# Organization and Performance of World Food Systems: NC-194

## OCCASIONAL PAPER SERIES



The work reported herewithin contributes to the objectives of North Central Regional Project NC-194, a joint research project of state agricultural experiment stations and the U.S. Department of Agriculture

# EXPERIMENTAL ECONOMICS: AN INTRODUCTION FOR APPLICATIONS TO INTERNATIONAL TRADE

DOUGLAS D. DAVIS AND SHANNON K. MITCHELL

OP-25 APRIL 1991



DEPARTMENT OF ECONOMICS, VIRGINIA COMMONWEALTH UNIVERSITY

PAPER PRESENTED AT THE WORKSHOP ON INDUSTRIAL ORGANIZATION AND INTERNATIONAL TRADE, NC-194, COLUMBUS, OHIO, APRIL 24-25, 1991

# EXPERIMENTAL ECONOMICS: AN INTRODUCTION FOR APPLICATIONS TO INTERNATIONAL TRADE

by
Douglas D. Davis and Shannon K. Mitchell*

## 1. INTRODUCTION

The use of laboratory methods to evaluate economic propositions has blossomed in the last two decades from being an isolated and curious novelty in the early 1970's to a widely accepted research method. At present, literally hundreds of economic experiments are reported each year, conducted in some two dozen experimental economics laboratories in the United States and around the world. Experimental methods have been used to investigate a wide variety of economic propositions in a diverse array of subfields. To the best of our knowledge, however, laboratory methods have not been used to investigate issues specific to problems in international trade. The absence of attention to trade issues is probably a consequence of the predominantly industrial organization and finance backgrounds of the early experimental economists. Certainly, there is no inherent justification for this oversight.

This paper is written as an introduction to experimental economics for trade theorists interested in considering laboratory techniques as an empirical method for evaluating trade issues. We proceed as follows. First, economics has traditionally been viewed as an inquiry not amenable to laboratory techniques. This perception remains among many economists, particularly in areas where experimental methods have not been used. For this reason, section 2 is devoted to a discussion of our view of the methodological role of experimentation in economics. Section 3 describes how a simple market experiment is constructed and administered. Market experiments are by no means the only type of experiment useful in economic analysis and, in fact, some of the more obvious applications of experimental techniques to international trade issues involve the evaluation of predictions in non-cooperative game theory. A market experiment is used for introduction because of the obvious nature of equilibrium predictions. Section 4 presents some guidelines for effective experimentation, while section 5 reviews some of the chief results of experimental inquiry. Possible

---

* Department of Economics, Virginia Commonwealth University, Richmond Virginia 23284-4000. Financial support from the National Science Foundation is gratefully acknowledged.

applications of experimental techniques to issues in international trade are discussed in section 6. Finally, some concluding remarks are offered in a brief 7th section.

Prior to proceeding, we note that a healthy dose of skepticism is appropriate, and indeed useful. Although there are a number of empirical issues in virtually every area of economics that can be usefully addressed with laboratory methods, there are also a number of other empirical issues where experimental methods are *not* likely to offer particularly useful insights. Even if we succeed in convincing the reader that experimental methods are a potentially useful means of collecting information relevant to some economic propositions, a critical eye toward what could be learned from any particular laboratory investigation can serve as an invaluable guide toward the fruitful application of this empirical method.

## 2. EXPERIMENTAL ANALYSIS AS AN EMPIRICAL TECHNIQUE.

### Experimental Analysis and the Social Sciences

Almost universally, scientists construct theories to organize and explain behavior. The typical expectation is that these theories are grounded in empirical reality, in the sense that they are evaluated on the basis of their capacity to predict outcomes. For example, a chemical theory predicting the explosive volatility of fixed proportions of gypsum, charcoal and sulfur in the presence of heat could be tested by lighting a firecracker.

Notably, although actually generating and evaluating theories requires considerable cleverness, the *process* of theory evaluation is at once both methodologically straightforward and intuitive. As a first matter, one would like to know if there is any circumstance under which the theory works. By progressively controlling for auxiliary assumptions, the theory can be given a "best shot." Suppose, for example, that the firecracker does not explode. The failure immediately suggests a series of additional tests which control for auxiliary factors that may have prevented the predicted response. For example, was heat actually present? Perhaps the fuse did not burn all the way into the gunpowder. Again, other chemicals (such as water) might impede the process. Was the gunpowder wet? Further tests that control for these concerns could be conducted in an effort to evaluate the theory. If the theory fails to work even under the most carefully controlled circumstances conceivable, then it tends to be rejected.[1]

On the other hand, if the predictions of a theory are confirmed in initial experimentation then investigation typically proceeds in the other direction, with a series of "stress tests" or boundary experiments, designed to examine the range of the theory's application. A junior high school chemist, for example, might want to know the sensitivity of the explosion to variations in the proportions and mix of the chemicals. Finally, in the course of evaluating a theory's range of application, persistent empirical regularities are often observed which are not part of the theory, but which suggest theory modification. For example, in the process of evaluating the limits of application for the above gunpowder theory, it may be observed that the rate of successful explosions varies inversely with the size of the firecracker encasement. For the junior high school chemist, this might represent an anomaly to be explored in its own right. Perhaps it will lead to a refinement of the young chemist's theory, e.g., the gunpowder theory might ultimately be modified to suggest that oxygen is also critical to an explosive chemical reaction. This process of identifying a baseline, exploring the limits of application for a theory, and noting empirical regularities in controlled conditions lies at the very heart of theory development, evaluation, and refinement in the physical sciences.

Although economics shares with the physical sciences the property of a rich and elegant theoretic structure, economic theory has been developed, refined, and evaluated in a different manner, almost exclusively outside the laboratory. The traditional position among economists is that laboratory methods are not applicable to economics due to the subject and scope of economic study: humans interacting in social institutions. Unlike physical processes, it is argued, economic events occur in circumstances so complex as to be spatially and temporally unique. By this reasoning, assumptions auxiliary to economic theories cannot be progressively controlled, and the only way to evaluate economic propositions is through the application of statistical techniques to data from the natural world.

Experimental economists contest the distinction of economics from the physical sciences, arguing that the difference is one of degree, rather than kind. While it is quixotic to attempt to evaluate an

[1] There rarely (if ever) exists an ultimate rejection of a theory. Rather, scientists tend to give up on one theory when it fails to work in any environment of practical concern.

economic theory in a laboratory context that retains the richness of the parallel natural context, it is often possible to construct a very sparse laboratory environment that meets the assumptions of a particular theory. If the theory fails to organize data in the simple laboratory environment, then it cannot be expected to motivate behavior in the more complicated natural world.

Notably, the comparative sparseness of the economics experiment limits the kinds of conclusions that can be drawn from experimental investigation. In many standard examples from the physical sciences, the distinction between observation in the laboratory and in natural contexts is often minor. We would, for example, expect the laboratory conditions for the gunpowder experiment to be substantially the same as conditions outside the lab. Therefore, we would expect the results of investigations into the limits of the "gunpowder theory" to have direct relevance to the natural world. Economists, however, must generally be much more circumspect in making claims about behavior in the natural world. For example, the predictions of static Nash equilibrium theory could be tested in the laboratory by presenting a pair of participants with a simple prisoner's dilemma game in normal form, and offering them the opportunity to make simultaneous action choices under the condition that they would earn the monetary amount listed in the game representation as the outcome of their actions. Even a laboratory result most supportive of the Nash equilibrium prediction would allow little to be said about performance in the natural world. One would generally not expect to see agents facing anything even vaguely resembling the normal-form game structure in any natural application of a prisoners' dilemma. Moreover, extenuating circumstances in the natural context would undoubtedly invite debate as to whether the agents actually faced the dilemma.

While the increased distance between the laboratory and the natural world limits the claims that can be made about the natural world from economics experiments, it does not impair the usefulness of experimentation in economics. To repeat: If the theory fails to organize behavior in very simple environments, it cannot be expected to organize behavior in the more complicated natural world. It is principally (but by no means exclusively) this role of evaluating theory that experimental techniques are useful in economics.

**The Value of Experiments in Economics**

The value of experimental techniques as tests of economic theories is highlighted by the difficulty of testing many economic propositions with data from the natural world. A surprising variety of very basic economic propositions are difficult or impossible to evaluate with naturally occurring data. Consider, for example, the competitive markets hypothesis fundamental to any economics principles course: Given sufficient information and barring market power, the price and the quantity of a traded private good is determined by the intersection of market supply and market demand curves.

Both practical and inherent data constraints impede the identification of market supply and demand curves in natural markets. Consider, for example, the problem of estimating market supply. Competitive concerns often provide incentives for firms to disguise the cost information. Moreover, firms are generally not interested in same data as economists, and (as would be attested by anyone who has ever attempted to calculate firm costs for a predatory pricing investigation) relevant cost information is often simply not collected. Even more fundamental data problems exist on the demand side. Consumers make decisions on the basis of perceived utility, which is impossible to directly observe.

Spphisticated statistical and econometric techniques have enabled researchers to make remarkable claims from limited available data from natural markets. These techniques, however, necessarily rely on yet other assumptions which are not directly observable in natural markets. Simultaneous equations methods, for example, may be used with transactions price data to circumvent the above-mentioned problems associated with obtaining natural cost and utility information to estimate market supply and demand relationships. But this technique presumes that markets equilibrate. Using this method, demand is estimated as a collection of equilibrium points determined by a series of supply shifts. Conversely for supply. It would, of course, be circular to subsequently use the estimated supply and demand curves to evaluate the proposition that markets equilibrate.

Data collection and availability problems of this type are pervasive in economics. The inability to test even basic propositions has led theorists to evaluate models on the basis of factors other than predictability, such as internal consistency and elegance. For this reason, wide classes of economic models are constructed without even the thought of verifiability. The careful use of laboratory methods can help provide an empirical basis for refining and discriminating among these economic theories.

## 3. A SIMPLE MARKET EXPERIMENT

It is instructive to consider how a researcher interested in evaluating the predictions of competitive price theory might construct a simply laboratory market.[2] Suppose the researcher enters a classroom full of volunteer participants, and passes out a number of cards, which are either green or yellow. On each card a number appears, denominated in dollars and cents. The green cards are seller cards, and the printed number represents the cardholder's sales cost for selling a "unit." A participant holding a seller card could earn as profits the difference between a negotiated unit sales price and the selling cost. Thus, the individual cost figure represents a lower bound on acceptable sales prices for the card holder. A market supply schedule is induced by providing each seller with different sales costs. Arraying costs from lowest to highest generates a supply curve, like that shown on the left side of figure 1.

