Submitted to *Management Science*
manuscript (Please, provide the mansucript number!)

Shake It Up Baby: Scheduling with Package Auctions

Kan Takeuchi
Department of Economics, University of Michigan, Ann Arbor, Michigan 48109, ktakeuch@umich.edu,

John C. Lin
School of Information, University of Michigan, Ann Arbor, Michigan 48109, johnclin@umich.edu,

Yan Chen
School of Information, University of Michigan, Ann Arbor, Michigan 48109, yanchen@umich.edu,

Thomas A. Finholt
School of Information, University of Michigan, Ann Arbor, Michigan 48109, finholt@umich.edu,

In many disciplines success depends on access to scarce resources, such as unique instruments. Research on computer-supported cooperative work (CSCW) has contributed to the development of technologies, such as collaboratories, that broaden access to scarce scientific and engineering resources. While broader access is often applauded, little attention has been focused on the problem of equitable and efficient resource allocation in the face of increased demand created through collaboratory use. This paper, then, applies concepts from the economic discipline of mechanism design to compare different resource allocation schemes (RAD, Vickery, and knapsack) within a hypothetical collaboratory. Experimental results show that knapsack achieves a more equitable distribution of resources than RAD or Vickery, but that RAD and Vickery are both more efficient than knapsack. The findings highlight the need for systematic exploration of allocation mechanisms within collaboratories, where simple optimization (e.g., knapsack) is likely to produce a sub-optimal match of resources to needs. More generally, the findings illustrate the utility of economic approaches in understanding issues that emerge in large-scale collaborations, such as entire scientific and engineering communities, that have not typically been the subject of CSCW research.

History: January 5, 2007 version

1. Introduction

The National Science Foundation (NSF) and other federal agencies are making significant investments in expanding the ability of geographically distributed groups of scientists to conduct research via the Internet (e.g. Atkins et al. 2003). One means to enhance the work of dispersed teams is the "collaboratory." First proposed in the late eighties, a collaboratory is a center without walls, in which researchers can perform their research without regard to physical location interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries (Wulf 1993). The collaboratory idea is a descendent of earlier notions, such as Vannevar Bush's 1945 proposal for the "memex;" Douglas Engelbart's 1968 demonstration of graphical user-interfaces for computer-supported meetings; and the creation in 1969 of the ARPAnet. There are currently several hundred collaboratories in use within multiple scientific communities, including space physics, medicine, software engineering, and neuroscience (Finholt 2002).

The George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) is a ten-year NSF program focused on accelerating and improving earthquake engineering research through the use of a collaboratory (it.nees.org). The NEES collaboratory connects earthquake engineering researchers to sixteen state-of-the-art laboratory facilities distributed around the United States. These facilities were built with the explicit intent of combining their capabilities through the Internet. For example, users of the NEES collaboratory are able to collect data across instrument modalities spanning key earthquake engineering sub-disciplines. In this approach a test of a bridge pier system might include performance data on: a) the soil the pier sits on (e.g., from physical geotechnical simulations run using specimens in a centrifuge); b) the pier structure (e.g., from physical structural simulations run using specimens on a shake table); and c) the bridge deck (e.g., from numerical simulations). Through NEESit the results of collaboratory studies will be accessible through shared data repositories, remote participants will be able to observe experiments, and under highly controlled circumstances remote participants will be able to 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 diminish the advantages of physical collocation. Specifically, proximity to unique instruments has frequently created differential access, both to the instruments and to the community of scientists who use the instruments (Hagstrom 1965, Traweek 1992). In NEES, a critical goal of making the sixteen new sites accessible over the Internet is to broaden use of the facilities, particularly among researchers at institutions that lack earthquake engineering research equipment. To ensure that broader use occurs, NSF stipulated the creation of a non-profit corporation, NEES Inc. (nees.org), as an additional element of the NEES program intended to operate the NEES collaboratory over the period 2004 through 2014.

NEES Inc. has responsibility for governing the use of NEES facilities, as well as for allocating maintenance and operation funds. A central concern of NEES operations is to ensure shared use of the sixteen new laboratories. Specifically, a special committee of the NEES Inc. board of directors has oversight of shared use policies and mediates disputes arising from conflicts in shared use. Unfortunately, beyond the requirement that NEES accomodates shared use, NSF has not offered specific guidelines about how to achieve this goal. At an abstract level, the problem confronting NEES Inc. is similar to a class of problems in economics described as mechanism design. That is, NEES Inc. controls instrument time designated for shared use at the sixteen NEES laboratories (i.e., use by researchers in the broader earthquake engineering community, not just those collocated with the NEES equipment). This instrument time, presumably, has value for researchers, such that demand for instrument time will exceed supply. For example, in astronomy there is intense competition for telescope time (**?**Olson and Porter 1994).

Understanding how to achieve efficient and equitable allocation of scarce resources is a key focus of research on mechanism design. For instance, allocation solutions in other domains (e.g., the sale of frequency spectrum by the Federal Communications Commission) involve various forms of auctions. There are two reasons why it is important to explore similar approaches for resource allocation in collaboratories. First, in the absence of mechanisms it seems likely that collaboratory users will default to allocation through a scheduling czar or by insiders (e.g., instrument owners). This approach runs the risk that facilities will be under-utilized and that allocation decisions may be subject to bias (e.g., political influences). Second, absent any specific allocation mechanisms, collaboratory users may arrive at naive pricing schemes for resources, such as flat unit costs versus costs that reflect actual demand (and hence value). For example, the unit cost assumption breaks down when a user places a premium on allocation of contiguous operating intervals (i.e., ten hours together versus ten one hour segments dispersed over several days).

Collaboratory operators, such as NEES Inc., will be the immediate beneficiaries of an effort to design and demonstrate mechanisms for allocation of instrument time within collaboratories. However, more generally, research into the problem of efficient and equitable resource allocation within large-scale collaborations is important. For instance, if the NEES collaboratory is a precursor to the organization of infrastructure across a wide variety of scientific and engineering communities then the problems confronted by NEES are likely to emerge in these other settings. That is, if sharing of instruments is to become the norm, then there must be a corresponding set of mechanisms to support this sharing. It seems unlikely that simple altruism, particularly in highly competitive scientific and engineering fields, will yield satisfactory solutions. Instead, the time is right to conduct research that will produce a toolbox of candidate mechanisms, with proven characteristics, so that when communities confront breakdowns or bottlenecks in the use of instruments, or any shared resource, these difficulties can be overcome quickly or avoided completely.

The rest of the paper is divided into four sections. In the first of these sections we identify three allocation mechanisms and describe their features. In the next section we describe our experimental design including our assumptions about the hypothetical collaboratory. The next section presents our results and in the final section we discuss the results.

2. Allocation Mechanisms

In this section we outline the technical details of three allocation mechanisms. 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. For example, 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. Therefore, package auctions might be an important mechanism in achieving efficient allocation of equipment time. We consider two package auction mechanisms, Vickrey and RAD, compared with an ordinal ranking mechanism, knapsack, selected as a best-case representation of how allocation is currently accomplished.

2.1. The Vickery Auction

The Vickrey auction (Vickrey 1961, Clarke 1971, Groves 1973) is an important standard for nearly all mechanism design work, and for auctions in particular. It is dominant strategy incentive compatible, i.e., bidding one's true valuation is always optimal regardless of others' strategies. Furthermore, it implements the efficient outcome.

A Vickrey auction with package bidding is an extension of the more familiar second-price auction. 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.

Next, once all bidders have submitted their bids, the auctioneer will choose the combination of submitted bids that yields the highest sum of bids. The set of bidders winning a package are the winning bidders.

After determining the winning bidders, the auctioneer then, one at a time, chooses each winning bidder as a pivotal bidder. The auctioneer examines the bids again, but ignores the bids of the pivotal bidder. The auctioneer determines the allocation of goods that maximizes the sum of bids, using the same rules as before, but not considering any bids placed by the pivotal bidder. Once this new allocation has been determined, the auctioneer compares the sum of bids generated by this allocation with those generated when no bids are excluded.

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

Previous laboratory experiments indicate that, in environments with a small number of packages, the Vickrey auction can achieve high allocation efficiency in 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 excellent performance in more complex environments. This study addresses this issue by expanding the number of packages to a more realistic level.

