $24.5 billion worth of goods.[1] The success of Amazon depends in part on the product reviews its community members write. Studies have shown that product reviews increase sales. When the same book is sold on two different website (Amazon and Barnesandnoble), the site that has more reviews for the book sold more of it (Chevalier and Mayzlin, 2006). Helpful reviews lead to more sales, especially for less popular books (Chen et al., 2006). For feature-based goods (such as electronics), reviews that contain objective information with a few subjective sentences receive more helpful votes, where as for experience-based goods (such as movies and books), reviews that are highly personalized receive more helpful votes (Ghose and Ipeirotis, 2007). The helpfulness of a review is determined by the larger context of the other reviews that have been written for the same product (Danescu-Niculescu-Mizil et al., 2009).

Product reviews are a type of public good. A public good is a good that is non-rival and non-excludable. Product reviews are non-rival because the value of the reviews to a person is not diminished when another person views the reviews. Product reviews are also non-excludable because Amazon has chosen to make the reviews publicly viewable regardless of whether a person has contributed

[1] 2009 SEC Filing

reviews or not. Free-riding is a standard problem with voluntary contribution of public goods (Samuelson, 1954). The focus of this chapter is not on characterizing the levels of free-riding, since we do not observe the utility and cost functions of the reviewers. Instead, the focus is on evaluating the effectiveness of Amazon's effort in encouraging review contribution through the use of leaderboards. Studies have shown that the display of social information, that is, information about the contribution behavior of other community members, can lead to an increase in the contribution to the public good. People contribute more money to a charitable cause if they learn that others in the community have also contributed (Frey and Meier, 2004). People conform to the median level of movie ratings when they are given information about the median contribution level of other reviewers. This shows that the median information can encourage those below the median contribution level to contribute more reviews (Chen et al., 2010).

Before October 23, 2008, Amazon recognized the contributions of its product reviewers through a publicly viewable leaderboard that ranked the reviewers based on the number of reviews that have been written and the number of helpful votes generated for the reviewers. This is the Classic leaderboard. On October 23, 2008, Amazon added a new ranking system, called the New leaderboard. In announcing the New ranking system, Amazon explained the three main differences between the Classic and the New leaderboards.[2] First, the New leaderboard would weight recent reviews more heavily. Second, the New leaderboard will weight review helpfulness more than the Classic leaderboard. This means writing unhelpful reviews will not move a reviewer up the ranking ladder. Third, the New leaderboard will discourage "ballot box stuffing". Ballot box stuffing describes the phenomenon where the same set of people rate the reviews written by a reviewer favorably. To get a sense of the differences between the two ranking systems, we can look at the top reviewer on the two leaderboards on August 4, 2009. Harriet Klausner is the number one reviewer under the Classic ranking system, with $19,657$ reviews and a helpfulness percentage of 71%. However, under the new system, her ranking is 607. In contrast, Mark, the number one reviewer under the New ranking system, has reviewed only 535 products, but has a helpfulness percentage of 95%. Mark's Classic rank is 303.

The main research question we address in this chapter is do reviewers respond

[2] `http://www.amazon.com/tag/top%20reviewers/forum?_encoding=UTF8&cdForum=Fx2Z5LRXMSUDQH2&cdThread=Tx1R05YOJUVFC7L`. Accessed on October 25, 2010.

to the ranking systems? We found that reviewers do respond to the New ranking system in that reviewers respond to the review-writing behavior of their neighbors in ranking under New but not under Classic. This is because under the New ranking system, reviewers are more likely to improve or lose their current ranking because the New ranking system weights recent activities more than the Classic ranking system. The rest of the chapter is organized as follows. We first provide a description of how we obtained the data and report the summary statistics. We then present a description of the two ranking systems on Amazon and introduce some notations. We discuss the results before concluding.

4.2 Data

We have collected data related to reviewers ranked under the Classic and the New leaderboards from February 19, 2010 to June 30, 2010 (131 days). Each day, we downloaded the relevant html pages from Amazon's website at 4am EST. It took about 45 minutes to download the complete set of data, which consists of 2000 html pages: 1000 html pages for the reviewers ranked in the Classic leaderboard and 1000 html pages for those ranked in the New leaderboard. Each html page contains the information for 10 reviewers, yielding 10,000 total reviewers per ranking per day. For each reviewer, we have the ranking of that reviewer for that day, as well as the number of lifetime reviews and the number of lifetime helpful votes up to that date. The same reviewer may appear on multiple days. The same reviewer may also be ranked under both the Classic and the New leaderboard on the same day. While we do not know how often Amazon update the rankings, we do know that rankings changed from day to day, so rankings are calculated at least once every 24 hours.

Given that there are two ranking systems listed per day and 10,000 reviewers per ranking system over 131 days, we should have 2.62 million records. However, we only have 2.553 million records. The missing 67,000 records are due to the fact that some of the rankings were not available when we downloaded the data. There are four dates in which the number of missing reviews are especially high (2010-3-7, 2010-3-24, 2010-3-31, and 2010-5-27). Most of the other days are missing between 10 to 30 records per ranking system. Table D.1 in Appendix D records the missing reviews by date. Table D.2 summarizes the range in ranks in which we have missing data over the remaining 127 days.

Even though we have missing data, the actual problem is not as severe as it first appears. This is because most of the results presented in this paper are based on the information on the first and the last day of the observation period. On the first day (2010-2-19), we are missing only 10 records from the Classic leaderboard. On the last day (2010-6-30), we are missing 20 records from the Classic leaderboard and 20 records from the New leaderboard.

We also present results that make use of the data from the full range of the dates in which we have collected the data. However, since the missing data resulted from the Amazon server being unavailable, there is no reason to believe that there is systematic correlation between the data that are missing across days. This can be seen in Table D.2.

Since most of the results are based on data from the first and last day of the observational period, it is useful to know the number of people that were ranked on those 2 days and the summary statistics of their rank, the number of lifetime reviews, and the number of lifetime helpful votes. There are three categories in which a reviewer can belong to in any given day: those that were ranked only under Classic leaderboard, those that were ranked only under New leaderboard, and those that were ranked under both. The first two columns of Table 4, "1st Day" and "131st Day", summarize the number of people who were ranked in those three categories on those 2 days. On the first day, there were 4994 people ranked only under the Classic leaderboard, 5004 people who were ranked only under the New leaderboard, and 4996 people ranked under both Classic and New leaderboards. On the last day, there were 5042 people ranked only under the Classic leaderboard, 5042 that were ranked only under the New leaderboard, and 4938 people ranked only under both. The last column, "Remained on the 131st Day", is defined in a more complex way. These are the people who were ranked on the first day who remained on either or both of two leaderboards on the last day. We will focus on this group of people in our data analysis. Note that the numbers generated in this column are not equivalent to the intersection between the numbers from the first two columns. For example, a person who is ranked under only Classic on the first day, but who on the last day is ranked under New, will be counted in the "Remained on the 131st Day" column but not in the "131st Day" column. The number of people who have left their respective category can be calculated by taking the difference between the columns "1st Day" and "Remained on the 131st Day". The number of people who have entered their respective category can be calculated by taking the difference between the columns "131st Day" and "Remained on 131st Day".

# of People	1st Day	131st Day	Remained on the 131st Day
Ranked Only Under Classic	4994	5042	4742
Ranked Only Under New	5004	5042	4324
Ranked Under Both Classic and New	4996	4938	4992

Table 4.1 Number of People Ranked

We report the key summary statistics of people by the three categories. Table 4 contains information for those that are ranked under Classic only, Table 4 contains information for those that are ranked under New only, and Table 4 contains information for those that are ranked under both Classic and New.

Classic Only		1st Day	131st Day	Remained on the 131st Day
# of people		4994	5042	4742
Classic Rank	Mean	6024.94	5962 .78	6002.22
	Std. Dev	2517.89	2534.94	2501.49
	Min	46	46	46
	Max	9940	9999	9999
# of Lifetime Reviews	Mean	113.67	118.84	117.77
	Std. Dev	160.57	165.91	168.32
	Min	0	0	0
	Max	7278	7343	7343
# of Lifetime Helpful Votes	Mean	804.25	854.69	838.98
	Std. Dev	737.77	784.17	765.42
	Min	0	0	0
	Max	13880	14009	14009

Table 4.2 Summary Statistics of People Ranked Under Classic Only

New Only		1st Day	131st Day	Remained on the 131st Day
# of people		5004	5042	4324
New Rank	Mean	6333.99	6329.57	5796.37
	Std. Dev	2380.91	2383.67	2400.95
	Min	248	221	167
	Max	9999	10000	9997
# of Lifetime Reviews	Mean	67.87	72.80	80.16
	Std. Dev	71.61	75.83	83.45
	Min	1	3	3
	Max	2164	2246	2246
# of Lifetime Helpful Votes	Mean	365.61	389.80	432.92
	Std. Dev	823.56	880.38	941.96
	Min	5	37	45
	Max	46859	47007	47007

Table 4.3 Summary Statistics of People Ranked Under New Only

Classic and New		1st Day	131st Day	Remained on the 131st Day
# of people		4996	4938	4992
Classic Rank	Mean	3923.76	3963.66	3767.99
	Std. Dev	2810.81	2837.90	2677.10
	Min	1	1	1
	Max	9940	9914	9914
New Rank	Mean	3664.87	3632.22	3595.56
	Std. Dev	2727.62	2714.71	2697.57
	Min	1	1	1
	Max	10000	9999	9994
# of Lifetime Reviews	Mean	291.04	305.78	303.09
	Std. Dev	495.95	517.39	514.96
	Min	0	0	0
	Max	21146	22161	22161
# of Lifetime Helpful Votes	Mean	2116.18	2209.57	2226.46
	Std. Dev	3502.40	3646.45	3625.72
	Min	0	0	0
	Max	80788	84424	84424

Table 4.4 Summary Statistics of People Ranked Under Classic and New

4.3 A Description of the Amazon Ranking System

This section provides a qualitative description of the Amazon ranking system.

The total population of reviewers is N. We do not observe N directly, though in any given period we are able to observe the top 10,000 reviewers ranked under the Classic ranking system and the top 10,000 reviewers ranked under the New ranking system. There are an infinite number of periods. We will denote the current period as t. A reviewer can be ranked under the Classic ranking system only, the New ranking system only, under both ranking systems, or neither. Thus the union of all reviewers in any given period can be as few as 10,000 (everyone is ranked under both systems) or as many as 200,000 (everyone is ranked only under one system).

Amazon has chosen two scoring functions, one called Classic and one called New. The reviewers are only given a qualitative description of the two scoring functions. From the qualitative description by Amazon, we know both scoring functions take into account the total number of reviews and the quality of reviews. The main difference between Classic and New is that the New scoring function weights recent reviews and helpful votes more. At the beginning of each period, the ranking for all reviewers are calculated under the Classic ranking formula and under the New ranking formula.

The reviewer can choose the number of reviews he wants to write, and the effort he will put in writing the reviews. The quality of the reviews is measured indirectly through the number of helpful votes a review has received.

4.4 Results

We present two main results. The first main result is we found evidence that Amazon implemented the scoring functions that it described. In particular, the New leaderboard weights recent activities more heavily than the Classic leaderboard. The second main result is we show that reviewers respond to the review writing behavior of their neighbors in New ranking, but not to their neighbors in Classic ranking.

Three pieces of supporting evidence show the New leaderboard weights recent activities in reviews and helpful votes more heavily. First, reviewers that were ranked under the Classic ranking system wrote fewer recent reviews and generated fewer recent helpful votes than those who were ranked under the New leaderboard were. This is consistent with the fact that the New leaderboard weights recent activities more heavily. Second, the ranking on the last day in our observation period is better explained by the ranking on the first day in the Classic leaderboard compared with the New leaderboard. This implies that the correlation between the first and last day ranking is higher under the Classic leaderboard. This is true because recent activities play less of a role in the ranking of those under the Classic ranking. Third, the marginal effect of an additional review and an additional helpful vote on the change in ranking is higher under the New ranking than under the Classic ranking. This means a recent review or a recent helpful vote has more impact on those ranked on the New leaderboard than those ranked under the Classic leaderboard do.

