Experimental Economics: Methods, Problems, and Promise

Douglas D. Davis and Charles A. Holt[*]

Abstract

Although economics is not commonly thought to be an experimental science, laboratory methods are being increasingly used as economic theories become more sophisticated. It is often possible to structure payoffs and procedures so that a theory is tested on its domain, with all structural assumptions satisfied. If the theory fails, individual components can be tested; if the theory works, it can be stressed with procedures that violate the structural assumptions in carefully chosen directions. Applications of this experimental method have been motivated by three distinct sets of issues: the relative efficiencies of alternative market institutions, the predictions and applications of game theory, and the behavioral validity of expected utility theory. This paper presents specific examples from each of these areas to highlight the usefulness of experimentation. Common objections to this methodology are also discussed, as are some of the principal lessons that have been learned, both about economic behavior, and about how to evaluate theoretic propositions in the laboratory.

1. Introduction

Economics is a study of practical affairs. Economists spend their time trying to find reasonable solutions to complex social problems such as debt crises, pollution, inflation, and bureaucratic stagnation. Despite this penchant for the practical, economists also see themselves as scientists. Like other scientists economists observe events in nature, devise theories to explain their observations, and evaluate their theories in light of additional evidence.

But not all commentators share the view that economics is both a policy-oriented study and a science. As the *Encyclopedia Britannica* (1991, p. 395) observes,

> "Economists are sometimes confronted with the charge that their discipline is not a science. Human behavior, it is said, cannot be analyzed with the same objectivity as the behavior of atoms and molecules. Value judgements, philosophical preconceptions, and ideological biases must interfere with the attempt to derive

---

[*] Virginia Commonwealth University and University of Virginia. This paper was prepared for a lecture at the XXVI meetings of the Latin American Econometric Society Meetings, in Mexico City, Mexico, September 1, 1992. Research support from the National Science Foundation (grants SES 9012694 and SES 9012691) is gratefully acknowledged.

> conclusions that are independent of the particular economist espousing them. Moreover, there is no laboratory in which economists can test their hypothesis."

Of course, most economists disagree with this view. It is not difficult, however, to appreciate its source. Economists often have very definite preconceptions about the desirability of market outcomes. Moreover, the bewildering collection of markets and activities that constitute the "economic jungle" is a rather difficult place to corner facts which would prove, dispel or even alter preconceived notions about the marketplace.

Matters are not improved by modern economic methods. Economic theories are often highly abstract mathematical models that can look incredibly arcane to an outsider. The standard empirical procedures used by the economist are similarly complex. Moreover, when applied to typical data sources from natural markets, the economists' empirical tools often constitute a rather cumbersome camera, a camera that rarely allows definitive focus on a theoretical or policy issue.[1]

Economics is far from unique among sciences in terms of theory complexity. To the contrary, theoretical physicists, chemists and molecular biologists all routinely rely on mathematical techniques that rival even the most complex methods of the economist. Rather, it is the limited capacity for theory evaluation that usually distinguishes economics from the traditional "hard" sciences, where abstract theoretical propositions have been evaluated and refined through the use of controlled laboratory experimentation. In contrast, economists have generally relied exclusively on data from events as they occur in the complex natural economy.

The scientific community has immense respect for the power of laboratory techniques, so much so, that one synonym for a "hard science," might be a "laboratory science." This respect is rightfully placed, as the interchange between theory and evidence made possible by laboratory investigation is the very foundation of modern scientific method. But there is no reason why economics cannot also become a laboratory science. Although humans are neither chemical processes nor molecules, human behavior can be observed objectively and replicated in a

[1] This is not always true. The extreme economic changes that we seen recently in Mexico, for example, provide a natural experiment in which economic predictions can be isolated.

controlled context. In fact, although laboratory techniques are neither widely known, nor (until recently) widely used by economists, the economics laboratory most definitely exists. Granted, when viewed in light of the elaborate equipment of the experimental physicist or chemist, the economics laboratory is quite crude. The authors' laboratories, for example, are just classrooms with banks of networked personal computers, separated by laminated styrofoam sheets.[2] Moreover, the economic situations investigated in the economics laboratory bear little resemblance to any natural market. The experimental environments are highly simplified variants of more complicated models or markets. Finally, the laboratory decision makers are far different from the decision makers that are the subject of many economic models. Laboratory decision makers are usually college students, who are motivated by relatively small financial rewards, who often know little about economics in general, and who usually know nothing about the particular context being investigated. Nevertheless, after observing the results of many interactions among students in these simplified environments, we have come to believe that experimental investigation represents an empirical tool that can help economists isolate the explanatory power of a variety of theories. In particular, laboratory investigation has given us a lot of confidence in our basic intuition that unregulated, decentralized markets can solve complex production and allocation problems in a remarkably efficient manner.

The purpose this paper is to discuss the growing importance of experimentation in economic analysis. In the sections that follow, we hope to give some sense of the variety of economic issues that have been explored with laboratory techniques. We also address some common objections to experimentation, as well as some of the principal lessons that have been learned. It is useful to begin with a specific example of an economics experiment.

2. A Laboratory Market

Consider an experiment designed to evaluate competitive price, quantity and efficiency predictions in a simple partial-equilibrium setting. Evaluation of such predictions requires specification of cost and value conditions for some good. Rather than using some specific good,

---

[2] We are referring to laboratories at Virginia Commonwealth University and the University of Virginia. More sophisticated economics laboratories do exist, and in a variety of places as diverse as Arizona, Barcelona, and California.

it is convenient to design the market in terms of an abstract commodity that has value to the buyers and sellers only in terms of cash profits or earnings obtained from trade. Use of an abstract commodity of this sort facilitates experimental control, as it avoids the unobservable preference differences that may exist among participants for specific goods such as candy bars, records or sardines, which could also be traded in a laboratory setting.

The preferences and technology for the laboratory commodity are induced by explaining to subjects how they can earn money. For example, one simple laboratory market set-up might involve passing out a number of cards to buyers and sellers. A redemption value would be printed on each of the cards passed to buyers, who would be told that they can keep the difference between this printed redemption value and a price that they negotiate with a seller in the market. Thus, if the number printed on a buyer card is $1.90 and if the buyer negotiates a transactions price of $1.00, then the buyer earns 90 cents. Because earnings are paid in cash, $1.90 represents that buyer's maximum willingness to pay, or unit demand.

Similarly, unit costs are printed on the cards passed to sellers, who are told that they will earn the difference between the price they negotiate and the printed cost number. For example, if $.80 is printed on a seller card and the seller negotiates a transactions price of $1.00, then the seller would earn 20 cents. Since earnings are paid in cash, the number printed on the card represents the seller's minimum willingness to sell, or unit cost.

Market demand and supply functions can be created by varying the numbers on cards given to different buyers and sellers. Ordering value cards from highest to lowest generates a market demand curve, and ordering cost cards from lowest to highest generates a market supply curve, as shown in figure 1. The intersection of these curves allows identification of the standard competitive equilibrium predictions. In the figure 1 design, for example, a competitive equilibrium allocation involves a quantity of seven units, and any price between $1.30 and $1.40. The trading surplus, or maximum gain from trade is $3.70. This maximum surplus, of course, corresponds to the area above the supply curve, below the demand curve and to the left of the equilibrium quantity prediction.

These predictions exhibit a number of desirable features. Perhaps most importantly, they are unambiguously specified *ex ante* by the researcher, and may be evaluated in light of subsequently collected data. Moreover, these cost and value conditions may be easily altered to

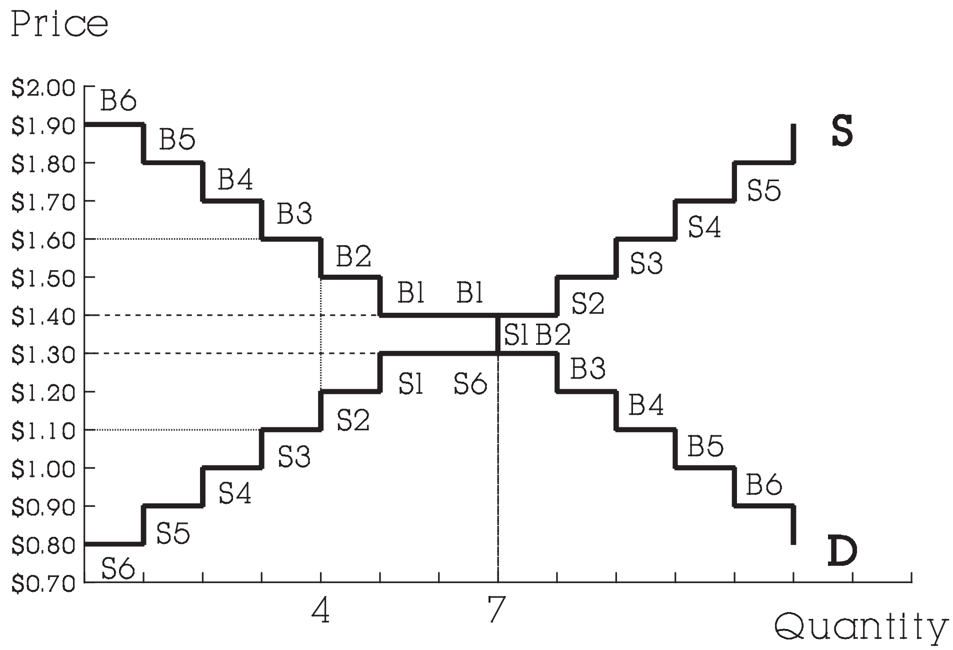

Figure 1. Supply and Demand Structure for a Market Experiment

generate predictions in nearly any desired market structure. Combined, these features allow a great deal of flexibility regarding the investigation of the competitive tendencies of markets. This is in stark contrast to the analysis of data from most naturally occurring markets, where a market structure is given, and where both market structure and performance must be assessed from the data, *ex post*.