2.2. Resource Allocation Design (RAD)

The RAD mechanism is an iterative ascending bid package auction proposed by Kwasnica et al. (Kwasnica et al. 2005). The RAD mechanism borrowed features from the AUSM mechanism (Banks et al. 1989) and the Milgrom FCC Design (Milgrom 2004). The features that RAD borrowed from the Milgrom FCC Design are: the auction is iterative instead of one shot; the use of an eligibility rule; and a price improvement rule. The feature that RAD borrowed from AUSM is the use of a package bidding language. RAD, in addition, calculates the prices on the individual months that underlie the packages. RAD is shown to outperform both parents (Kwasnica et al. 2005).

The details of the RAD mechanism are explained in (Kwasnica et al. 2005). Here, we will briefly explain the different features on a global level. An iterative auction is one where the bidding activities are separated into different rounds. Within a round, bidders can place bids. At the end of a round, bidders are given feedback information that can be used before bids are placed in the next round. In the case of RAD, bidders are told at the end of a round if the bids they placed were provisionally winning or losing. At the end of the last round, there is no provisional winning bid. A winning bid at the end of the last round is the binding winning bid.

The effect of the eligibility rule can be briefly summarized as a "use it or lose it rule". Roughly, the more packages a bidder bids on, the higher is the eligibility number for that round. In the subsequent round, the number of packages a bidder can bid on depends on the eligibility number from the previous round. The higher the eligibility number, the more packages a bidder can bid on.

The price improvement rule specifies a minimum price for each package based on the bids that were submitted in the previous round and a price improvement factor. The price improvement rule, in conjunction with the eligibility rule, helped to drive an auction to a close.

The difference between RAD and the FCC Design is RAD allows the placement of package bids. In the hypothetical collaboratory context, this means the ability to say a bidder wants a three month package starting in the first month, or a two month package starting in the second month, but not both. Note that the bidder who placed the above two bids can only win at most one of the two placed packages, not both. Furthermore, if a bidder wins a package, the bidder is guaranteed to win the complete package, not part of a package. With the FCC design, bidders can only place bids on single months. So even though bidders can place bids that are equivalent to a two month package starting in the second month, it is possible for the bidder to win only part of the package, namely, the second or the third month, instead of the package containing both months. Thus bidders can be exposed to the risk that they may overpay. This is known as the "exposure problem" in the economic literature.

Lastly, the novel aspect of RAD is its ability to calculate prices for the underlying months that the packages are constructed out of.

Single unit iterative ascending auctions, such as the English auction, tend to outperform their sealed bid counterpart in terms of efficiency. Therefore, in the package bidding context, we expect that RAD will generate higher efficiency than Vickrey.

2.3. Knapsack with Ordinal Ranking

In many scientific communities, allocation of instrument time can be determined by a scheduling czar, a committee (Olson and Porter 1994) or some formal or informal optimization procedure which uses the ordinal ranking information from potential instrument users. The knapsack mechanism is an idealized representation of the latter. In this mechanism, everyone submits ordinal rankings of packages, from the top choice (# 1) to the last choice (# n for n packages). The top choice is awarded n points, second choice $n-1$ points, ..., and the last choice 1 point. and the computer will allocate goods in such a way that maximizes the total number of points.

This class of knapsack mechanisms with ordinal ranking has two problems. First, truth-telling might not be a dominant strategy. One can easily construct examples to show that truthful ranking based on one's valuations for the packages might not be optimal. Second, bidders cannot express the intensity of their preferences. Therefore, we expect that the knapsack mechanism will generate lower efficiency than Vickrey and RAD. Because of the prevalence of knapsack-like mechanisms in scheduling, we use it as a useful benchmark of typical allocation approaches in order to evaluate the two auction mechanisms.

3. Method: Experimental Design

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

3.1. The hypothetical collaboratory

The hypothetical collaboratory was designed to capture the essential aspects of a scientific or engineering collaboratory, using the NEES collaboratory as a model. In some areas, simplifications were made to keep the problem tractable in the experimental setting, where subjects were only available for a limited amount of time in each session. Participants were told that they represented research groups seeking instrument time via a collaboratory within a 24-month window. We defined participants in terms of their resource level (i.e., small vs. big projects) and in terms of their time preference (i.e., early, indifferent, or late in the 24-month window). Each participant was asked to consider sixty-three (for a big project) or sixty-six (for a small project) packages of instrument time. The packages were defined by the starting month and the duration of the use. For example, a package that started on the first month and lasted for three months contained month 1, month 2, and month 3. For simplicity, packages were denoted in the form of an ordered pair (starting month, length of package) or (1,3) for the example above.

A small project bidder had values for packages that lasted for 2 to 4 months. A big project bidder had values for packages that lasted for 3 to 5 months. As a result, a small project bidder was able to bid on packages that started on month 1 to month 23 (i.e., the first number in the ordered pair is a number between 1 to 23), with lengths of 2, 3, or 4 (i.e., the second number in the ordered pair is a number between 2 to 4). Big project bidders were also able to bid on packages starting from month 1 to month 22 (i.e., the first number in the ordered pair is a number between 1 to 22), but with lengths of 3, 4, or 5 (i.e., the second number in the ordered pair is a number between 3 to 5). Note a big project bidder was bidding on sixty-three packages, while a small project bidder was bidding on sixty-six packages.

The value of a package was determined by the type and the time preference factor of a bidder, as well as a v parameter that specified the value of an additional month for the participant. As a first step, we determined the value of a base package, (1, 2) for a small project bidder and (1, 3) for a big project bidder. For a small project bidder the value of the base package was drawn from the uniform distribution [20, 100]. For a big project bidder the value of the base package was drawn from the uniform distribution [20, 150].

Once we determined base package values, we determined the value of packages that start in the first month. A v parameter was drawn from the uniform distribution of [10, 20] for all the participants. For example, for a small project bidder, where the value of the base package (1, 2) was 50 and the v drawn for that particular player was 15, the values for the packages (1, 3) and (1, 4) were 65 and 80, respectively. We determined the values for the other packages from the packages that start in the first month. As an illustration, the value of the (2, 2) package was the value of the (1, 2) package times the time preference factor. So if a bidder had the time preference factor of 1.2, the (2, 2) package was more valuable than the (1, 2) package (in general, this means the participant prefers packages that start later rather than earlier). If the participant had the time preference factor of 1, the (2, 2) package was worth the same as the (1, 2) package (in general, this means the participant was indifferent between the starting time of the package). If the participant had the time preference factor of 0.9, the (2, 2) package was less valuable than the (1, 2) package (in general, this means the participant prefers packages that start earlier rather than later).

In summary, the values of the packages were determined by the value of the base package, the time preference factor of the bidder, and the value of an additional month, beta. The value of the base package in turn was determined by the bidder's type.

We settled on nine bidders in each experimental session to mimic the number of projects funded to use the NEES collaboratory in the first year of operation (2004-05). Of these nine there was a distribution of big and small projects, which we operationalized as three big projects and six small projects. Of the three big project bidders all three time preference factors were represented among the bidders. Of the six small project bidders, one had the time preference of 1.2, two had the time

Bidder ID	Project Type	Time Preference	Package Size
1	Big	Prefer later	3 to 5 months
2	Big	Indifferent	3 to 5 months
3	Big	Prefer earlier	3 to 5 months
4	Small	Prefer later	2 to 4 months
5	Small	Indifferent	2 to 4 months
6	Small	Indifferent	2 to 4 months
7	Small	Prefer earlier	2 to 4 months
8	Small	Prefer earlier	2 to 4 months
9	Small	Prefer earlier	2 to 4 months

Table 1 Design Parameters: Bidder Preferences

preference of 1, and three had the time preference of 0.9. Earlier months are more competitive than later ones, to capture the stylized fact in most scientific communities that earlier experiment, and hence, potential discoveries, are more valuable to scientists. Table 1 summarizes the main feature of bidder preferences.