Demand may be symmetrically induced with the yellow buyer cards. The number printed on each buyer card represents a redemption value. An agreement to pay a buyer the difference between a negotiated sales price and the printed redemption value implies that the printed value represents a maximum willingness to pay. Varying the unit valuation steps and arraying units from highest to lowest valuation generates a demand schedule, such as that shown on the left side of figure 1. Notice that construction of supply and demand allows clear *a priori* identification of competitive price and quantity predictions. In this case the competitive price prediction is $4.70 and the competitive quantity prediction is either 6 or 7 units.[3]

The process of trading units requires construction of a *market institution*, or a set of trading rules. A

---

[2] This method of supply and demand inducement very closely resembles that used in the first market experiments. See, e.g., Chamberlin (1948), and Smith (1964, 1965).

[3] Due to the discrete number of exchangeable units, competitive predictions are characterized by some indeterminacy in either price or quantity. Early experiments were typically constructed with a unique price prediction, and some quantity indeterminacy as shown in figure 1. In this sort of design, sales of the marginally profitable units may be induced via the payment of a small, but nonnegotiable per unit sales commission. (e.g.,in the figure 1 session, buyers and sellers each received $.05 per trade.) Sales commissions are generally viewed today as somewhat undesirable, as they essentially shift the demand curve up, and the supply curve down by the amount of the commission, creating a "price tunnel" where final trades can only be struck at a unique price. Most current designs are instead characterized by a vertical overlap, admitting some price ambiguity but predicting a unique quantity.

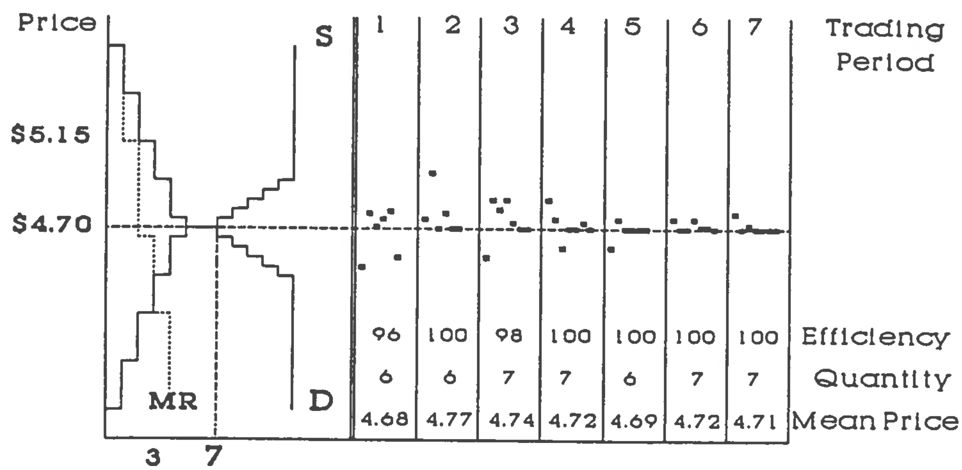


**Figure 1.** Induced Supply and Demand Arrays, and the Sequence of Contracts for a Double Auction Experiment. (Source: Session IIpda 24; Smith and Williams, 1983.)

variety of alternatives exist. Consider, for example the following *double auction* or "stock-market" rules. At the beginning of a trading period, of some fixed and specified duration, sellers may submit offers to sell, which are publicly written on a chalkboard at the front of the room. Similarly buyers may submit purchase bids, which are also displayed on the chalkboard. At any time a seller is free to make a sale at the price suggested by a buyer, or to post a new sales price. Buyers are free to accept any sellers' offer price or to submit a new bid. Each time either a buyer or seller accepts the displayed terms of exchange, a transaction is executed and written on the board, the buyer and seller hand in their cards, and they record their earnings. Trade resumes as other buyers and sellers propose new terms of trade for remaining units, and continues until the expiration of the time period.

Trades struck under these conditions in a representative trading period are illustrated as dots on between the double line and the first vertical line to the right of the double line in figure 1. The process may be repeated by again passing out the green and yellow buyer and seller cards, and starting a new fixed-duration trading period. The temporal sequence of contracts for six additional trading periods are summarized by the information presented between the last 7 vertical bars shown on the right side of figure 1.

The drawing power of competitive price and quantity predictions in the 7 trading periods shown on the right side of figure 1 is striking. Participants had no previous experience with the trading institution in this session, yet the mean contract price is within $.04 of the competitive prediction in all but one period (trading period 2). Similarly, transacted quantities were within the competitive range in each trading period.

In addition to price and quantity predictions, neoclassical price theory suggests that competitive markets are efficient in the sense that they maximize the possible gains from trade. Efficiency is typically measured in experimental markets as a surplus extraction rate, "E," which is the proportion of the realized gains from trade to the possible gains from exchange under the competitive prediction.[4] E is bounded between 0 and 100%. Under the competitive prediction E = 100%. Efficiency numbers for each trading period of the market shown in figure 1 are illustrated at the bottom of the figure. As with competitive price and quantity predictions, market efficiency extraction rates are very near the competitive prediction, at nearly 100% in every period.

## 4. SOME GUIDELINES FOR EFFECTIVE EXPERIMENTATION

*Replicability* and *control* are the principle advantages of experimentation. Replicability refers to the capacity to independently verify results with a new data set.[5] Opportunities for unilateral verification enhance the credibility of data collection, providing both a "carrot" of professional accolades for the generation of new data, as well as a "stick" of professional embarrassment if others are unable to duplicate ones' findings.[6] Control is the capacity to refine the environment in a way that abstracts out auxiliary behavioral motivations. By controlling participant incentives and information flows, a laboratory market

---

4 This efficiency measure was developed by Plott and Smith (1979).

5 Replication in an experimental context should be distinguished from the concept of replication in econometrics. In econometrics, replication refers to the capacity to reproduce one's results with a given data set. As Roth (1990) notes in an experimental context, replication is the capacity to create an entirely new set of observations.

6 Smith (1989) cogently argues that the desirability of placing professional credibility on data collection is not restricted to data collected in the laboratory. Much better data would also be generated from natural markets if it was collected by economists (rather than businessmen, or bureaucrats), who had a professional stake in the data collection process.

vastly reduces the number of auxiliary assumptions necessary to evaluate a hypothesis.

To maximize the benefits of an experimental investigation, it is critical that the investigation be structured in a way that fully exploits the advantages of replicability and control. This section offers a series of five criteria that may assist in constructing a useful experimental test: *procedural regularity, motivation, unbiasedness, calibration, and parallelism*. Importantly, these guidelines should not be viewed as a laundry list of conditions to be satisfied before an experiment can be viewed as valid. While some characteristics (such as unbiasedness and calibration) may be fully satisfied *ex ante*, other features (such as motivation or parallelism) must be addressed, but cannot be fully verified. The satisfaction of these latter criteria is frequently the focus of debate over the legitimacy and relevance of an otherwise well-constructed and executed experiment. Consider these terms, in sequence.

a. *Procedural regularity.* Standardization of procedures is critical for replication, and thus, the guiding principle for standardizing and reporting instructions and procedures should be to permit a replication that both the researcher and outside observers would accept as being valid. The scope of standardization necessary for valid replication is wide. In addition to the text of instructions the experimenter should report any illustrative examples used, any tests of subjects' understanding, the protocol for answering subjects' questions (e.g., no information beyond instructions), the nature of monetary or other rewards and the presence of "trial" or practice periods with no rewards. Procedural standards that should be adopted and reported include information regarding the subject pool, the method of recruiting subjects, the number and experience levels of subjects, procedures for matching subjects and roles, the location and typical duration of experimental sessions, any intentional deception of subjects, and any procedural irregularities in specific sessions that may require special interpretation. Journal space limitations will often preclude the publication of many of these aspects. But they should be made available to journal referees and to interested readers upon publication.

b. *Motivation.* Even under the most tightly controlled laboratory circumstances, the motivation for individual action cannot be directly observed. Motives for action in the laboratory may be diverse. In addition to "doing well" by maximizing otherwise meaningless points or hypothetical dollars in a laboratory experiment, for example, participant actions may be motivated by boredom, an eagerness to satisfy the perceived objective of the researcher, or a desire to minimize the time spent in the laboratory. To enhance the likelihood that participants are motivated in accordance with the incentives that are the subject of investigation, participants are usually paid a monetary reward for participation. Critically, the rewards must be salient, in the sense that rewards vary both directly and prominently with the reward medium specified in the laboratory. There is little reason to expect money payments to affect decisions if substantially identical earnings are realized regardless the decisions made.

This is not to say that all non-salient payments are undesirable. It is standard to pay participants an up-front appearance fee (usually $3.00) in addition to the salient rewards. This fee has the administrative advantages that it induces prompt arrival and provides participants some incentive to pay attention to instructions. An up-front participation fee can also help to assure participants that there are no "tricks;" e.g., they will be paid cash as promised.

The selection of a given level of financial motivation as adequate, is unfortunately somewhat arbitrary. Clearly there are diminishing benefits to increased payments. Although behavioral variability tends to diminish with increases in salient financial incentives, the change in behavior is not monotonic. Individuals are bound by constraints of cognition and in some instances social convention. Simply increasing financial incentives will not generate predicted responses. Just as the lure of $6 million dollars could not induce either of these authors to dunk a basketball (despite our best efforts), increased rewards in the laboratory will not enable participants to performance tasks beyond their cognitive skill level. As a general practice, laboratory earnings for optimizing decisions are set so as to slightly exceed the opportunity cost of the participants' time in the laboratory.

c. *Unbiasedness.* The drawbacks of explicitly suggesting responses to experiment participants are clear, and incentives to replicate unusual laboratory results should temper the creation of intentionally deceptive research. Rather, this section is written to highlight the ease of introducing unintended bias, through inadvertent behavioral references (such as "collusion" or "conspiracy") and through verbal emphasis on certain terms. The laboratory researcher must be quite careful to not inadvertently suggest particular types of behavior. Careful attention to wording in experimental instructions can mitigate the effects of possibly pejorative language. The use of computers to present instructions can facilitate control over "experimenter" effects in both the reading of instructions and in the administration of laboratory sessions.

