1987

Theory and Misbehavior of First-Price Auctions

Glenn W. Harrison

Citation of this paper:

Harrison, Glenn W.. "Theory and Misbehavior of First-Price Auctions." Department of Economics Research Reports, 8710. London, ON: Department of Economics, University of Western Ontario (1987).

ISSN:0318-725X
ISBN:0-7714-0913-3

RESEARCH REPORT 8710

THEORY AND MISBEHAVIOR OF FIRST-PRICE AUCTIONS

Glenn W. Harrison

Department of Economics

University of Western Ontario

London, Ontario, Canada

N6A 5C2

September 1987

THEORY AND MISBEHAVIOR OF FIRST-PRICE AUCTIONS

by

Glenn W. Harrison*

September 1987

---

* University of Western Ontario. I am grateful to the Social Sciences and Humanities Research Council of Canada for financial support, and to Jim Cox, Dan Friedman, Kevin McCabe, E.E. Rutström, Vernon Smith, Michael Veall, a referee, Orley Ashenfelter, and seminar participants at McMaster University, the Stockholm School of Economics, and the Universities of Melbourne, Pittsburgh, Western Ontario and Windsor for helpful suggestions.

ABSTRACT

A number of recent studies have proposed, refined, extended and tested alternative noncooperative Nash Equilibrium models of the behavior of First-Price sealed-bid auctions with independent and private values. These tests employ carefully controlled laboratory experiments designed to give the alternative models their "best shot". They find that observed subject behavior cannot be reconciled with models that assume either risk neutral agents or identically risk averse agents. The first class of models is rejected by evidence that subjects tend to bid "significantly" higher than the risk neutral prediction. This overbidding is consistent with risk aversion, but subject bidding behavior is too heterogeneous to be consistent with a model that assumes identical (albeit risk averse) bidders. We consider the saliency of the experimental evidence leading to these conclusions.

The basis of our reconsideration of the experimental evidence is quite simple. All of the existing tests have concentrated on deviations of subjects from predictions in the <u>message space</u> of auction: bid deviations. We suggest that it is more natural to evaluate subject behavior in <u>expected payoff space</u>. Certainly the latter is the appropriate metric to evaluate the incentives that subjects face in any experiment. What are indeed "statistically significant" deviations in terms of bids are <u>not</u> "statistically (or perceptually) significant" deviations in terms of foregone expected payoff. In brief, then, our reconsideration alters the metric of evaluation of the theoretical hypotheses in terms of the experimental results. We find that the experiments were not salient in terms of the inferences one would like to draw from them. We therefore conclude that the evidence against the proposed models is not significant enough to warrant their rejection.

<u>Correspondence to</u>:
Glenn W. Harrison
Department of Economics
University of Western Ontario
London, Ontario N6A 5C2
Canada
Phone: (519) 661-3493(Office)
(519) 661-3500 (Messages)

1. INTRODUCTION

In a series of classic studies Cox, Roberson and Smith [1982] and Cox, Smith and Walker [1983a] [1983b] [1985a] [1986] have proposed, refined, extended and tested alternative noncooperative Nash Equilibrium models of the behavior of First-Price sealed-bid auctions with independent and private values.[1] They find that observed subject behavior cannot be reconciled with models that assume either risk neutral agents (Vickrey [1961]) or identically risk averse agents (Holt [1980], Harris and Raviv [1981], Riley and Samuelson [1981]).[2] The first class of models is rejected by evidence that subjects tend to bid "significantly" higher than the risk neutral prediction. This overbidding is consistent with risk aversion, but subject bidding behavior is too heterogeneous to be consistent with a model that assumes identical (albeit risk averse) bidders. In this paper we reconsider the saliency of the experimental evidence leading to these conclusions.

The basis of our reconsideration of the experimental evidence is, at one level, quite simple. All of the existing tests have concentrated on deviations of subjects from predictions in the <u>message space</u> of the auction. We suggest that it is more natural to evaluate subject behavior in <u>expected payoff space</u>. Certainly the latter is the appropriate metric to evaluate the incentives that subjects face in any experiment. It may be that what are indeed "statistically significant" deviations in terms of bids are not "statistically (or perceptually) significant" deviations in terms of foregone expected payoff. In brief, then, our reconsideration alters the metric of evaluation of the theoretical hypotheses in terms of the experimental results.

A simple illustration may help motivate the virtues of the proposed alternative metric. Consider a First-Price auction with three bidders in which valuations are drawn at random from a uniform density on the closed interval [\$0.10, \$4.90]; these are the parameters adopted in Cox, Roberson and Smith [1982]. Assume that a bidder receives valuations of \$3.50, \$2.30 and \$0.50 in different trading periods. The corresponding risk neutral bid predictions are \$2.37, \$1.57 and \$0.37, respectively.[3] Let us then assume that this bidder is a degenerate mis-behaver, and always bids \$0.13 above these predictions. By any statistical measure, his <u>bids</u> significantly differ from the predictions. However, his expected income from bidding the predicted bid in each case is \$0.569, \$0.154 and \$0.001, respectively. Does it make sense to treat the first and the last <u>bid</u> as equally significant observations? Surely he has "nothing to lose" by bidding zero or his valuation in the last case, whereas he loses an expected income of \$0.569 from bidding zero or his valuation in the first case.[4] From the subject's perspective, at least, his first bid is obviously more important. The proposed metric explicitly and naturally weights the experiment observations from that perspective.

Under certain simplifying assumptions, detailed below, it is possible to calculate an <u>approximate</u> values for the expected income that is <u>foregone</u> by any given bid. We propose two ways of operationalizing the concept and metric of foregone income. Obviously this foregone expected income is zero (by construction) when the actual bid equals the predicted bid. What is interesting is that it is generically <u>much smaller</u> than the bid deviations. For the three hypothetical valuations and bid deviations given above, the foregone expected income using one of our operationalizations is only \$0.006, \$0.004 and \$0.001, respectively! These are <u>very</u> small numbers.

One might respond that any deviations of expected income from the predicted value, however small, should lead us to reject that theory. We regard this inferential position as methodologically untenable, implying a scientific nihilism that is not widespread. But an important issue now arises: what is the perceptive threshold of experimental subjects? If it is deemed to be a penny (per observable action) then the above bids are not <u>perceptually</u> significant deviations from theory using our proposed metric (and the particular operationalization used in those calculations). If it is deemed to be $0.005, then the first bid is perceptually significant and the other two are not. We do not propose any definitive answer to this question. Rather we present a mapping from alternative priors (about this threshold level) into a set of decisions as to the significance of subject misbehavior.

In Section 2 we present a Nash equilibrium model of bidding behavior due to Cox, Roberson and Smith [1982]. In Section 3 we develop two operational definitions of the notion of foregone expected income. These two alternatives employ different sets of assumptions about the expectations that an individual bidder might have about his competitors. In Section 4 we evaluate observed bidding behavior from a series of laboratory experiments. The broader methodological implications of our approach are discussed in Section 5.

2. A NASH EQUILIBRIUM MODEL OF BIDDING BEHAVIOR

Following the notation of Cox, Roberson and Smith [1982], hereafter CRS, consider a subject that receives a valuation v_i in a First-Price (FP) auction. This v_i is drawn at random with replacement from a uniform distribution defined over the nonempty closed interval $[\underline{v}, v']$, $\underline{v} \geq 0$. A single object is to be auctioned. If subject i bids b_i and wins the object he receives the income $y^i = v_i - b_i$; if he does not win the object he receives nothing. Let the number of bidders be denoted N.