3.2. Experimental Procedures

Each experimental session required exactly nine participants. The participants were recruited from the University of Michigan, including both the graduate and undergraduate population. The participants had declared majors in the fields of science, math, or engineering. Once the participants arrived they reviewed and signed an informed consent form. At the beginning of each session, each participant was given printed instructions. Before the instructions were read aloud by one of the experimenters, the participants drew from a deck of index cards for a player ID that determined their project type (big or small) and their time preference factor (1.2, 1, or 0.9). The instructions were then read aloud. Participants were encouraged to ask questions during and after the experiment. The instruction period averaged around 30 minutes. Then participants took a quiz designed to test their understanding of the mechanism. At the end of the quiz, the experimenters collected, graded, and returned the quiz to each of the participants. The experimenters then reviewed the answers with the participants. The participants were given 10 minutes to do the quiz for the RAD treatment, and 7 minutes to do the quiz for the Vickrey treatment and the knapsack treatment.

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

Table 2 Features of Experimental Sessions

There were no practice auctions in any of the experimental conditions. In the Vickrey and knapsack condition, participants had seven minutes per auction to input their bids, while in the RAD condition, the participants had four minutes for the first round and two minutes for all subsequent rounds in each auction. There were a total of eight auctions in each of the Vickrey and knapsack sessions, and three to five auctions in each of the RAD sessions.[1]

At the conclusion of auctions in each condition participants tallied their cumulative earnings, filled out a short demographic survey, and wrote down the strategies that they used. We use the induced value method (Smith 1982). Participants were paid based on their experimental profits.

Table 2 presents the relevant features of the experimental sessions, including mechanisms, number of experimental sessions, the number of participants in each condition and the exchange rates. Overall, 15 independent computerized sessions were conducted in the RCGD lab at the University of Michigan from July 2005 to February 2006. No participant was used in more than one session, yielding a total of 135 participants across all treatments. Each session lasted approximately two and a half hours. In addition to their auction earnings, participants could win or lose money based on their quiz answers. A participant with fully correct answers gets up to $5. The average earning (including quiz award and a $10 showup fee) was $34.44, and the standard deviation was $16.21. Data are available from the authors upon request.

4. Results

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

[1] Recall that RAD is an iterative auction, therefore, the number of rounds in each auction is endogenous and varies from auction to auction.

4.1. Efficiency

To derive the efficiency for an auction outcome, we needed to find the *optimal allocation* of packages. At the optimal allocation, the sum of the value of allocated package that each participant receives is maximized. Then the efficiency is the ratio of realized total bidder value and maximum potential total bidder value. It is formally defined as follows,

$$\text{Efficiency} = \frac{\sum_i v_i(x_i) - \sum_i v_i(x_i^{SD})}{\sum_i v_i(x_i^*) - \sum_i v_i(x_i^{SD})} \tag{1}$$

where x_i is the package that is actually allocated to participant i in the outcome, x_i^* is the package that would be allocated in the optimal allocation and x_i^{SD} is the package that would be allocated from running serial dictatorship mechanism. [2] The efficiency is 100% if the auction results in the optimal allocation.

For example, suppose we have two packages, A and B, and two bidders, John and Tom. Package A is worth $30 to John and $10 to Tom, while package B is worth $20 to John and $30 to Tom. The efficient allocation is to give A to John and B to Tom, which yields a total value of $60 and 100% efficiency. If, for some reason, John gets B and Tom gets A, the total value is only $30, and the efficiency is 30/60=50%. Efficiency requires that the package goes to the bidder who values it the most.

Figure 1 presents the average time series efficiency and standard deviation (error bars) of the three mechanisms. Consistent with individual learning under the Vickrey auction, the average efficiency under Vickrey increased over time and was higher than knapsack. RAD also generated higher efficiency than knapsack. Note that RAD only had 3-5 auctions in each session, while both Vickrey and knapsack had eight auctions.[3] While RAD generates higher efficiency in the earlier auctions, Vickrey catches up by auction #4.

[2] The serial dictatorship allocation is calculated by averaging 10,000 trials per auction, based on the values of the packages drawn for each participant in each of the auctions.

[3] To avoid laying the error bars on top of each other, there is a slight shift for each mechanism, which has no numerical significance.

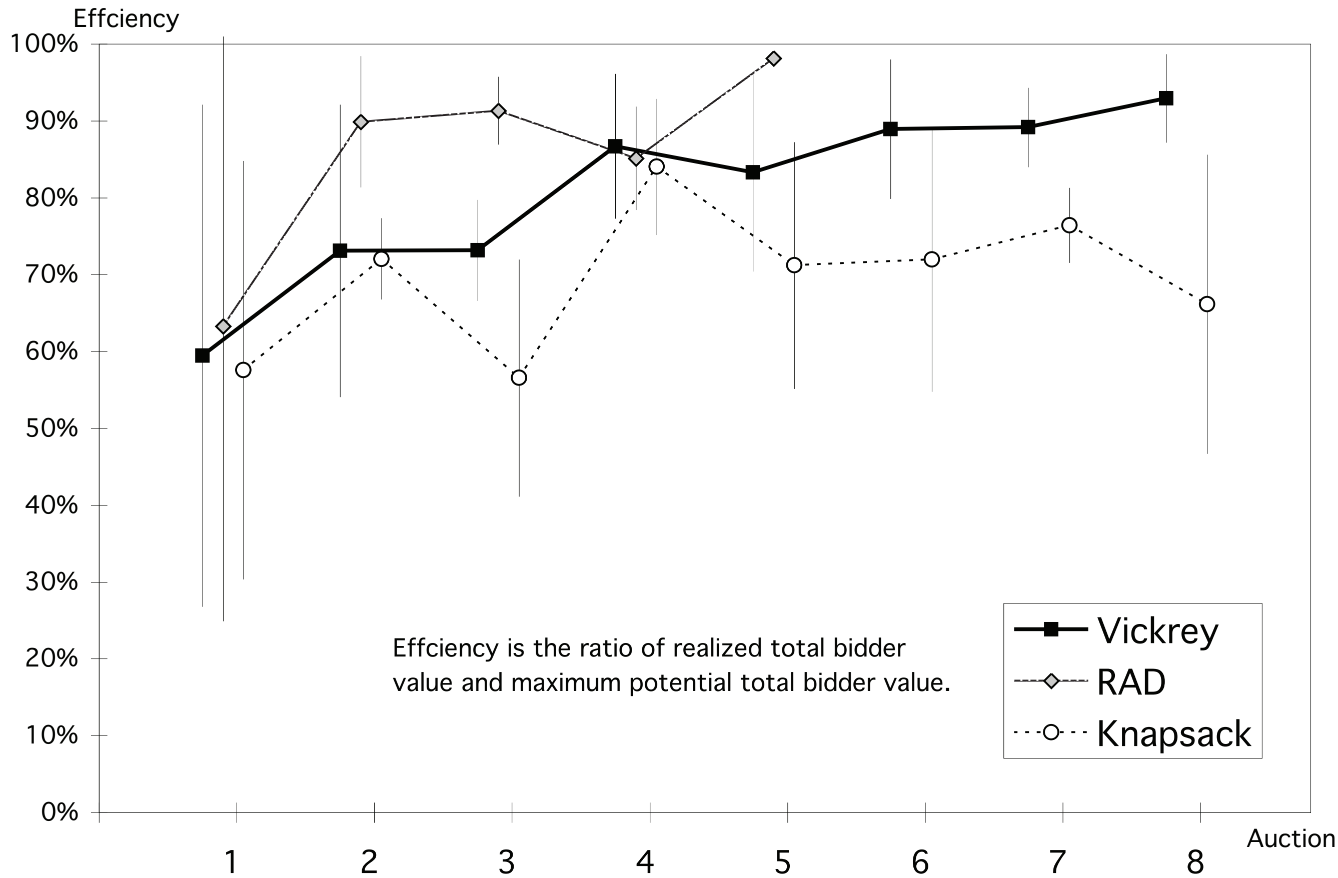

Figure 1 **Efficiency comparison across mechanisms**

Statistical tests confirmed our impressions from Figure 1. We computed the session average, average of the last two auctions in each session, as well as auction by auction efficiency comparisons across mechanisms. We used the permutation tests[4] to compare the performance of the mechanism, where the number of independent observations was the number of the sessions. For session average efficiency comparisons, we had the following results, where the p-value for each one-sided test is written under the inequality sign.

$$\text{Knapsack} \underset{0.04}{<} \text{Vickrey} \underset{0.38}{<} \text{RAD},$$

$$\text{Knapsack} \underset{\mathbf{0.05}}{<} \text{RAD}.$$

[4] The permutation test, also known as the Fisher randomization test, is a nonparametric version of a difference of two means t-test.