First, we find that people who stayed on the Classic leaderboard wrote fewer recent reviews and generated fewer recent helpful votes than those ranked under the New leaderboard. We show this by running two OLS regressions. In Equation (4.1), we test for differences in the number of reviews between those that are ranked under Classic versus those that are ranked under New. In Equation (4.2), we test for differences in the number of helpful votes between the two leaderboards.

In Equation (4.1), the dependent variable is the difference in the number of reviews written between the last and the first day of the observation period. The independent variable, RankedUnderClassic, is a dummy variable indicating whether a person is ranked under Classic or New on the first and last day. The dummy variable takes on the value of one if the person is ranked only under Classic, and takes on the value of zero if the person is ranked only under New. Equation (4.2) is similar except the dependent variable is the number of helpful votes generated between the first and last day.

$$(\mathrm{r}_{i,131} - \mathrm{r}_{i,1}) = \beta_0 + \beta_1 * \mathrm{I}(RankedUnderClassic) + \varepsilon_i \tag{4.1}$$

$$(\mathrm{h}_{i,131} - \mathrm{h}_{i,1}) = \beta_0 + \beta_1 * \mathrm{I}(RankedUnderClassic) + \varepsilon_i \tag{4.2}$$

where

$$RankedUnderClassic = \begin{cases} 0 & \text{for reviewers ranked under} \\ & \text{New only on first and last day,} \\ 1 & \text{for reviewers ranked under} \\ & \text{Classic only on first and last day} \end{cases}$$

$$
\begin{aligned}
r_{i,t} &= \text{\# of lifetime reviews written by reviewer i at time } t \\
h_{i,t} &= \text{\# of lifetime helpfulness votes of reviewer i at time } t
\end{aligned}
$$

The data used in the analysis are as follows. First, we included only those who were ranked on the first and the last day in our observation period. For example, a person who was ranked under Classic on the first day but not the last day was not included in the analysis.[3] Second, we excluded people who were ranked in both Classic and New on the first day. We did this to focus our analysis on the difference in behavior between reviewers who were ranked only in the Classic leaderboard against those were ranked only in the New leaderboard. This is equivalent to the people who were described under the column "Remained on the 131st Day" in Table 4 and Table 4.

Amazon's own description of the qualitative differences between the Classic ranking and the New ranking is that the Classic ranking formula weights the lifetime (cumulative) number of activities more than the recency of the activities. If this were true, we would expect that people ranked under Classic in the first day would write fewer reviews and generate fewer helpful votes. In turn, we would expect β_1 in Equations (4.1) and (4.2) to be negative.

	$r_{i,131} - r_{i,1}$ (4.1)	$h_{i,131} - h_{i,1}$ (4.2)
Constant	8.87^{***}	59.50^{***}
RankedUnderClassic	-7.82^{***}	-43.21^{***}
N	9066	9066
R^2	0.0752	0.0219

*** significant at 1% level

Table 4.5 Differences Between Classic and New Leaderboards

Table 4 summarizes the results from the regressions. There are 4742 people who were ranked under Classic only and 4324 ranked under New only, yielding a total of 9066 observations. The coefficients from Equation (4.1) and Equation (4.2) show that people who stayed on the Classic leaderboard wrote 7.82 fewer reviews and generated 43.21 fewer helpful votes than those who stayed on the New leaderboard. There are two ways to interpret these results, which are not mutually exclusive. We

[3] It is possible for a reviewer to leave a leaderboard in the interim and return to the leaderboard in the last period.

can say that the New ranking system tends to select people who are more active in the observation period. We can also say that reviewers need to work harder to stay on the New leaderboard.

For the second supporting evidence, we look at three factors that could predict the rankings on the last day of the observation period: the number of reviews written between the first and the last day, the number of helpful votes generated between the first and the last day, and the ranking on the first day. All three explanatory variables are statistically significant predictors of the ranking on the last day, though the ranking on the first day explains the most substantial amount of variation.

For each of the two ranking systems, we regressed the ranking on the last day against the first day ranking only (Model I), and then against the reviews written and the helpful votes generated in the past 131 days in addition to the first day ranking (Model II). Equation 4.3 is the specification for Model II under the Classic leaderboard. Equation (4.4) is the specification under the New leaderboard.

$$\begin{aligned} \text{Classic Rank}_{i,131} &= \beta_0 + \beta_1 * (\mathrm{r}_{i,131} - \mathrm{r}_{i,1}) + \beta_2 * (\mathrm{h}_{i,131} - \mathrm{h}_{i,1}) + \\ &\quad \beta_3 * \text{Classic Rank}_{i,1} + \varepsilon_i \qquad (4.3) \\ \text{New Rank}_{i,131} &= \beta_0 + \beta_1 * (\mathrm{r}_{i,131} - \mathrm{r}_{i,1}) + \beta_2 * (\mathrm{h}_{i,131} - \mathrm{h}_{i,1}) + \\ &\quad \beta_3 * \text{New Rank}_{i,1} + \varepsilon_i \qquad (4.4) \end{aligned}$$

Table 4 reports the coefficients for the two models specified in Equation (4.3) and in Equation (4.4). There are three things of note. First, the R^2 value does not change substantially between Model I and Model II for either ranking system. This means the first day ranking explains a substantial amount of the variation in the last day ranking. In other words, there is high correlation between first day and last day ranking under both ranking systems. Second, the number of reviews and the number of helpful votes generated in the 131 days have more explanatory power under the New ranking compared with the Classic ranking. Third the ranking on the last day has less explanatory power under the New ranking compared with the Classic ranking.

The Model II specification also allows us to understand the marginal effect of an additional review written and an additional helpful vote generated. Writing one additional review in the 131-day period increases the last day Classic rank by 2.51 and the last day New rank by 3.86. One additional helpful vote increases the Classic rank by 0.28 and the New rank by 0.18. These observations imply that the New ranking weights the recent activities more than the Classic ranking. These results also show that reviews are weighted more heavily than the helpful votes. These observations are consistent with the qualitative description of the two ranking systems provided by Amazon, and provide an approximation to the scoring rule that Amazon does not disclose publicly.

Dependent Variable	$Classic_{i,131}$	$Classic_{i,131}$		$New_{i,131}$	$New_{i,131}$
	Model I	Model II		Model I	Model II
$Constant$	76.88^{***}	175.63^{***}	$Constant$	55.58^{***}	190.61^{***}
$r_{i,131} - r_{i,1}$		-2.51^{***}	$r_{i,131} - r_{i,1}$		-3.86^{***}
$h_{i,131} - h_{i,1}$		-0.28^{***}	$h_{i,131} - h_{i,1}$		-0.18^{***}
$Classic_{i,1}$	0.94^{***}	$0.93^{(}***)$	$New_{i,1}$	1.07^{***}	1.05^{***}
N	4634	4634	N	4634	4634
R^2	0.9783	0.9804	R^2	0.9274	0.9300

*** significant at 1% level
The coefficients for the number of reviews and number of helpful votes are negative because a lower rank number implies a higher rank.

Table 4.6 Predictors of Last Day Ranking

For the third supporting piece of evidence, we look at the differential impact on ranking under the New leaderboard and the Classic leaderboard given the same level of effort. We do this by reformulating Equations (4.3) and (4.4) by subtracting the first day ranking from both sides of the equation, yielding the following transformed equations:

$$\text{Classic Rank}_{i,131} - \text{Classic Rank}_{i,1} = \beta_0 + \beta_1 * (\mathrm{r}_{i,131} - \mathrm{r}_{i,1}) + \beta_2 * (\mathrm{h}_{i,131} - \mathrm{h}_{i,1}) + + \varepsilon_i \quad (4.5)$$

$$\text{New Rank}_{i,131} - \text{New Rank}_{i,1} = \beta_0 + \beta_1 * (\mathrm{r}_{i,131} - \mathrm{r}_{i,1}) + \beta_2 * (\mathrm{h}_{i,131} - \mathrm{h}_{i,1}) + + \varepsilon_i \quad (4.6)$$

Table 4.7 summarizes the coefficients from running OLS regression specified by Equations (4.5) and (4.6). Writing a single review improves the rank by 2.59 under Classic and 4.27 under New. Having one more helpful vote improves the rank by 0.084 under Classic and 0.36 under New. We see that for the same effort level, there is more mobility under New than under Classic. This once again is consistent with the qualitative description of the ranking systems by Amazon.

Table 4.7 Predictors of Change in Ranking

Dependent Variable	$Classic_{i,131}$ -$Classic_{i,1}$	$New_{i,131}$ -$New_{i,1}$
	Equation (4.5)	Equation(4.6)
Constant	-95.83^{***}	388.70^{***}
$r_{i,131} - r_{i,1}$	-2.59^{***}	-4.27^{***}
$h_{i,131} - h_{i,1}$	-0.084^{***}	-0.36^{***}
N	4634	4634
R^2	0.0465	0.0623

*** significant at 1% level
The coefficients for the number of reviews and number of helpful votes are negative because a lower rank number implies a higher rank.

We now turn to see if people are competing against their neighbors in ranking, which would provide evidence of competitive behavior. In particular, we are looking to see whether there is correlation between the number of reviews written in one period with the number of reviews written in the previous period by the same reviewer and the number of reviews written in the previous period by the reviewer's higher and lower ranked neighbors in Classic and in New.

In OLS regression 4.7, the dependent variable is the difference in the number of reviews written between the current period (t) and the previous period ($t-1$) by a reviewer. The independent variables are the differences in the number of reviews written by the various neighbors between the previous period ($t-1$) and the period before that ($t-2$). In each ranking, there is an upper neighbor and a lower neighbor, thus yielding four neighbors.[4] In our data analysis, we only include observations where a reviewer has neighbors in all four positions in the previous period. In other words, we only include reviewers that were ranked under both Classic and New leaderboards in the same period. Since the same reviewer can appear multiple times, we cluster the errors on the level of each individual reviewer.

[4]It is possible that there are multiple reviewers who are a reviewer's neighbor due to the presence of ties. For example, if three people are tied at the rank of three, the next ranked person has a rank of four and has three higher ranked neighbors. In these cases, we randomly picked one of the neighbors that are tied in ranking.

$$
\begin{aligned}
(\mathrm{r}_{i,t} - \mathrm{r}_{i,t-1}) \quad = \quad & \beta_0 + \\
& \beta_1 * (\mathrm{r}_{i,t-1} - \mathrm{r}_{i,t-2}) + \\
& \beta_2 * (\mathrm{r}_{\text{i's C higher ranked neighbor},t-1} - \mathrm{r}_{\text{i's C higher ranked neighbor},t-2}) + \\
& \beta_3 * (\mathrm{r}_{\text{i's C lower ranked neighbor},t-1} - \mathrm{r}_{\text{i's C lower ranked neighbor},t-2}) + \\
& \beta_4 * (\mathrm{r}_{\text{i's N higher ranked neighbor},t-1} - \mathrm{r}_{\text{i's N higher ranked neighbor},t-2}) + \\
& \beta_5 * (\mathrm{r}_{\text{i's N lower ranked neighbor},t-1} - \mathrm{r}_{\text{i's N lower ranked neighbor},t-2}) + \\
& \varepsilon_{i,t} \qquad (4.7)
\end{aligned}
$$

If people are paying attention and responding to the review-writing behavior of their neighbors, we would expect to see a correlation between the number of reviews written by a reviewer in one period to the number of reviews written by her neighbors in ranking in the previous period. That is, β_2 (higher ranked Classic neighbor) and β_3 (lower ranked Classic neighbor) should be positive for the Classic ranking neighbors and similarly, β_4 and β_5 should be positive for the New ranking neighbors. If there is correlation between a person's number of reviews written in one period to her number of reviews written in the previous period, we would expect β_1 to be positive.