Assume that each subject is drawn from a population of constant relative risk averse (CRRA) agents. Each subject is characterized by an Arrow-Pratt CRRA parameter $1-r_i$ such that $r_i=1$ denotes risk neutrality; specifically, assume that utility for agent i is $U_i(y,r_i) = y^{r_i}$. Moreover, assume that each bidder knows that his opponents will each have an r_j drawn from an <u>arbitrary</u> probability distribution on [0,1]. He does not know the specific individual attitude to risk of each of his opponents, but does know the distribution that "Nature" has used to draw them from. This is also common knowledge, along with the distribution used to generate each v_i.

Assuming noncooperative bidding, CRS show that the Nash Equilibrium (NE) bid function, for a given v_j, is

$$(1)\quad b_j^o = \underline{v} + \frac{N-1}{N-1+r_j}(v_j - \underline{v}).$$

This bid function satisfies the NE condition in that it is the best reply for agent j given that he believes that all other agents use the same function.[5] CRS also show that the expected utility function of bid b_j, given v_j, is

$$(2)\quad U(b_j) = \gamma^{N-1}(b_j - \underline{v})^{N-1}(v_j - b_j)^{r_j}$$

where

$$\gamma = \frac{N - 1 + E(r)}{(N-1)(v' - \underline{v})}$$

and E(r) is the expected value of r_i.

3. THE COST OF MISBEHAVIOR

We propose two ways of making operational the measurement of the opportunity cost of "sub-optimal" behavior. The first method follows directly from the conceptual experiment used in the derivation of the Nash bidding function (1). The second method addresses the question of how to allow for possible (and observed) "off-equilibrium" behavior when evaluating such costs.

3.1 Metric 1: The Nash Conceptual Experiment

Consider the (correct) derivation procedure used by CRS to demonstrate (1). First they suppose (p.11) "that bidder j believes that the bid of each of his rivals satisfies bid function" (1). Second, they demonstrate that the expected utility of a bid b_j is given by (2), allowing for the probability that a rival has bid higher (because the rival received a valuation inducing him, by (1), to so bid). Finally, the first-order condition for a maximum of (2) is shown to be equal to (1).

Metric 1 employs the first two stages of this derivation but not the third. We assume that bidder i acts as if his rivals are diehard followers of the bid function (1). It follows that (2) describes the expected utility of alternative bids for this bidder, including $b_i <> b_i^o$. To be sure, the maximum level of expected utility, $U^* = U(b_i = b_i^o)$, is attained when $b_i = b_i^o$. The question considered here, however, is how U differs from U* as b_i deviates from b_i^o. To answer this question we simply evaluate (2) for various (arbitrary) b_i, in particular the observed bid. Clearly (2) is only an approximation to the extent that other bidders do not follow (1) exactly.

It is convenient to assume that all bidders are risk-neutral when applying this matric. We consider the possibility of bidder risk aversion in our second metric. Assuming risk neutrality here provides a useful benchmark measure of opportunity cost. Moreover, several of the empirical experiments examined later correspond literally to the conceptual experiment proposed here: namely, evaluating expected utility from (2) for arbitrary bids by a risk neutral bidder who knows that his competitors are risk neutral and "mechanically" employ bid function (1).

We want to emphasize that our approach does not involve the computation of "alternative" NE to the one proposed in Section 2. This point should be obvious: allowing b_i to deviate arbitrarily from b_i^o, especially when b_i^o has been shown to be unique, precludes a NE.[6] On the other hand, our approach does constitute a natural sensitivity analysis with respect to the NE prediction. Recall that the general definition of a NE is that it is a set of (possibly mixed) best reply strategies for all agent such that no agent would want to unilaterally change his strategy. If any agent had an incentive to change his strategy then the initial set of strategies would not be a NE. We are concerned here with the strength of his incentives to bid exactly b_i^o. Conceptual and numerical experiments such as we undertake are precisely the sort that subjects might go through in order to decide upon their best reply bids (whether or not such bids constitute an NE). Evidence that differences between observed b_i and b_i^o depend systematically on the size of incentives is consistent with subjects behaving "as if" undertaking our proposed conceptual experiment.

Figures 1 and 2 illustrate the computation of foregone expected income using this metric. We assume, as in the experiments considered later, N=4, $\underline{v}$ = \$0.01 and v' = \$10.00. In Figure 1a we trace out the expected income from alternative bids for a (relatively lucky) agent that received a valuation of \$8.00. This curve evaluates equation (2) for r = E(r) = 1 and the values of b shown on the bottom axis. The optimum bid is seen to be \$6.00, which is exactly as predicted by equation (1). This bid earns the agent an expected income of \$1.024. Figure 1b is a simple re-drawing of Figure 1a, showing the difference between the expected income if the agent bids alternative values and the optimal expected income of \$1.024: the foregone expected income is zero at the NE bid.

Figure 2 shows a series of foregone expected income curves for alternative valuations of \$2.00, \$4.00, \$6.00 and \$8.00. The optimal bids vary with the valuation, hence we use bid deviation $(b_i - b_i^*)$ on the bottom axis for comparability. The flatness of these curves in the region of the optimal bid is evident. Irrespective of the valuation, bid deviations of <u>30 cents</u> imply that the agent is foregoing expected income of less than <u>2 cents</u> (in fact, less than 1.4 cents).

3.2 <u>Metric 2: Rationalizable Opportunity Cost</u>

One difficulty with metric 1 is that it assumes that all bidders, except the one whose opportunity costs are being evaluated, follow bid function (1). Although it is possible to control for this with computer-simulated opponents in laboratory experiments, this assumption is grossly violated when we have human opponents. We now consider a metric that allows for "off equilibrium" behavior by opponents when measuring opportunity cost.