A second critical advantage of specifying a laboratory market design is that it forces a very detailed consideration of the way the market is administered. Although largely ignored in standard economic theory, trade is governed by a series of conventions that have evolved or have been externally imposed. A set of such trading rules, or *trading institution*, must also be specified in the laboratory. One of the primary results of market experimentation is that performance can be critically affected by variations in the trading institution.

The importance of trading rules became evident at the outset of market experimentation. For example, the very first reported market experiment was conducted by Chamberlin (1948), who, upon observing the failures of markets in the Depression, believed that there were features inherent in the competitive process that would tend to generate inefficient outcomes. To evaluate his hypothesis, Chamberlin conducted a classroom market under a *negotiated price* institution, where participants walked around the room and completed contracts via a series of private, unstructured negotiations. In this laboratory market, Chamberlin did observe his predicted efficiency losses.[3]

Subsequent experimentation was motivated by the idea that competitive price theory would be given a better test in a market where bids and asks were publicly tendered from a centralized location, as in a stock exchange. Vernon Smith (who incidentally was a participant in Chamberlin's market) devised an alternative set of trading rules, now known as the *double-auction* trading institution. In the double auction, buyers are free to accept the price "ask" of sellers, or to propose alternative terms in the form of "bids." Symmetrically, sellers may accept the bids proposed by buyers, or counter with asks. All bid and ask information is public, and is typically submitted under a bid/ask spread-reduction rule, or a condition that only proposals that improve upon the best standing terms are admissible.

Outcomes in Smith's double-auction markets conformed much more closely to competitive equilibrium predictions.[4] In fact, competitive predictions turn out to be remarkably robust in double-auction markets. Competitive allocations are generated in double-auction markets under extreme structural conditions (sometimes even under monopoly), and under unusual supply and demand configurations (for example in situations where the distribution of earnings goes entirely to one side of the market).[5] For this reason, the double auction has become a performance

---

[3] The inefficiencies arose from a tendency for privately negotiating agents with *infra*marginal units (e.g., units to the left of the intersection of aggregate supply and demand) to complete contracts with agents holding *extra*marginal units (to the right of the equilibrium prediction). In terms of figure 1, for example, such an inefficiency loss would arise if B6 completed a contract for a first, high-value unit with S5, who was selling a second high-cost unit. This contract would preclude a seller with a low-cost inframarginal unit from making a sale with an inframarginal buyer. In Chamberlin's market, one contract involving an infra- and an extra-marginal unit tended to be offset by another, similarly inefficient contract. Thus, Chamberlain observed a combination of inefficient outcomes, and an excess of traded units over the competitive prediction.

[4] Smith (1963, 1964) also differed from Chamberlin in two procedural respects. First, rather than motivating participants with hypothetical cash payments, Smith soon began to use real financial incentives. Second, Smith's markets were repeated -- that is, after the expiration of an initial period of trading, buyers and sellers were reendowed with unit values and costs for additional trading. Both of these features may also affect behavior, and are now standard experimental procedures.

[5] See, e.g, Smith, 1982 and Smith and Williams, 1990. In fact, competitive predictions are so pervasive in the double auction that deviations are of considerable interest. See e.g., Holt, Langan and Villamil (1986).

standard against which other institutions are evaluated.

Table 1. Negotiations in Double-Auction Trading

(Value)	Buyer ID	Bid	Ask	Seller ID	*(Cost)*
			$2.00	S4	*($1.00)*
			$1.90	S2	*($1.20)*
($1.80)	B5	$1.00			
($1.60)	B3	$1.10			
			$1.80	S3	*($1.10)*
			$1.70	S5	*($ .90)*
			$1.60	S1	*($1.30)*
($1.40)	B1	$1.25			
($1.80)	B5		**$1.35**	S4	*($1.00)*

It is instructive to consider double-auction trading in more detail. Table 1 illustrates negotiations for a single double-auction contract in a laboratory market we recently conducted.[6] The motivation for "double-auction" terminology should be clear from this table. Starting with seller S4's opening ask of $2.00, sellers compete by offering progressively lower terms. Similarly, buyers compete with each other by increasing bids from buyer B5's opening bid of $1.00. A contract occurs at a price between the initial ask and bid proposals, when B5 accepts S4's ask of $1.35. Thus, the double-auction institution is like two auctions at once, an ascending-price bid auction that one of the sellers stops with a sale, and a descending ask-auction that one of the buyers stops with a purchase. It is worth emphasizing that these negotiations are conducted under conditions of an extreme privacy. Although bids and asks are centrally displayed, unit values and costs for the negotiating buyers and sellers (displayed in parentheses

[6] This session was conducted in the summer of 1992 at the University of Virginia. The market was conducted orally (e.g., without computers), with inexperienced subjects. The market consisted of 6 buyers and 6 sellers, and used the supply and demand structure in figure 1.

in the extreme left and right columns of table 1) are known only to the negotiating agent. Following completion of an initial contract, trading resumes in the double auction, as agents begin negotiations for a second unit. Trading continues in this way for the second and subsequent units, usually until the expiration of a preannounced time limit, at which time the trading *period* ends.

The results of an initial market trading period are illustrated by the left-most pair of vertical bars in figure 2. In the figure, the dashed horizontal lines illustrate the competitive price band, while the dots illustrate the sequence contracts completed in the period. An experimental market session typically consists of a sequence of trading periods. After expiration of the time-limit for an initial period, buyers and sellers are re-endowed with costs and values, and the second trading period begins. The market shown in figure 2 consisted of three trading periods, and the dots representing contract prices in each period are separated by vertical bars. Trading efficiency is calculated as the total of all subjects' earnings, expressed as a percentage of the maximum total surplus. This efficiency is shown under the price dots for each period. The transactions quantity for the period is shown in parentheses.

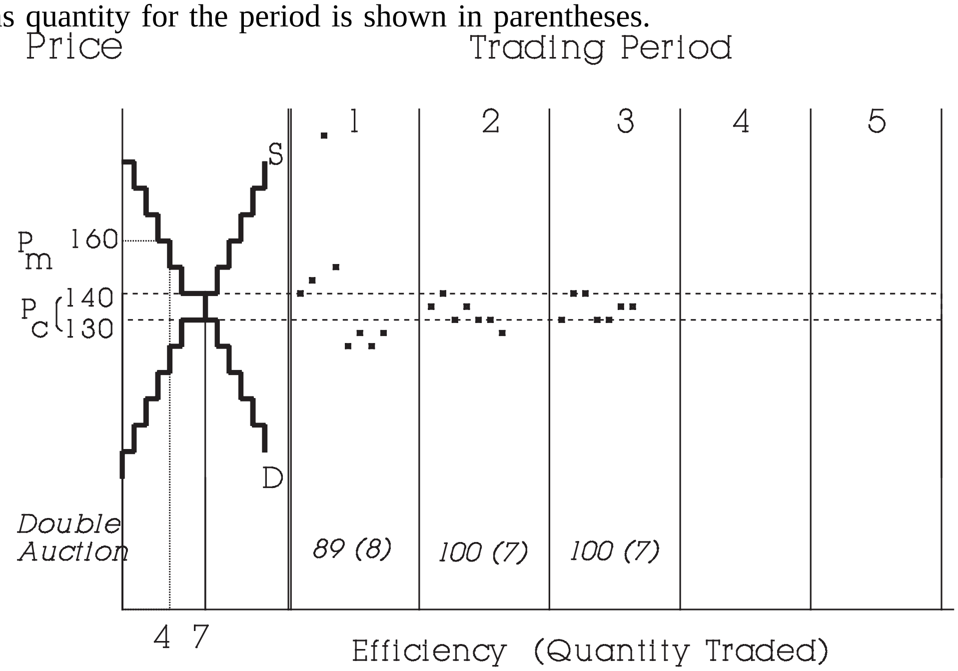

Figure 2. Sequence of Contract Prices for a Double Auction

The accuracy of competitive predictions in figure 2 deserves comment. By the second trading period, seven (of seven predicted) units trade, and the market is 100% efficient. Moreover, in the second and third periods combined, all but one of the contracts are struck in the competitive price band. This occurs despite a complete absence of information regarding either equilibrium price predictions, or aggregate supply and demand functions. Rather, individuals see only their private cost and value information, and the public messages of the market. As suggested above, this pattern is typical of double-auction markets.