An interesting issue regarding bias and instructions has to do with the amount of context-related language to use in instructions, for there is a clear tradeoff. For example, in evaluating oligopoly market performance, obvious potential biases are presented by mentioning "cartels," "collusion" or "defections." On the other hand, the task of explaining the structure of the market to the participant is greatly facilitated by using market-specific references to "buyers," "sellers," and "contracts." To fully appreciate the benefit of these latter terms, consider for a moment the problem of explaining to participants the workings of the double auction described in the section 3 without any references to buyers, sellers, unit costs, unit valuations and purchases or sales. Just what are participants doing if they are not buying and selling units of a commodity?

Unfortunately, we can offer no clear guidance as to the location of the line between eludicatory and pejorative language. As an alternative, we recommend starting the construction of any new experiment with standard instructions from published sources. Experienced and reputable experimenters have carefully considered sources of unintended bias in the creation of their instructions, and by their example they may help a new investigator avoid many common undesirable instruction effects.

d. *Calibration.* Consider the sample market experiment discussed in section 3, and illustrated in figure 1. Although observed price, quantity and efficiency extraction rates appear to conform quite nicely to competitive predictions in figure 1, it would be rather difficult to decide just what sort of alternative pricing performances would also be "close enough" to warrant a conclusion that a market was competitive. In fact, from a statistical perspective we cannot conclude that even the very stable price series observed in figure 1 conforms to competitive predictions. Rather, statistical inquiry is a process of falsification. With only the competitive hypothesis as a standard we would, at best, be unable to reject the hypothesis that the market illustrated in figure 1 generated competitive predictions.

Much stronger statistical support for a hypothesis can be generated through the process of rejecting rival predictions. For example, monopoly and monopsony predictions represent natural alternatives to the competitive markets hypothesis. As can be seen by the marginal revenue curve (MR) in figure 1, the optimal price and quantity choices for a monopolist (or cartel) are $5.15 and 3 units, respectively. Sufficient replication of markets conducted in the figure 1 design with different groups of participants might allow rejection of the monopoly hypothesis. The empirical consequences of a monopsony hypothesis could similarly be evaluated with the data relative to the competitive markets hypothesis. This is the practical problem of calibration. Theories are much more meaningfully evaluated in light of alternatives. Rejection of alternative behavioral motivations, combined with a failure to reject the maintained (research) hypothesis provides strong support for the research hypothesis.

There invariably exists some behavioral "noise" in any experiment. The residual price variability in the final periods of the market shown in figure 1, for example, is a general characteristic of the robustly competitive double auction. In light of this behavioral noise, we offer two additional comments on the issue of calibration. The first is a design issue. Rival theoretical predictions are a necessary but not a sufficient condition for a well-constructed experiment. In particular, it is extremely desirable to construct an experiment so that the behavioral consequences of rival theoretic predictions are sufficiently distinct that behavioral differences can be distinguished from inherent performance variability. It would be very difficult to distinguish seller-profit-maximization from competitive performance in the market experiment illustrated in figure 1, for example, if the demand curve was so elastic that the price difference separating the predictions was only $.05.

The second issue has to do with anticipated performance variability that is outside the domain of the theory. Although some behavioral variability is pure, irreducible noise there exist other theoretically irrelevant factors that quite regularly affect performance, such as experience with the experimental environment, group effects, and the order in which treatments are presented. In order to draw legitimate statistical claims regarding performance, it is important to control for these anticipated sources of variability. This can be done with the careful use of blocking, which is systematic variation across expected sources of variability. For example, the market shown in figure 1 used 4 sellers and 4 buyers. Suppose we were interested in evaluating the effects of reducing to one the number of sellers on performance in the market design illustrated in figure 1. A variety of predictable but theoretically irrelevant sources of variability might affect observed behavior. Some groups, for instance, may simply be more competitive than others. The effects of either particularly rivalistic or particularly cooperative groups may be controlled by conducting both the 4 and 1 seller treatments in a single session. Similarly, monopolists may price more (or less) aggressively after having been in a competitive market. Alternating the order of the treatments could control for sequence effects of this sort. Independent of a treatment sequence effect, some individuals may be more intent on manipulating prices favorably than others, particularly as monopolists. Thus, there could also be individual effects. If it is not possible to construct a design where each seller holds the monopoly position, then the method of selecting the monopolist should be completely random (determined, for example, by the roll of a die).

e. *Parallelism.* Not all experiments are (or need to be) constructed to test formal theory. Useful insights into the likely performance effects of newly created trading institutions, for example, have been generated via laboratory experimentation. Relevant examples include the construction of auctions to allocate slots for commercial and scientific projects on the NASA space station (e.g. Banks, Ledyard and Porter, 1987), and auctions for transportation rights to natural gas pipelines (McCabe, Rassenti and Smith, 1988; Plott, 1988). These examples share the characteristic that demand for one good is dependent on the price and availability of related goods. Demand for transportation rights over a given pipeline segment, for example, is sensitive to the availability of transportation links connecting the segment to the wellhead and to the consumer. Similarly, the demand for a physical slot on a space station is importantly affected by the availability of astronauts to service and monitor the installed hardware and other characteristics, such as exposure of the physical slot to the sun. Simply identifying a set of reasonable equilibria in complex environments of this type often strains the modelling tools of the skilled theorist. Theory, moreover, can often provide little insight as to the dynamic performance of such markets, or to the selected equilibrium. A laboratory experiment that incorporates features parallel to the relevant problem in natural market can provide some insight regarding these issues of dynamic performance and equilibrium selection.

The usefulness of even well-constructed experiments of this sort lies in the inclusion of parallel elements perceived as critical to the natural context under consideration. As with the issue of subject motivation, parallelism is an issue that cannot indisputably satisfied. Rather, debate over the applicability of the study will rest on the success with which the relevant problem was induced in the laboratory.[7]

A final issue in parallelism regards subject selection. College undergraduates are typically selected a participants, due to their availability, and to the low opportunity cost of their time. Of course, these participants are less sophisticated, and know less about the natural context than the actual agents involved. As a general matter, we do not think that it is necessary to use participants from the natural parallel circumstance. The costs of conducting experiments of this type can be exceedingly high due to the opportunity cost of time to skilled professionals. Moreover, the additional benefits of using such participants is often rather low. A variety of experiments have been conducted with "relevant" participants, and generally there are no significant behavioral differences (see e.g., Mestalman and Feeny, 1988, DeJong, Forsythe, Lundholm, and Uecker, 1985, and Smith, Suchanak and Williams, 1988).[8] Given a sufficiently simple design and sufficiently motivated participants, the behavioral aspects theory can be given a "best shot" even without participants from the parallel natural context.

This is not to say that concerns regarding subject sophistication are unwarranted. There are a

---

7 Charles Plott has likely done more experimental investigations with public policy implications than anyone, and does a commendable job addressing this issue of parallelism. See in particular Plott (1987).

8 Some differences between behavior of the undergraduates and "relevant professionals" in the laboratory have been observed. Dwyer, Kagel and Levin (1989) and Burns (1985) for example, find that businessmen involved in markets that are relevant to the experiments sometimes attempt to apply certain rules of thumb. Although these rules of thumb are useful for dealing with uncertainty in the naturally occuring markets, these patterns of behavior are generally ineffective in the laboratory.

number of studies that indicate that performance varies with proxies for the aptitude of participants, such as the undergraduate institution (e.g., Davis and Holt, 1990) or using graduate instead of undergraduate students. In some instances it may be useful to appropriately restrict the participant pool. The motives for placing restrictions of this type on the subject pool should be carefully considered, and explicitly reported.

## 5. SOME CENTRAL RESULTS OF ECONOMIC EXPERIMENTATION

The experimental literature within industrial organization alone is extensive, and any attempt to categorize results would clearly exceed the scope of this paper. A number of excellent surveys of experimental findings are available.[9] Rather than attempt to present a compendium of results, this section briefly discusses the origins of two types of economic experiments, and then presents some of the central lessons learned.[10] We illustrate these conclusions by showing the results of sample laboratory session to give the reader some feel for designs, and the data generated in the laboratory. The discussion is divided into three parts. In a first subsection, we discuss principle results and applications of market experiments, which were introduced in the initial example. The second subsection reviews the origins and some central findings of experimental examinations of the predictions of noncooperative game theory. The third subsection contains a brief summary.

### Market Experiments

The experimental investigation of markets was initiated by Chamberlin (1948) who believed that aspects of the Great Depression could be explained by certain deficiencies in neoclassical price theory. Frustrated by the absence of relevant data from the natural world, Chamberlin induced costs and values in a classroom market, in a fashion very similar to that explained above (see figure 1). Unlike the market

---

9 Standard references include Plott (1982, 1989), Roth (1988) and Smith (1982).

10 We ignore a third type of economic experiments, e.g, individual decision-making experiments. These experiments grew from efforts to evaluate the credibility and consistency of assumptions in Savage/von-Neuman Morgenstern expected utility theory, and this literature has generated a number of very provocative findings. (See e.g., Thaler, 1987) We omit discussion of these important experiments, as we feel that they are less applicable to trade problems.

discussed above, however, Chamberlin, allowed participants to mill about the classroom and privately negotiate contracts.

Chamberlin's markets did not generally achieve competitive outcomes. In particular the quantity exchanged persistently exceeded the competitive prediction. Vernon Smith, who observed the experiments as a graduate student, thought that the privacy of communications in Chamberlin's "negotiated price" trading rules failed to give competitive price theory a "best shot," and structured an alterative experiment under the double auction rules described above. Smith's (1964, 1965) support for competitive price theory generated little more initial interest from the profession at large than Chamberlin's rejections. But Smith became intrigued by the effect of trading rule manipulations on market outcomes and initiated a stream of experimental research into the performance of markets. The robustness of competitive price predictions and the sensitivity of markets to alterations in the trading rules defining the institution of exchange represent perhaps the central themes of this research stream. We elaborate upon these broad conclusions below.

a) *Competitive predictions generally tend to organize experimental market data very well -- particularly those organized under double auction trading rules.* As mentioned above, the tight adherence of the contracts to the competitive price and quantity predictions shown on the left side of figure 1, is characteristic of a wide body of experimental markets. To illustrate this point, it is instructive to review results of two double auction market sessions reported in studies designed to assess the boundaries or limits of application for competitive price predictions for double auction markets. The dimensions considered are: (a) the distribution of producers' and consumers' surpluses and (b) the number of sellers necessary to reliably generate competitive predictions as outcomes.