Figure 1a
EXPECTED INCOME

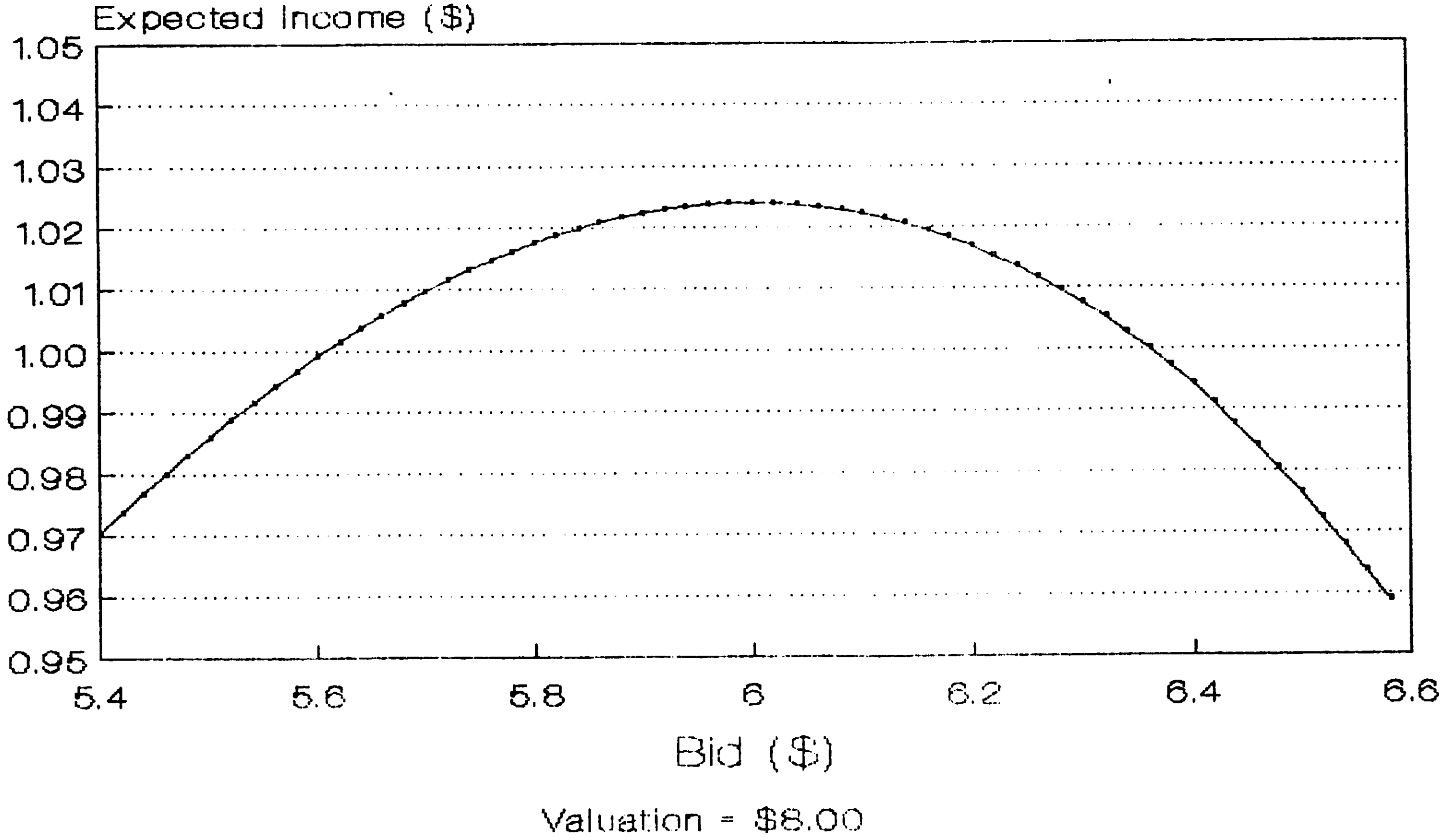


Figure 1b
FOREGONE EXPECTED INCOME

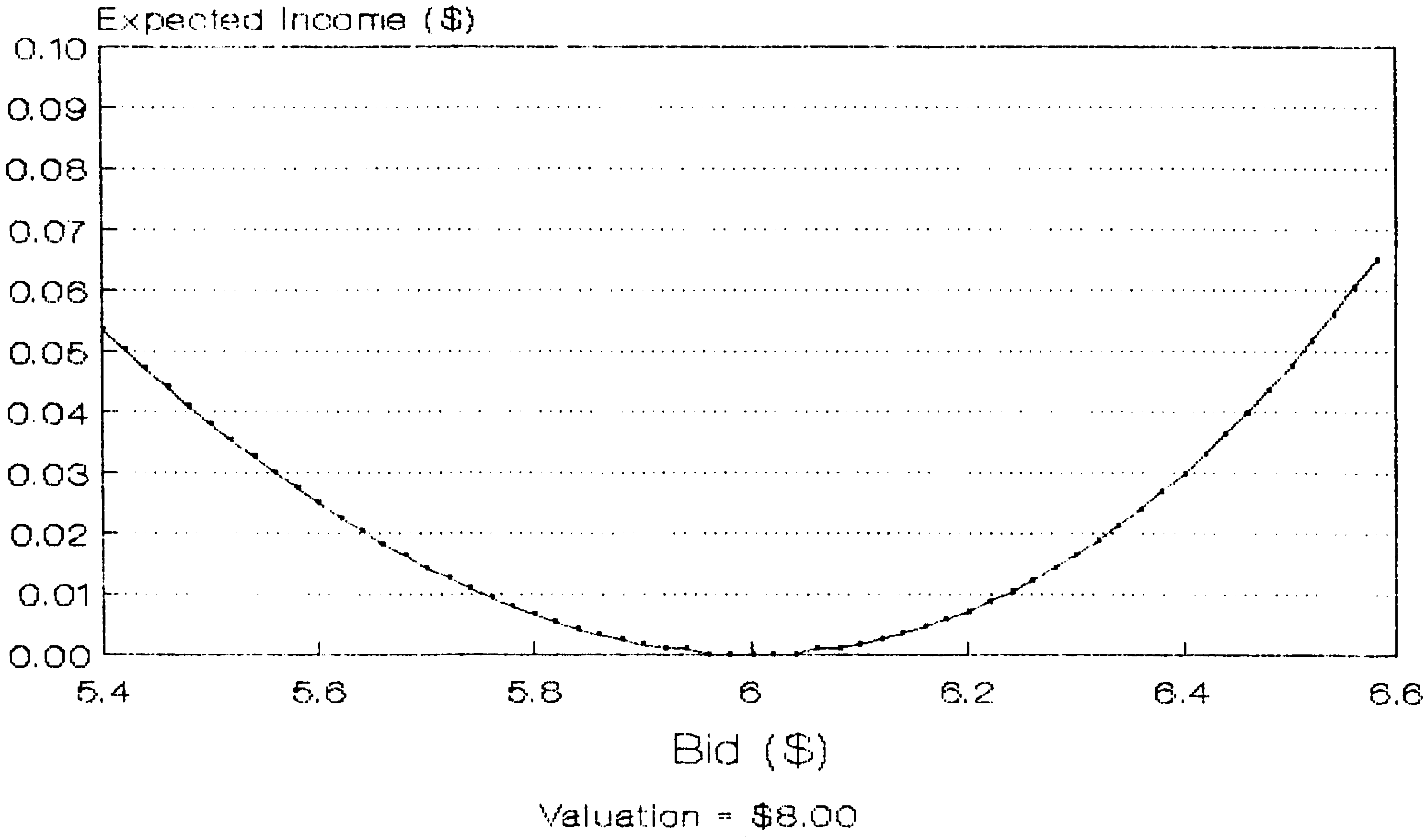
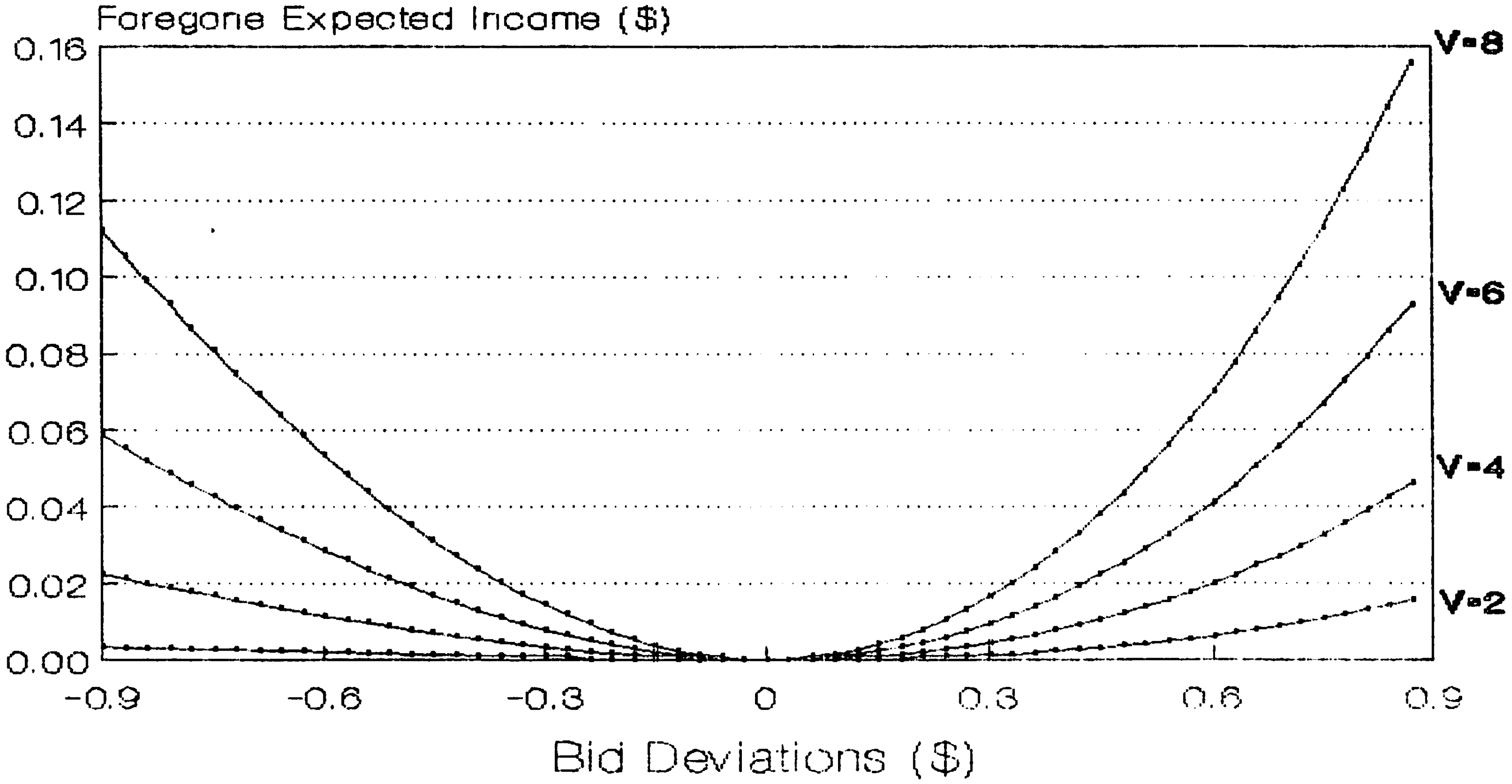