Still other trading institutions are possible. One very natural alternative, for example, is a *posted-offer* institution that parallels many of the important features of retail-type exchange. A posted-offer trading period proceeds as a two-step sequence. First, sellers select prices independently.[7] These prices are collected and then publicly posted, as shown in table 2. Second, after all prices are posted, buyers are randomly drawn from a waiting mode, and are given the opportunity to make purchases at the posted prices, on a take-it-or-leave-it basis.

Table 2. Price Information in a Posted-Offer Trading Period

Seller ID	S1	S2	S3	S4	S5	S6
Price	$1.55	$1.55	$1.35	$1.35	$1.60	$1.40
(Cost)	*($1.30)*	*($1.20)*	*($1.10)*	*($1.00)*	*($0.90)*	*($0.80)*

Trading in the posted-offer institution generates much less public information than in the double auction. In a posted-offer trading period, the only public information is the price postings of sellers. This contrasts sharply with the information-rich double auction. Recall, for example, that the sequence of double-auction negotiations shown in table 1 was for a single unit, rather than for an entire trading period. Yet more price quotes were generated in that negotiation process that for the entire posted-offer trading period shown in table 2.

This limited flow of information, together with the asymmetry between buyers' and sellers' activities in a posted offer market, has definite consequences on market performance.

[7] Sellers also make a private quantity-limit selection. Except for the initial periods of a market when learning is incomplete, sellers routinely offer all units that may profitably sell at their selected price.

Figure 3 shows the price sequence for a posted-offer market.[8] This market was conducted under the same induced cost and value conditions as the double-auction market just discussed, and the figure is formatted in the same manner as figure 2, except that contracts are denoted by "*" markings rather than dots, and the "+" signs indicate price-postings where no units were sold.

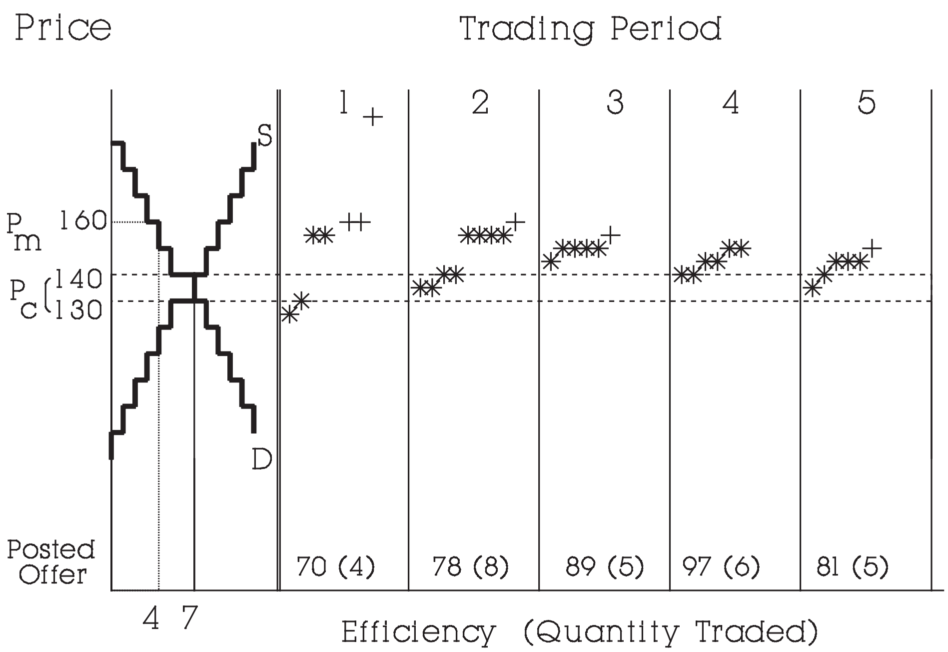

Figure 3. Price Sequence for a Posted-Offer Auction

Notice the pattern of contract prices for the posted-offer market shown in figure 3. Relative to the scattered price cluster of the double auction, posted trading appears to follow a definite structure: Within trading periods, posted-offer contracts follow an upward trend, reflecting the predominant tendency for buyers to follow a simple full-revelation strategy of purchasing all available units, starting with the least expensive units first. Across periods, the prices tend to start high and then ease down toward the competitive level. These features are general characteristics of posted-offer markets (Davis and Williams, 1986).

But more importantly, notice the comparatively weak performance of the posted-offer

8 This market was also conducted at the University of Virginia in the summer of 1992. As with the double auction discussed above, the market was conducted orally, and used participants who were inexperienced with the posted-offer trading institution.

market, relative to competitive predictions. By the second double-auction trading period, shown in figure 2, for example, all contracts but one are in the competitive equilibrium price tunnel, and 100% of the possible gains from trade are extracted. In contrast, only four of eight posted-offer contracts are within the competitive price tunnel in the second posted-offer trading period, and only 78% of the possible gains from trade are extracted. The pattern continues in trading period 3, where all double-auction contracts are in the competitive price range and efficiency is 100%, while all posted-offer contracts are outside the competitive range, and efficiency is only 89%. This performance difference is commonly observed: Posted-offer markets tend to the competitive equilibrium more slowly than double-auction markets, and they extract less of the available surplus in the process (Ketcham, Smith and Williams, 1984; Plott, 1986). It also turns out that posted-offer markets are more susceptible to conspiracy and the exercise of market power than double-auction markets (Isaac, Ramey and Williams, 1984; Davis and Williams, 1991).

Experimentalists were not alone in observing the importance of institutional rules. This same lesson, for example, is a primary result of the "new" game-theoretic industrial organization literature. But the laboratory data add a critical dimension to the investigation, in that they provide a direct link between theory and data. The posted-offer environment, for example, closely parallels the assumptions of a Bertrand price-setting game. A comparison of equilibrium predictions and observed outcomes allows unique insight into the behavioral relevance of alternative equilibrium concepts.[9]

The capacity to test theories in the laboratory is a decisive advantage of experimentation, and bears emphasis. Although simple laboratory environments miss much of the texture of a natural market, they can be made to conform to the assumptions of theoretical models in a manner simply not possible with natural data. In a laboratory market, for example, demand, supply and equilibrium predictions are directly induced, and are specified *ex ante*. In contrast,

[9] The institution-specific models in the new industrial-organization literature typically involve much simpler structures than a posted-offer market. However, Holt and Solis-Soberon (1992) have recently begun to use the tools of noncooperative game theory to explain observed posted-offer market outcomes. One can intuitively begin to see how a posted-offer market could be analyzed as a game by considering a single-period posted-offer market with one buyer, one seller, and a capacity-constraint of a single unit for the seller. In such a setting, it should be obvious an equilibrium for a game should involve a seller posting, and the buyer accepting, a limit price equal to the buyers' maximum willingness to pay for a unit.

in a natural market, we can only estimate supply and demand, with the use of a number of more or less problematic auxiliary assumptions about functional forms, the nature of cost conditions, and preferences.

3. Kinds of Experiments

Market predictions are only one of the many kinds of economic propositions that can be evaluated with experimental methods, and experimentalists have examined a wide variety of economic issues in the laboratory. Although it is beyond the scope of this presentation to exhaustively review the types of experiments that have been conducted, it is useful to convey some feeling for the range of applications.[10] One way to illustrate the diversity of applications is to consider the historical roots of experimentation. Interest in laboratory methods in economics arose from three, more or less distinct sources: market cooperation and competition, game theory, and individual decision theory. We consider these different sources in turn, and discuss briefly how research has evolved in each area.

a. As indicated above, *market experiments*, have led to an investigation of the effects of alterations in the rules defining the trading institution. Some of these studies have a traditional industrial organization focus, such as the investigation of mechanisms for regulating and restraining monopoly power (e.g., Coursey, Isaac and Smith, 1984; Harrison and McKee, 1985). Other investigations have focused on allocations in financial markets (Smith, Suchanek and Williams, 1988; Forsythe and Lundholm, 1990). Interest in laboratory financial markets has been spurred by the immanence of electronic stock exchanges that bypass dealers, especially after the traditional exchanges have closed. Still other market investigations have involved assessing the likely performance characteristics of new trading institutions in situations where allocations have been traditionally determined by direct regulation. Examples range from proposed markets for pollution permits (Kruse and Elliott, 1990), to proposed markets for transportation rights on gas pipelines and electrical-power networks (McCabe, Rassenti and Smith, 1990a, 1990b).

b. A second class of experiments, *game experiments*, were initiated in the late 1950's and

[10] Standard reviews include Plott (1982, 1989), Smith (1982) and Roth (1988). We have attempted to be fairly comprehensive in our book (Davis and Holt, 1992). See also Roth and Kagel (forthcoming).

early 1960's by sociologists and social psychologists who were unconvinced that rational individuals would stumble into the jointly undesirable outcomes predicted in the famous "prisoners' dilemma" game. Due to the close link between the structure of this game and oligopoly coordination problems, the results of prisoner's dilemma and related game experiments were received with considerable interest by economists.