The coefficients are: β_0 = 0.19 ($p < 0.001$); β_1 = 0.074 ($p = 0.152$); β_2 = 0.0030 ($p = 0.283$); β_3 = 0.0052 ($p = 0.177$); β_4 = 0.013 ($p = 0.012$); and β_5 = 0.020 ($p = 0.004$) with a R^2 value of 0.0058 (N=53966, 1079 clusters by reviewer). The fact that β_2 and β_3 are not statistically significant but that β_4 and β_5 are means that reviewers are responding to the number of reviews written by their New neighbors but not their Classic neighbors. The correlation between the observations from the same reviewer has been taken into account by clustering the errors on the level of reviewers.

On average, when the higher ranked New neighbor writes 1 review, the reviewer would respond by writing 0.013 review. When the lower ranked New neighbor writes 1 review, the reviewer would respond by writing 0.020 review. The difference in magnitude implies that reviewers are more concerned with being overtaken by their lower ranked New neighbors than trying to overtake their higher ranked New neighbors.

4.5 Conclusion

In this chapter, we found evidence that show Amazon implements the New and Classic leaderboards where the New leaderboard weights recent activities of reviewers more than the Classic leaderboard. We then show that reviewers respond to the review-writing behavior of their ranked neighbors on the New leaderboard but not the Classic leaderboard. This provides evidence that the ranking on the New leaderboard is more motivating: exerting effort is more likely to lead to change in ranking under the New leaderboard. This shows that presenting the right type of social information can be an effective way to encourage users of online communities to contribute toward a public good from by harnessing the competitive nature of the users.

Chapter 5

Conclusion

The three chapters show that institutions matters, but in designing the institutions, it is important to validate the assumptions that we make in modeling the behavior of the individuals. The importance of understanding individual behavior is seen clearest in the laboratory studies of allocating access to a large scientific equipment and in finding the optimal structure of prizes in contests. One of the challenges the individuals faced in these two studies is the complexity of the decision problem they face. While it is not surprising that people would use heuristics in the decision making process, it is crucial to gain an understanding of exactly the type of heuristics that they do use. The two experimental studies presented in this dissertation are a step forward in understanding individual decision making in complex settings.

This dissertation also shows the interplay between theory, laboratory experiments, and field studies. One connection between the first and the second chapter is methodology. Both studies rely on results from auction theory, and those results are evaluated in the laboratory. The Amazon field study is of interest because leaderboards are a kind of contest for status. We did not modeled the dynamics of the leaderboard contest using tools from auction theory because of the complexity of the game. However, despite the absence of a model that can provide a closed-form prediction, it is still of great value to evaluate the effectiveness of the two different leaderboards in encouraging reviewers to write product reviews.

A natural extension of the package auction study is to extend the auction mechanism so that bidders can bid with chits instead of money, since it may be awkward to ask scientists to pay to use the scientific equipment from their grant since the money came from the funding agency in the first place. An open question from the contest study is why is there more variance in the behavior of subjects in the Concave-Complete treatment. It is of interest to explore this question because the higher variance is correlated with a higher effort level. Understanding why there is more variance in the behavior of subjects in this particular treatment could help us build a more complete model of individual behavior of decision making. One natural extension to the empirical study on the product review-writing behavior of Amazon reviewers is we were not able to make use the day-to-day data but aggregated the data over the whole observation period. We needed to do this due to the complex interactions between the behavior of reviewers and the selections of reviewers according to the different scoring rules. This would be a rich and interesting area to explore for the field of econometrics.

Appendices

Appendix A

Experiment Instructions for the NEES Experiment

Experiment Instructions for VCG

Player ID: _______________
Player Type: ________
Computer #: ________

Instruction

You are about to participate in an economics experiment in which you will earn money based on the decisions you and others make. All earnings you make in the experiment are yours to keep. Please do not talk to each other during the experiment. If you have a question, please raise your hand and the experimenter will come and help you.

Overview

- All values and prices will be stated in francs. Each franc you earn can be converted into US currency at the rate specified by the experimenter. The exchange rate **is 12 francs (₣) per dollar ($)**.
- In this experiment, you will participate in a series of auctions that allocates time use of a major piece of scientific equipment. The computer will be the auctioneer and you will compete against 8 other people in the room.
- If you win a time slot at the end of an auction, your profit will be the difference between your value of that time slot and the price you pay for it. If you do not win any time slot, your profit will be zero. Therefore, for you:

  $$\textbf{Profit} = \textbf{value} - \textbf{final price}.$$

  Please note that your profit, which will be used to determine a portion of your payment for the experiment, depends on the **difference** between your value and the final price, **not** your value alone.
- At the end of each auction, you will fill out your Auction Worksheet and a monitor will verify your earnings.

Background

You are one of nine scientists who are bidding for access time to a piece of major scientific equipment. Three researchers are conducting big projects, which require at least 3 months of equipment time to complete, while six researchers are conducting smaller experiments that require at least 2 months to complete. The time slots (also called packages) on which you will be bidding are composed of consecutive months within a 24-month timeframe. You only have one experiment that you want to run. No matter how many packages you bid on, you will never be allocated more than one package.

Value Determination

You will have a unique value for each time slot (package) depending on when you are able to start using the equipment as well as the length of time for which you can use your equipment.

The values of the various time slots to different researchers depend on several factors.

1. *Value for each minimum package for a small researcher* (2 months) is randomly drawn from the set of integers between 20 and 100, inclusive, where each integer is equally likely to be drawn. A package of fewer than 2 months is worth zero to a small researcher.
2. *Value for each minimum package for a big researcher* (3 months) is randomly drawn from the set of integers between 20 and 150, inclusive, where each integer is equally likely to be drawn. A package of fewer than 3 months is worth zero to a big researcher.
3. *Value for each additional month* is randomly drawn from the interval between 10 and 20, inclusive. Both sets of researchers get added value from using the equipment for more time than the absolute minimum. Small researchers derive more value from using the equipment for 3 months instead of 2, and the most value from using the equipment for 4 months. Using the equipment for five or more months, however, gives them no more value than just using it for four months. Similarly, large researchers get more value from 4 months of use of the equipment as opposed to 3, and they get the most value from using the equipment for 5 months. More than 5 months of time, however, does not give them any additional value.
4. *Starting month*: If your player ID starts with the digit 3, 7, 8, or 9, you prefer to use the equipment earlier rather than later. If your player ID starts with the digit 2, 5, or 6, you are indifferent between starting earlier and later. Finally, if your player id starts with the digit 1 or 4, you prefer to start later rather than earlier.

For each participant, the various components of his or her value will be randomly drawn in each auction. Whether you are a big or small researcher, and whether you prefer to start earlier or later will already be taken into account in the values for packages that you see on your screen.

Auction Process

The experiment may have several auctions. For each auction, you may bid on packages. At the end of each auction, you will see if the bids that you submitted were winning or losing.

Months and Packages

In the experiment, you can only bid on packages, but packages are constructed from individual months. In this case, there are 24 different months. The packages are all the possible ways in which the months can be put together in consecutive order with lengths of 2, 3, 4 for a small researcher and lengths 3, 4, 5 for a big researcher. The starting month and the length of the package uniquely identify a package. For example, if a package starts on month 1 and lasts for three months, it consists of month 1, month 2, and month 3.

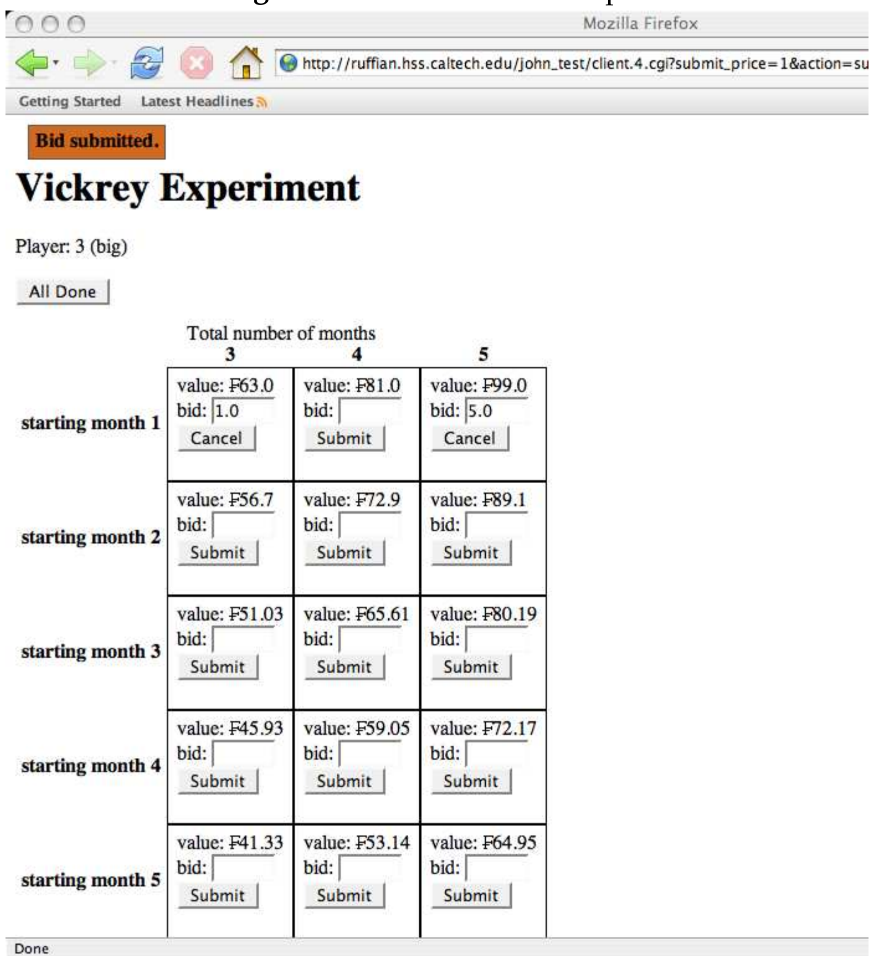
Submitting an Order

The packages will be presented to you in a table format. The different rows will represent different starting months, and the different columns will represent different durations. For example, to find a package that starts on the 4th month, and lasts for 3 months, you will first look for the row that is labeled 4, and find the column that is labeled 3.

In each cell, you will see the value of that package to you. If you wish to bid on that package, you can submit a price in the textbox and then click on the submit button. Note that you will need to click the submit button for each time that you submit a bid. Refer to Figure 1 below to see how your screen will look during the submission process.

You are allowed to submit a price for a package that is higher than your value for that package, but if you end up winning the package, you may make a negative profit. **If you earn a negative profit for an auction, that amount will be deducted from your experiment pay. If your cumulative earning becomes negative, that is, if the sum of your show up fee, quiz score, and profit from participating in the experiment is negative, you will be asked to leave the experiment**.

Figure A.1 Submission Sample



Canceling an Order

You can always cancel a bid before you click on the "All Done" button. You can do this by clicking on the "Cancel" button for the bid that you want to cancel.

All Done Button

Once you have submitted all the bids for a round, please click on the "All Done" button. You do not have to wait until time runs out before pressing the button. You will be instructed to wait until all the other players have clicked on the "All Done" button.

The Solver

At the end of each auction the solver will determine which bids are the winning bids. The solver selects the combination of bids that maximizes the revenue to be received. For simplicity, in all subsequent examples, we assume that there are only three months, and the minimum package is one month.

> Example: At the end of an auction, the following bids have been submitted:
>
> Bidder 1: {month 1, month 2, month 3} = 200
> Bidder 2: {month 1} = 100
> Bidder 3: {month 2, month 3} = 75
>
> The solver can choose to allocate the months in two different ways.
>
> 1. Give month 1, month 2, and month 3 to Bidder 1. The total revenue is 200.
> 2. Give month 1 to Bidder 2 and month 2 and month 3 to Bidder 3. The total revenue is 175.
>
> Since the first allocation gives a higher total revenue, the solver will make Bidder 1 the winning bidder of month 1, month 2, and month 3.

The Rebate

The amount that the winning bidders are required to pay at the end of the auction depends on the additional revenue that each bidder generated, which is calculated by comparing the revenue obtained by the auctioneer, versus the revenue obtained by the auctioneer when the given bidder was ignored.