Consider first the question of the range of surplus distributions between buyers and sellers that will behaviorally generate competitive price and quantity outcomes. This boundary may be tested in a market where a number of sellers all have common and constant costs, and where a number of buyers all share constant and common unit values. By varying the number of units available to buyers vis-a-vis sellers the

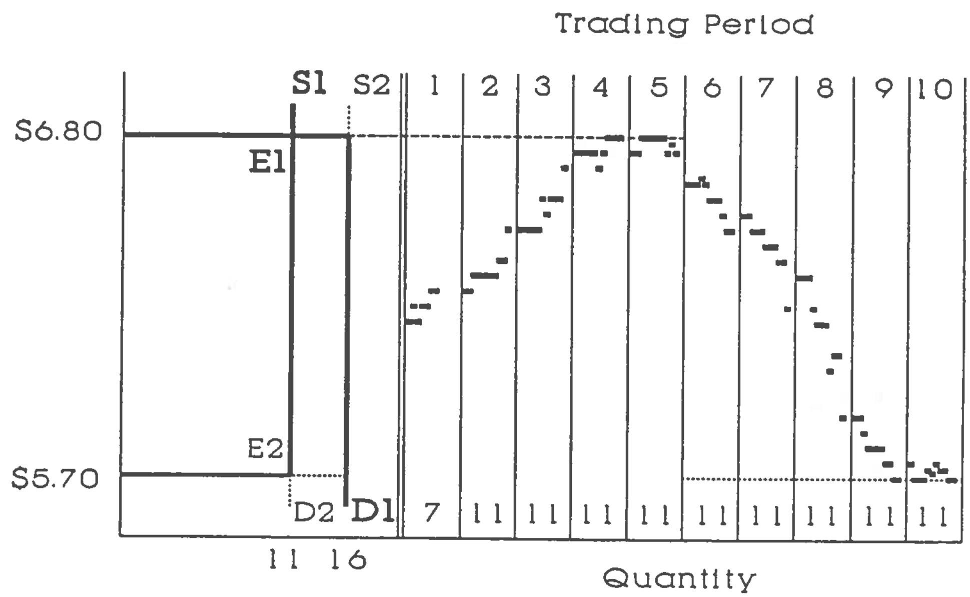


**Figure 2.** A Double Auction Market Session Conducted Under Severe Earnings Inequities (Source: Holt, Langen and Villamil, 1986).

entire surplus will go either to buyers or to sellers in the competitive prediction. This situation is illustrated in the supply and demand arrays shown in figure 2. This market is composed of 4 sellers and 4 buyers. Equilibrium E1 is characterized by excess demand: demand curve D1 is characterized by 16 units at a constant unit value of $6.80 per unit, while supply curve S1 is composed of 11 units are at a constant unit cost of $5.70 per unit. Given the excess demand and the cost unit valuation, all the surplus goes to sellers at a competitive price of $6.80. The surplus distribution condition is reversed by creating excess supply in equilibrium E2. In E2 demand is reduced to 11 units and supply is increased to 16 units.

The results of an oral double auction market conducted in this design reported by Holt, Langen and Villamil (1986) is illustrated by the sequence of contracts shown on the right side of figure 2. In the first five 5-minute trading periods, supply and demand conditions S1 and D1 were in effect. After period 5, some sellers were given an extra unit and units were taken away from some buyers in an unannounced shift to supply and demand conditions S2 and D2. As suggested by the sequence of contracts, the drawing power of competitive predictions are robust to extreme surplus distribution predictions. Competitive outcomes are generated even when one side of the market receives none of the available surplus. The results of this session are representative of other experiments in similar designs by Smith (1965), Smith (1980) and Smith and Williams (1989).

As a second boundary issue, consider the minimum number of sellers necessary to ensure a competitive price/quantity outcome. Figure 3 illustrates the results of a monopoly double-auction session, reported by Smith (1981). Notice from the sequence of contracts in the figure that the monopolist was singularly unsuccessful in raising the price to the monopoly prediction of 110. In fact, prices in later periods were below the competitive prediction of 80. Some double auction monopolists fare better than the seller shown in figure 3. Many, however, do not. Markets with two sellers generate competitive predictions even more regularly, and some commentators (e.g. Smith and Williams, 1989) suggest that double-auction markets with as few as three sellers are robustly competitive. In any event, a large number of sellers are clearly unnecessary to ensure competitive outcomes.

Competitive predictions have been robust to a variety of other boundary examinations, including between-period shifts in market supply and demand arrays (Williams, 1979, Williams and Smith, 1983, Davis, Harrison and Williams, 1991), multiple commodities (Smith, Williams and Ledyard, 1985) and even explicit

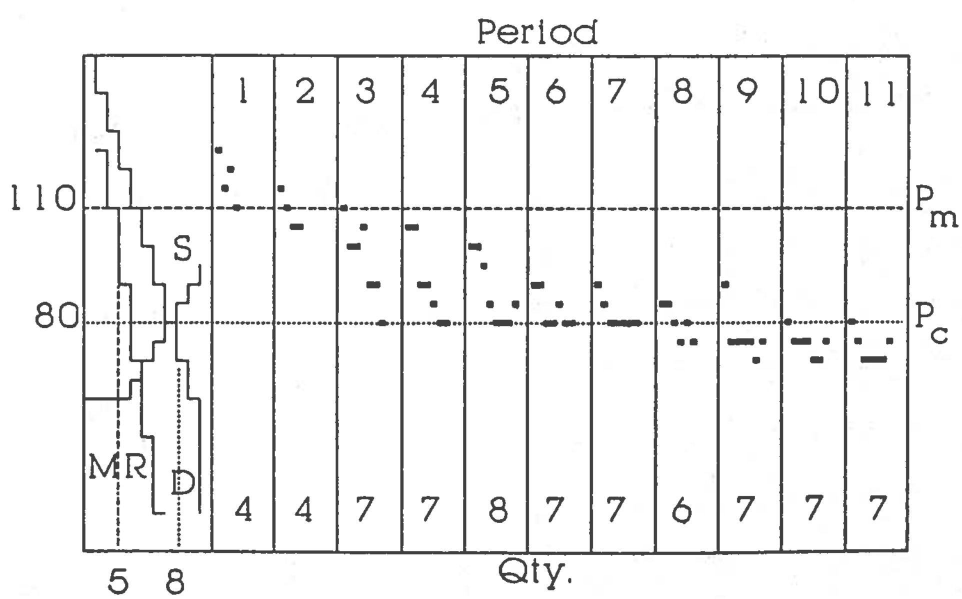


Figure 3. A Double Auction Monopoly. (Source: Smith, 1981)

conspiracy (Isaac and Plott, 1981). Competitive predictions appear to break down in double auctions only occasionally, and only under very extreme circumstances, such as completely random supply and/or demand shocks, or markets characterized by a single seller.

b) *The trading rules defining the institution of exchange matters a great deal.* Markets may well be organized under institutional trading rules other than the double auction, and these alternative structures can importantly affect market performance. Consider, for example, the effects of shifting from double auction trading rules to a posted offer, or retail-type institution. The predominant characteristic of the posted offer is that sellers publicly submit binding price offers that remain in effect for some period of time.

The standard laboratory implementation of the posted offer proceeds as a series of 2-step trading periods (see e.g., Ketcham, Smith and Williams, 1984). Within each period, sellers first privately and simultaneously make a price and quantity posting decision. Prices (but not quantities) are then publicly displayed to the market. After all sellers have posted prices a shopping sequence begins, where buyers are drawn one at a time from a waiting mode to make purchases. Each buyer may make as many purchases as desired from sellers on a take-it-or-leave-it basis, at the posted price. The period ends when all buyers are out of units.

Markets organized under posted-offer trading rules are behaviorally similar to double auction markets in many respects. In particular, posted-offer markets tend to competitive outcomes in a wide variety of circumstances. Posted-offer markets, however, tend to the competitive equilibrium prediction from above, and they converge somewhat more slowly than comparable double auction markets. Moreover, the range of application for competitive predictions, is much more limited in posted offer markets that in double auctions. Consider, for example, the sequence of contracts for the posted-offer monopoly market, shown in figure 4. This market, reported by Smith (1981), was conducted under the same supply and demand conditions used in figure 3. In stark contrast to the double auction monopoly illustrated in figure 3, the posted offer monopolist achieved both high and stable prices at the monopoly level at the end of the session.

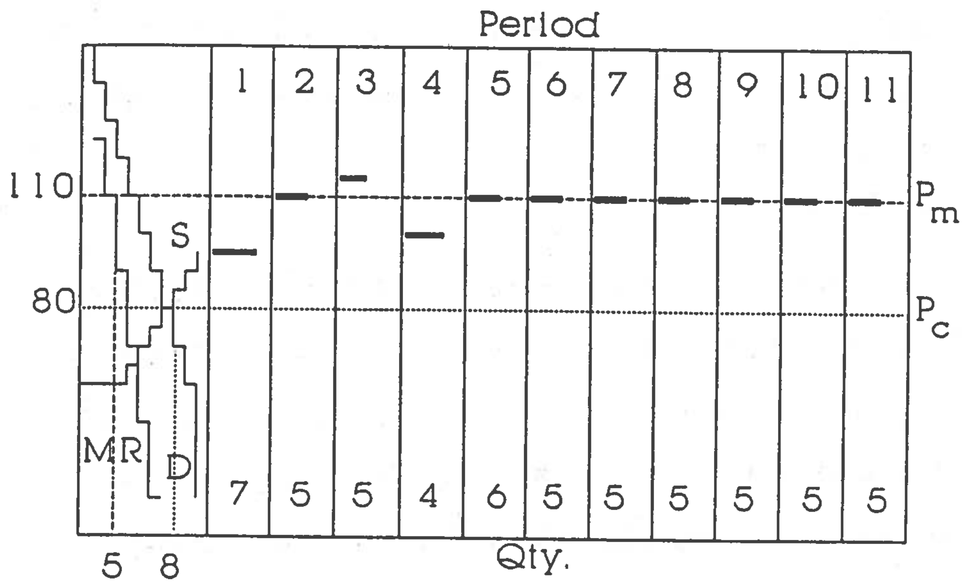


**Figure 4.** A Posted Offer Monopoly. (Source: Smith, 1981)

We would be sorely remiss if we failed to comment that not all posted offer monopolists are as successful at extracting the monopoly price prediction as the monopolist shown in figure 4. Figure 4 is representative, however, in the sense that posted offer monopolists quite generally extract supra-competitive prices (see, e.g., Isaac, Ramey and Williams, 1984). Posted-offer markets have also been shown to be susceptible to explicit conspiracy (Isaac, Ramey and Williams, 1984), tacit conspiracy in duopolies of long duration (Alger, 1987, Bensen and Faminow, 1988 and Stoeker, 1980) and market power (Davis and Williams, 1991, Davis, Holt and Villamil, 1990). Markets organized under posted offer trading rules also respond poorly to unannounced between-period shifts in market supply and particularly market demand curves (Davis, Harrison and Williams, 1991).