Table 3. A Prisoner's Dilemma Game

		Column Player	
		HIGH	LOW
	HIGH	800, *800*	0, *1000*
Row Player	LOW	1000, *0*	350, *350*

The pricing problem faced by duopolists who cannot explicitly conspire, for example, is a classic prisoner's dilemma problem, as can be seen by considering the game in table 3. The row and column players must either post a HIGH price (top or left) or a LOW price (bottom or right). The (row, *column*) payoff consequences for these decisions are printed in the matrix entry for the decisions made. Notice that each seller earns a profit of 800 at the joint-maximizing, HIGH price, as indicated in the upper left box. But it is risky for a seller to post the joint-profit maximizing price, since profits would drop to 0 if the other seller did not reciprocate. At the same time, any seller would find it tempting to shade on a joint profit-maximizing price posting, since earnings would increase from 800 to 1000 by being the only seller to post the competitive price. The temptation to shade on the cooperative price to increase earnings, combined with the incentive to post a LOW price as a means of protection against a defection by the other, makes posting the LOW price a Nash equilibrium, despite the fact the both players would prefer a situation where they both posted HIGH prices.

Despite the attractiveness of the joint-maximizing solution, experimental evidence suggests that the Nash equilibrium has considerable drawing power. For example, Cooper et al. (1991) conducted a prisoner's dilemma experiment using the incentives in table 3. In the experiment, participants made a series of thirty decisions, facing a different, anonymously assigned opponent in each period. Results of this experiment are summarized in table 4, in the form of the proportion of HIGH decisions. Notice that HIGH is selected less than half the time in matchings 1-5, and in all subsequent groups of matchings. Notice also, that the propensity for participants to select LOW increases as they become experienced with the game: The rate at which HIGH is selected decreases from an average of 43% for matchings 1 - 5 to 20% for matchings 25 - 30.

Table 4. One-Stage Prisoners' Dilemma Game Outcomes (Source: Cooper et al., 1991)

Matching Number	Percentage of Cooperative HIGH-PRICE Choices
1 - 5	43%
6 - 10	33%
11 - 15	25%
16 - 20	20%

Hundreds of experiments involving games with a prisoner's dilemma structure were conducted in the 1960's and 1970's.[11] Typically, these experiments involved extensive repetition of a single-stage structure like that shown in table 3. Repetition tends to increase the incidence of cooperation. The increase, however, is not as much as you might expect. In particular, one popular theoretical conjecture is that repetition increases cooperation through the use of trigger strategies. That is, participants might support the cooperative HIGH price choice via a threat to punish defection (a LOW price posting), with a punishment consisting of several periods of Nash equilibrium play. There is little evidence for the notion that subjects regularly support cooperation with the use of trigger strategies, at least in a prisoners' dilemma context.[12]

Simple games of this sort overlook many of the institutional elements that have been shown to affect behavior not only in the natural world, but even in laboratory markets. These simple structures are desirable, however, in that they facilitate the identification and evaluation of alternative game-theoretic predictions. This point is more clearly seen in a second example.

[11] Coleman (1983), for example, cites 1500 papers.

[12] But subjects do appear to use punish/reward strategies in other game contexts. See, Davis and Holt (1990, 1991a).

Unlike the static prisoners' dilemma, which has a single Nash equilibrium, many games exhibit multiple equilibria. As a general matter, there are a variety of ways to discriminate among multiple equilibria. But in the special case that the equilibria are Pareto ranked (e.g., some equilibria involve higher payoffs to every player than other equilibria), theorists often assume that the players coordinate on the equilibrium outcome that is best for all, or the *Pareto-dominant* equilibrium. This assumption is useful for making comparative statics or welfare statements, and is the type of assumption that can be tested in the laboratory.

For an intuitive appreciation of the Pareto-dominance equilibrium selection criterion, consider the following problem. Suppose a standing university committee consists of fifteen members. The committee meets once a week, on Friday afternoons. Meetings are scheduled to start each Friday at 3:00 p.m., but all members must be present before the meeting can start. Finally, regardless of when a meeting starts, it lasts two hours.

Consider each committee member's incentives to appear at a particular meeting. On the one hand, everyone would prefer to start on time, and go home promptly at 5:00 p.m. On the other hand, no one wants to arrive before all of the other committee members are present, since the meeting does not start until everyone is present. This situation is characterized by multiple Nash equilibria, since no single committee member can unilaterally improve on any common appearance time by all the members. Moreover, the equilibria are Pareto ranked: Assuming that everyone would like to leave as soon as possible, the equilibria that involve starting closer to the announced 3:00 p.m. are preferred to those with later starting times.

Van Huyck et al. (1990) conducted an experiment involving a game with this structure. In the experiment 15 subjects independently chose an X value numbered 1, 2, ..., 7. In terms of our committee-meeting story, a choice of 7 corresponds to appearing at the meeting on time, while smaller numbers indicated progressively later appearances. Table 5 provides payoffs determined by an individual's own choice and the choice of the last person to arrive. The payoffs are denominated in pennies. Consider a common decision of 7, which yields 130 per person, the Pareto-dominant outcome. By reducing your choice from 7 to 6, you also reduce the minimum choice to 6, so the outcome is moved up along the diagonal to a common payoff of 120. As is clear from inspection of these payoffs, all of the decisions involving a common choice (e.g., on the diagonal) are Nash equilibria, since payoffs decrease with a unilateral

Table 5. A Coordination Game with Multiple, Pareto-Ranked Nash Equilibria: Your Payoff in Pennies

		Smallest value of *X* chosen						
		1	2	3	4	5	6	7
	1	70	-	-	-	-	-	-
Your	2	60	80	-	-	-	-	-
	3	50	70	90	-	-	-	-
choice	4	40	60	80	100	-	-	-
	5	30	50	70	90	110	-	-
of *X*	6	20	40	60	80	100	120	-
	7	10	30	50	70	90	110	130

deviations from any diagonal element. Each experiment session consisted of a sequence of ten stages. In each stage participants made decisions privately. Then the lowest numbered choice was announced, and payoffs were determined, and the process repeated, until after the tenth stage, at which time earnings were totaled, and participants were paid.

The results of a representative session are shown in figure 4. The stage number and choices are listed on the two axes defining the horizontal plane, while the vertical axis represents the proportion of participants selecting a choice. In initial stages of the game, many participants selected the choice involving the Pareto-dominant outcome. For example, in the first stage, many participants chose 7, and the modal decision was 5. Stage 1 earnings, however, were determined by the single participant who chose 4. These low payoffs clearly damped incentives to appear "on time," in subsequent stages, and the portion of "late arrivals" increased as the periods passed. By stage 10, almost all participants chose decision 1.

Van Huyck et al. did observe more cooperation in games with the same structure, but with smaller (2 or 3 person) committees. The primary result of the experiment, however is clear from figure 4: Pareto dominance alone does not determine the equilibrium selected.

Much of the current interest in game experiments stems from a desire to assess the empirical properties of equilibrium selection criterion such as Pareto dominance. A critical problem with many noncooperative games is that they exhibit an embarrassingly large array of

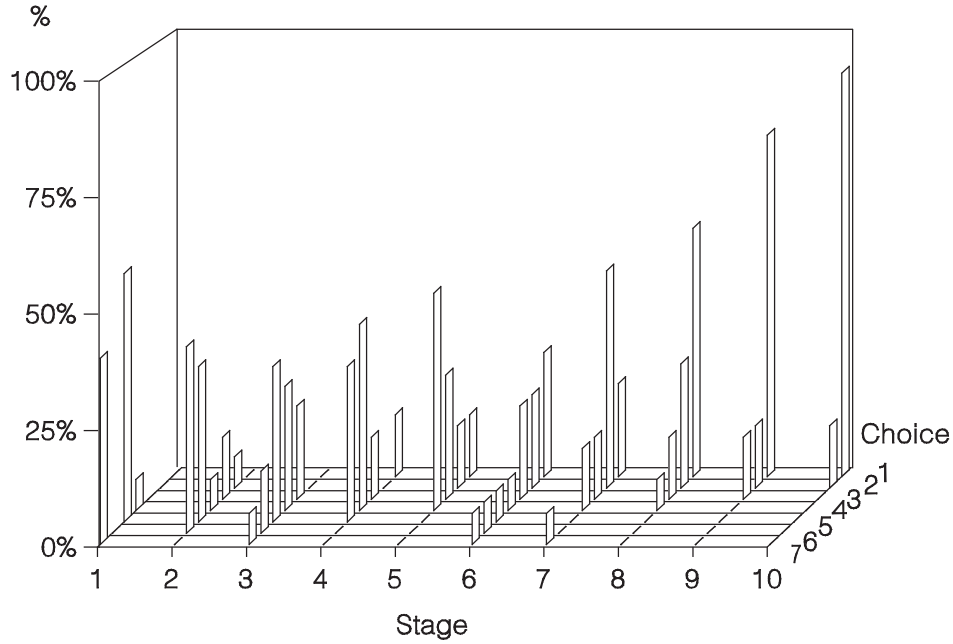

Figure 4. Results of a Coordination Game Session (Source: Van Huyck et al., 1990, Session 4)

Nash equilibria. The subsequent effort among theorists to prune away the less plausible of these equilibria has spawned an entire "refinements" literature, characterized by a series of highly suggestive adjectives such as "perfect," "intuitive," "strategically stable," "divine," and "universally divine," which modify the Nash equilibrium concept. If game theory is to avoid becoming a branch of mathematics (or theology), it is essential to empirically distinguish among these definitions. The laboratory is particularly useful for this purpose, since these refinements often differ in very subtle ways, such as differing assumptions regarding the nature of beliefs off the equilibrium path, or in portions of the game that are never played.