> Example: At the end of an auction, the following bids have been submitted:
>
> Bidder 1: {month 1, month 2, month 3} = 200, {month 1} = 80
> Bidder 2: {**month 1**} **= 100**, {month 1, month 2}=120
> Bidder 3: {**month 2, month 3**} **= 150**, {month 3} = 80
>
> The winning bids are Bidder 2's bid and Bidder 3's bid, because they generate the highest revenue $100 + 150 = 250$.
>
> However, the auctioneer does not ask Bidder 2 to pay 100. Suppose that Bidder 2 were ignored. The winning bids then become Bidder 1's bid on (month 1) and Bidder 3's bid on (month 2, month 3).
>
> Bidder 1: {month 1, month 2, month 3} = 200, {**month 1**} **= 80**
> ~~Bidder 2: {month 1} = 100, {month 1, month 2}=120~~
> Bidder 3: {**month 2, month 3**} **= 150**, {month 3} = 80
>
> Then the auctioneer calculates the revenue that those winning bids would generate, which is $80 + 150 = 230$. Thus, the additional revenue that Bidder 2 makes is 20, since $250 - 230 = 20$. This 20 is the rebate for Bidder 2. Therefore, Bidder 2 pays 100 and receives 20 back. His final price is 80.
>
> Similarly, Bidder 3 might not have to pay 150. When Bidder 3 is ignored, the winning bid is Bidder 1's bid on (month 1, month 2, month 3).
>
> Bidder 1: {**month 1, month 2, month 3**} **= 200**, {month 1} = 80
> Bidder 2: {month 1} = 100, {month 1, month 2}=120
> ~~Bidder 3: {month 2, month 3} = 150, {month 3} = 80~~
>
> The revenue to the auctioneer would be 200. Therefore, the additional revenue that Bidder 3 generates is 50, since $250 - 200 = 50$. The rebate for Bidder 3 is 50. Bidder 3 pays 150 and receives the rebate of 50. Thus, Bidder 3's final price is 100.

Notice that your final price depends on other bidders' bids, which you cannot observe during the auction.

Timing

For each auction, you will be given up to 7 minutes to submit your bids. An experimenter will alert you when your time for each auction is up and will ask you to click the "All Done" button if you have not already.

Auction Results

At the end of each auction, please record your winning package, if any, the value you have for that package, and your profit.

Special Notes

We do not offer any guarantees that the software will not crash. When the computer crashes there may be some excitement, but there is no need to panic.

Worksheet

Please fill out the worksheet completely as instructed. Your values are private information. Please do not reveal them to anyone.

At the end of the auction the owner of each package will receive his or her profit. Again, this is calculated in the following way:

$$\textbf{Profit} = \textbf{value} - \textbf{final price}.$$

There will be a total of 8 auctions. Your total profit will be the sum of profit in each of the 8 auctions.

Player ID: ________________
Player Type: Big or Small (circle one)
Computer #: ________

Review Questions:

(You will be **paid** the specified amount for each correct answer.)

Q.1 Number of winning bids. ($0.50)

a) ($0.25) What is the minimum number of **winning bids** you may have at the **end** of the auction?

b) ($0.25) What is the maximum number of **winning bids** you may have at the **end** of the auction?

Q.2 Winning Bids ($1)

Suppose the bids submitted by four bidders are as listed in the table below. Which set of bids will be the winning bids?

	Package	Bid
Bidder 1	1–2	₣30
Bidder 2	1–2	₣20
Bidder 3	1–4	₣60
Bidder 4	3–5	₣40

Q.3 Rebate ($1)

Suppose the bids submitted by four bidders are as listed in the table below.

	Months	Bid
Bidder 1	1–3	₣30
Bidder 2	1–3	₣50
Bidder 3	2–5	₣70
Bidder 4	4–5	₣50

The winning bids are Bidder 2's bid and Bidder 4's bid, because they generate the highest revenue $50 + 50 = 100$. To compute the rebate for Bidder 2, let us ignore Bidder 2's bid.

a) ($0.25) Which set of bids will be the winning bids if Bidder 2's bid was ignored?

b) ($0.25) How much is the revenue that those winning bid(s) would generate?

c) ($0.25) How much is the additional revenue that Bidder 2 generates?
(Hint: The highest revenue is 100 when Bidder 2's bid is not ignored.)

d) ($0.25) How much is the final price for Bidder 2?
(Hint: The final price = Bid – Rebate)

Player ID: _______________

Player Type: Big or Small (circle one)

Computer #: ________

Earning Work Sheet:

Auction #	Winning Package (starting month, package length)	Value	Profit
Example	(2,3)	100	20
Auction 1			
Auction 2			
Auction 3			
Auction 4			
Auction 5			
Auction 6			
Auction 7			
Auction 8			
Total		Auction Profit=	

(______ / ______) + _____ + ______ = ______

Auction Profit I   Exchange Rate   Quiz Score   Show Up Fee   Total Earning

Experiment Instructions for Knapsack

Player ID: _______________
Player Type: ________
Computer #: ________

Instruction

[The same as above.]

Overview

- All values and prices will be stated in francs (₣). Each franc you earn can be converted into US currency at the rate specified by the experimenter. The exchange rate **is 20 francs (₣) per dollar ($)**.
- In this experiment, you will participate in an allocation process that allocates time use of a major piece of scientific equipment. The computer will be the coordinator of the process and you will compete against 8 other people in the room.
- If you are allocated a time slot, your earnings will be the value of that time slot. If you do not receive any time slot, your earnings will be zero.

Profit = **value of the package you get**.

- After each allocation is completed, you will fill out your Worksheet and a monitor will verify your earnings.

Background

[The same as above.]

Value Determination

[The same as above.]

Allocation Process

In the experiment, you can only rank packages, but packages are constructed from individual months. [The remaining part is the same as above]

Submitting an Order

The packages will be presented to you in a table format. The different rows will represent different starting months, and the different columns will represent different durations. For example, to find a package that starts on the 4th month, and lasts for 3 months, you will first look for the row that is labeled 4, and find the column that is labeled 3.

To submit your rankings, you will rank packages using integers between 1 and 200. You will be asked to rank fewer than 200 packages, but you are given extra integers so that you have flexibility in how you assign rankings. Specifically, you will assign an integer to each package that you think you want to get. Since these numbers are rankings, if you assign a package "1", the coordinator will interpret this as your first-choice package. You can choose and rank packages in any order of your preference.

In each cell, you will see the value of that package to you. If you wish to submit a rank for that package, you can type a number in the textbox and then click on the submit button. Note that you will need to click the submit button for each time that you submit a ranking. Refer to Figure 1 below to see how your screen will look during the submission process.

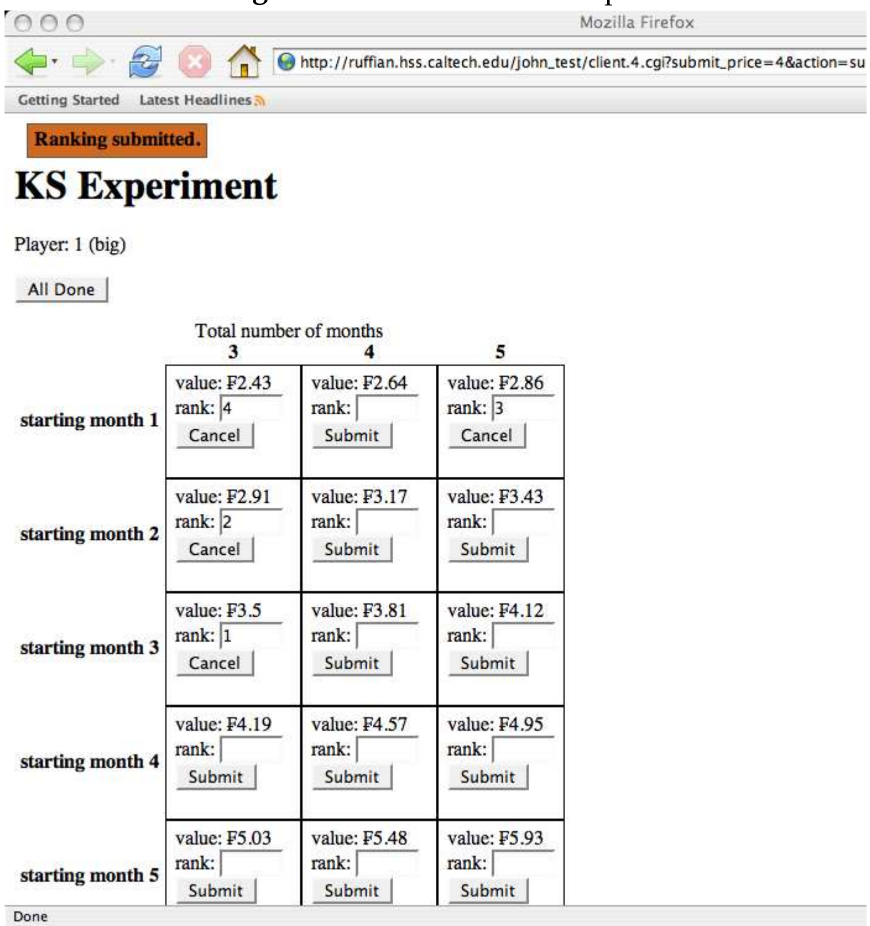
Canceling an Order

You can always cancel a ranking before you click on the "All Done" button. You can do this by clicking on the "Cancel" button for the ranking that you want to cancel.

All Done Button

Once you have submitted all your rankings, please click on the "All Done" button. You do not have to wait until time runs out before pressing the button. You will be instructed to wait until all the other players have clicked on the "All Done" button.

Figure A.2 Submission Sample



The Solver

After all participants have submitted their rankings, the solver will determine how to allocate the months. The solver goes though the following three steps to do so:

1. translate ranks submitted by a participant into consecutive numbers,
2. convert the consecutive numbers into points, and
3. find the allocation that maximizes the total points.

1. Translation

For each participant, the solver changes the submitted ranks into consecutive numbers.

> Example: Suppose that a participant has submitted the following ranks shown in the left column of the table below.
>
> | Package | Rank: Before translation | Rank: After translation |
> |---|---|---|
> | {month1, month2, month3} | 5 | 2 |
> | {month2, month3, month4} | 2 | 1 |
> | {month2, month3} | 6 | 3 |
> | {month1, month2} | 10 | 4 |
>
> This participant ranks 4 packages with the numbers 5, 2, 6 and 10, or $2 < 5 < 6 < 10$ when sorted in increasing order. These ranks are translated into 1, 2, 3 and 4 based on the order. Thus, the ranking $(5, 2, 6, 10)$ is changed to $(2, 1, 3, 4)$.

When the solver translates submitted ranks, it will not assign two packages the same number. If a participant assigns the same rank to more than one package, the solver will randomly break the tie and translate the rank into two different numbers.

> Example: Suppose that a participant has submitted the following ranks shown in the left column of the table below.

Package Package	Rank Before translation	After translation (possibility 1)	After translation (possibility 2)
{month1, month2}	10	4	4
{month2, month3}	5	2	3
{month3, month4}	5	3	2
{month2, month3, month4}	1	1	1

As seen in the table, this participant ranks 2 packages as rank 5. The rank of the package {month2, month3, month4} remains 1, as there is no other package ranked with 1. The solver, however, randomly selects the two packages {month2, month3} and {month3, month4} in sequence and changes the rank into 2 and 3. The table above shows two possible results of this randomization process, and each of the results is equally likely to happen. Finally, the solver changes the rank for package {month1, month 2} into 4.

2. Conversion

The solver converts all numbers into points for each participant based on the translated rank. The package labeled as 1 in the translated rank will get 66 points, and the next package labeled as 2 will get 65 points and so forth. The solver gives points in this way to all of the packages for which a participant has submitted a ranking.

Example: Suppose that a participant has submitted ranks for 4 packages as shown in the table below.

Package	Rank Before translation	After translation	Points
{month1, month2}	5	2	65
{month2, month3}	2	1	66
{month3, month4}	6	3	64
{month2, month3, month4}	10	4	63

The solver assigns 66 points through 63 points to the packages according to the translated ranks.