Figure 2
FOREGONE EXPECTED INCOME FOR ALTERNATIVE VALUATIONS



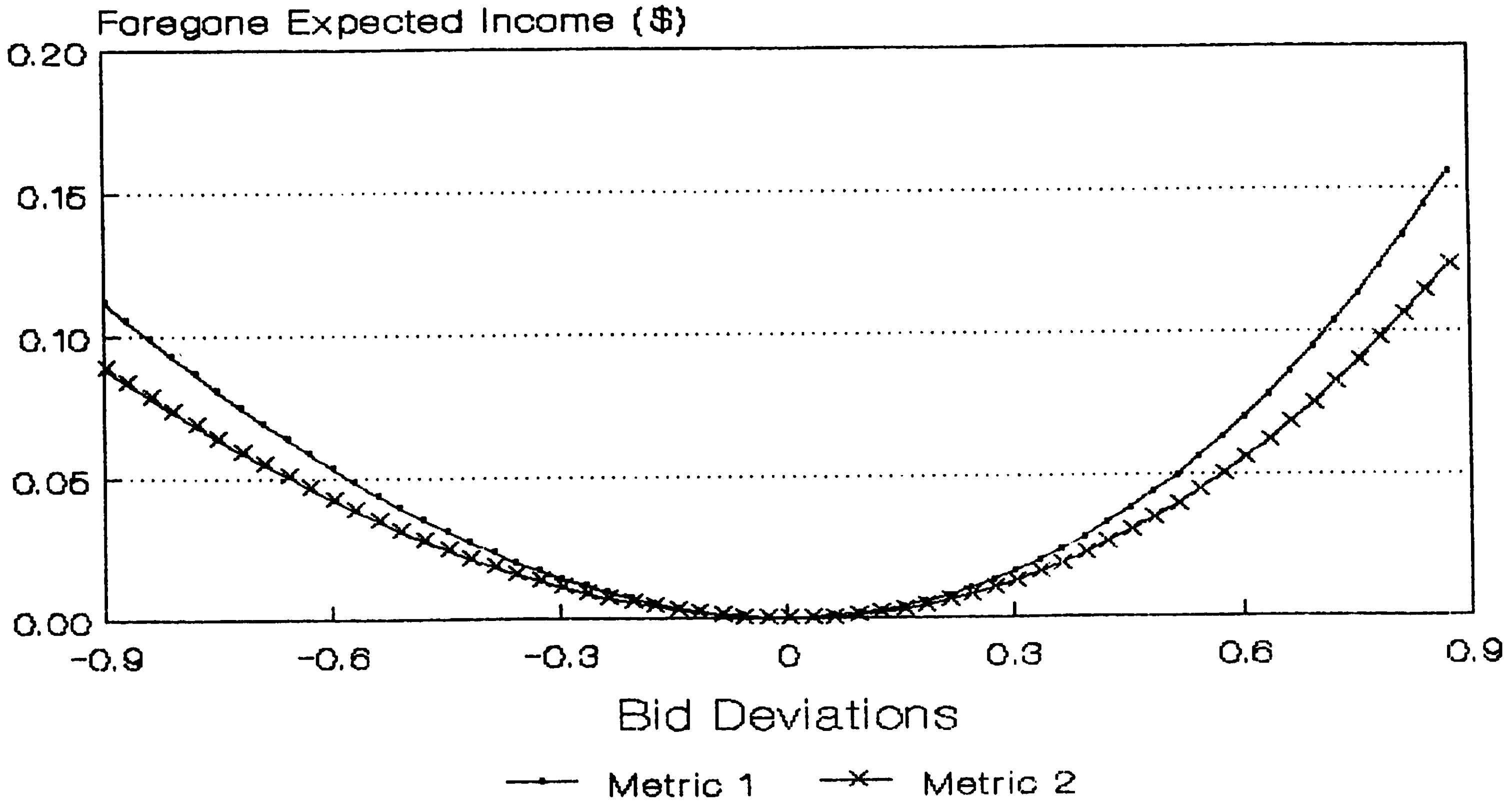
Assuming that my competitors will follow the bidding rule (1) generates an expectation as to the bids that they will enter, conditional on my expectation of their valuations. If we deny this assumption we must replace the (Nash) expectation with some other expectation.[6]

In terms of the NE bid function (1), a natural rationalization of observed bids that exceed the NE risk neutral bid is that the bidder involved is risk averse.[7] It can be shown that allowing for the empirical distribution of risk attitudes of typical experimental subjects allows one to account for observed bidding in excess of risk neutral predictions: see Harrison [1986b] and section 4.2 below. There are other plausible explanations for the observed bidding behavior, but "bidder risk aversion" is the most obvious from the perspective of received NE bidding theory.

Thus it is plausible for a given bidder to behave as if his competitors are risk averse. By doing so, he will condition his own bid on expected bids by competitors that more closely corresponds with the empirical distribution of bids than if he assumed that his competitors were all risk neutral. Even if he is risk-neutral himself, so that $r_j = 1$ in (1), this means that we evaluate the opportunity cost of sub-optimal bids using (2) as in Metric 1, but with $E(r) < 1$ rather than $E(r) = 1$. It is convenient, for the purpose of comparing the two metrics, to again assume that the bidder whose behavior is being evaluated is risk-neutral.