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

$$\text{Knapsack} \underset{\mathbf{0.01}}{<} \text{Vickrey} \underset{0.65}{<} \text{RAD},$$
$$\text{Knapsack} \underset{\mathbf{0.01}}{<} \text{RAD}.$$

We concluded that efficiency was significantly higher under RAD than under knapsack. In the last two auctions, Vickrey and RAD each generated significantly higher efficiency than knapsack. The efficiency comparison between RAD and Vickrey was insignificant.

4.2. Equity: Gini Coefficient

While the efficiency index measures whether time slots are allocated to the bidders who value them the most, we used the Gini coefficient to measure the distributional equality among bidders. The Gini coefficient is widely used to measure income inequality in a country and has been used by CSCW researchers to measure equality of participation in group discussions (e.g. **?**)). We derived the Gini coefficient for each auction outcome based on the value of the allocated package to each participant. Note that a higher Gini Coefficient corresponds to greater inequality. In the extreme cases, it is 0 if every participant gets the same value from the assigned package (perfect equality), while it is approximately 1 if one participant receives some months while other participants receive nothing (perfect inequality).

Figure 2 presents the average time series Gini coefficients and the standard deviation (error bars) for each of the three mechanisms. One striking feature is that knapsack was roughly the lower envelope of the observations, indicating more equitable distribution of time slots than the other two mechanisms. Permutation tests of session average and last two auction average across mechanisms confirmed this impression.

$$\text{Knapsack} \underset{\mathbf{0.01}}{<} \text{Vickrey} \underset{0.55}{<} \text{RAD},$$

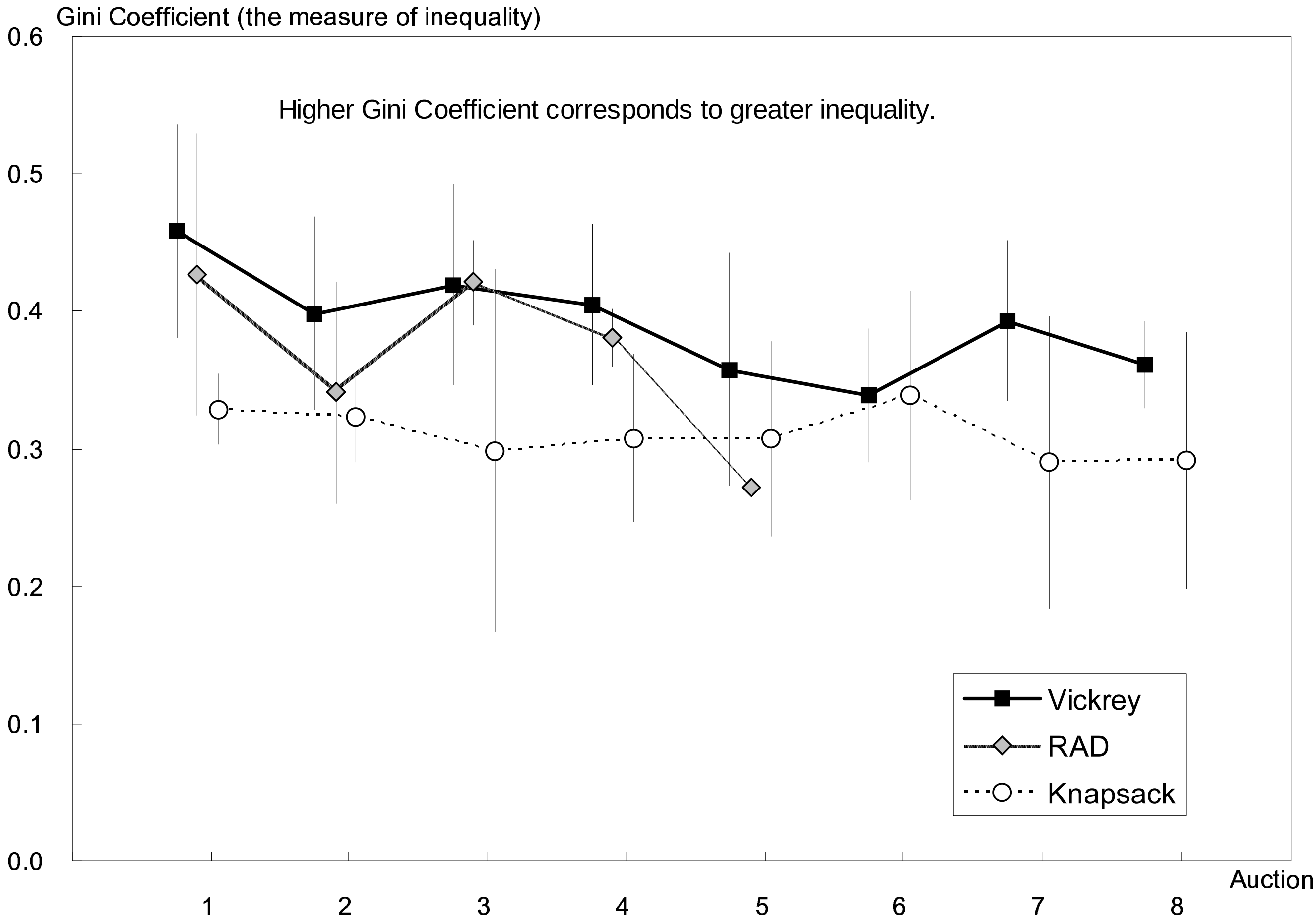

Figure 2 **Equity comparison across mechanisms**

$$\text{Knapsack} \underset{\mathbf{0.01}}{<} \text{RAD}.$$

Taking into account the effects of learning, we looked at the average Gini coefficient of the last two auctions in each session, and found the following result:

$$\text{Knapsack} \underset{0.03}{<} \text{Vickrey} \underset{\mathbf{0.27}}{<} \text{RAD},$$
$$\text{Knapsack} \underset{0.02}{<} \text{RAD}.$$

Therefore, we concluded that the Knapsack mechanism was significantly more equitable than the Vickrey or RAD mechanism. Vickrey, in turn, was weakly more equitable (i.e., at the 10% significance level) than RAD.

4.3. Bidder Earnings Comparison by Type

	Comparison by Package Size	Comparison by Time Preference		
	Big > Small	Indifferent > Early	Indifferent > Late	Late > Early
RAD	0.34	<0.01	<0.01	0.47
Vickrey	0.06	< 0.01	< 0.01	0.04
Knapsack	0.07	< 0.01	< 0.01	0.33

Table 3 p-values of Bidder Profit Comparison Based on Types

Another way to compare between the mechanisms is to look at the earnings by the bidders conditional on their types. Recalled that there are two dimensions in defining the types of the bidders: package size and time preference. Table 3 reports the p-values from running permutation tests that compare the profits earn by the bidders condition on their package preference and time preference. The results suggest that the types of bidders do not effect the bidder's earning in RAD, but do effect the bidders' earnings in Vickrey and Knapsack. In particular, big type bidders earn significantly more than small type bidders in Vickrey and Knapsack in comparison with RAD. In addition, bidders who are indifferent to the time that experiments are run earn significantly more than bidders who prefer to run their experiments earlier or later under Vickrey and Knapsack.

This is a curious and interesting empirical result that beckons more investigation. Is this result particular to the particular valuation environment explored in this paper, or would this result be robust across different valuation environments? Do the differences arise directly from the differences in the mechanisms, or do they come indirectly from how the mechanisms shape the bidding strategies of the bidders?

4.4. Active Bids in Vickrey and Knapsack and Bid-Value Ratio in Vickrey

We first examine two aspects of the bidding behavior in Vickrey, the proportion of active bids and the bid/value ratio for those active bids. Recall it is a weakly dominant strategy for each bidder to bid on all packages, and to bid their true value on each package.