3. Maximization of Points

After the solver goes through the two steps above for each participant, it selects the allocation that maximizes the aggregate points. For simplicity, in all subsequent examples, we assume that there are only three months, and that the minimum package is one month.

Example: The following rankings have been submitted:

Participant	Package	Points
Participant 1	{month1, month2, month3}	66
	{month1, month2}	65
	{month1}	64
Participant 2	{month2, month3}	66
Participant 3	{month1}	66
	{month2, month3}	65

Note that this table shows only points, omitting the rankings submitted by the participants. The solver can choose to allocate the months in five different ways.

1. Give month 1, month 2, and month 3 to Participant 1. The total points are 66.
2. Give month 1 and month 2 to Participant 1. The total points are 65.
3. Give month 1 to Participant 1 and month 2 and month 3 to Participant 2. The total points are 130 = 64 + 66.
4. Give month 1 to Participant 1 and month 2 and month 3 to Participant 3. The total points are 129 = 64 + 65.
5. Give month 1 to Participant 3 and month 2 and month 3 to Participant 2. The total points are 132 = 66 + 66.

Since the last allocation gives the highest total points, the solver will allocate month 1 to Participant 3 and month 2 and month 3 to Participant 2.

[The remaining part of the instruction is almost identical to the one for VCG.]

Player ID: ________________
Player Type: Big or Small (circle one)
Computer #: ________

Review Questions:

(You will be **paid** the specified amount for each correct answer.)

Q.1 Number of packages. ($0.50)

a) ($0.25) What is the minimum number of packages you may receive in a single allocation?

b) ($0.25) What is the maximum number of packages you may receive in a single allocation?

Q.2 Translation and Conversion ($1)

Suppose the rankings submitted by a participant are as listed in the table below.

	Rank		
Package	Before translation	After translation	Points
{month1, month2}	10		
{month2, month3}	102		
{month3, month4}	7		
{month5, month6}	4		

a) ($0.5) Translate the submitted rank into consecutive numbers and fill in the column labelled "After translation" in the table above.

b) ($0.5) Convert the translated ranks into points and fill in the column labeled "Points" in the table above.

Q.3 Conversion ($0.5) Suppose a participant assigns a rank of 1 to two different packages. Choose one of the following situations that correctly describes the traslation and conversion process for those rankings.

a) Both of the packages get 1 pt.

b) Both of the packages get 66 pts.

c) One of the packages is randomly selected and given 66 pts, and the other gets 0 pts.
d) One of the packages is randomly selected and given 66 pts, and the other gets 65 pts.
e) One of the packages is randomly selected and given 132 pts, and the other gets 0 pts.

Your Answer ____

Please turn over

Q.4 Allocation ($1.5)

Suppose the rankings submitted by two participants are as listed in the table below. The following rankings have been submitted:

Participant	Package	Points
Participant 1	{month1, month2, month3}	66
	{month1, month2}	65
	{month1}	64
Participant 2	{month2, month3}	66
	{month1, month2}	65
	{month3}	64

The solver can choose to allocate the months in four different ways.

- Give {month1, month2, month3} to Participant 1. The total points are 66.
- Give {month1} to Participant 1 and {month3} to Participant 2. The total points are 64 + 64 = 128.
- Give { } to Participant 1 and { } to Participant 2.

  The total points are { }+{ }={ }.
- Give { } to Participant 1 and { } to Participant 2.

  The total points are { }+{ }={ }.

a) ($0.5) Find the other two ways of allocation (in any order) and fill in the blank spaces above.

b) ($0.5) Compute the total points for each of the two allocations and fill in the blank spaces above.

c) ($0.5) Find the allocation that maximizes the total points and fill in the blank space below.
Participant 1 gets { } and
Participant 2 gets { }.

Experiment Instructions for RAD

Player ID: ________________
Player Type: ________
Computer #: ________

Instruction

[The same as above.]

Overview

[The same as above.]

Background

[The same as above.]

Value Determination

[The same as above.]

Auction Process

[The same as above.]

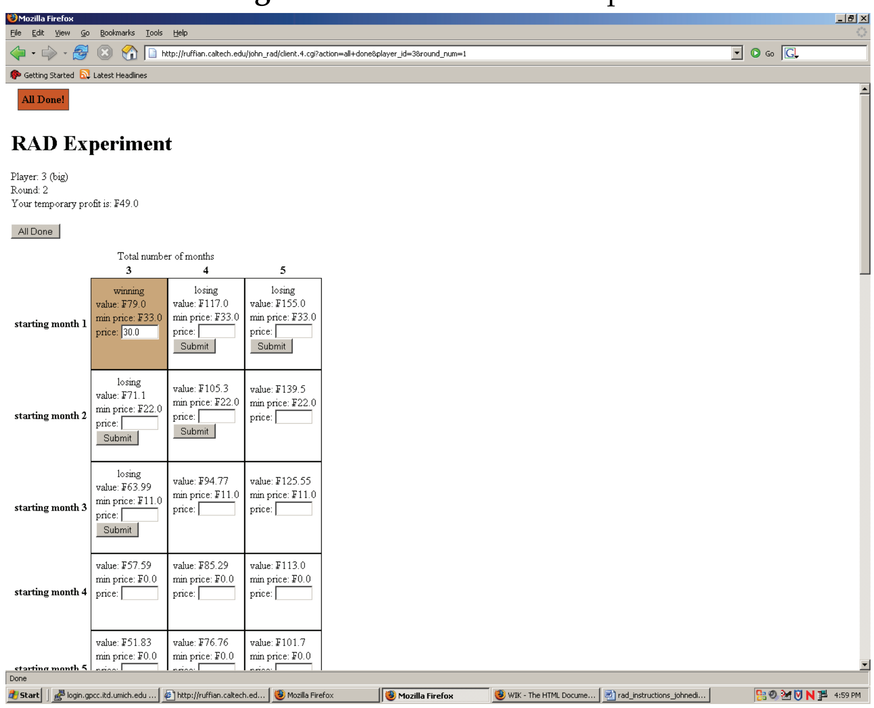
Submitting an Order

The packages will be presented to you in a table format. The different rows will represent different starting months, and the different columns will represent different durations. For example, to find a package that starts on the 4th month, and lasts for 3 months, you will first look for the row that is labeled 4, and find the column that is labeled 3.
In each cell, you will see the value of that package to you. If you wish to bid on that package, you can submit a price in the textbox and then click on the submit button.

Note that you will need to click the submit button for each time that you submit a bid. Refer to Figure 1 below to see how your screen will look during the submission process.

If your submitted price is lower than the minimum price, your bid will not be accepted. You are allowed to submit a price for a package that is higher than your value for that package, but if you end up winning the package, you will be making a negative profit. **If you earn a negative profit for an auction, that amount will be deducted from your experiment pay. If your cumulative earning becomes negative, that is if the sum of your show up fee, quiz score, and profit from participating in the experiment is negative, you will be asked to leave the experiment**.

Figure A.3 Submission Sample



Canceling an Order

[The same as above.]

All Done Button

[The same as above.]

Beginning of Next Round

Once all the players have submitted the bids for a round and clicked on "All Done", the computer will determine, for each of the submitted bids, if it was winning or losing. The word "winning" or "losing" will appear in the corresponding cells.

Eligibility

To encourage active bidding there is a "use it or lose it" rule. In the first round, you can bid for any and all packages. In each subsequent round, you will be allowed to bid only on as many months you have placed bids on in the previous round. The maximum number of months you are allowed to have in all of your bids in a round is called your eligibility. Thus, your eligibility at the beginning of the auction is all the months that you can bid on, which is 24.

> Example: Suppose that in the first round a bidder makes three bids. These are for:
>
> Bid 1: A package beginning in month 3 and lasting for three months
> Bid 2: A package beginning in month 4 and lasting for four months
> Bid 3: A package beginning in month 13 and lasting for four months
>
> The months that the bidder has placed bids on are:
>
> Bid 1: month 3, month 4, month 5
> Bid 2: month 4, month 5, month 6, month 7
> Bid 3: month 13, month 14, month 15, month16
>
> Therefore, the unique months that the bidder has placed bids on are: month 3, month 4, month 5, month 6, month 7, month 13, month 14, month 15, and month 16. This makes a total of nine distinct months. Therefore, the bidder will have an eligibility of 9 in the next round; she will be able to place bids on 9 distinct months.
>
> Note that the eligibility will be computed on your behalf at the start of each round from the bids that you submitted in the previous round. If you are eligible to place a bid on a package, the Submit button will be available to you. The Submit button for a package will not be available to you if placing that package will result in you exceeding your eligibility.
>
> Note that the submission of one bid may affect your eligibility for other bids in the same round.

The Solver

At the end of each round the solver will determine which bids are the temporary winning bids. The solver selects the combination of bids that maximizes the revenue to be received. For simplicity, in all subsequent examples, we assume that there are only three months, and the minimum package is one month.

Example: :

At the end of round 1, the following bids have been submitted:

Bidder 1: month 1, month 2, month 3 = 200
Bidder 2: month 1 = 100
Bidder 3: month 2, month 3 = 75

The solver can choose to allocate the months in two different ways.

1. Give month 1, month 2, and month 3 to Bidder 1. The total revenue is 200.
2. Give month 1 to Bidder 2 and month 2 and month 3 to Bidder 3. The total revenue is 175.

Since the first allocation gives a higher total revenue, the solver will make Bidder 1 the temporary winning bidder of month 1, month 2, and month 3.

If you have a temporary winning bid at the end of a round, that bid will automatically be resubmitted. You may not withdraw a temporary winning bid.

The Calculation of Prices

While the computer is calculating the temporary winning bids, it will also compute prices for each month. In turn, these prices will determine the minimum prices of the packages in the next round. This process takes place in two steps.

First, the computer will first try to calculate prices, one for each month, so they sum up to the temporary winning bids and are greater than any losing bids whenever possible. When this is not possible, the computer will find prices that come as close as possible to meeting the rules. So it is possible that you will have a bid that exceeds the minimum price and yet is not a winning bid.

In the following two examples, a * in front of a bid indicates that it is a temporary winning bid.

Example: 1:

*Bidder 1: month 1 = 30
*Bidder 2: month 2 = 10
*Bidder 3: month 3 = 21

Month 1: price 30
Month 2: price 10
Month 3: price 21

Example: 2:

Bidder 1: month1, month 2 = 30
Bidder 2: month 3 = 8
*Bidder 3: month 1, month 2, month 3 = 39

Month 1: price 15
Month 2: price 15
Month 3: price 9

The temporary winning bid is bidder number 3's bid for month1, month 2, and month 3. Choosing it yields the highest revenue. The computer will calculate prices such that:

1. The prices for month 1, month 2, and month 3 add up to 39
2. The prices for month 1 and month 2 add up to at least 30
3. A price for month 3 that is at least 8

If there are several possible solutions then the computer will try to equalize the prices.

Next, after the solver has computed the prices, it sets the minimum prices for the next round by increasing the computed prices by 10%. So, following the result of Example 1, the minimum prices that would be displayed in the next round would be:

Month 1: minimum price 33
Month 2: minimum price 11
Month 3: minimum price 23.1

In the next round, Bidder 1's temporary winning bid of 30 for month 1 will be resubmitted at that price. However, the minimum price for anyone else who wishes to bid on this package will be 33.

Minimum Bids

At the start of each round, a price for each package will be displayed by the computer. Your bid for any package must be greater than or equal to the minimum price posted.

Timing

For each auction, you will be given up to four minutes to submit your bids in the first round. In rounds 2-9 of that auction, you will be given two minutes to submit your bids. In rounds 10 and higher, you will be given one minute. An experimenter will alert you when your time for each round is up and will ask you to click the "All Done" button if you have not already.

Auction Results

At the end of each auction (not the end of each round), please record your winning package, if any, the value you have for that package, the price you paid for the package, and your profit (value – paid price).

Stopping the Auction

Each auction will consist of a series of rounds. The auction will be closed at the end of round T if there is no change in the temporary ownership between the end of round T-1 and the end of round T.

Special Notes

We do not offer any guarantees that the software will not crash. When the computer crashes there may be some excitement, but there is no need to panic.