Figure 3 compares the foregone expected income implied by our two metrics. We use the same illustrative parameters as in Figures 1 and 2. For concreteness, we assume that $E(r) = 0.7$ for metric 2 (this specific value is justified later). The effect of the alternative assumption about opponent bidding behavior is to "flatten" the foregone expected income curve slightly.

Figure 3
FOREGONE EXPECTED INCOME:
METRICS 1 AND 2 COMPARED



The common-sense of this result is obvious: ceteris paribus valuations, if my opponents raise their bids the probability of any arbitrary bid of mine winning declines. Since metric 2 differs from metric 1 only by assuming that my opponents are risk averse rather than risk neutral, it has the effect of leading me to expect them to bid higher (recall equation 1).

This qualitative result is important, because it indicates that the results obtained with metric 1 will tend to overstate the opportunity cost of sub-optimal bids. Thus if we find that those costs are "small" under metric 1 we can be certain that they are even smaller under metric 2 for all values of $E(r) < 1$. This means that nothing crucial hinges on the choice of any particular value of $E(r)$ when evaluating metric 2.

4. OBSERVED BIDDING BEHAVIOR

4.1 The Experiments

Six experimental sessions were conducted, using the design indicated in Table 1. The general procedures followed those introduced by CRS and Cox, Smith and Walker [1985b][8]. Design features common to all experiments were the use of 4 bidders, lower and upper valuations of $0.01 (or 1 point) and $10.00 (or 1000 points), and a horizon of 20 periods. All subjects were economics undergraduates at the University of Western Ontario and received $3 just for showing up to an experimental session.

Three treatments are considered. The first is the level of subject experience. We deem a subject "experienced" if he has participated in one, and only one, previous FP experiment (in which N=2,3 or 4) with the same payoff condition (discussed below). Experiments 1 and 1P use inexperienced subjects, whereas the remaining four experiments use only experienced subjects.

TABLE 1

Experimental Design

Common design features:
$N = 4$
$\underline{v}$ = \$0.01 or 1 point
v' = \$10.00 or 1000 points
20 periods

Experiment	Level of Experience	Payoff in Dollars or Lottery Points?	Simulated Nash Opponent?	Number of Replications Per Period?	Total Number of Human Bids?
1	Inexperienced	Dollars	No	4	320
1P	Inexperienced	Points	No	4	320
2	Experienced	Dollars	No	5	400
2P	Experienced	Points	No	4	320
3	Experienced	Dollars	Yes	14	280
3P	Experienced	Points	Yes	16	320

The second treatment is the use of dollars or "lottery ticket points" to reward subjects for any earned profits. The points procedure is designed to induced risk-neutral bidding behaviour, following Roth and Malouf [1979], Berg, Daley, Dickhaut and O'Brien [1986] and Cox, Smith and Walker [1985b].

One point corresponds to one cent in the dollar payoff condition.[9] At the end of each experiment subjects had their total points earnings converted one-for-one to lottery tickets. Each subject then played an individual lottery, in which they either won $0.00 or $10.00. The lottery was of the following form: the computer randomly drew a number from the interval [0,1000], each number having equal probability; if the subject had earned more lottery tickets (i.e., points) than that number, he received the larger dollar prize; otherwise he received the smaller dollar prize. Subjects were explicitly told that the more points they earned, the more lottery tickets they would earn, and hence the better were their chances at winning the larger dollar prize. Experiments 1P, 2P and 3P employed this points procedure, and are paired with experiments 1, 2 and 3, respectively, that used dollars directly.

The final treatment is the use of computer-simulated "Nash opponents". In this case we had the computer enter risk-neutral NE bids for the 3 opponents that each human bidder faced in each period. Moreover, subjects were told exactly what bidding rule the computer would use: it would bid 75% of the valuation that it draws for each of the 3 simulated bidders.[10] This condition reduces the bidding game from a game against 4 human opponents and Nature to a game against Nature.[11] As such, it controls for the ability of the human subject to judge how rational his opponent's bidding rule is. Experiments 3 and 3P employ computer-simulated Nash opponents. Experiment 3P corresponds <u>exactly</u> with the Nash Conceptual Metric proposed earlier.

One feature of our experiments is the random assignment of opponents (human or simulated) in each period. Previous FP experiments have kept the same N players together for repetitions of the game. Although there is nothing in the game form to justify this, subjects may be led to adopt multi-period strategies even though they are irrational from a static NE perspective.[12] In order to mitigate these possibilities the computer randomly shuffles opponents from period to period.

The number of replications of each bidding game in each period is shown in Table 1. The total number of (human) bids in each experiment ranges from 280 to 400, with most experiments generating 320 bid observations. The valuations for experiments that are paired by the payoff procedure (e.g., experiment 1 and 1P) are identical for each player and period. Valuations obviously vary across players in a given replication and across period, however. Each replication in a given period also employs the same N valuations, since replications occur simultaneously in a given experiment.

4.2 Bidding Behavior

Table 2 presents some statistics describing the observed bid deviations (observed bids minus the NE predicted bid) in our experiments. The NE bid is conditional on an assumed value for r, the individual bidders' coefficient of relative risk aversion. We consider two values for r: the risk-neutral case (r=1.0) and a risk-averse case (r = 0.7). The selection of this particular risk-averse case is discussed momentarily.

TABLE 2

Observed Bid Deviations

Experiment	Assumed r	Median	Mean	Standard Deviation	Kolmogorov-Smirnov test (Probability)
1	1.0	$0.790	$0.733	$0.682	0.172 (0.0002)
	0.7	0.480	0.421	0.613	0.116 (0.0277)
1P	1.0	0.835	0.833	0.675	0.203 (0.0000)
	0.7	0.560	0.521	0.600	0.153 (0.0011)
2	1.0	0.655	0.599	0.627	0.158 (0.0001)
	0.7	0.340	0.287	0.577	0.112 (0.0127)
2P	1.0	0.525	0.521	0.748	0.162 (0.0004)
	0.7	0.270	0.209	0.711	0.091 (0.1444)
3	1.0	0.391	0.367	0.685	0.196 (0.0000)
3P	1.0	0.680	0.600	0.486	0.253 (0.0000)

Note: multiply all values for experiments 1P, 2P and 3P by 100 to obtain the equivalent in "points".

Using the risk-neutral NE bid for comparison, we see from Table 2 that observed bids are generally much higher than predicted. This result is completely consistent with previous experimental findings. The median bid deviation is in excess of 50 cents with human opponents, and is still in excess of 39 cents with simulated opponents.

It is important to note that virtually all of the sample distributions considered here, and in Table 3 and 4 below, are severely non-Gaussian. The bid deviations in Table 2 are (negatively) skewed and leptokurtic (i.e., sharp peaked, with a positive kurtosis).[13] We therefore focus on the Median as a more reliable measure of location than the Mean, and employ the non-parametric Kolmogorov-Smirnov test when comparing two sample distributions. In Table 2 we test the null hypothesis that the sample observed bids and the sample of predicted bids have the same distribution. The critical probability values reported test this null hypothesis against the alternative. Low probability values indicate that the distributions are likely to be different.