This "posted-offer effect" has important policy implications. For example, as an experiment by Hong and Plott (1982), suggests barge operators on the Mississippi had good reason to be very skeptical of proposals (by the railroads) to limit barge freight-rate negotiations and require public, binding, price postings. In a laboratory setting that paralleled the demand and concentration characteristics of the market for barge transportation services on the Mississippi river, Hong and Plott found that public price posting tended to generate higher prices, making alternative transportation services, such as the railroads more competitive. Plott (1986) similarly reports that concerns about the effects of binding public price postings discouraged the Civil Aeronautics Board from imposing "posting rules" in a market for airline landing slots. There are many other examples. Rather than highlight the effects of this particular trading rule manipulation, however, our purpose is to document the importance of trading rule adjustments on market performance in general. The effects of institutional or trading rule adjustments on performance stands as one of the most prominent results of experimental inquiry. Manipulations in the structure of voluntary contributions mechanisms for public goods provision, for example, have been shown to dramatically alter public goods provision levels (e.g., Isaac and Walker, 1988), and changes in the structure of single seller auctions importantly affect both the revenue and efficiency characteristics of auctions (e.g., Cox, Roberson and Smith, 1982).

In summary, the examination of market behavior in the laboratory is largely a study of the effects of institutional adjustments on market performance. The extreme resiliency of competitive predictions in markets organized under the double auction trading institution has led to use of double auction markets as the behavioral standard against which performance of rival institutions is evaluated. Double-auction trading rules are interesting in their own right, due to parallels of double auction rules to characteristics of many modern financial markets. Deviations from double auction rules can importantly affect market performance. In particular, competitive predictions are consistently generated in a much narrower range of circumstances in markets organized under a posted-offer trading institution.

**Tests of Noncooperative Game Theory**

For ease in presentation, most of the discussion to this point has been in terms of market experiments. A very large body of economics experiments, however, have been conducted to evaluate the predictions of noncooperative game theory. Experimental tests of game-theory originated with psychologists, sociologists and political scientists who were skeptical that individuals would jointly reason themselves to Pareto-inferior outcomes in a prisoners' dilemma context. The relationship of simple generalizations of the prisoners' dilemma game to problems of oligopoly cooperation quickly sparked interest among economists desiring to evaluate the predictions of Cournot and Bertrand models of noncooperative behavior. (e.g., Fouraker and Seigel, 1963, Murphy, 1966, and Dolbear et al., 1968). These experimental investigations focused on behavioral issues under conditions that more nearly matched the conditions of noncooperative game theory. For this reason, elements of natural markets (such as pricing rules, and buyer decisions) were subsumed and participants were presented with more information than is usually made available in a market experiment regarding, for example, information about demand as well as the costs of other sellers. Thus these investigations are even closer to pure-theory tests than market experiments, and the value of such experiments lies, in large part, in the relationship between the environment created in the laboratory and the relevant theory.

There has been a resurgence of interest in experiments testing equilibrium predictions of noncooperative game theory lately, with much of the current focus on the behavioral properties of Nash equilibrium refinements, the resolution of coordination problems, and on noncooperative bargaining theory. Again, although it is well beyond the scope of this paper to comprehensively review the experimental game literature, we offer a few broad conclusions from the laboratory analysis of noncooperative games.

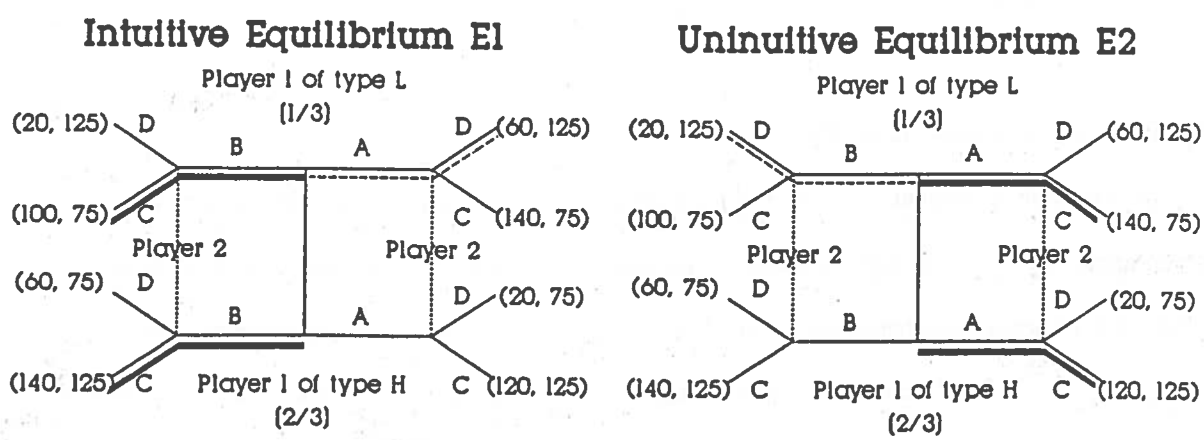


**Figure 5.** A Signalling Game (Source: Brandts and Holt, 1991)

1) *Laboratory researchers have identified baseline environments in which a surprising variety of equilibrium predictions organize data very well.* This conclusion may be illustrated by considering decisions made by laboratory participants in an asymmetric information signalling game reported by Brandts and Holt (1991). The extensive form of this game is shown in both the left- and right-hand panels of figure 5. The game is composed of two player-types, player 1 and player 2, who sequentially make single decisions. Player 1 has an attribute, L or H, which is known to player 1, but is unobservable *ex ante* to player 2. Player 2 knows only that player 1 is from a population where 2/3 of its members are type H. The game proceeds in two steps: Player 1 picks action A or B, then player 2 selects a response, C or D. In either panel of figure 5, actions taken by a player 1 of type L are shown in the upper portion of the panel. Action choices for a player 1 of type H are shown in the lower portion of the figure. Player 2 responses, and (player 1 payoff, player 2 payoff) payoffs are shown about the perimeter of each panel. The vertical dotted lines represent an information set for player 2, e.g., player 2 makes a decision with only the knowledge that player 1 has selected either action A, or action B.

The left and right panels of figure 5 illustrate the two sequential equilibria for this game. Plays along the equilibrium path are demarked by bolded lines in each panel. There are two bolded lines for each strategy, one for the case where the type 1 player is of type L, and the other for the case where the type 1 players is of type H. As is readily apparent from inspection of the panels in figure 5, both equilibria involve "pooling" in the sense that player 2 will not be able to discriminate among L and H type player 1's on the basis of the signal player 1 sends with the decision to play A or B.

Consider equilibrium E1, shown in the left panel of figure 5. In E1 player 1 selects action B, regardless of type, and player 2 responds by picking C. It can be verified that this is a sequential equilibrium by considering the gain to a unilateral deviation by either player. For player 2, action C is an equilibrium response because it is more likely that a player is of type H than type L, making the 125 payoff more likely than the payoff of 75. Verifying the optimality of action B for player 1 involves a more subtle calculation, as it depends on player 2's response to a deviation by player 1, which in turn depends on the assumption player 2 makes about the player 1 type most likely to deviate. In equilibrium E1, player 2 assumes that type L player 1's are more likely to select action A. Player 2's assumption about who deviates, as well that players best response (action D) is illustrated by the dashed line. Given player 2's response, player 1 will find deviation from action B unprofitable, regardless of type. Hence E1 is an equilibrium.

By similar reasoning, equilibrium E2, shown on the right side of figure 5, is also a sequential equilibrium; action C maximizes player 2's expected income, given both H and L type players select action A. These outcomes are highlighted by the bolded lines. As indicated by the dashed line in the right panel of figure 5, it is critical that player 2 assume that a player 1 of type L is most likely to deviate from action A. Given this assumption, response D is an equilibrium response by player 2, and neither the H nor the L type player 1 would find deviation profitable.

The laboratory investigation of this game presents nested behavioral issues. First, does the principle of sequentiality have any behavioral appeal, e.g., do humans tend to select either or both of the equilibria just discussed? Second, given that sequentiality has some behavioral appeal, which of the two equilibria tend to be selected? An "intuitive criterion" equilibrium refinement proposed by Cho and Kreps (1987) suggests that E1 would be selected. This intuitive criterion discriminates among equilibria on the basis of the plausibility of beliefs off the equilibrium path. In the present context, equilibrium E1 is "intuitive" because it is reasonable for player 1 of type L to deviate from a B action. Other things constant, a unilateral deviation by player 1 to action B would increase earnings for player 1 of type L from 100 to 140. Equilibrium E2, on the other hand, is "unintuitive" in the sense that it is not plausible to suppose that a player 1 of type L would most likely deviate from an action A choice. Even under the best of circumstances earnings for player 1 of type L would fall from 140 to 100. Rather, it is more reasonable to believe that player 1 of type H would deviate from action A. But such a belief would break the equilibrium, as both player 2, and player 1 type H would find unilateral deviation profitable.

The point of this exercise is that despite the labelling of this equilibrium definition as "intuitive" it relies on rather subtle assumptions about anticipated actions. Yet, surprisingly, after some experience with the game, participants (who had no training in game theory at all) tended not only to coordinate on one of the two sequential equilibria, but that they overwhelmingly selected "intuitive" equilibria E1. Of 112 signalling games conducted in the last two-thirds of each session, 101 matched equilibrium E1, and only 7 corresponded to E2. Baseline environments supporting these and even more subtle equilibrium refinements are reported by Camerer and Weigelt (1988), and by Banks, Camerer and Porter (1988).