Laboratory investigation in this area has already provided some very concrete lessons. The approach of most refinements is to consider an equilibrium, and to think about what would be reasonable beliefs about behavior following a deviation. This approach is especially speculative for a deviation that takes players to a part of the game that is never reached in the process of equilibrium play. But subjects in experiments have played the same game with a sequence of different partners, and disequilibrium decisions in early matchings may have taken subjects to parts of the game tree that are not encountered in later matchings after some equilibrium is reached. What is important here is that subjects' beliefs about what would happen off of the equilibrium path are influenced by what *actually* happened during the disequilibrium adjustment process in previous matchings. Brandts and Holt (1992) show that the observed patterns of naive behavior during early matchings can correspond to beliefs that are ruled out by all of the standard refinements (equilibrium dominance, divinity, and strategic stability). As a consequence, outcomes in some of their sessions approximated the outcomes for equilibria that are ruled out by these refinements. These data highlight the importance of studying adjustment patterns instead of constructing deductive arguments about subjects' beliefs.

c. A third stream of experiments, called *individual decision theory* experiments, grew from a skepticism among psychologists and others regarding the behavioral relevance of the expected utility theory proposed by von Neumann and Morgenstern (1944). The resulting experiments were of obvious interest to theorists, as expected utility theory forms the basis of the way that economists' model allocations in stochastic environments.

A typical experimental investigation of this type evaluates one of the axioms of expected utility theory by eliciting responses to a question that cleanly distinguishes behavior that is inconsistent with the axiom, from behavior that is consistent. Such questions typically involve a choice among pairs of lotteries. Consider, for example, two lotteries with payoffs (denominated in British pounds) determined by a roulette wheel. The wheel has 100-stops, which are numbered consecutively 1, 2, ..., 100. Payoffs for the two lotteries are illustrated in figure 5. As is clear from the upper row, the "safe" lottery S1 yields a payoff of £7 with certainty, while the "risky" lottery R1 yields a payoff £10 with a probability of 20%; £7 with a probability of .75; and £0 with a probability of .05.

	1—----5	6---80	81----------------100
Lottery S1	£7	£7	£7
Lottery R1	£0	£7	£10

Figure 5. Representation of the Choice Between a Lottery Pair R1 and S1

The format of figure 5 highlights an important relationship between the 2 lotteries: they share the common consequence of a £7 payoff for outcomes 6 - 80. Under the independence axiom of expected utility theory, the preferences of an expected utility maximizer are independent of a change in a common consequence. The independence axiom could be evaluated by changing the payoff for the common consequence from £7 to £0. This alteration generates the second lottery pair, S2 and R2 shown in figure 6. If the independence axiom is valid, an expected utility maximizer given a choice of both lotteries would choose either S1 and S2, or R1 and R2.

	1-—- 5	6-- ----80	81---------------- 100
Lottery S2	£7	£0	£7
Lottery R2	£0	£0	£10

Figure 6. Lotteries R1 and S1 Transformed to Lotteries R2 and S2 by a Common Consequence

Laboratory investigations of this type of lottery choice reveal rather persistent deviations from expected utility theory. Starmer and Sugden (1991), for example, report an experiment where participants were presented with the (S1, R1) and (S2, R2) lottery pairs. Results of their experiment are summarized in table 6. As is clear from the table, participants do not tend to view these lottery choices as equivalent. Given a choice between S1 and R1, 68% of the participants selected the relatively safe lottery. But preferences for the relatively safe lottery were damped when given the choice of S2 and R2. In this case, only 42% of the participants selected "safe" lottery S2.

Violations of this type have been observed in numerous instances. The Starmer and Sugden design is remarkable in the care that the authors used in controlling for rival behavioral motivations. For example, unlike many studies of individual decision theory by psychologists (e.g., Kahneman and Tversky, 1979), decisions are financially motivated. Moreover, Starmer and Sugden avoid potential complications due to wealth effects by having separate samples individuals drawn from the same pool make the (S1, R1) and (S2, R2) lottery choices.[13]

Table 6. Allais Paradox Data (Source: Starmer and Sugden, 1991)

	Percent of Subjects Choosing	
	S1 over R1	S2 over R2
40 subjects (with single financially motivated choice)	68%	~
40 subjects (with single financially motivated choice)	~	42%

Anomalous decisions of this type are mitigated somewhat by factors such as experience and increases in payoffs (Kagel, MacDonald, and Battalio, 1990). Such factors, however, do not eliminate the inconsistent choice patterns.[14] The persistence of such anomalies raises a series of rather difficult theoretical questions. For example, exactly which axiom is violated? Is it the independence axiom, as suggested above, or the somewhat less problematic reduction of compound lotteries axiom? Should expected utility theory be generalized to accommodate the observed behavior, and if so, how?

Rather than moving toward a consensus, researchers in this area appear to be sorting themselves into three rather distinct camps. One group appears to reject expected utility theory altogether, and proposes founding economic theory on an alternative basis (e.g., Kahneman and Tversky, 1979). Another group proposes retaining expected utility theory, but generalizing it in a way that accommodates the observed behavior (Machina, 1982; Loomes and Sugden, 1982). These generalized theories have the disadvantage that they involve assumptions that are more cumbersome to incorporate into complex models than the assumptions underlying a simple von

[13] Wealth effects could become a problem if participants played a sequence of lotteries, due to changes in earnings within the session.

[14] One experimental design with particularly salient rewards involves the use of rats, who make choices over levers that yield random payoffs in food pellets (Battalio, Kagel and MacDonald, 1985; Kagel, MacDonald and Battalio, 1990). Although rats presumably do not have the same cognitive process as humans, it is interesting to note that rats' exhibit some of the same choice patterns as humans.

Neumann-Morgenstern utility function. Finally, a third group suggests that it is possible that the pattern of violations could be explained within expected utility theory, but allowing for the possibility that humans make mistakes, particularly when the difference in expected values are small (Hey, 1991). This "theory of errors" is clearly the least radical of the alternatives, but it is not clear at this time which alternative will ultimately be more useful.

Finally, it is worth observing that not all individual-decision experiments are designed to assess the axioms of expected utility theory. A variety of other issues involving individual decisions have been investigated in the laboratory including eliciting values for nonmarketed goods such as clean air or a scenic view (Coursey, Hovis and Schultze, 1987; Knetsch and Sinden, 1987), the predictions of sequential search theory (Schotter and Braunstein, 1981; Cox and Oaxaca, 1989), and forecasting (Williams, 1987; Smith, Suchanek and Williams, 1988).

4. Some Common Objections to Experimentation

Experimentation, of course, is no panacea, and there are a variety of legitimate reasons for avoiding experimental methods in some contexts. Experimentation, for example, could render little useful information about the values of specific parameters outside of the laboratory, parameters such as the average costs of a firm or individual bequest motives. But carefully designed and administered experiments can generate valuable insights in a very wide variety situations. Although experimental methods have in fact been broadly applied in economics, there are a number of reservations regarding the use of experiments that we feel are not legitimate, which have inhibited experimentation. In this section we identify and respond to four of the most common such reservations. These objections fall into two basic classes. The first class, summarized in (a) and (b) below, stems from the perception that logical status of theory makes experimentation unnecessary. The second class of reservations, articulated below in (c) and (d), regards misgivings about the way experiments are conducted.

a. *Theory is logically true.* A first reservation is that experimentation is unnecessary, because a theory is "correct" as long as it is internally consistent. But more is, or at least should be, required of a theory than that there are no errors in the underlying mathematics. Internal consistency is only a first step. Economic theories should have some explanatory power as well.