When we assume that bidders are risk-averse (r = 0.7), bid deviations are smaller but still significantly positive. The specific value of 0.7 is chosen on the basis of an empirical distribution of estimated CRRA values derived in Harrison [1986b]. This distribution is generated by obtaining the maximum likelihood estimate of r for each of the 46 (experienced) subjects[14] who participated in the experimental test for risk aversion described in Harrison [1986a]. This sample was then truncated by deleting all estimates outside the unit interval, which is consistent with the presumption in the NE model of section 2 that Nature is selecting bidders with r values in that range. The value 0.7 is finally the mean of this (normalized) truncated empirical distribution. This value has a natural operational interpretation: if I had been trying, as an experimentalist, to mimic Nature's role in selecting bidding agents, then I could well have followed the above procedure (i.e., used the risk aversion pre-test to screen subjects for the bidding game).

The importance of the risk-averse case here is that "risk aversion" implies a distribution of NE bids that is more consistent with the observed distribution of bids than is the assumption of risk-neutrality. Bid deviations are still significantly positive, but significantly less than with risk-neutrality. Thus, assuming that bidders behave "as if risk-averse" provides a better characterization of observed behavior than assuming that they behave "as if risk-neutral".[15] Hence, by conditioning on the former assumption when evaluating foregone expected income for a bidder, we obtain a metric (Metric 2) that is conditioned on an empirically better approximation of his opponent's behavior ("better" than the metric, Metric 1, that is conditioned on the latter assumption).

4.3 The Costs of Misbehavior

Table 3 summarizes the observed cost of misbehavior in our experiments according to Metric 1 ($r = E(r) = 1.0$) and Metric 2 ($r = 1$, $E(r) = 0.7$). We report the median, mean and standard deviation, along with the median bid deviation from Table 2 for comparison purposes. Again, we emphasize the non-Gaussian shape of these distributions: each of the distributions of foregone expected income are significantly positively skewed and leptokurtic. Accordingly it is appropriate to focus on the median as a measure of location.

The results in Table 3 are striking. Even though median bid deviations range from 39 cents up to 83.5 cents, median measures of foregone expected income are no higher than 6 cents. Experiments 3 and 3P provide the most telling evidence, since they control for the uncertainty about opponent

TABLE 3

Foregone Expected Income: Metrics 1 and 2

Experiment	Assumed r	Assumed E(r)	Median	Mean	Standard Deviation	Median Bid Deviation (Table 2)
1	1.0	1.0	$0.054	$0.178	$0.275	$0.790
	1.0	0.7	0.043	0.141	0.217	0.790
	0.7	0.7	0.038	0.126	0.201	0.480
1P	1.0	1.0	0.060	0.206	0.316	0.835
	1.0	0.7	0.047	0.163	0.250	0.835
	0.7	0.7	0.039	0.148	0.235	0.560
2	1.0	1.0	0.033	0.116	0.200	0.655
	1.0	0.7	0.026	0.091	0.158	0.655
	0.7	0.7	0.018	0.078	0.145	0.340
2P	1.0	1.0	0.023	0.118	0.217	0.525
	1.0	0.7	0.019	0.093	0.171	0.525
	0.7	0.7	0.009	0.082	0.158	0.270
3	1.0	1.0	0.019	0.073	0.153	0.390
3P	1.0	1.0	0.033	0.084	0.143	0.680

behavior. Here we observe median bid deviations of 39 cents and 68 cents and median foregone expected incomes of only 1.9 cents and 3.3 cents, respectively.

For completeness, we also report in Table 3 the relevant values if we assume that all four bidders are risk-averse (i.e., $r = E(r) = 0.7$). Although bid deviations are lower when we assume $r = 0.7$ instead of $r = 1.0$, so are our measures of the opportunity cost of misbehavior.

The overwhelming impression from Table 3 is that the foregone expected income from bidding misbehavior is generally very small. Table 5 provides measures of the "significance" of subject misbehavior in this experiment. These measures are <u>conditional</u> on assumed values for the perceptive threshold of subjects. We provide a range of values between 1 and 25 cents. We do not want to be dogmatic on the question of "the" perceptive threshold, although there is some experimental literature on this issue.[16] Conditional on a given threshold, we apply a non-parametric Sign test in Table 5 of the statistical significance of observing the indicated percentage of "failures" in the implied number of "trials". A failure occurs when the observed foregone expected income <u>exceeds</u> the given threshold value; a success is any outcome within that threshold value. In Experiment 3, for example, the percent of observations of foregone expected income that exceed 1, 2, 3, 4 and 5 cents, respectively, is 60%, 50%, 42%, 35% and 29%. As we increase the given threshold level it clearly becomes easier for the theory to "succeed" when confronted by the given set of observations.

TABLE 4

How Significant is Foregone Expected Income?

Threshold Price in Cents	Experiment						
	1	1P	2	2P	3	3P	All
				(a) Metric 1			
1	1.000	1.000	1.000	0.996	1.000	1.000	1.000
2	1.000	1.000	1.000	0.766	0.524	0.998	1.000
3	0.999	1.000	0.853	0.120	0.005	0.829	0.987
4	0.919	0.999	0.125	0.002		0.003	0.001
5	0.693	0.919	0.002				
6	0.390	0.522					
7	0.234	0.234					
8	0.081	0.099					
9	0.015	0.015					
10	0.003	0.004					
11	0.001	0.002					
				(b) Metric 2			
1	1.000	1.000	1.000	0.934	–	–	1.000
2	1.000	1.000	0.773	0.269			1.000
3	0.985	1.000	0.171	0.003			0.697
4	0.799	0.856	0.001				0.002
5	0.522	0.390					
6	0.307	0.144					
7	0.052	0.015					
8	0.006	0.003					
9	0.004	0.001					

The specific test procedure used in Table 4 is described in Conover [1980; p. 96ff.] and DeGroot [1975; p. 482]. Our test is one-tailed, with the null hypothesis that the probability of a failure on any trial is less than or equal to one-half. The alternative hypothesis is that the probability of a failure on any trial is greater than one-half. We report the critical (minimal) probability level required to reject the null hypothesis. Low probability values are therefore "good" for the theory, indicating that the observed failure rate of the theory was probably a chance event.

To continue the example of Experiment 3, the 50% failure rate with a threshold of 2 cents implies a critical probability level of 0.524. We can therefore not reconcile the observations and the theory if we require that the former be within 2 cents of the latter. However, with a threshold of 3 cents the failure rate drops to 42%, which implies a critical probability of only 0.005; thus we can reconcile the observations and the theory if we only require that the former be within 3 cents of the latter.

The results in Table 4 confirm the intuition gained from our discussion of Tables 2 and 3. There is no statistical doubt that the theory is unable to explain the behavior of these subjects in message space. However, when we turn to look at the foregone income metric, we see something entirely different. Pooling over all experiments, at a threshold level of only 4 cents the level of significance on the null hypothesis has dropped to less than 0.002 for either Metric. Given these threshold levels, there is little doubt that the observed failure rate of the theory is indeed just a chance event.

It is important to note again that there exist threshold levels at which the null hypothesis could not be rejected (at conventional levels of significance). At a 10 percent critical level, for example, one would have to entertain a threshold level of 3 cents or less in order to maintain that the observed foregone incomes are significant deviations from the theoretical predictions (again, pooling over all experiments). An observer of this data would have to insist, dogmatically, that foregone income be less than or equal to 3 cents before the data would lead one to reject the theory. In our opinion such dogmatism is excessive and unrepresentative of the scientific community.