Worksheet

Please fill out the worksheet completely as instructed. Your values are private information. Please do not reveal them to anyone.

At the end of the auction the owner of each package will receive his or her profit. Again, this is calculated in the following way:

$$\textbf{Profit} = \textbf{value} - \textbf{final price}.$$

Player ID: ________________
Player Type: Big or Small (circle one)
Computer #: ________

Review Questions:

(You will be **paid** the specified amount for each correct answer.)

Q.1 Number of temporary winning bids and winning bids. ($1)

a) ($0.25) What is the minimum number of **temporary winning bids** you may have at the **beginning** of a round?
b) ($0.25) What is the maximum number of **temporary winning bids** you may have at the **beginning** of a round?
c) ($0.25) What is the minimum number of **winning bids** you may have at the **end** of the auction?
d) ($0.25) What is the maximum number of **winning bids** you may have at the **end** of the auction?

For questions 2 and 3, refer to the Figure 1: Submission Sample on page 3 of the instruction.

Q.2 ($1)

a) ($0.25) Which package is the temporary winning package for player 3?
b) ($0.25) What is the value of the temporary winning package for player 3?
c) ($0.25) What is the price player 3 is paying for this package?
d) ($0.25) What is the temporary profit for player 3? (Hint: Profit = Value - Price)

Q.3 ($1) For the package that starts on month 2 and lasts for 3 months:

a) ($0.25) Is it possible to submit a price of ₣10?
b) ($0.25) Is it possible to submit a price of ₣25?
c) ($0.25) Is it possible to submit a price of ₣100?
d) ($0.25) If player 3 submits a price of ₣80, and this package becomes the temporary winning bid in round 3, what would be the temporary profit?

For questions 4 and 5, refer to this table.

	Package	Bid
Bidder 1	1–2	₣50
Bidder 2	1–2	₣20
Bidder 3	1–4	₣60
Bidder 4	3–5	₣50

Q.4 Winning Bids ($1) Suppose the bids submitted by four bidders are as listed in the table above. Which set of bids will be marked temporarily winning in Round 2?

Q.5 Eligibility ($1) What is the eligibility for each bidder at the beginning of Round 2?

($0.25) Bidder 1:
($0.25) Bidder 2:
($0.25) Bidder 3:
($0.25) Bidder 4:

Player ID: ________________
Player Type: Big or Small (circle one)
Computer #: ________

Demographics Information

Age: ________________

Gender: ________________

Major: ________________

Undergrad:
First Year, Second Year, Third Year, Fourth Year, Fifth Year, Six or more
Or
Grad:
First Year, Second Year, Third Year, Fourth Year, Fifth Year, Six or more

Number of Game Theory Classes Taken:________________

Have you discussed auction strategy in any of your courses? Yes or No.
If so, which courses?

Appendix B

Experiment Instructions for the Contest for Status Experiment

Player ID: _______________
Computer #: ________

Instruction

[For Weibull[1.5, 10] and TS. [Date of Expt.]]

You are about to participate in an economics experiment in which you will earn money based on the decisions you and others make. All earnings you make in the experiment are yours to keep. Please do not talk to each other during the experiment. If you have a question during the experiment, please raise your hand and the experimenter will come and help you.

Overview

- All values and prices will be stated in francs. Each franc you earn can be converted into US currency at the rate specified by the experimenter. The exchange rate is **3.33 francs per dollar (or 0.3 dollar per franc)**.
- In this experiment, you will participate in 20 periods of a game. The details of the game will be explained below.

Background

You are one of twelve participants in this experiment. There will be 20 periods in the experiment. In each period, you will be randomly matched with three other people to form a group of four people. Therefore, in each period, there will be three groups formed. Your earnings in each period will be based on the decisions that you and others in your group make.

In this game, you will choose how much effort to put toward a task. The amount of effort that you choose, in conjunction with the effort choosen by others in your group, will determine your ranking within the group. Your earning will be determined by the value for achieving your ranking minus the cost in achieving the ranking.

Determination of Earning

The amount of money you will make in each game is:

Earning = Value from achieving a ranking - Cost in achieving the ranking

Value from achieving a ranking

The table below shows the value associated with each ranking:

Table B.1 Value Table

Ranking	Status Category	Value
1	Top	8 francs
2	Bottom	4 francs
3	Bottom	4 francs
4	Bottom	4 francs

There are two status categories: top and bottom. There is only one spot in the top category, and three spots in the bottom category. The person with the highest effort will be in the top category and earn a value of 8 francs. The other three people will be in the bottom category and each will earn a value of 4 francs.

If more than one person choose the same highest effort level, the experimenter will randomly break the tie using a random device. There will never be a tied in the top category.

Cost in achieving the ranking

The cost is the effort level you choose divided by your ability factor:

Cost in achieving a rank = Your Effort / Your Ability Factor

The effort that you choose must be a number between 0 and 700. You should choose your effort level in such a way that will make you the most amount of money. While it is true that the higher the effort level you choose, the more likely you are to be in the top category and earn the higher earning, you should also keep in mind that you will incurred a higher cost. Thus, the higher the effort level you choose, the more likely you are to get a negative net earning. In choosing your effort level, you need to take both the benefit and cost into consideration. It is possible to earn a negative earning, so you want to choose your effort level carefully!

Ability Factor

Everyone's ability factor is drawn from the following distribution:

$$\text{Cumulative Distribution of Ability Factor} = 1 - e^{-0.03x^{1.5}}$$
$$\text{Probability Distribution of Ability Factor} = 0.05e^{-0.03x^{1.5}}x^{0.5}$$

However, the realization for each draw might be different. A graph of the probability density function and the cumulative distribution function can be seen in Figure B.1 below.

What Figure B.1 illustrates is the likelihood of drawing different values of the ability factor. It is important to understand this because this will help you make better choices which in turn will help you make more money in this experiment. The likelihood of drawing an ability factor is important because your earning is determined by the effort you choose relative to others' effort level, and the effort level that you and others choose will be determined in part by the ability factors that you and

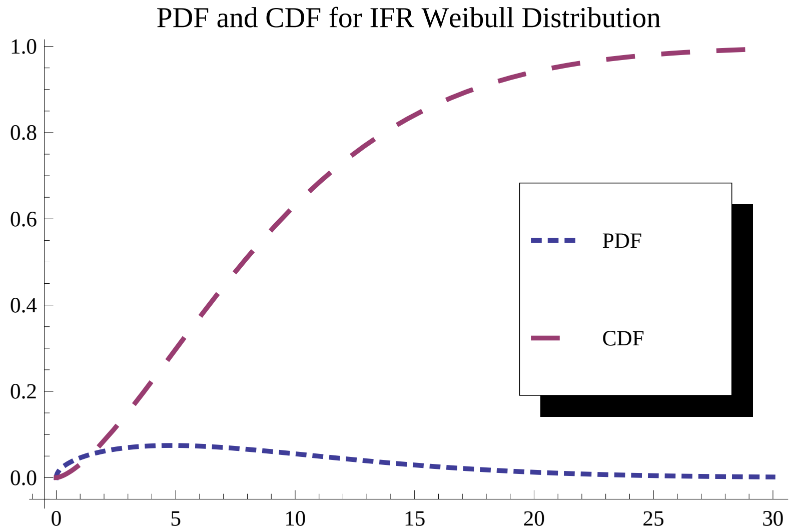


Figure B.1 The Probability Distribution of the Ability Factor.

others have drawn.

Table B.2 is another way to interpret the figure above. The table shows you how likely other people would have a lower ability value than you if you were to draw a specific value. For example, let us take a look at the first row of the table. If you draw an ability factor of 1 (as indicated by the left column), then 3.1% of the time, people will draw a lower ability factor than you (as indicated by the right column). Note that the higher the ability factor you draw, the more likely that other people will have a lower ability factor than you.

Timing and Matching of Players

You will play 20 periods. You will have 2 minutes in each period to make a decision. The ability values for everyone will be redrawn every period. In each period, the group members in your team will be randomly chosen from the pool of all players. In other words, each period we will randomly divide the 12 people into three groups of four people.

Table B.2 Cumulative Distribution Table of Ability Factor

Value of Your Ability Factor	How likely other people have a lower ability value
1	3.1%
2	8.6%
3	15%
4	22%
5	30%
6	37%
7	44%
8	51%
9	57%
10	63 %
11	68%
12	73%
13	77%
14	81%
15	84%
16	87%
17	89%
18	91%
19	93%
20	94%
21	95%
22	96%
23	97%
24	98%
25	98%
26	98%
27	99%
28	99%
29	99%
30	99%
...	.....
40	100%
50	100%
60	100%
70	100%
80	100%
90	100%
100	100%
...	.....
200	100%

Results

After everyone has submitted his or her effort level in each game, you will be told the status category you are in and your earning. You have 20 seconds to view this page. Remember, you can also view your earnings history at any time during the experiment.

Total Earnings from Experiment

Your total earnings will be your show up fee ($5 or ₣16.65) plus the money you earn from the experiment.

Should your total earnings (including the show up fee) become negative in a period, we will stop the experiment for everyone and you will leave the experiment with no earnings. Others will leave the experiment with the earnings they have made up to that period.

Remember, you can avoid negative earning if you choose your effort level carefully by taking into consideration how likely it is for you to achieve a ranking and the cost of achieving a ranking. The experimenter has provided you with a "What-If Scenario Analyzer" that will help you determine your potential earnings. The "What-If Scenario Analyzer" will be explained in more details below.

User Interface

A screenshot of the user interface can be found in Figure B.2. The main interface is divided into three areas: the place you can enter your effort level on the upper-left hand corner; the "What-If Scenario Analyzer" on the upper-right hand corner; and the history panel on the lower part of the screen. We will go through each of the three sections in turn.

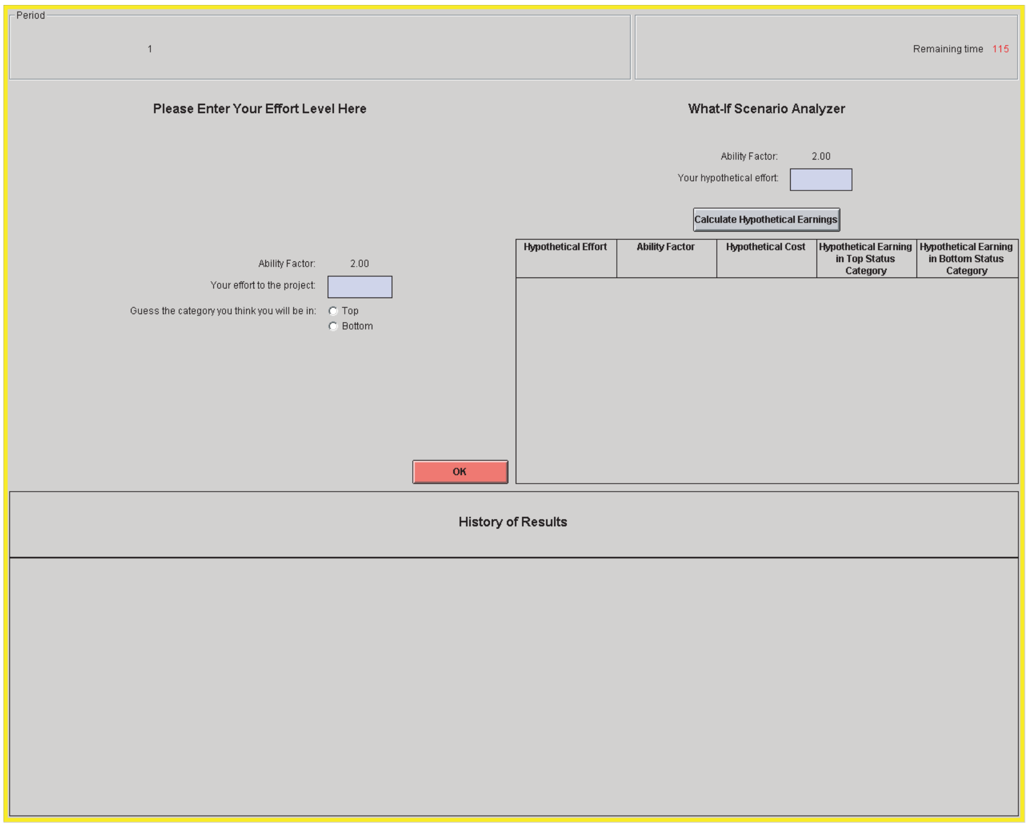


Figure B.2 The User Interface.