2) *In some respects, however, game theoretic predictions do not explain human behavior well at all.* A recent series of experiments investigating performance of Nash equilibrium predictions in the Rubenstein/Stahl noncooperative bargaining game provide two convenient examples which illustrate this conclusion. First, fairness or equity concerns impede the attainment of Nash equilibrium predictions in some contexts. Consider the following ultimatum game: Two players bargain over a fixed monetary prize, For convenience, assume a dollar is divided. Bargaining takes place in a two-step sequence. First, one player (the "chooser") proposes a division of the dollar. Second, the other player (the "decider") elects whether or not to accept the proposal. In the unique Nash equilibrium for this ultimatum game, the chooser offers a division that gives the decider only $.01, and takes the remainder. The decider, faced with options of either $.01 or nothing, accepts the offer. Experimental investigation of this game reveals that the Nash equilibrium does not organize behavioral data well at all. Deciders tend to frequently reject "unfair" offers in excess of about a 60/40 split.[11]

A second instance where human behavior and Nash equilibrium predictions persistently diverge is seen in a finitely repeated version of this game with discounting. Consider, for example, a three-stage version of the above ultimatum game. The first stage proceeds as described above, except that in the event that the decider elects to reject the proposed division, the dollar prize is reduced in value to $.50, the players switch roles, and a second stage begins. A rejection by the decider in the second stage will again prompt a reversal of roles, cause the prize to shrink again by one half (to $.25), and initiate a third stage. The game terminates after the third stage, with both players receiving payoffs of 0 in the event of a rejection after the

---

[11] Roth, Prasnikar, Okuno-Fujiwara and Zamir (1990) find that the notion of a minimum acceptable division varies across cultures. These authors conducted bargaining experiments of the type mentioned in the text in the United States (Pittsburgh), Israel, Japan and Yugoslavia. Participants in Israel and Japan made much less equitable proposals than did participants in the U.S. or Yugoslavia, but there were no significant differences in decider acceptance rates across the countries.

third stage. The unique, subgame-perfect equilibrium solution to this game is readily apparent via backward induction:[12] In the third stage, the chooser could ask for (and get) slightly less than $.25. Given this baseline return, proposed divisions in excess of a $.25/$.25 split should be rejected in the second stage, implying that a $.75 /$.25 proposal in first stage constitutes an equilibrium for the game.

A variety of experiments have been conducted to evaluate performance in variants of this game with differing discount rates and differing numbers of periods. (See e. g., Binmore, Shaked and Sutton, 1985, Neelin, Sonnenschien and Speigel, 1988, and Ochs and Roth, 1989). The general result of these experiments is that humans do not tend to be good "backward inducters." Rather, the participants tend to settle on a division of the prize in the first stage that gives the decider the portion of the prize that discounts in going to the second stage. In the current example, this implies a $.50/ $.50 split of the dollar.

Importantly, the principle of backward induction is not devoid of empirical content in this context. Rather, as Harrison and McCabe (1988) report, subgame perfect Nash equilibrium divisions of the pie tend to occur, but only after participants have had considerable experience with the game as a whole, and particularly with the latter stages of the game. In what has come to be termed the "swing-back" hypothesis, participants appear to learn the Nash equilibrium recursively, appreciating first the Nash equilibrium for the final stage game, and then the equilibrium for the second to last stage, and etc. on to the first stage. This "swing-back" effect has been observed in a variety of other laboratory contexts, including a series of thirty 10-stage repeated prisoner dilemma game, reported by Selten and Stoeker (1986), and in a series of iterated 3-period asset market games, reported by Anderson, Johnston, Walker and Williams (1991). The effect suggests that equilibrium predictions which rely on backward induction are most likely to apply in contexts where experienced participants have repeatedly observed outcomes in all subgames of the relevant game.

*3) Various institutional adjustments can very prominently affect outcomes in particular circumstances.* This third conclusion, in some sense, is an analogue to the effects of the trading institution selection on

---

[12] There also exist an infinite number of Nash equilibria that are not subgame perfect. These equilibria depend on threats of plays off the equilibrium path that are not credible, in the sense that they are not equilibrium plays.

market performance. We illustrate this conclusion by discussing a series of experiments designed to evaluate factors that help resolve equilibrium selection in games with multiple equilibrium. The well-known "Battle-of-the-Sexes" game, illustrated in figure 6 is a standard example of a single stage game characterized by multiple equilibria. In figure 6, the row and column players simultaneously choose actions 1 or 2. From inspection of the payoffs, it is readily apparent that there are 2 pure-strategy equilibria in figure 6; [R2, C1], with (row, column) payoffs of (600, 200); and [R1, C2] with payoffs of (200, 600).[13] Now, although the row and column players prefer different equilibria in a single-stage game, both players are better off coordinating on either of the equilibria than they would be in either the upper-left or the lower-right corners. The multiplicity of equilibria presents a formidable behavioral problem. Abdalla, Cooper, DeJong, Forsythe and Ross, 1989 (hereafter "ACDFR") report that disequilibrium outcomes occurred in 59% of games conducted in the last half of their experimental session.

| | | Column Player's Strategy 1 | Column Player's Strategy 2 |
|---|---|---|---|
| Row Player's Strategy | 1 | 0, 0 | 200, 600 |
| | 2 | 600, 200 | 0, 0 |

**Figure 6.** A Battle of the Sexes Game (Source: Abdalla, Cooper, DeJong, Forsythe and Ross, 1989)

A number of institutional modifications have been proposed to resolve this equilibrium selection problem. For example, suppose that prior to making a decision for the figure 6 game the row player is given an outside option of taking a certain payment of 450 in lieu of playing the game. With the addition of this outside option, equilibrium [R2,C1] uniquely satisfies a principle of *forward induction*, proposed by Kohlberg-

[13] There also exists a mixed strategy equilibria where each player randomly picks among their strategy choices with probability of .5. Expected payoffs in the mixing equilibria are (400, 400). The presence of the mixing equilibrium does not affect the capacity of an outside payoff option of 450 for row (described below in the text) to make [R2, C1] a unique equilibrium by forward induction, because payoffs to the row player in the mixing equilibrium are less than earnings from taking the option.

Mertens (1986): The row player would only forgo the outside option if earnings were increased by playing the figure 6 game, e.g., if the equilibrium [R2, C1] with payoffs of (600, 200) was selected. ACDFR find that the outside option very effectively resolves the equilibrium selection problem in this context as predicted by forward induction. Over the last half of the games conducted in each experiment session, the addition of the outside option increased selection of the [R2, C1] equilibrium from 19% to 72% of all outcomes.

The addition of an outside option, however, only resolves coordination problems in certain contexts. Consider, for example, the 3x3 coordination game illustrated in figure 7. Incentives are somewhat more complex in figure 7 than in the Battle-of-the-Sexes game. From inspection of figure, it is evident that there exist two pure-strategy Nash equilibria this game: [R1, C1] and [R2, C2]. Both players strictly prefer equilibrium [R2, C2] to equilibrium [R1, C1], but [R1, C1] is less risky, in the sense that the row (column) player is certain to earn 350 by playing R1 (C1).

As with the Battle-of-the-Sexes game, ACDFR (1989) report that this game presents a behavioral coordination problem. Players overwhelmingly select equilibrium [R1, C1] in this game: Over the last half of their reported trials [R1, C1], was chosen 70% of the time, while [R2, C2] equilibrium was played in only 2% of the time. However, the addition of an outside option for row that makes equilibrium [R2, C2] unique by forward induction, does little to improve the selection of that equilibrium. With an outside option, [R2,C2] was chosen in only 3% of the trials in the last half of their experiment. Rather, row typically selected the outside option (73% of the time).

Following theoretical analyses of the effect of communications on equilibrium selection by Farrell and Maskin (1990) and others, ACDFR find another alteration in the structure of the game substantially resolves the coordination problem in both the Battle-of-the-Sexes and the coordination games. Prior to a play of either game, ACDFR allowed the row player to communicate a nonbinding intended play to the column player. The richer message space allowed by these "one-sided, nonbinding communications" very effectively improved performance in both games. Over the last half of the games in a session, the [R2, C1] equilibrium was selected 96% of the time in a Battle-of-the-Sexes game. Similarly, the selection rate for the Pareto-dominant [R2,C2] equilibrium improved to 67% in the last half of sessions.

| | | Column Player's Strategy | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Row Player's Strategy | 1 | 350, 350 | 350, 250 | 1000, 0 |
| | 2 | 250, 350 | 550, 550 | 0, 0 |
| | 3 | 0, 1000 | 0, 0 | 600, 600 |

**Figure 7.** A Coordination Game (Source: Abdalla, Cooper, DeJong, Forsythe and Ross, 1989)

Still richer communications are necessary to resolve equilibrium selection problems in other environments. For example, Isaac and Walker (1988) found that face-to-face discussions dramatically improves contributions in a public goods experiment. Similarly, Isaac, Ramey and Williams (1984) report that sellers effectively capitalize on opportunities for explicit conspiracy in posted offer markets. Even face-to-face communications fail to always coordinate behavior, however. Isaac and Plott (1981), for example, report that sellers in a double auction market were generally unable to implement and maintain a conspiracy despite communications between market periods.

**Summary.**

Central experimental results may, for the present purposes, be divided into two basic types: market experiments and tests of non-cooperative game theory. Market experiments were initially constructed to evaluate the predictions of competitive price theory, and competitive price and quantity predictions organize experimental data in a very robust set of laboratory circumstances. Manipulations in the trading rules defining the institution of exchange importantly affect the performance of laboratory markets.

Experimental analysis of the predictions of noncooperative game theory initially grew from a desire to investigate the behavioral properties of the Nash equilibrium prediction in the paradoxical prisoner's dilemma. Similar to market experiment results, baseline game environments can be found which support surprisingly subtle equilibrium predictions. There exist, however, some notable differences and persistent discrepancies between human behavior and the predictions of game theory. Finally, many of the more interesting issues (both behavioral and theoretic) involve examining institutional features surrounding a particular game. As will be discussed in the next section, it is our perception that many of the more immediate applications of experimental techniques to trade issues will involve consideration of issues specific to games played in a trade context.