The first two columns of Table 4 present the proportion of active bids and the bid/value ratio under the Vickrey auction. Bidders in a Vickrey auction, on average, bid 73.00% of their true value. Overall, both measures are smaller than the theory predicts. However, both the proportion

	Vickrey		Knapsack
	Active Bids%	Bid/Value	Active Bids%
Auction 1	0.291	0.551	0.510
Auction 2	0.434	0.574	0.591
Auction 3	0.441	0.582	0.580
Auction 4	0.480	0.708	0.597
Auction 5	0.499	0.726	0.599
Auction 6	0.537	0.734	0.585
Auction 7	0.565	0.742	0.580
Auction 8	0.587	1.038	0.585

Table 4 Bidding Under Vickrey and Knapsack

of active bids and the Bid/Value ratio are increasing, indicating that bidders were learning the weak dominant strategy as they repeated the auctions.

The third column of Table 4 presents the proportion of active bids under the knapsack mechanism. Unlike the Vickrey mechanism, the proportion of active bids remained in the $(0.5, 0.6)$ range during the eight auctions. Not bidding on every package can sometimes be an optimal strategy under knapsack.

4.5. Level of Under-Bidding, Truthful-Bidding, and Over-Bidding in Vickrey

We also classified Vickrey bidders into three categories: Under bidder, Truthful bidder and Over bidder. Specifically, we ran the following simple OLS regression on active bids for each bidder with robust clustering at the auction level.

$$\text{Bid}(x_i) = \beta \text{Value}(x_i) + \epsilon_i. \tag{2}$$

Then, we tested the null hypothesis: $\hat{\beta} = 1$. Based on the result, we classified each bidder into one of the following.

1. Underbidder: If we can reject the hypothesis of truthful bidding at the 5% level and the coefficient is below 1.

2. Truthful Bidder: If we cannot reject the hypothesis of truthful bidding at the 5% level.

3. Overbidder: If we can reject the hypothesis at the 5% level and the coefficient is above 1.

Analysis shows that bidders in a Vickrey auction, on average, bid 76.8% [5]of their true value. Of the 45 participants, 73.3% were classified as underbidders, 20.0% as truthful bidders and 6.7% as

[5] This is calculated by averaging the $\hat{\beta}$ for each of the 45 participants.

overbidders[6]. This is surprising, as most previous laboratory studies of single-unit Vickrey auctions find that bidders tend to overbid in such environments (Kagel 1995). In multi-unit uniform price auctions, bidders tend to overbid on the first unit and underbid on the second unit, which is consistent with the theoretical prediction of demand reduction (Kagel and Levin 2001). Our finding that most bidders either underbid or bid their true value in Vickrey auctions is consistent with (Chen and Takeuchi 2005).

Figure 3 presents the scatter plot of raw bids under Vickrey for small and big project bidders. We notice earlier months are much more competitive, and therefore, most of the bids which were over value (above the 45 degree line) were from researchers who preferred earlier months.

Individual behavior under RAD was straightforward. Bids were mostly between the minimal price and value for a package.

These individual behavioral patterns had direct implications on the aggregate performance of the mechanisms.

We used two indices to compare the aggregate performance of the three mechanisms: efficiency and the Gini coefficient. These indices were calculated for each auction outcome, which was characterized by the allocation of months and the value of those months as a package.

4.6. Bidding Behavior in RAD

We classified the bidding behavior in RAD on the levels of the individual bids. Figure 4 plots the bid-value ratio of a bid, collected into bins, on the x-axis and the frequency in which the bids in a particular bin occurs on the y-axis.

In Figure 4, the aggregate result is reported for all the bids that were submitted. This includes bids that were submitted and were losing, as well as bids that were submitted and were winning. Recalled that in RAD there is the notion of a provisional winning bid. A provisional winning bid is a bid that is winning at the end of a round, but where the round is not the final round of the auction.

[6] 3 of the truthful bidders (6.7% of all bidders) have $R^2 < 0.70$

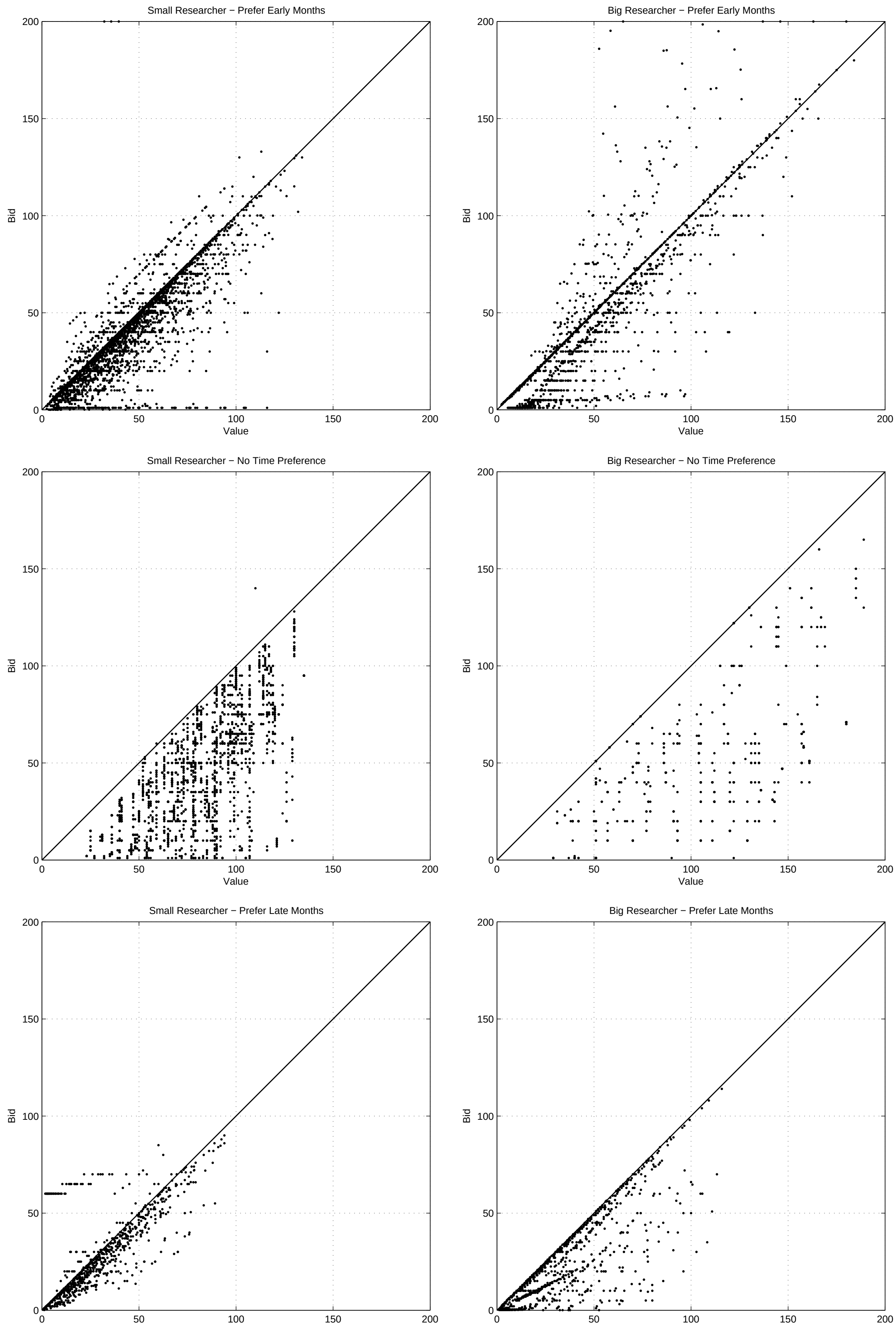

Figure 3 **Bidding Behavior in Vickrey Auction**

$$\text{Bid-Value Ratio}(x_i) = \frac{\text{Submit Price}(x_i) - \text{Market Price}(x_i)}{\text{Value}(x_i) - \text{Market Price}(x_i)} \tag{3}$$

The bid-value ratio of bids are collected into seven bins. Five of the seven bins collect bids that have a ratio between 0 and 1. Of the other two bins, one bin collect bids with the bid-value ratio less than 0 and another bin collect bids with the bid-value ratio greater than 1. The bid-value ratio could be less than 0 or greater than 1 when the submit price is greater than the value. When the submit price is greater than the value and when the value is above the market price, the bid-value ratio is greater than 1. When the submit price is greater than the value and when the value is less than the market price, than the bid-value ratio is less than 0.