Economic theories are based on two kinds of assumptions, *behavioral* and *structural*. The primary purpose of experimentation is to evaluate behavioral assumptions. In the laboratory, we match a theory's incentive and structural conditions, and then observe the validity of the behavioral assumptions. If humans don't match our theoretical behavioral presumptions, then the theory is incorrect, and it must be altered, despite its internal consistency.

Critically, humans are not necessarily stupid or foolish if they fail to make the decisions predicted by our theories, and it is generally *not* useful to tell the participants how to behave in such instances. If behavioral suggestions of this type are enforced, they convert an experimental investigation into a simulation. This is not to denigrate the importance of simulations as an analytic tool. Rather our purpose is to distinguish experimental investigation, an empirical technique, from simulation, which is a theoretical device for extending the range of application for theories that do not have analytic solutions. Simulations impose *both* behavioral and structural assumptions on parameterized versions of a model, and then rely on a computer algorithm to generate outcomes. Experiments are a means of evaluating behavioral assumptions.

b. *Theory is normative*. A second objection to experimentation is that it is unnecessary because theory tells us how humans should behave, rather than predicting how they do behave. This attitude is not at all unusual for a decision theorist. Operations research theorists, for example, have traditionally considered their job as prescriptive. The mindset, however, is considerably more problematic for economists, who have traditionally viewed their science as a descriptive inquiry. For example, we find the recently espoused view that game-theory is normative (Rubinstein, 1991), rather bizarre. The application of game theory to industrial organization issues, for example, is presumably not done primarily as a corrective service for imperfectly operating cartels. Rather, we are interested in predicting when stable, supra-competitive prices will or will not be observed.

c. *The laboratory is too simple.* This third reservation regards the way experiments are conducted. The world is a complicated place, filled with complex, multi-dimensional interactions. Laboratory environments, in stark contrast, examine human decisions in extremely simple environments. How can decisions made under such simple, streamlined conditions hope to tell us anything useful?

Upon reflection, it should be clear that this reservation is a criticism of economic theory as much as it is a criticism of experimentation. Theories are necessarily extremely simplified characterizations of the complicated natural world. If the complications of the natural world are expected to systematically affect outcomes, then a more complex theory should be constructed. If not, then the theory should be evaluated. Moreover, this evaluation process should ideally begin, not in the domain of the complex natural world, where numerous confounding events may impinge on variables of interest, but strictly on the domain of the theory, where all structural assumptions can be implemented. The laboratory is an ideal and unique environment for evaluating a theory on its ow n domain. Of course, observation of the theory "working" in the laboratory does not imply that it explains behavior in the natural world. But the failure of a theory under the "best shot" circumstances of the laboratory suggest that the theory is not a good explainer of behavior. It is perhaps in this role of theory rejection that experimentation is most useful.

Importantly, it is not the view of experimentalists that other empirical methods, particularly econometrics, are without merit. Econometrics very usefully allows theory to be evaluated in light of data from the natural world, via the use of a series of auxiliary assumptions. Experimentation allows more direct evaluation of theory in a simplified environment, but without the need of auxiliary assumptions.

d. *Subjects are too naive.* A final reservation regards the subjects typically used in laboratory research. Even if it is desirable to evaluate a theoretical prediction in a very simple environment, the critic may contend that the environment is inappropriate because the laboratory decision makers (typically college students) are less sophisticated that the decision makers in the relevant natural environment.

This final reservation is not an objection to experimentation *per se*, but rather, an objection to the way economists have come to conduct experiments. Certainly, if "relevant" professionals do not behave like student subjects, then the appropriate subject pool should be composed of relevant professionals. Of course, providing salient rewards to relevant professionals would increase the costs of experimental investigations, often by a dramatic amount. But this would just mean that experimentation is more expensive than previously thought.

For obvious reasons, experimentalists have been interested in the relationship between decisions made by participants in "standard" subject pools, and the decisions of relevant professionals (e.g., Ball and Cech, 1991). While the subject pool issue is a matter that must be addressed on an experiment-by-experiment basis, it is interesting to note that in a variety of instances where the laboratory behavior of both students and relevant professionals has been examined, performance has generally not varied substantially across subject pools. Businessmen traders, for example, tend to generate speculative price bubbles, as do college sophomores. (Smith, Suchanek and Williams, 1988). Similarly, building contractors are as susceptible to a "winner's curse" as college students (Dyer, Kagel and Levine, 1989),[15] and one group of ecologically conscious environmentalists were observed to free ride in a manner very similar to college sophomores (Mestelman and Feeny, 1988).

In summary, there a number of reasons for not doing experiments. Although some of these claims are not meritless, they do not outweigh the critical advantages of replicability and control allowed by careful experimental investigation. For these reasons, the use of experimentation as a means of evaluating economic theory propositions both has grown and should continue to grow.

5. What We've Learned: How to Ask a Question

If experimentation is valuable, and economists have been doing it for several decades, what have we learned? Certainly, we've learned something about economic theory. These are the most expected lessons of experimentation, and we will consider them in the next section. But we have also learned a great deal about how to conduct experiments. It is these lessons we which to consider first. Importantly, although some of these lessons may appear a bit specific to a general reader, as a group they constitute far more than just a series of narrow procedural issues. To the contrary, successful experimentation is a process of learning how to pose questions, get answers, and then develop further questions. This process is critical to the development of any empirical science.

a. *The Details Matter.* First, we've learned that physical procedures are important, and that close attention must be given to a number of issues. At the heart of experimentation is the

---

[15] The winner's curse arises in *common value* auctions, where an item has the same value to all bidders, but where that value is unknown until after the purchase. Under such conditions, the agent that submits the highest bid is likely to be the one who makes the largest mistake. Unless the bidders make an appropriate downward adjustment in their bids, the winner of such an auction may be cursed with a loss.

capacity for independent verification. Independent verification not insures the honesty of results, but allows examination of the extent to which outcomes are the result of particular parameters or procedures. Without careful attention to detail, verification is impossible.

Thus, an experiment must be administered meticulously. Subjects must be called, instructed, monitored, and paid in a manner that is both uniform and standard. Similarly, experiment instructions and administrative details must be spelled out, followed, and reported in detail. The test that any experimenter should use as a guide in administering and reporting results of an experiment is an affirmative answer to the following question: Can an outside observer replicate the experiment in a manner that the original researcher and other observers would accept as being valid?

It is easy to find examples of published and unpublished papers with "fatal" procedural errors, such as carelessness with instructions or procedures, and low or nonsalient financial incentives. Those familiar with experimental methods simply will not take the results of an experiment seriously unless it satisfies some basic standards of replicability.

b. *The Importance of Design.* Second, we've learned that attention to the design of treatments is essential to drawing unambiguous conclusions from laboratory evidence. Parameters must be selected carefully, to avoid extra, unwanted equilibria, or to avoid having someone explain away the results as due to some overlooked, extraneous factors (such as focalness or symmetry). Experimentation is a little like computer programming; a seemingly small design error can render the output difficult to interpret, or even useless.

As with procedural matters, there are a number of common "fatal" design errors typically made by a new researcher in designing an experiment. The most common of these errors are: (1) failure calibrate results with a baseline treatment; (2) failure to restrict focus to a few treatments of interest; and (3) failure to choose the degree of institutional complexity appropriate to the problem being investigated. These design errors are less obvious than the procedural issues just discussed, and for this reason we consider them separately, below.

1. *Calibration*. In order to draw a conclusion that a given variable (or combination of variables) elicits a particular kind of behavior, it is necessary to conduct baseline sessions that parallel the treatment sessions in every relevant respect except for hypothesized variable(s) of interest. This is made clear with an example. Suppose we are interested in the effects of market

power on prices in posted-offer markets. (By market power, we mean designs where the competitive equilibrium is not a Nash equilibrium, because one or more sellers can profit by unilaterally increasing prices above the competitive level.) To evaluate this question, suppose we conduct a series laboratory sessions in markets where some sellers have market power, and find that prices hang high above competitive levels. Regardless of the number of sessions we conduct, we cannot conclude from this treatment alone that market power increases prices. Although the observed high prices could be due to market power, they could also be a consequence of a variety of other design features, such as the number of sellers or the low excess supply at supracompetitive prices. Rather, to determine the effects of market power, two treatments must be conducted: one with power and one without, holding all other things (such as the number of sellers, and aggregate supply and demand) constant. Price increases can be attributed to market power only if price increases were observed in the power sessions, but not in the no-power sessions.

2. *Focus*. When designing an experiment, it also important to confine attention to only a few control variables. Altering a variable to explore an additional effect is seductively easy. This temptation, however must be avoided, since the number of necessary treatment cells increases exponentially with increases in the number of treatment variables. Suppose, for example, that we were interested in examining the effects of a change in the number of sellers in the market power experiment discussed above. This issue increases the number of treatment combinations from 2, to (2^2 =) 4. A third control variable, say the effects of live rather than simulated buyers, would again double the number of possible treatment combinations, to (2^3 =) 8 cells. Granted, not all treatment combinations are always of equal interest, and under certain conditions it is possible to reduce the number of treatment combinations for a given number of variables. But the lesson remains: Trying to look at too much in a single investigation can easily make it impossible to learn anything.