Entering Effort Level

In this panel, you will be shown the ability factor that has been drawn for you for this period. You will in turn need to make a decision on how much effort you want to choose. In addition, you will be asked to pick the status category that you think you will achieve given the effort level that you have picked. Note that once you click on the “OK” button, then you are done for the period. Your earnings will not be affected by your guess of the status category.

In order to help you make your decision, we have also included a “What-If Scenario Analyzer” that is described in the next section. Note that you are free to use the analyzer to try out different effort levels before you decide on the amount of effort you want to contribute for the period.

“What-If Scenario Analyzer”

The “What-If Scenario Analyzer” is provided as a tool for you to make better decisions. On the top of this panel, you will be shown the ability factor that have been drawn for you. You can then enter a hypothetical effort in the input box. After you click on the “Calculate Hypothetical Earnings” button, the table below will display the earning that you will make for the effort level that you inputted in the two different status categories. Note that these earnings are hypothetical earnings, and you do not know which ranking you will achieve. What the hypothetical earnings tell you is the range of earnings that you can achieve, but you need to also consider how likely you will be able to achieve a given status category.

You can use the analyzer as many times as you like before you make your actual decision.

History of Results

In this panel, you will be able to look at the history of the choices that you made, the status categories that you achieved, and your earnings for the previous periods.

Player ID: ________________
Computer #: ________

Review Questions:

Q.1 How many times will you be assigned an ability factor?

a) once every period
b) once every 2 periods
c) once every 4 periods
d) once every 30 periods
e) randomly decided

Q.2 Which of the values below cannot be a valid ability factor?

a) -1
b) 0
c) 0.7
d) 4
e) 50

Q.3 Is it more likely to get an ability factor between 0 and 1 or between 1 and 2? Hint: You may want to consult Table B.2 or Figure B.1.

a) between 0 and 1
b) between 1 and 2

Q.4 If your ability level is 2, and you choose an effort level of 10, calculate your possible range of earnings for the different ranks.

Ranking	Status Category	Value	Cost	Utility
1	Top	8		
2	Bottom	4		
3	Bottom	4		
4	Bottom	4		

Q.5 Under which of the following scenerio will you go bankrupt and that you will leave the experiment with no earnings at all (i.e. you lose your show up fee)?

a) money earn from the experiment (excluding the show up fee) is 20 francs
b) money earn from the experiment (excluding the show up fee) is 10 francs
c) money earn from the experiment (excluding the show up fee) is 0 francs
d) money earn from the experiment (excluding the show up fee) is -10 francs
e) money earn from the experiment (excluding the show up fee) is -20 francs

Appendix C

Contest Laboratory Experiment Figures

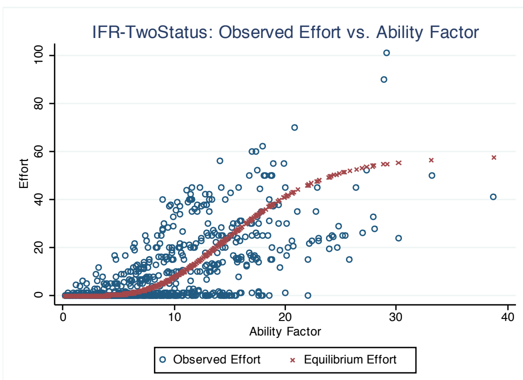


Figure C.1 IFR-TwoStatus: Observed Effort vs. Ability Factor.

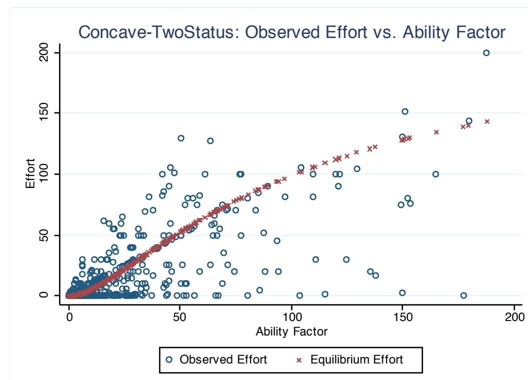


Figure C.2 Concave-TwoStatus: Observed Effort vs Ability Factor.

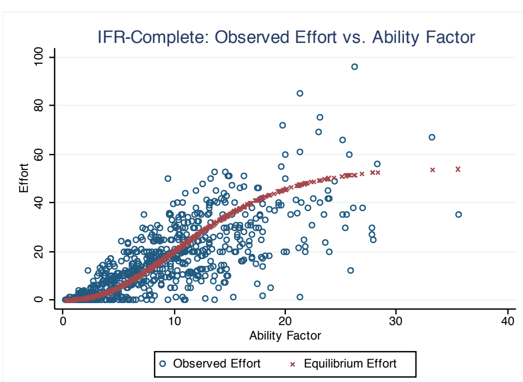


Figure C.3 IFR-Complete: Observed Effort vs Ability Factor.

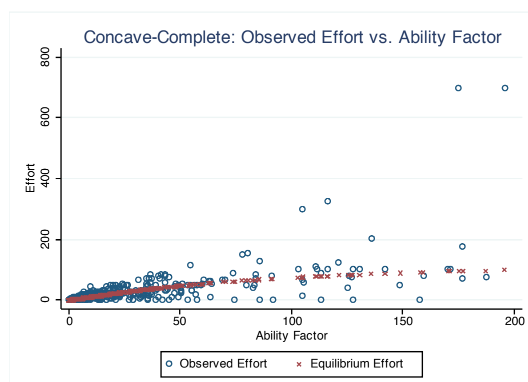


Figure C.4 Concave-Complete: Observed Effort vs Ability Factor.

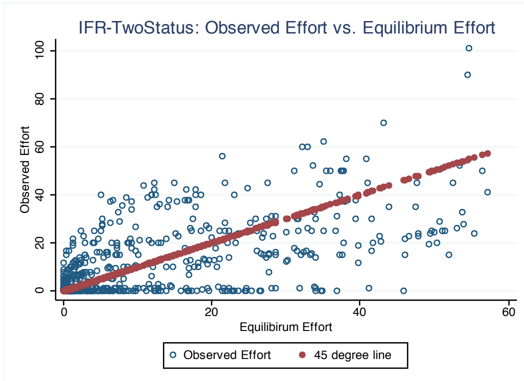


Figure C.5 IFR-TwoStatus: Observed Effort vs. Equilibrium Effort.

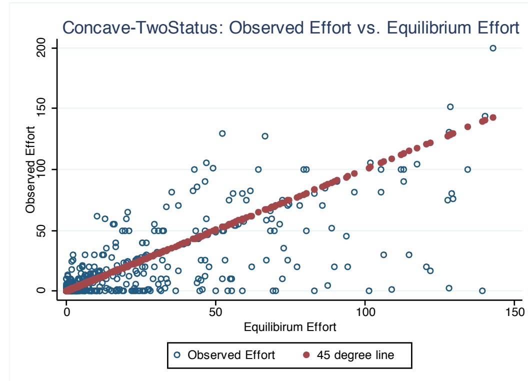


Figure C.6 Concave-TwoStatus: Observed Effort vs. Equilibrium Effort.

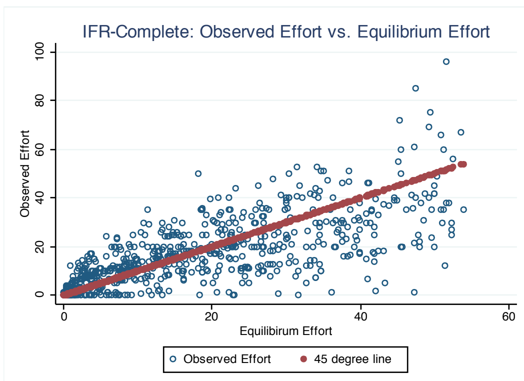


Figure C.7 IFR-Complete: Observed Effort vs. Equilibrium Effort.

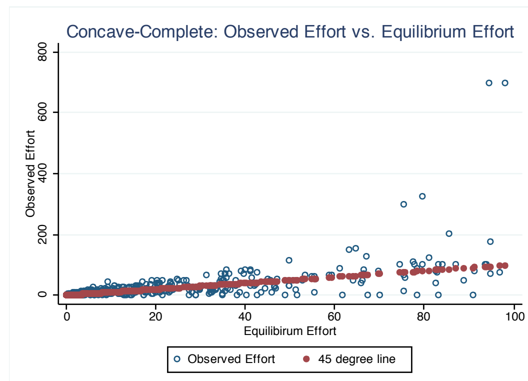
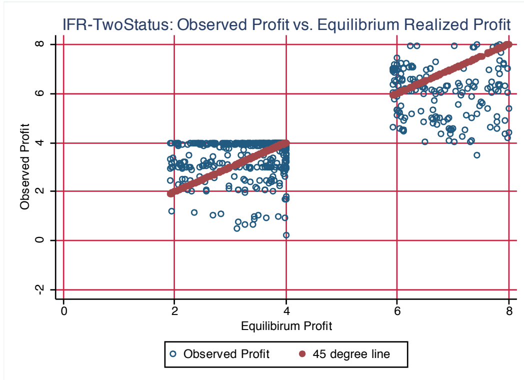


Figure C.8 Concave-Complete: Observed Effort vs. Equilibrium Effort.



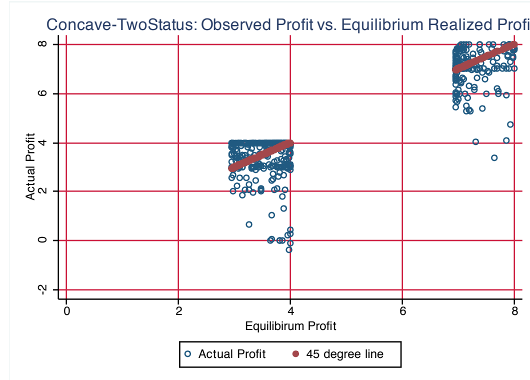
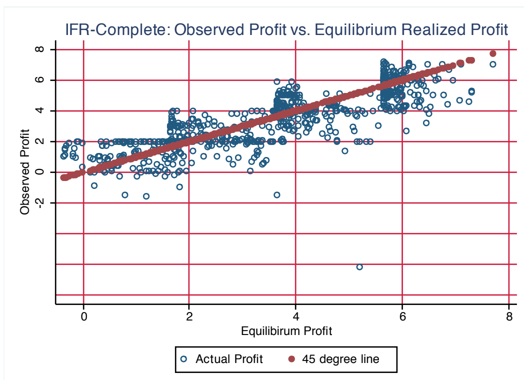


Figure C.9 IFR-TwoStatus: Observed Profit vs. Equilibrium Profit.

Figure C.10 Concave-TwoStatus: Observed Profit vs. Equilibrium Profit.



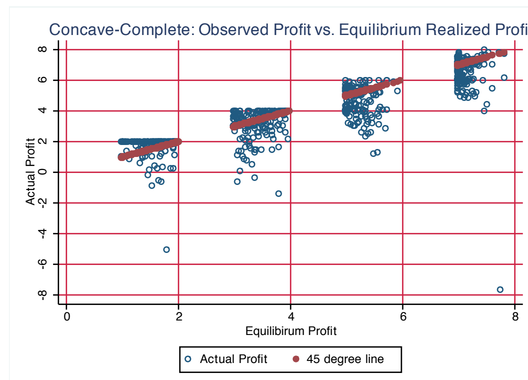


Figure C.11 IFR-Complete: Observed Profit vs. Equilibrium Profit.

Figure C.12 Concave-Complete: Observed Profit vs. Equilibrium Profit.