First, note that from Figure 4, one can see that for bids that are losing, the percentage of bids decrease as the bid-value ratio increases. This means that more bids are placed closer to the market price then the value. One reason people may be willing to bid closer to the market price is because RAD is an iterative auction, so even if a bidder loses a package because the submit price is too low, the bidder will have a chance to bid on that package again in the next round. However, it is also important to note that many of the bids have a bid-value ratio that is greater than 20%. One possible behavioral explanation is that bidders are willing to bid more than the minimum value to move the auction along. Another possible behavioral explanation is that by bidding closer to the true value of the package, the bidder can quickly find out the likelihood that a specific package can be won. If it turns out that someone else out bids the bidder on that package, the bidder will be able to focus on other packages. Since there is no dominant strategy in RAD, it is not possible to compare the bidding behavior of human subjects against a theoretical dominant strategy benchmark.

Second, note that also from Figure 4, the winning bidders' bid-value ratio tend to be in the negative bin. This is the case because a bidder with a provisional winning bid is able, and required, to submit a bid at the same price as when it was submitted in the previous round. This is a feature of the RAD mechanism. In those situations, the bid-value ratio are negative because the submit price is lower than the market price.

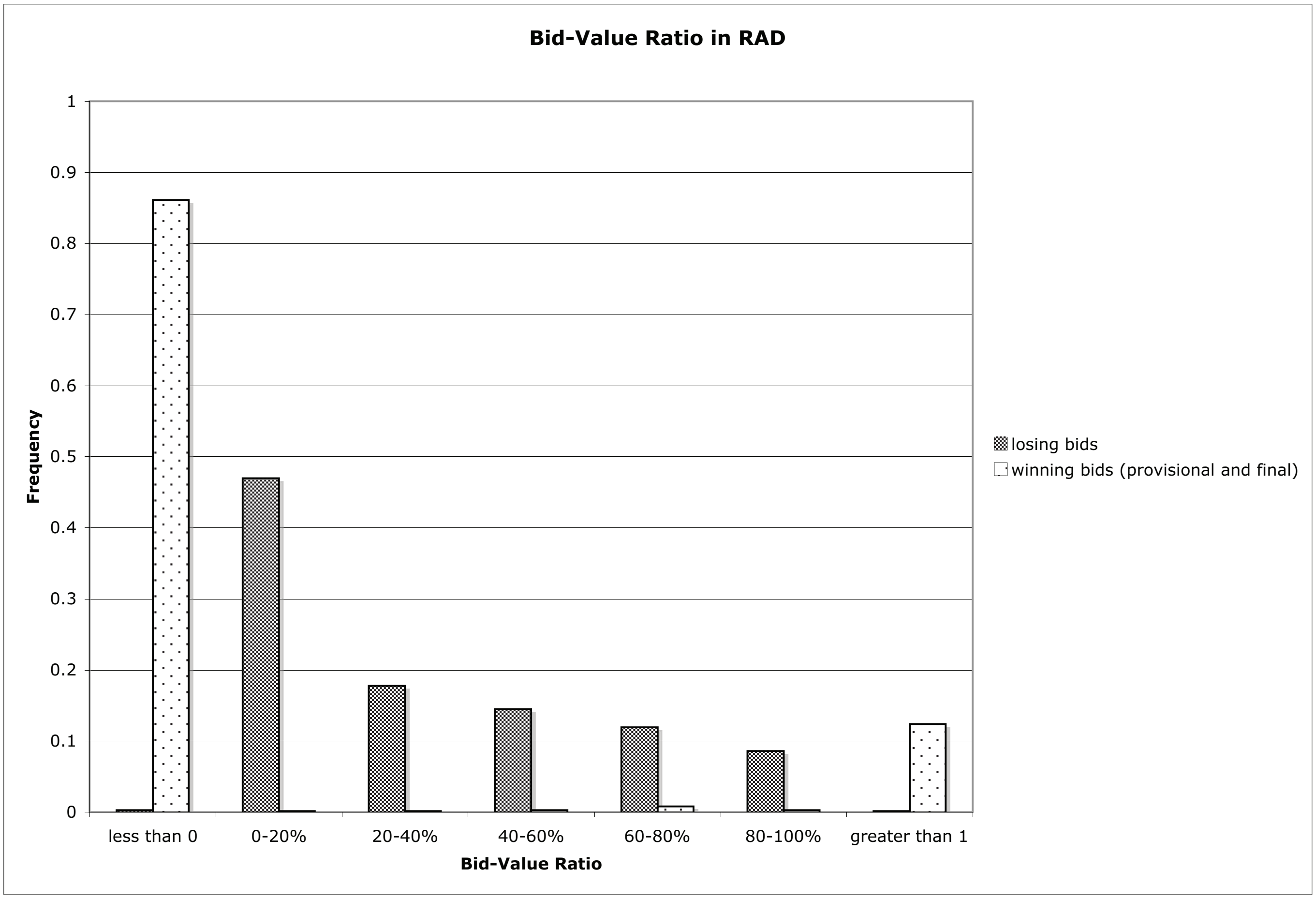

Figure 4 Bid-Value Ratio in RAD (Aggregate)

Figure 5 presents the scatter plot of raw bids under RAD for small and big project bidders further broken down by time preferences. This is similar to Figure 3. We note that one difference between the RAD bidders and the Vickrey bidders is that we don't see as much over-bidding in RAD as we do see in Vickrey. This makes sense because RAD is an iterative mechanism. Bidders will always have a chance to bid at a package that they lost in the next round. We note that just as it is the case in Vickrey, the bidders who are indifferent can often win a package without bidding their true value for that package.

4.7. Bidding Behavior in Knapsack

Since there is no dominant strategy in the knapsack mechanism proposed, we will present the empirical findings as they stand. We will present two types of bidding behavior in the Knapsack mechanism. One is how often do the bidders misrepresent their rankings, and another is how