5. CONCLUDING REMARKS

The conclusions one draws from our analysis of the experimental data depend entirely on the metric that is believed to be appropriate for these experiments. If one adopts the message space (i.e., bids) as the relevant metric then subjects appear to deviate "significantly" from the theoretical predictions of the NE hypothesis. However, alternative metrics based on notions of foregone expected income due to subject misbehavior are statistically negligible (by any commonsense notion of subjective opportunity cost).

Frankly, our results pose a serious dilemma for the interpretation of these experiments. In terms of the precepts proposed by Smith [1982], these experiments do not satisfy the requirement that rewards <u>dominate</u> the subjective costs of the activity. As such, one has lost "control" over the experimental microeconomy under study. It is the purpose of an experimental design to make the monetary payoffs salient enough to be used as a surrogate for subjects' utilities. If we cannot conclude that this is the case for observed deviations from predicted behavior, then we are unable to reject the theory generating those predictions.

It should finally be noted that this methodological point is potentially a problem in all empirical evaluations of the predictions of optimizing models.[17] The optimal demands from a model of utility maximizing agents, for example, are invariant to monotone utility transformations. Thus there is no way to know what "utility loss" would be associated with sub-optimal demands.[18] In principle the experimental method has an advantage over other empirical techniques if one can design <u>controlled</u> experiments that induce subjects' utility functions. In this case the experimentalist can say just "how large" the discrepancies are between observed and predicted behavior in a natural metric from the perspective of theory.

FOOTNOTES

1. Comparable research has also been undertaken for multiple-unit discriminative auctions by Cox, Smith and Walker [1984] [1985b], and for common value auctions by Levin, Kagel, Battalio and Meyer [1985]. Important earlier experimental studies include Coppinger, Smith and Titus [1980], Miller and Plott [1985] and Smith [1967]. Kagel, Harstad and Levin [1985] and Kagel and Levin [1985] also report several First-Price sealed bid experiments with independent and private values.

2. They are concerned with many issues other than these (e.g., tests of Dutch auction behavior and alternative models of bidder behavior).

3. Actually the optimal <u>feasible</u> bids corresponding to these valuations are \$2.30, \$1.50 and \$0.30, respectively, since subjects were forced to "round" bids in the Cox, Roberson and Smith [1982] experiments.

4. One actually observes "throw away" bids of zero in these experiments (cf. Cox, Smith and Walker [1985a; p.161]). Less frequently, one also observes bidding in excess of valuation (cf. Cox, Smith and Walker [1985b; pp. 201-2]).

5. This bid function is defined analytically only for bids that do not exceed $b' = \underline{v} + [(N-1)/N](v' - \underline{v})$, which is the NE bid that a risk neutral agent who received the highest possible valuation would enter. The number of observed bids that exceed b' in the experiments considered here is extremely small. The bid function (1) is a direct extension of the Vickrey [1961] bid function to allow for valuations on an

arbitrary interval (rather than the unit interval) and to incorporate the single-parameter CRRA family of utility functions (rather than the risk neutral case).

6. One obvious and tractable candidate might be the correct expectation. In other words, we could ask what the best-response bid would be if the agent knew exactly what each of his opponents has bid. The observed bid could then be compared with this best-response bid to determine an upper-bound on foregone expected income. In the present case we have a tractable model of why bidders may deviate from the risk-neutral NE bid (namely, a model that allows my opponents to be risk averse), hence we do not need to pursue the degenerate case of "correct expectations". Harrison and McCabe [1987] do make use of this degenerate assumption when evaluating foregone expected income from observed behavior in an experiment which does not have such tractable alternative models available.

7. The term "rationalizable" is borrowed, deliberately, from Bernheim [1984] and Pearce [1984].

8. These experiments were conducted in an IBM-PC microcomputer laboratory rather than on the PLATO system employed by CRS and Cox, Smith and Walker [1985b]. The instructions and raw data are freely available on request. Although I have replicated most of their experimental results, I do not claim to have replicated the procedures of CRS. I am very grateful to Kevin McCabe for allowing me to use software that we had developed jointly for other experiments.

9. Unlike CRS, but like Cox, Smith and Walker [1985b], valuations and bids could be in units of one cent (or one point).

10. Walker, Smith and Cox [1986] have conducted experiments in which individual bidders face computer-simulated opponents and were explicitly told the simulated bidding rule. Unfortunately that rule is not (1), but is a rule based on the empirical distribution of bids in previous "all-human" experiments. In any case, observed bids with this treatment do not differ significantly from those observed in previous experiments.

11. This condition does not alter the speed of the experiment (which would influence the expected payoff per unit time). The computer waits for all human subjects to enter their bids in each period before informing anybody of the outcome. Thus the first subject to imput his bid must wait for all other human subjects to enter their bids before learning the outcome of his auction. To the extent that we used the same number of human subjects in each session (typically 16), each experiment would be paced approximately the same from the perspective of computer logistics. Inexperienced subjects are typically slower at entering bids for the first 5 or so periods.

12. I have observed such behavior in an anecdotal form in an experiment for teaching purposes (and hence with no financial rewards) using the exact procedures of CRS. The subject explained his strategy in the following fashion: "Whenever I get a valuation less than one-half of the top valuation, I bid zero. This lulls my opponents into a false sense of security by letting them win more often. In those cases I would only win a few pennies anyway, even if I did win. Where I make up is the greater chance that I will win big when I have higher valuations, since my opponent's are not so aggressive now." Whether or not such strategies are rationalizable is an open issue.

13. The values for the Skewness Statistic are -1.9, -1.6, -2.8, -2.8, -1.2 and -0.4 for each of the six experiments, respectively. The reference Skewness values for a Gaussian distribution lie between 0.122 and 0.146, implying that these distributions are significantly skewed. The Kurtosis statistics are 12.6, 8.9, 25.6, 19.5, 3.4 and 0.6, respectively. The reference Kurtosis value for a Gaussian distribution is zero. All statistics reported here were computed using procedures taken from Press, Flannery, Teukolsky and Vetterling [1986].

14. These subjects are not the ones that participated in our sealed-bid experiments. However, they were drawn from the same population of subjects (viz., economics undergraduates at the University of Western Ontario).

15. This is true even in the experiments (1P and 2P) in which we employed a procedure that is supposed to induce "as if risk-neutral" behavior. How do we justify assuming $r < 1$ in these experiments? First, one could simply re-interpret our assumption to include the failure of the lottery procedure: thus we are assuming that subjects behave "as if they receiving one cent for every pont earned <u>and</u> as if they are risk-averse". Second, one could argue that risk aversion is just one interpretation for the parameter r that Nature selects and which characterizes agents in some way. An alternative interpretation, that r parameterizes a family of subjective expectations about bidding behavior, is proposed in a subtle argument by Cox, Smith and Walker [1985b]. What is profound about this argument is that the

formal NE bidding models (and hence the theoretical predictions) that follows from one interpretation or the other are identical. For expositional simplicity we will continue to refer to the $r < 1$ assumption as "the risk-averse case", even if alterantive interpretations are equally valid logically.

16. See Green and Swets [1974] on signal-detection experiments, Myers, Fort Katz and Suydam [1963] on binary choice experiments, and Siegel [1961] on Bernoulli trials experiments, for example. Heiner [1985] and Winter [1982] review some of this literature for economists. Decision theorists and practitioners have long been aware of related "flat maxima" problems: see Von Winterfeldt and Edwards [1986; Ch. 11]. In experimental economics there is one recent study that bears on the perceptive threshold question: Smith and Williams [1983] vary the level of commissions in a double auction context.