Appendix D

Tables

Table D.1 Missing Ranking Data By Date

Ranking Date	# of Missing Records Classic	# of Missing Records New
2010-2-19	10	0
2010-3-3	0	10
2010-3-7	4630	4640
2010-3-12	20	30
2010-3-13	10	20
2010-3-14	10	20
2010-3-15	10	20
2010-3-16	10	20
2010-3-17	10	20
2010-3-18	10	20
2010-3-19	10	20
2010-3-20	10	20
2010-3-21	10	20
2010-3-22	10	20
2010-3-23	10	0
2010-3-24	2110	2110
2010-3-25	10	20
2010-3-26	20	20
2010-3-27	20	20
2010-3-28	10	20
2010-3-29	10	20
2010-3-30	10	20
2010-3-31	9460	9470
2010-4-1	10	20
2010-4-2	10	20
2010-4-3	10	20
2010-4-4	10	20
2010-4-5	10	20
2010-4-6	20	20
2010-4-7	20	20
2010-4-8	20	20
2010-4-9	30	20

2010-4-10	10	20
2010-4-11	10	20
2010-4-12	10	20
2010-4-13	20	20
2010-4-14	10	20
2010-4-15	10	20
2010-4-16	10	20
2010-4-17	10	20
2010-4-18	10	20
2010-4-19	10	20
2010-4-20	10	20
2010-4-21	10	20
2010-4-22	10	20
2010-4-23	10	20
2010-4-24	20	20
2010-4-25	10	20
2010-4-26	10	20
2010-4-27	10	20
2010-4-28	20	20
2010-4-29	10	20
2010-4-30	10	20
2010-5-1	10	20
2010-5-2	20	20
2010-5-3	20	20
2010-5-4	10	20
2010-5-5	10	20
2010-5-6	10	20
2010-5-7	10	20
2010-5-8	10	20
2010-5-9	10	20
2010-5-10	20	20
2010-5-11	10	20
2010-5-12	20	30
2010-5-13	10	20
2010-5-14	10	20
2010-5-15	10	20

2010-5-16	10	20
2010-5-17	10	20
2010-5-18	20	20
2010-5-19	10	20
2010-5-20	10	20
2010-5-21	10	20
2010-5-22	10	30
2010-5-23	10	30
2010-5-24	20	20
2010-5-25	10	30
2010-5-26	10	20
2010-5-27	9770	9780
2010-5-28	10	20
2010-5-29	10	20
2010-5-30	10	20
2010-5-31	10	20
2010-6-1	20	20
2010-6-2	10	20
2010-6-3	10	20
2010-6-4	20	20
2010-6-5	10	20
2010-6-6	10	20
2010-6-7	10	20
2010-6-8	10	20
2010-6-9	30	20
2010-6-10	10	20
2010-6-11	20	20
2010-6-12	10	20
2010-6-13	10	20
2010-6-14	10	20
2010-6-15	10	20
2010-6-16	20	20
2010-6-17	10	20
2010-6-18	60	30
2010-6-19	10	20
2010-6-20	10	20

2010-6-21	20	20
2010-6-22	10	20
2010-6-23	10	20
2010-6-24	10	20
2010-6-25	10	20
2010-6-26	10	20
2010-6-27	10	20
2010-6-28	10	20
2010-6-29	20	20
2010-6-30	10	20

Table D.2 Missing Ranking Data By Range

Ranking Range (23 means ranks 231 to 240 are missing)	# of Missing Records Classic	# of Missing Records New
23	0	10
25	10	0
52	10	0
54	0	10
91	10	0
120	0	10
130	10	0
152	10	0
175	10	0
182	0	10
188	10	0
204	10	0
228	10	0
280	10	0
282	10	0
284	10	0
347	10	0
393	10	0
406	10	0
439	10	0
452	10	0
461	10	0
464	10	0
474	10	0
483	10	0
489	10	0
500	10	0
527	0	10
537	0	10
637	10	0
640	10	0
667	10	0

671	0	20
672	0	20
673	10	0
685	0	30
689	0	30
692	10	0
695	0	20
700	0	30
704	0	10
705	0	30
711	0	20
712	0	30
716	0	60
717	0	50
719	0	30
721	0	30
722	0	30
723	0	30
724	0	30
725	0	30
726	0	30
727	0	60
729	0	70
730	0	20
731	0	20
734	0	10
736	0	30
737	0	60
738	0	30
739	0	30
741	0	30
742	0	60
743	0	70
745	0	40
747	0	10
748	0	50

751	0	90
755	0	30
766	0	30
768	0	50
791	10	0
799	20	0
800	40	0
801	10	0
802	10	0
803	70	0
804	40	0
805	60	0
806	80	0
807	150	0
808	90	0
811	60	60
812	70	70
813	60	30
814	40	0
815	50	0
816	10	0
817	80	30
818	40	60
820	20	0
822	30	60
823	50	30
824	10	0
825	0	20
835	0	30
836	10	0
837	0	30
838	0	30
840	0	30
842	0	40
843	0	50
844	10	0

846	0	30
853	0	10
854	0	30
858	10	0
876	0	30
885	0	30
887	0	30
893	0	90
895	0	30
921	0	10
923	0	10
974	10	0

Bibliography

Mohammed Abdellaoui. Parameter-free elicitation of utility and probability weighting functions. Management Science, 46(11):1497–1512, 2000.

E. Amann and W Leininger. Auctions and contests: An experimental investigation. Wirtschaftstheoretische Diskussionsbeiträge, Nr.98-06, Essen University, 1998.

James Andreoni and William T. Harbaugh. Unexpected utility: Experimental tests of five key questions about preferences over risk. April 2010.

Lawrence M. Ausubel and Paul R. Milgrom. The lovely but lonely Vickrey auction. In Peter Cramton, Yoav Shoham, and Richard Steinberg, editors, Combinatorial Auctions. MIT Press, Cambridge, 2006.

Jeffrey S. Banks, John O. Ledyard, and David P. Porter. Allocating uncertain and unresponsive resources: An experimental approach. Rand Journal of Economics, 20:1–25, 1989.

Yasar Barut, Dan Kovenock, and Charles N. Noussair. A comparison of multiple-unit all-pay and winner-pay auctions under incomplete information. International Economic Review, 43(3):675–708, August 2002.

Christoph Brunner, Jacob K. Goeree, Charles A. Holt, and John O. Ledyard. Combinatorial auctioneering. Caltech Working Paper, 2006.

Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hiearchy model of games. The Quarterly Journal of Economics, pages 861–898, 2004.

Estelle Cantillon and Martin Pesendorfer. Auctioning bus routes: The London experience. In Peter Cramton, Yoav Shoham, and Richard Steinberg, editors, Combinatorial Auctions. MIT Press, Cambridge, 2006.

Peiyu Chen, Samita Dhanasobhon, and Michael D. Smith. All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon.com. Carnegie Mellon University, July 2006.

Yan Chen and Yuri Khoroshilov. Learning under limited information. Games and Economic Behavior, 44:1–25, 2003.

Yan Chen and Kan Takeuchi. Multi-object auctions with package bidding: An experimental comparison of iBEA and Vickrey. University of Michigan Manuscript, 2005.

Yan Chen, Max Harper, Joseph Konstan, and Sherry Xin Li. Social comparisons and contributions to online communities: A field experiment on movielens. American Economic Review, forthcoming, 2010.

Judith A. Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. American Marketing Association, 2006.

Edward H. Clarke. Multipart pricing of public goods. Public Choice, 11:17–33, 1971.

Cristian Danescu-Niculescu-Mizil, Gueorgi Kossinets, Jon Kleinberg, and Lillian Lee. How opinions are received by online communities: A case study on amazon.com helpfulness votes. In World Wide Web Conference. ACM, 2009.

Douglas D. Davis and Robert J. Reilly. Do too many cooks always spoil the stew? an experimental analysis of rent-seeking and the role of a strategic buyer. Public Choice, 95:89–115, 1998.

Pradeep Dubey and John Geanakoplos. Grading exams: 100, 99, ...1 or a, b, c? incentives in games of status. Cowles Foundation Discussion Paper No. 1467., June 2004.

Thomas A. Finholt. Collaboratories. In B. Cronin, editor, Annual Review of Information Science and Technology, volume 36, pages 74–107, Washington, D.C., 2002. American Society for Information Science.

Urs Fischbacher. z-tree: A toolbox for readymade economic experiments - experimenter's manual. Experimental Economics, 10(2):171–178, 2007.

Bruno S. Frey and Stephan Meier. Social comparisons and pro-social behavior - testing 'conditional cooperation' in a field experiment. American Economic Review, 94:1717–1722, December 2004.

Qiang Fu and Jingfeng Lu. The beauty of "bigness": On optimal design of multi-winner contests. Games and Economic Behavior, 66:146–161, 2009.

Anindya Ghose and Panagiotis G. Ipeirotis. Designing novel review ranking systems: Predicting the usefulness and impact of reviews. In International Conference on Electronic Commerce. ACM, 2007.

Uri Gneezy and Rann Smorodinsky. All-pay auctions – an experimental study. Journal of Economic Behavior & Organization, 61(2):255–275, 2006.

Theodore Groves. Incentives in teams. Econometrica, 41(4):617–31, 1973.

Warren O. Hagstrom. Scientific Community. Basic Books, New York, 1965.

R. Mark Isaac and Duncan James. Robustness of the incentive compatible combinatorial auction. Experimental Economics, 3(1):31–53, June 2000.

John H. Kagel. Auctions: A survey of experimental research. In J. Kagel and A. Roth, editors, Handbook of Experimental Economics. Princeton University Press, Princeton, New Jersey, 1995.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. Ecoometrica, 47(2):263–292, 1979.

Ajay Kalra and Mengze Shi. Designing optimal sales contests: A theoretical perspective. Management Science, 20(2):170–193, 2001.

Elena Katok and Alvin E. Roth. Auctions of homogeneous goods with increasing returns: Experimental comparison of alternative Dutch auctions. Management Science, 50(8):1044–1063, August 2004.

Anthony M. Kwasnica, John O. Ledyard, David Porter, and Christine DeMartini. A new and improved design for multi-object iterative auctions. Management Science, 51(3):419–434, March 2005.

John O. Ledyard, Charles Noussair, and David Porter. The allocation of a shared resource within an organization. Economic Design, 2(1):163–192, December 1996.

John O. Ledyard, Mark Olson, David P. Porter, Joseph A. Swanson, and David Torma. The first use of a combined-value auction for transportation services. Interfaces, 32(5):4–12, 2002.

Paul Milgrom. Package auctions and exchanges. Econometrica, 75(4):935–965, 2007.

Paul R. Milgrom. Putting Auction Theory to Work. Cambridge University Press, New York, 2004.

Benny Moldovanu and Aner Sela. The optimal allocation of prizes in contests. The American Economic Review, 91(3):542–558, January 2001.

Benny Moldovanu and Aner Sela. Contest architecture. Journal of Economic Theory, 126:70–96, 2006.

Benny Moldovanu, Aner Sela, and Xianwen Shi. Contest for status. Journal of Political Economy, 115(2):338–363, 2007.

Wieland Müller and Andrew Schotter. Workacholics and drop outs in optimal organizations. C.E.S.S. Working Paper Series, Working Paper #2007-07, New York University, January 2007.

Mark Olson and David Porter. An experimental examination into design of decentralized methods to solve the assignment problem with and without money. Economic Theory, 4:11–40, 1994.

Jan Potters, Casper G. de Vries, and Frans van Winden. An experimental examination of rational rent-seeking. European Journal of Political Economy, 14:783–800, 1998.

Paul A. Samuelson. The pure theory of public expenditure. Review of Economics and Statistics, 36(4):387–389, 1954.

Roman M. Sheremeta. Contest design: An experimental investigation. Purdue University Manuscript, 2008.

S.T. Traweek. Beamtimes and lifetimes: The world of high energy physics. Harvard University Press, Cambridge, MA, 1992.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5:297–323, 1992.

William S. Vickrey. Counterspeculation, auctions and competitive sealed tenders. Journal of Finance, 16:8–37, 1961.

Wm A. Wulf. The collaboratory opportunity. Science, 261(5123):854–855, 1993.