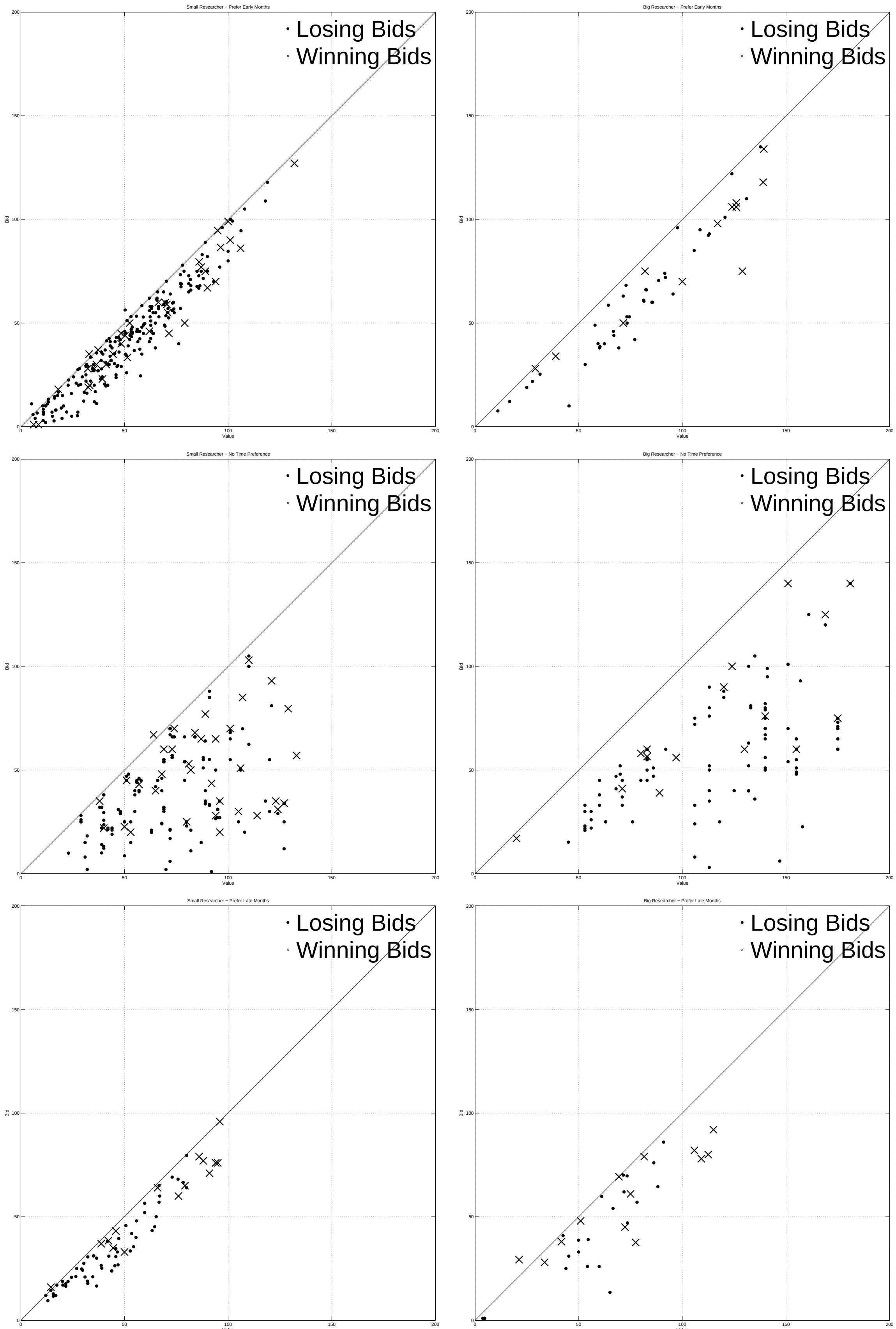

Figure 5 Bidding Behavior in RAD Auction – Winning and Losing Bids in Last Round

many packages do the bidders bid on. Recall that in the Knapsack treatment, we ran five independent experiments, with 9 subjects in each of the five independent experiments, where there are eight auctions per experiment. Instead of categorizing the behaviors on the level of the individual participant (45 of them), we opt to report the behavior of each individual participant in each of the auctions. As such, the data that we report is based on 360 data points (45 players multiplied by 8 auctions per player). The rationale for doing this is because if we report the behavior based on the level of the 45 individual players, we lose granularity on behavior. Given the lack of theoretical results on ranking mechanisms, we opt to provide the behavioral information in more detail in the hope that this information may be of use to other researchers.

In figure 6, we report the number of bids ranked by the bidders. The first bar shows that about 26% of the bidders ranked between 0-20 packages. The second bar shows that about 32% of the bidders ranked between 21 to 40 packages, etc.

In figure 7, we report the percentage of bidders that misrepresented their bids. For each of the 360 player-auction pairs, we determine the percentage of bids that are ”misrepresented” in the following way. First, we sort all the bids that a player has ranked in a particular auction by the values the packages in decreasing order. So the highest valued ranked bid is the on the top of the list, and the lowest valued ranked bid is on the bottom of the list. Next, we look at the rankings that were assigned for each of the bid, starting from the top. The rule is that if the participant ranked a higher-valued bid lower than a lower-valued bid, than that counts as a instance of misrepresentation. Below is an example.

Imagine that a player ranked on three packages A, B, and C. Package A is worth 300, package B is worth 200, and package C is worth 100. The player ranked package B as first, package A as second, and package C as third. When we sort the list in decreasing value, the order would be package A followed by B and then C. Now, we see that package B has a higher ranked than package A, even though package B is worth less than package A. If the player ranked according to the values of the packages, the player would have ranked package A as the number one package.

As such, we would say that 1 out of the 3 bids was misrepresented by the bidder, and that 33% of the bids made by this player in this particular auction were misrepresented.

Returning to figure 7. The x-axis is the level of misrepresentation and the y-axis is the fraction of the bidders that have misrepresentation level in a particular category . For example, the first bar shows that about 46% of the bidders misrepresent 0 to 20% of their bids. The next bar shows that about 31% of the bidders misrepresent 20-40% of their bids, and so on. From the graph, we see that most of the bidders reported their bids truthfully.

It is important to note that there may be other rules that one can use to measure the level of misrepresentation. The measure that was used in the calculation has the property that higher-valued bids should be ranked higher by a player if that player is truthful. There are other measures that have this property.

5. Discussion

This paper represents an exploration of aspects of collaboration at a larger scale than is typically studied in CSCW research. Specifically, the emergence of community-level technologies such as collaboratories (i.e., serving tens to thousands of users), introduces new kinds of phenomena that are not easily captured using the techniques commonly employed in CSCW studies (e.g., ethnomethodology or social psychology). As an alternative we proposed a hypothetical collaboratory where we experimentally compared competing economic approaches for allocation of scarce resources, such as instrument time.

Our hypothetical collaboratory was designed to mimic key aspects of actual collaboratories, such as the NEES collaboratory. That is, we captured variation in the resources of collaboratory users (i.e., big vs. small projects) and in the time preferences of users (i.e., early, middle, or late in a 24-month period). Experimental participants were then assigned randomly to project types and asked to bit for instrument time using one of three mechanisms – two that were auction-based (Vickery and RAD) and one based on ordinal ranking (knapsack).

Results showed that knapsack was more equitable than either Vickery or RAD. However, both Vickery and RAD were more efficient than knapsack. Intuitively, by using ordinal ranking, the