## 6. POTENTIAL APPLICATIONS OF EXPERIMENTAL METHODS TO TRADE ISSUES

In this section we suggest two trade topics that might benefit from the application of experimental methods. These topics are merely research suggestions. Although it is not clear that either topic will ultimately merit investigation, we discuss them as a means of evaluating the elements necessary for a productive application of experimental techniques.

Notably, each of the topics mentioned involve issues in international trade negotiations rather than tests of the standard trade theorems. While a variety the standard theorems may well merit laboratory investigation, two practical design considerations limit the range of useful projects of this nature. First, is task complexity. An important element of control is that the experiment is simple enough that subjects can understand it in a small portion of the typical 2 to 3 hour laboratory session. A variety of trade models are poor candidates for laboratory investigation for this reason, because they involve overwhelmingly complex environments. It is difficult, for example, to imagine participants mastering cost-minimization, profit-maximization, and trade even in a 2-factor, 2-good, 2-country general equilibrium model.

The absence of institutional detail in some theory represents a second concern. Much theory is written at a level of generality which ignores the institutional structure of an economy. But if anything has been learned from laboratory investigations, it is that these institutional details can very importantly affect behavior. Laboratory designs based on *ad hoc* assumptions about the nature of exchange in an economy constitute joint tests of the relevant theory and institutional assumptions. It would be impossible to determine from such a test whether observed behavior was a consequence of the theory or the institutional assumptions. In light of these considerations, consider the following research suggestions.

a) *Self-Enforcing Agreements and Bilateral Trade Negotiations.* Consider the problem of tariff negotiations between two countries. Since trade liberalization is voluntary the relevant game is noncooperative. A question that has been the focus of considerable theoretical interest has to do with the kinds of agreements that are *self-enforcing* in repeated games of this sort. In a self-enforcing equilibrium the parties must not only agree on a particular outcome, but they must also be able to support that outcomes with credible punishment threats for deviating from the agreement.

An issue of considerable interest among theorists regards the minimal conditions necessary for threatened punishments to be credible in the sense that they can sustain a self-enforcing agreement. Some commentators, for example, would argue that credibility requires that punishments be subgame-perfect, e.g., that the punishments must themselves be equilibria. Others (e.g. Farrell, 1987) would argue that credible punishments must satisfy an even stronger condition of "renegotiation proofness," that the punishment equilibrium is not less desired than the reward equilibria for the punisher. Finally, it may be the case that a variety of elements which are independent of the structure of the equilibria determine the credibility of threats.

Davis and Holt (1990) report an experiment conducted in a simple two-stage game design that initiates examination of this issue. The authors' design consists of a prisoner's-dilemma-type game in a first stage, followed by a 2x2 battle-of-the-sexes game in the second stage. The two pure-strategy equilibria in the second stage are structured so that a "cooperative" equilibrium for the two-stage game as a whole involves a play of the preferred but non-equilibrium "cooperative" outcome in the first stage, followed by a "reward" of coordination on the more desirable of the pure-strategy equilibria in the Battle-of-the-Sexes game. This cooperative equilibrium is supported by a "threat" to coordinate on the less desired of the two pure-strategy equilibria in the Battle-of-the-Sexes game. Davis and Holt report that, given some experience with the game, participants do in fact show some propensity to engage in punish/reward behavior, relative to a baseline design where second-stage payoffs were invariant to choices. Further, although the authors collected only limited evidence regarding the effects of alterations in the structure of punishments on cooperation rates, they find little evidence that alterations in the structure of punishment/reward equilibria affect the incidence of cooperative outcomes in their design.

The relevant baseline environment for applying these issues to problems in trade differs in some important respects from the environment examined by Davis and Holt. Perhaps most prominently, repeated and explicit communications between the countries is very much a part of the negotiation process in a relevant trade game. As discussed above, communications can powerfully affect equilibrium selection. In fact, a pair of (unreported) pilot experiments by Davis and Holt provide two suggestions regarding the effects of communications on outcomes. First, the results of a pilot session which allowed participants the opportunity to submit one-sided nonbinding communications of the type discussed by ACDFR, suggest that this type of "cheap talk" most likely will not improve either cooperation rates or the administration of rewards. The addition of this restrictive form of communications generated virtually no change in behavior. In a second pilot session, participants were given the opportunity to discuss possible strategies. Results of this second session suggest that the richer message space allowed by face-to-face communications may very likely improve at least the incidence of cooperation.

Other alterations in the structure of the game relevant to trade issues may affect the selection of a cooperative equilibrium. For example both negotiations and agreements between countries occur continuously. The incidence of punish/reward outcomes may evolve more reliably in longer, or even indefinite repetitions versions of a game of this type.[14]

This proposed experimental agenda has a number of advantages which deserve comment. First, the results may easily be calibrated, by comparing results against a baseline treatment where the payoffs in the second stage are invariant to choices. Second, the research questions remain close enough to purely theoretic propositions that the relevant questions may be addressed in a relatively simple environment, unencumbered by complexity that would make it difficult to distinguish behavior from learning. Finally, despite the theoretic nature of the investigation, the agenda examines the effects of institutional modifications specific to issues in trade.

---

[14] Of course, the equilibrium set explodes with increases in the length of the game. But the an equilibrium supported by the threat of punishment in a shorter game will remain an equilibrium in a longer version of the game.

b) *Agenda Formation in Multilateral Trade Negotiations.* Agenda formation in multilateral trade negotiations represents a second trade topic where experimental methods might be used. While GATT negotiations could include discussions on a variety of issues, including quotas, intellectual property rights or even environmental concerns, for specificity, let us confine our attention to tariff-reduction negotiations.

Given that tariffs impede trade flows it is not unreasonable to assume that there exists a set of tariff-reduction agreements that would make all countries better off. Countries, however, undoubtedly do not share a common most-preferred agreement within that set, and the diversity of preferences serves as the impetus for bargaining. A bargaining *agenda*, or the order and terms under which various issues are discussed may importantly influence negotiated outcomes.

Mayer (1981) developed a theoretical model in the context of bilateral tariff-negotiations, which showed that the rules for negotiating tariff-reductions were important. He took his cue from the procedures used in the Kennedy and Tokyo Rounds, in which participants agreed first on a tariff-reduction rule and then negotiated the actual cuts. One rule he considered was cutting all tariffs by an equal percentage and then the negotiations would center around the magnitude of the common percentage reduction. The alternative was a harmonizing rule, under which higher tariff rates were to be cut more in order to make rates more equal. The second stage of negotiations would then determine the exact nature of the cuts. Mayer showed that the selection of the rule affected the set of possible equilibria. Casual empiricism suggests some concern among countries regarding agenda changes. In the Kennedy and Tokyo rounds, for example, the U.S. favored the former rule, as it was beginning negotiations with higher average tariffs than the E.E.C., which favored the harmonizing formula.

A variety of components other than the terms of discussion may also affect negotiated outcomes. For example, the decision to open all tariffs to negotiations or to restrict negotiations to a class of products in a round of talks may affect both the probability and the nature of an agreement. Similarly, the decision to ratify sections of the agreement one-by-one, as they are successfully negotiated, or to ratify the entire agreement after all negotiations are concluded may affect the agreement. The former method ensures that all will not be lost if future negotiations on different sections are unsuccessful. The latter method has an advantage in that concessions in one area of discussion can be bartered for advantages in another area of discussion without risk. The Uruguay Round was conducted using the latter method and the entire round of negotiations is now in jeopardy due to the U.S.-E.E.C. squabble over European agricultural policies.

There already exists an experimental literature on the effects of agenda manipulation on voting outcomes. Results of voting experiments by Plott and Levine (1978) suggest that, given sufficiently diverse preferences, voting agenda may be manipulated to yield virtually any outcome. Further results by Eckel and Holt (1989) indicate that strategic behavior by experienced participants can counteract the effects of strategic agenda manipulations.

Multilateral negotiations are distinguishable from voting in that compliance with the agreement is voluntary. With voting, the minority opinion is generally compelled to accept majority rule. By contrast, an unsatisfied participant in trade negotiations has the option of withdrawing from the agreement. Since the value of trade liberalization increases with the number of countries participating, there is some incentive to build a consensus broader than a simple majority.

These questions are, of course, preliminary, and are sketched far too broadly to begin experimentation. A program constructed along these lines, however, has several promising characteristics. Agenda manipulations in multilateral negotiations represent an institutional characteristic perhaps peculiar to trade. Further, these institutional modifications are of the sort that can be cleanly examined in a relatively parsimonious laboratory setting. Keep in mind that the value of an experiment is enhanced if it can be founded on some rival behavioral assumptions or predictions from the relevant theoretical literature. Maintaining control over incentives becomes a further design consideration in experiments involving face-to-face bargaining since anticipated dealings with the other subjects outside of the laboratory could dominate financial incentives.

## 7. CONCLUSIONS

Experimental methods represent a complement to the other, more standard, empirical methods used by economists. Continuing developments in econometric methods have yielded impressive gains in the amount that can be learned from natural data. There are, however, some issues fundamental to economic theory that simply cannot be evaluated with natural data. The laboratory provides an excellent environment for directly evaluating some of these propositions.

There are undoubtedly a variety useful applications of experimental methods to problems in trade. As potentially interesting initial possibilities, we suggest investigation of the effects of some trade-specific institutional modifications in problems that have previously been examined in the laboratory. We discussed an analysis of the effects of direct communications on the selection of self-enforcing equilibria, and an examination of the effects of agenda manipulations in a multilateral negotiation context. Undoubtedly there exist a variety of other examples. As a closing recommendation, we suggest some reading of the literature prior to initiating an experimental study. It is wasteful to reinvent the wheel, and while there are, to the best of our knowledge, no experiments specifically framed as "trade experiments," there do exist a wide variety of papers that address behavioral issues that may be of central concern to problems in trade.

## REFERENCES


Abdalla, Adil., Russell Cooper, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross, (1989) "Forward Induction in Coordination and Battle of the Sexes Games: Some Experimental Results", working paper #89-22, the University of Iowa, Department of Economics.

Alger, Daniel, (1987) "Laboratory Tests of Equilibrium Predictions with Disequilibrium Data," *Review of Economic Studies*, 54 (1), 105-146.

Anderson, Scott, David Johnston, James M. Walker, and Arlington W. Williams, (1991) "The Efficiency of Experimental Asset Markets: Empirical Robustness and Subject Sophistication," in R. Mark Isaac (ed.) *Research in Experimental Economics, Vol. 4*, Greenwich, Conn.: J.A.I. Press (forthcoming).