A related issue pertains to parameters that are in fact altered across treatments. It is surprisingly common for researchers to make inadvertent alterations along with intentional variable changes. In a lottery-choice experiment, for example, it is not uncommon to give participants an initial $10 balance when the choice involves an expected loss, but to consider such a payment unnecessary when the decision involves an expected gain. Comparison across decisions is muddied in this case by the difference in wealth.

This does not necessarily mean that only a single parameter can altered in a session. To the contrary, in order to hold the predictions of a theory constant, it is sometimes necessary to change more than one economic parameter. For example, to evaluate the effects of changing the number of sellers on market power, concentration must simultaneously be increased if market power (the underlying Nash equilibrium) is held constant.[16]

A more general representation of the relationship between theories, treatments and possible outcomes is presented in figure 7. In the figure, two variables under the experimenter's control (denoted *x* and *y*), are represented in the horizontal plane, where each point corresponds to a treatment. Observed outcomes are measured along the vertical axis. If the experimenter wishes to evaluate two alternative theories, treatments must be altered so that predictions, denominated in terms of the observable outcomes, allow a distinction between the two theories. In the figure, for example, one of the theories predicts no change in outcomes (denoted by the solid horizontal line) while the other (denoted by the dotted line) predicts a change. The critical element is to manipulate the control variables, either singly or in combination, in such a way that in one (baseline) treatment the predictions of two theories overlap and in another treatment, the predictions diverge. The former treatment allows assessment of presumably spurious alterations in sessions (such as group effects, and the inherent variability of outcomes), while the latter, allows evaluation of the relative performance of theories in light of the variability observed in the baseline.

3. *Appropriate Complexity.* Experimentation is a process of hypothesizing, designing an environment that allows evaluation of the hypothesis, collecting relevant data, and then starting the cycle again, using what was learned as a basis for further hypothesizing. Rarely is it the case that the "crucial experiment" is conducted in a first iteration of this process. (For that matter, such an experiment is rarely conducted in any iteration!) An important element in this repeated cycle of hypothesis, design and evaluation, is selecting the appropriate complexity. The natural economy is overwhelmingly complex. The process of theorizing is one of disregarding supposedly inessential details. Yet further simplification is possible if the researcher

---

[16] See, for example, Davis and Holt (1991b).

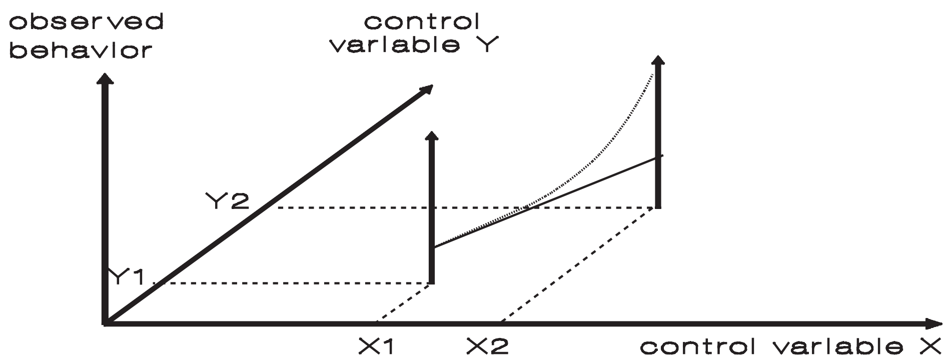

Figure 7. Evaluating a Theory: The Relationship Between Control Variables and Outcomes

intends to evaluate only certain elements of a theory. Where in this array of possibilities should testing begin? Given a starting point, in what direction should future investigation proceed? Finally, how are tests of differing underlying complexity related?

Although there is no necessary starting point for experimental investigation, we can say something about the directions research can take. To keep this discussion concrete, suppose we focus on just two dimensions: the *environment*, which includes the structure of preferences, technology, etc., and the *institution*, which defines the rules of exchange. A simple market theory would involve a particular environment (e.g., some number of buyers and sellers, with given value and cost incentives) and a given institution, such as the posted offer. These two dimensions are illustrated in figure 8.

Each point in the two-dimensional domain is a treatment. The domain of the theory is the set of points for which the assumptions of the theory are satisfied exactly (remember that this can be restrictive, even for general theories, if we need to make assumptions about functional forms to get comparative statics predictions that can be tested.) Often, a natural place to begin evaluation is at a point that is strictly on the domain of a theory. A "theory test" (TT) would involve a pair of points in this domain. For example, a theory test might be the resilience of competitive price predictions to institutional alterations I1 and I2 (e.g., from posted offer to double auction), within a given market structure (environment E1).

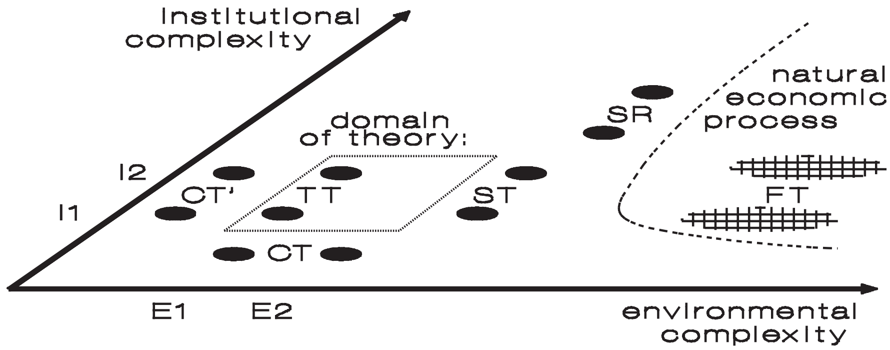

Figure 8. Types of Experiments (Key: CT – component test, TT – theory test, ST – stress test, SR – search for empirical regularities, FT – field test)

Regardless of the results of the experiment, further testing is usually called for. If the theory fails on its domain, the issue becomes a question of determining which component of the theory fails. This leads to *component tests* in simpler environments (points CT). Recent research involving ultimatum games illustrates the notion of a component test. Consider the following problem. Suppose two laboratory participants are presented a $10 pie, which they are to divide between themselves. The division is determined by a very simple process: One player proposes (as an "ultimatum") the terms of the division. The other player either accepts the proposed division, in which case the terms are as proposed, or rejects, in which case both players earn nothing. The unique Nash equilibrium for this game, involves an extremely inequitable division of the pie: The first player should ask for the whole pie, minus epsilon. This division should be accepted, since the epsilon offered to the other is greater than the zero that results from rejection. Behaviorally, however, quite different outcomes tend to be generated in laboratory ultimatum games. Results of a representative ultimatum game experiment reported by Forstyhe et al., 1988 are illustrated by the dark spike in the back of figure 9. As is evident from the spike, the median offer is half of the pie.

Given the failure of the Nash equilibrium prediction, the next step was a component test, to determine the questionable assumption of the theory. For example, are equal divisions observed because proposers are concerned about fairness, or because they are concerned about

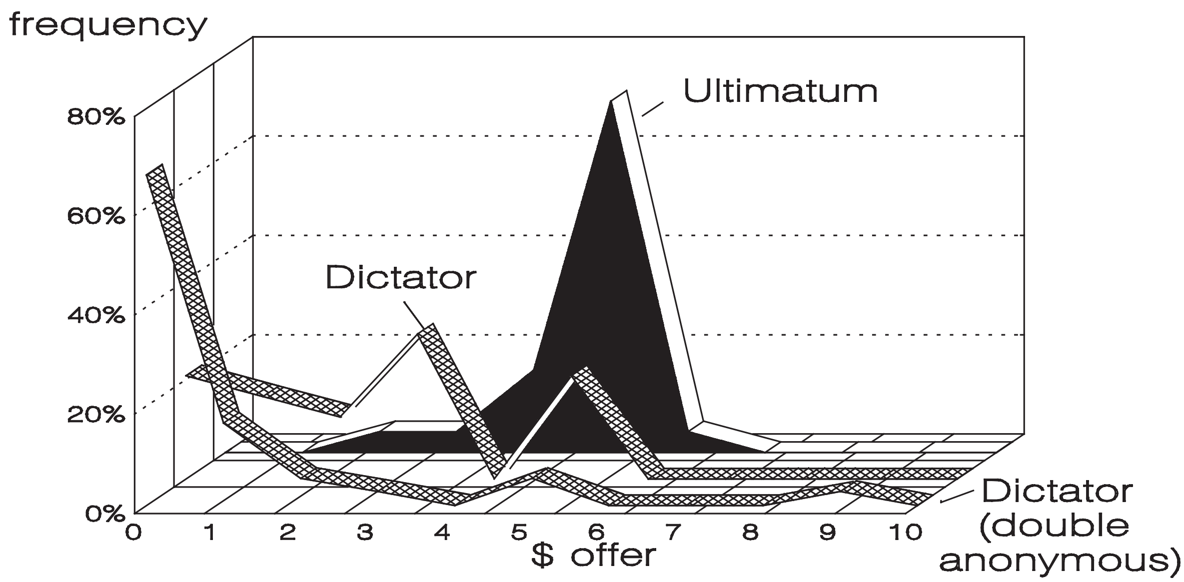

Figure 9. Frequency of Offers in Ultimatum and Dictator Games
(Sources: Forsythe et al., 1988, for ultimatum and dictator game data; Hoffman et al., 1991, for double-blind, dictator-game data)

the rejection of an “equilibrium” proposal. These possibilities may be distinguished via still simpler component tests. For example, fairness concerns can be distinguished from concerns about “irrational” buyer rejections in a “dictator” game, where one player proposes a split of the $10 pie. Unlike the ultimatum game, however, the proposal determines the outcome; the other player has no opportunity to accept or reject the proposed split.