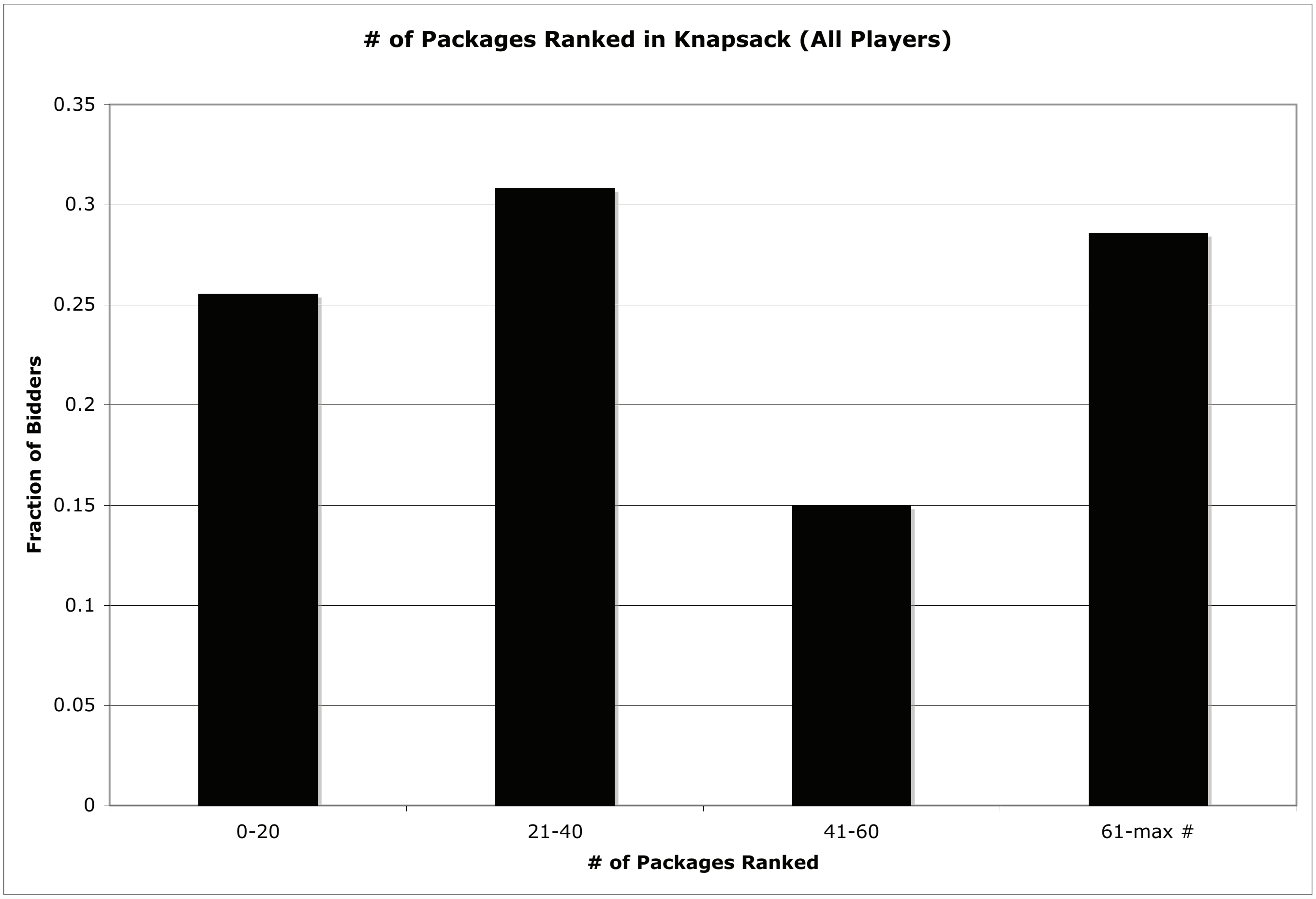

Figure 6 **Misrepresentation in Knapsack**

knapsack mechanism does not allow bidders to express the intensity of their preferences. In the hypothetical collaboratory, this favored small project bidders and made it easier for them to obtain slots. Therefore, the knapsack allocation was more equitable, to the extent that everyone got something. Knapsack was not good at giving slots to those who valued them the most, however, where both of the auction mechanisms were better at giving the right people the right slots, at the expense of equity. Therefore, a choice among these mechanisms roughly boils down to a tradeoff between efficiency and equity, and which has more weight in the designer's objectives.

There are a number of limitations to this study, such as the usual caveats about the external validity of laboratory experiments used to simulate real world phenomena. However, there is extensive evidence that experimental participants do accurately represent economic behavior within the domain of auctions – and this is the basis for a growing sub-discipline within eco-

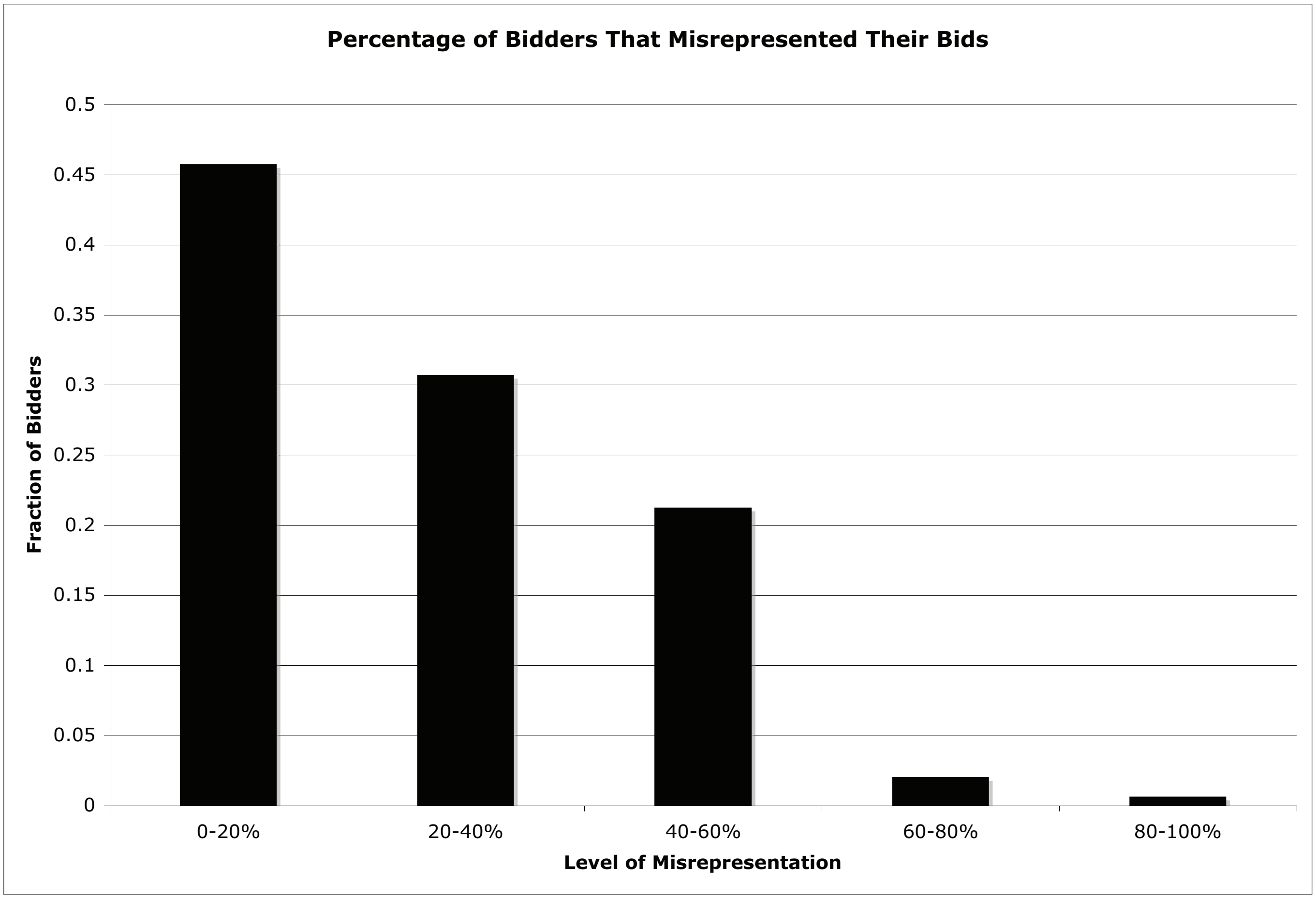

Figure 7 **Misrepresentation in Knapsack**

nomics, called experimental economics (recently recognized through the award of a Nobel prize to Vernon Smith, a pioneer of experimental economic methods). Similarly, while the hypothetical collaboratory can't model all the nuance of an actual collaboratory, it is possible to stylize key aspects of collaboratory use relevant to specific problems (e.g., allocation of scarce instrument time) and to operationize these aspects as parameters within an experimental design. Perhaps most important, it was not our purpose to use experimental methods to capture the full fidelity of collaboratory use. Rather, the experimental approach gave us a sufficiently realistic arena to examine particular aspects of collaboratory use with an eye toward future work that might focus on allocation in actual collaboratories.

We see four critical next steps in terms of advancing beyond this paper. First, we believe the results will have greater validity when participants are members of authentic scientific and engineering communities. That is, instead of using undergraduates, 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 better identify 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 on the Chandra X-ray observatory. 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, like NEES, and then study the consequences of the use of the allocation mechanism. Ideally we would like to find multiple communities to conduct quasi-experiments contrasting varying mechanisms. Finally, from a design perspective, an important goal of future work will be the creation of effective and usable interfaces to the various auction mechanisms. For instance, a limitation on RAD is that bidders can't quickly scan bids to maintain an intuitive sense of where overall bidding may stand.

Acknowledgments

We thank Tony Kwasnica, John Ledyard and Jeff MacKie-Mason for helpful discussions, Sasha Achen and Benjamin Taylor for excellent research assistance. Chen gratefully acknowledges the financial support from the National Science Foundation through grant no. SES-0339587. Any remaining errors are our own.

References

Atkins, Daniel E., Kevin K. Droegemeier, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschnmitt, Paul Messina, Jeramiah P. Ostriker, Margaret H. Wright. 2003. Revolutionizing science and engineering through cyberinfrastructure. *Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*.

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

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

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

Finholt, Thomas A. 2002. Collaboratories. B. Cronin, ed., *Annual Review of Information Science and Technology*, vol. 36. American Society for Information Science, Washington, D.C., 74–107.

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

Hagstrom, Warren O. 1965. *Scientific community*. Basic Books, New York.

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

Kagel, John H. 1995. Auctions: A survey of experimental research. J. Kagel, A. Roth, eds., *Handbook of Experimental Economics*. Princeton University Press, Princeton, New Jersey, 131–148.

Kagel, John H., Dan Levin. 2001. Behavior in multi-unit demand auctions: Experiments with uniform price and dynamic vickrey auctions. *Econometrica* **69** 413–454.

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

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

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

Smith, Vernon L. 1982. Microeconomic systems as an experimental science. *American Economic Review* **72** 923–955.

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

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

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