Banks, Camerer and Porter (1989) "Experimental Tests of Nash Refinements in Signaling Games," working paper, University of Arizona.

Banks L. S., Ledyard, H. O., and Porter, D. P. (1988). "The Design of Mechanisms to Allocate Space Station Resources," CALTECH, Social Science Working Paper 648.

Benson, Bruce L. and Michael D. Faminow, (1988), "The Impact of Experience on Prices and Profits in Experimental Duopoly Markets," *Journal of Economic Behavior and Organization*, 9, 345-365.

Binmore, K., A. Shaked, and J. Sutton, (1985), "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Review*, 75, 1178-80.

Brandts, Jordi, and Charles A. Holt, (1991), "An Experimental Test of Equilibrium Dominance in Signaling Games," *American Economic Review*, forthcoming.

Burns, Penny (1985) "Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction," in V. L. Smith (ed.) *Research in Experimental Economics*, vol. 3, Greenwich Conn.: JAI Press, 139 - 157.

Camerer, C. F., and Weigelt, Keith, (1988), "Experimental Tests of a Sequential Equilibrium Reputation Model, *Econometrica*, 56, 1-36.

Chamberlin, Edward H., (1948), "An Experimental Imperfect Market," *Journal of Political Economy*, 56, 95-108.

Cho, In-Koo and David M. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics* 102, 179-221.

Cox, James. C., B. Roberson, and Vernon L. Smith, (1982). Theory and Behavior in Single Object Auctions," in Vernon L. Smith (ed.), *Research in Experimental Economics vol 2*, Greenwich: JAI Press.

Davis, Douglas D., Glenn W. Harrison, and Arlington W. Williams (1991) "Convergence to Nonstationary Competitive Equilibria: An Experimental Analysis," working paper, Indiana University.

______________ and Charles A. Holt (1990a) "Equilibrium Cooperation in Two-Stage Games: Experimental Evidence", working paper, University of Virginia.

________________________________, and Anne P. Villamil (1990) "Supracompetitive Prices and Market Power in Posted-Offer Experiments," University of Illinois BBER Faculty Working Paper No. 90-1648.

____________ and Arlington Williams, (1991) "The Hayak Hypothesis in Experimental Auctions: Market Power and Institutional Effects," *Economic Inquiry*, forthcoming.

DeJong, Douglas V., Robert Forsythe, R. Lundholm, and W.C. Uecker (1985), "A Note on the Use of Businessmen as Subjects in Sealed Offer Markets," *Journal of Economic Behavior and Organization*, 9, 87-100.

Dolbear, F. T., L.B. Lave, G. Bowman, A. Lieberman, E. Prescott, F Rueter and R. Sherman (1968), "Collusion in Oligopoly: An Experiment on the Effect of Numbers and Information," *Quarterly Journal of Economics*, 82, 240-259.

Dwyer, Douglas, John Kagel and Dan Levin (1989), "Resolving Undertainty About the Number of Bidders in Independent Private-Value Auctions: An Experimental Analysis," *Rand Journal of Economics*, 20 (2) 268 - 279.

Eckel, Catherine, and Charles A. Holt, (1989), "Strategic Voting in Agenda-Controlled Committee Experiments," *The American Economic Review* 779, 4, 763-773.

Farrell, J., (1987) "Cheap Talk, Coordination, and Entry," *Rand Journal of Economics* 18, 34-39.

Farrell, Joseph and Eric Maskin, (1990) "Renegotiation in Repeated Games," *Games and Economic Behavior* (forthcoming).

Fouraker, Lawrence E. and Sidney Siegel (1963) *Bargaining Behavior*, New York: McGraw Hill.

Harrison, Glenn W. and McCabe, Kevin A. (1988) "Testing Bargaining Theory in Experiments," working paper, University of Western Ontario.

Holt, Charles A., Loren Langan, and Anne Villamil (1986) "Market Power in Oral Double Auctions," *Economic Inquiry*, 24, 107-123.

Hong, James T. and Charles R. Plott, (1982), "Rate Filing Policies for Inland Transportation, An Experimental Approach," *Bell Journal of Economics*, 13, 1-19.

Isaac, R. Mark and Charles R. Plott (1981) "The Opportunity for Conspiracy in Restraint of Trade," *Journal of Economic Behavior and Organization*, 2, 1-30.

____________, Valerie Ramey, and Arlington Williams (1984) "The Effects of Market Organization on Conspiracies in Restraint of Trade," *Journal of Economic Behavior and Organization*, 5, 191-222.

____________, and James Walker, (1988) "Communication and Free Riding Behavior in the Voluntary Contribution Mechanism," *Economic Inquiry*, 26, 585-608.

Ketcham, Jon, Vernon L. Smith, and Arlington W. Williams, "A Comparison of Posted-Offer and Double-Auction Pricing Institutions," *Review of Economic Studies*, 13, 1-19.

Kohlberg, E. and J Mertens, (1986), "On the Strategic Stability of Equilibria", *Econometrica*, 54, 1003 -1038.

Mayer, Wolfgang (1981) "Theoretical Considerations on Negotiated Tariff Adjustments," *Oxford Economic Papers*, 33, 135-153.

McCabe, Kevin, Steven Rassenti, and Vernon L. Smith (1988) "An Experimental Examination of Competition and 'Smart' Markets on Natural Gas Pipeline Networks," *Federal Energy Regulatory Commission, Technical Report 88-3.*

Mestalman, Stuart and D. H. Feeny (1988), Does Ideology Matter?: Anecdotal Experimental Evidence on the Voluntary Provision of Public Goods, *Public Choice* 57(3) 281-286.

Murphy, James L. (1966) "Effects of the Threat of Losses on Duopoly Bargaining," *Quarterly Journal of Economics*, 80, 296-313.

Neelin, Janet, Sonnenschein, Hugo and Spiegel, Matthew, "A Further Test of Noncooperative Bargaining Theory," *American Economic Review*, September 1988, 78, 824-36.

Ochs, J., and Alvin E. Roth (1989) "An Experimental Study of Sequential Bargaining," *American Economic Review*, 79 (3), 355-384.

Plott, Charles R., (1982), "Industrial Organization Theory and Experimental Economics," *Journal of Economic Literature*, 20, 1, 1485-1527.

________________, (1986), "Laboratory Experiments in Economics: The Implications of Posted Price Institutions," *Science*, 732-738.

________________, (1987), "Dimensions of Parallelism: Some Policy Applications of Experimental Methods," in *Laboratory Experimentation in Economics: Six Points of View* (A. Roth, ed.) Cambridge University Press: Cambridge, 193-219.

________________, (1988) "Research on Pricing in a Gas Transportation Network," *Federal Energy Regulatory Commission, Technical Report 88-2.*

________________, (1989) "An Updated Review of Industrial Organization: Applications of Experimental Methods," in *Handbook of Industrial Organization*, Volume II, (R. Schmalensee and R.D. Willig, eds.), Elsevier Science Publishers: New York, 1989.

________________, and Michael, E. Levine, (1978) "A Model of Agenda Influence on Committee Decisions," *American Economic Review*, 68, 146-60.

________________, and Vernon, L. Smith (1979) "An Experimental Examination of Two Exchange Institutions," *Review of Economic Studies*, 45, 113-153.

Roth, Alvin E., (1988) "Laboratory Experimentation in Economics: A Methodological Overview," *Economic Journal*, 98, 974-1031.

____________, (1990) "Let's Keep the Con Out of Experimental Economics: A Methodological Note," working paper, University of Pittsburgh.

____________, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, (1990), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," working paper, University of Pittsburgh, 1990.

Selten, Reinhard and Rolf Stoeker (1986) "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior and Organization*, 7,47-70.

Smith, Vernon L., (1964) "The Effect of Market Organization on Competitive Equilibrium," *Quarterly Journal of Economics*, 78, 181-201.

_____________, (1965) "Experimental Auction Markets and the Walrasian Hypothesis," *Journal of Political Economy*, 73, 387-393.

_____________, (1980) "Relevance of Laboratory Experiments to Testing Resource Allocation Theory," in J. Kmenta and J. Ramsey (eds.) *Evaluation of Econometric Models*, New York: Academic Press, 345-377.

_____________, (1981) "An Empirical Study of Decentralized Institutions of Monopoly Restraint," in J. Quirk and G. Horwich (eds.) *Essays in Contemporary Fields of Economics in Honor of E. T. Weiler*, West Lafayette: Purdue University, 1914-1979.

_____________, (1982) "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72(5), 923-955.

_____________, and Arlington W. Williams (1989) "The Boundaries of Competitive Price Theory: Convergence, Expectations, and Transactions Costs," in Leonard Green and John Kagel (eds.), *Advances in Behavioral Economics*, vol. 2, Norwood N.J.: Ablex Publishing.

_______________, G. Suchanak and Arlington Williams (1988) "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56(5) 1119-1151.

Thaler, Richard, (1987), "The Psychology of Choice and the Assumptions of Economics," in *Laboratory Experimentation in Economics: Six Points of View*, (A. Roth, ed., Cambridge University Press: Cambridge, 99-130.

Williams, Arlington W., (1979) "Intertemporal Competitive Equilibrium: On Further Experimental Results," in V.L. Smith (ed.) *Research in Experimental Economics, Vol. 1,* Greenwich, Conn.: J.A.I. Press, 255-278.

__________________, and Vernon L. Smith, (1983), "Cyclical Double-Auction Markets With and Without Speculators," *Journal of Business,* 57, 1-33.

__________________________________, and John Ledyard (1985) "Simultaneous Trading in Two Competitive Markets: An Experimental Examination," working paper, Indiana University.

Stoecker, Rolf (1980) "Experimentelle Untersuchung des Entscheidungsverhaltens," im *Bertrand-Oligopol, Wirtshaftstheoretische Entscheidungsforschung*, 4, Universitat Bielefeld, Bielefeld: Pfeffersche Buchhandlung Bielefeld.

This material is based in part on work supported by the U.S. Department of Agriculture, Cooperative State Research Service, under Agreement No. 89-34210-04238 and successor(s).



Additional information on NC-194 and a complete list of project publications can be obtained from:

*Executive Director, NC-194*
*Department of Agricultural Economics*
*The Ohio State University*
*2120 Fyffe Road*
*Columbus, Ohio 43210-1099*
*(614)292-2194*