17. I am indebted to Orley Ashenfelter for this argument.

18. This leads to the informal econometric presumption that "maximization errors" imply zero-mean error terms, but do not imply anything about their variance. Theil [1971] argues, however, that the relative variances of different demand functions could be related to compensated substitution effects. These effects are one measure of the flatness of the utility function in the neighbourhood of the optimum. He correctly notes that "... the idea in the case of one single decision variable is that, if the second-order derivative at the point of the extremum is close to zero, the loss incurred by a decision which deviates from the optimum to a moderate extent is very small, and the disturbance of the corresponding behavioral equation may then be expected to have a large variance." (p. 191).

REFERENCES

Berg, J.E.; Daley, L.A.; Dickhart, J.W., and O'Brien, J.R., "Controlling Preferences for Lotteries on Units of Experimental Exchange", Quarterly Journal of Economics, v. 101, May 1986, pp. 281-306.

Bernheim, B.D., "Rationalizable Strategic Behavior", Econometrica, v. 52, July 1984, pp. 1007-1028.

Conover, W.J., Practical Nonparametric Statistics (Second Edition, Wiley, New York, 1980).

Coppinger, V.; Smith, V.L.; and Titus, J., "Incentives and Behavior in English, Dutch, and Sealed-Bid Auctions", Economic Inquiry, v. 18, 1980, pp. 1-22.

Cox, J.C.; Roberson, B.; and Smith, V.L., "Theory and Behavior of Single Object Auctions," in V.L. Smith (ed.), Research in Experimental Economics (Greenwich, JAI Press, Volume 2, 1982, pp. 1-43).

Cox, J.C.; Smith, V.L., and Walker, J.M., "Tests of a Heterogeneous Bidders Theory of First Price Auctions", Economics Letters, v. 12, 1983a, pp. 207-212.

Cox, J.C.; Smith, V.L.; and Walker, J.M., "A Test That Discriminates Between Two Models of the Dutch-First Auction Non-Isomorphism," Journal of Economic Behavior and Organization, v. 4, 1983b, pp. 205-219.

Cox, J.C.; Smith, V.L., and Walker, J.M., "Theory and Behavior of Multiple Unit Discriminative Auctions," Journal of Finance, v. 39, September 1984, pp. 983-1010.

Cox, J.C.; Smith, V.L.; and Walker, J.M., "Expected Revenue in Discriminative and Uniform Sealed-Bid Auctions," in V.L. Smith (ed.), Research in Experimental Economics (Greenwich, JAI Press, v. 3, 1985a).

Cox, J.C; Smith, V.L.; and Walker, J.M., "Experimental Development of Sealed Bid Auction Theory: Calibrating Controls for Risk Aversion," American Economic Review (Papers and Proceedings), v. 75, 1985b, pp. 160-165.

Cox, J.C.; Smith, V.L.; and Walker, J.M., "Theory and Individual Behavior of First Price Auctions," Unpublished Manuscript, Department of Economics, University of Arizona, 1986.

DeGroot, M.H., Probability and Statistics (Reading: Addison-Wesley, 1975).

Green, D.M., and Swets, J.A., Signal Detection Theory and Psychophysics (New York: Robert Kriegur, 1974).

Harris, M., and Raviv, A., "Allocation Mechanisms and the Design of Auctions," Econometrica, v. 49, November 1981, pp. 1477-1499.

Harrison, G.W., "An Experimental Test for Risk Aversion", Economics Letters, v. 21, 1986, pp. 7-11.

Harrison, G.W., "Risk Attitudes in Private Value Auction Experiments," Unpublished Manuscript, Department of Economics, University of Western Ontario, 1986b.

Heiner, R.A., "Uncertainty, Signal-Detection Experiments, and Modeling Behavior," in R.N. Lanlois (ed.), Economics As A Process (New York: Cambridge University Press, 1985).

Holt, C.A., "Competitive Bidding for Contracts Under Alternative Auction Procedures," Journal of Political Economy, v. 88, June 1980, pp. 433-445.

Kagel, J.H.; Harstad, R.M.; and Levin, D., "Information Impact in First-Price Auctions With Affiliated Private Values: A Laboratory Study," Unpublished Manuscript, Department of Economics, University of Houston, 1985; Econometrica, forthcoming.

Kagel, J.H., and Levin, D., "Individual Bidder Behavior in First-Price Private Value Auctions," Economics Letters, v. 20, 1985, pp. 125-128.

Levin, D.; Kagel, J.H.; Battalio, R.C.; and Meyer, D.J., "First-Price, Common-Value Auctions: Bidder Behavior and the 'Winner's Curse'", Unpublished Manuscript, Department of Economics, University of Houston, 1985.

Miller, G., and Plott, C.R., "Revenue Generating Properties of Sealed-Bid Auctions: An Experimental Analysis of One-Price and Discriminative Process," in V.L. Smith (ed.), Research in Experimental Economics (Greenwich, JAI Press, v. 3, 1985).

Myers, J.L.; Fort, J.G.; Katz, L.; and Suydam, M.M., "Differential Monetary Gains and Losses and Event Probability in a Two-Choice Situation," Journal of Experimental Psychology, v. 66, 1963, pp. 521-522.

Pearce, D.G., "Rationalizable Strategic Behavior and the Problem of Perfection", Econometrica, v. 52, July 1984, pp. 1029-1050.

Press, W.H.; Flannery, B.P.; Teukolsky, S.A., and Vetterling, W.T., Numerical Recipes: The Art of Scientific Computing (Cambridge: Cambridge University Press, 1986).

Riley, J.G., and Samuelson, W.F., "Optimal Auctions," American Economic Review, v. 71, June 1981, pp. 381-392.

Roth, A. E., and Malouf, M. W. K., "Game-theoretic Models and the Role of Information in Bargaining," Psychological Review, v. 86, 1979, pp. 574-594.

Siegel, S., "Decision Making and Learning Under Varying Conditions of Reinforcement," Annals of the New York Academy of Science, v. 89, 1961, pp. 766-783.

Smith, V.L., "Experimental Studies of Discrimination Versus Competition in Sealed-Bid Auction Markets," Journal of Business, v. 40, 1967, pp. 56-84.

Smith, V.L., and Williams, A.W., "An Experimental Comparison of Alternative Rules for Competitive Market Exchange," in R. Engelbrecht-Wiggans, M. Shubik, and R.M. Stark (eds.), Auctions, Bidding, and Contracting: Uses and Theory (New York: New York University Press, 1983).

Theil, H., "An Economic Theory of the Second Moments of Disturbances of Behavioral Equations", American Economic Review, v. 61, March 1971, pp. 190-194.

Vickrey, W.S., "Counterspeculation, Auctions, and Competitive Sealed Tenders," Journal of Finance, v. 16, March 1961, pp. 8-37.

Von Winterfeldt, D.; and Edwards, W.; Decision Analysis and Behavioral Research (Cambridge: Cambridge University Press, 1986).

Walker, J.; Smith, V.L.; and Cox, J.C., "Inducing Risk Neutral Preferences: An Examination in a Controlled Market Environment," Working Paper No. 86-09, Department of Economics, University of Arizona, 1986.

Winter, S.G., "Binary Choice and the Supply of Memory," Journal of Economic Behavior and Organization, v. 3, 1982, pp. 277-321.