As in the ultimatum game, the proposer takes essentially the whole pie in a Nash equilibrium for the dictator game, and leaves nothing to the other player. This test is a component of the ultimatum game, however, in that “irrational” rejection opportunities, (along with the consequences of anticipating them) are forbidden.

Data for a dictator game experiment reported by Forsythe et al. (1988) is illustrated by the middle ribbon in figure 9. As suggested by inspection of the data, the possibility of rejection does not alone determine the tendency for sellers to lower prices. In the dictator game experiments, the median offer is $3.00, below the $5.00 equal-rent division, but well above the $0.00 division consistent with a Nash equilibrium.

The results of this test allow us to rule out “irrational” rejections as the primary cause of equitable divisions. Yet other procedures must be used, to determine whether fairness concerns, or some other procedural issue determines the divisions. A recent paper by Hoffman et al. (1991) proposes an alternative procedural motivation for the division: the embarrassment of facing an experimenter who knows you have taken everything from the participant with whom you are paired. Hoffman et al. evaluate this hypothesis with a "double anonymous" treatment, in which great care is taken to keep the experimenter from knowing who proposes what split, and in which the participants know that the experimenter is unaware of the splits. Results of an experiment conducted under this double-anonymous condition are illustrated by the band in the front of figure 9. Results clearly suggest that anonymity, rather than fairness may explain outcomes: under the double-anonymous condition, most dictators offer nothing.

Now consider a second direction for in an experimental research program. If the theory works on its domain in an initial investigation, then we would want to look at *stress tests*, or treatment points off of the domain of the theory, which stress the theory, points like ST. These points fall in the shadow of the theory. Even if a theory performs well on its domain, failure of the theory in such stress tests can cast serious doubts on a theory's usefulness. Theories of perfect competition are not very useful, for example, if an infinity of traders is needed. Similarly, the proposition that potential entrants or "contestants" can regulate natural monopolies if entry and exit is costless, is not particularly interesting if it fails under conditions of small entry or exit costs. In short, stress tests allow evaluation of the appropriateness of the abstractions made by the theory.

Successful stress tests may lead to tests that deviate even more substantially from the assumptions of the theory. In the extreme case, *field tests* (points FT) are conducted in natural markets. Finally, not all laboratory tests are conducted with exact reference to the domain of any theory. Experimentalists can look for interesting patterns of behavior, guided only by intuition or informal arguments. This process of *searching for empirical regularities* is illustrated by the points labelled SR in the upper right portion of the figure. Examinations of this type can provide useful information for theory development.

5. What We've Learned: The State of Theory

Finally, we turn to the question of what experimentation has taught us about economic theory. We are able to make discussion rather terse, because the examples used above were not selected randomly, but were chosen to illustrate some of the primary lessons of experimentation. In this section we summarize five general results of experimental investigation, and illustrate them in terms of the examples mentioned above.

a. *In at least some markets, competitive predictions are remarkably robust.* Markets organized under double-auction rules generate competitive predictions so pervasively, that any design that generates deviation is publishable. This result was illustrated by the double auction market session (figure 2).

b. *Institutions matter.* Performance in markets organized under posted-offer trading rules, for example, can be markedly different from performance in double-auction markets that similar in all respects except for the institution (as seen in figures 2 and 3 above). The specification of the institutional rules parallels work in the "new" game-theoretic industrial organization. For this reason, there has been a lot of interest in experimental work among both game and industrial organization theorists.

c. *Some of the predictions of game theory work.* In general, participants appear drawn to Nash equilibria when they exist, particularly in static games. This was illustrated in the one-shot prisoner's dilemma experiment, summarized in tables 3 and 4.

d. *Our game theories are sometimes incomplete.* The refinement of Pareto dominance, for example, appears to organize equilibrium selection in a two-person coordination game, but Pareto dominance fails when the number of participants is increased to fifteen (as shown in figure 5). This conclusion is seen in a number of other important applications. In public goods experiments, for example, a whole host of theoretically irrelevant variables appear to affect participant's tendencies to "free ride."

e. *In yet other instances we observe anomalies: the assumptions of our theories just seem to fail to work.* This problem is particularly noticeable in individual decision experiments, as illustrated by results of the Allais paradox experiment shown in figures 5 and 6, and in table 6. Such "anomalies" call for further investigation to determine their pervasiveness, as well as for a consideration of alternative theories.

What you might conclude about what we have learned depends on your perspective, and on where you start on the list. Some overview presentations of experimental work focus on the elements at the top of the list. These studies are of great comfort to theorist, as they make experimentation look like a codification of the theory economists have so skillfully crafted over the last two centuries. Others prefer to focus on the lessons mentioned at the bottom of the above list. From these, one gets the distinct impression that we should quit economics, or at least start over and study psychology.

Our conclusion is intermediate. Sometimes theory explains behavior well, sometimes it isn't rich enough, and sometimes it is simply wrong. Of course, the case where a theory is completely general, is palatable to the theorist. But we wish to emphasize the converse point. The possibility that theory can be wrong, or limited in application, does not imply that theory must be rejected, or that the clay feet upon which our science is founded will or even should lead to a toppling of the entire edifice. Anomalies simply are not devastating. Rather, they are a normal part of an empirical science, and are one of the signs that a paradigm should be modified. Things that work are an indication that we should continue with the same paradigm. The places where theory is insufficiently rich suggest an interchange between the theorist and the experimenter.

6. Conclusions

As mentioned above, experimentation is no panacea. Many important issues in economics simply cannot be addressed the lab. An antitrust case, for example, may hinge on the measurement of an average cost (to determine whether a seller was pricing in a predatory manner), and this is an issue that must be resolved by measurement outside of the laboratory. For similar reasons, experimentation cannot be expected to tell us much interesting about bequest motives or Ricardian equivalence. These are issues about parameters and preferences in nature.

But from the laboratory, we both have learned and can learn much about economic theory. Through trial and error, experimentalists over the last 40 years have cultivated the capacity to exploit the control allowed by the laboratory to evaluate economic theory precisely on its own domain. Similar data simply are not available in the natural world, and the importance such data will undoubtedly appreciate in value as the new, institution-specific theories become more refined. Careful laboratory testing may provide the discipline that prevents this new theory from collapsing under its own complexity.

There are anomalies, inconsistencies, and surprises in the lab. But this is troublesome only to the extent that we expect actual behavior to conform with all of the elegance and precision of the theoretical models that pervade the text books and journals. Indeed, a willingness to seek inconsistencies is useful. As Vernon Smith (1989) argues, the most productive attitude is to be skeptical of both the theory and the data. And we believe that both our understanding, and this healthy skepticism will increase as economics becomes more of an experimental science.

References

Ball, Sheryl B., and Paula A. Cech (1991) "The What, When and Why of Picking a Subject Pool," working paper, Indiana University.

Battalio, Raymond C., John Kagel, and Don N. MacDonald (1985) "Animals' Choices over Uncertain Outcomes: Some Initial Experimental Results," *American Economic Review*, *75*, 597–613.

Battalio, Raymond C., John Kagel, and Komain Jiranyakul (1990) "Testing Between Alternative Models of Choice under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty*, *3*, 25–50.

Brandts, Jordi, and Charles A. Holt (1992) "An Experimental Test of Equilibrium Dominance in Signaling Games," *American Economic Review*, *82*, .

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

Coleman, Andrew (1983) *Game Theory and Experimental Work.* London: Pergamon Press.

Cooper, Russell W., Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1991) "Cooperation without Reputation," working paper, University of Iowa.

Coursey, Don L., John L. Hovis, and William D. Schultze (1987) "On the Supposed Disparity between Willingness to Accept and Willingness to Pay Measures of Value," *Quarterly Journal of Economics*, *102*, 679–690.

Coursey, Don R., R. Mark Isaac and Vernon L. Smith (1984) "Natural Monopoly and the Contested Markets: Some Experimental Results," *Journal of Law and Economics*, *27*, 91-113.

Cox, James C., and Ronald L. Oaxaca (1989) "Laboratory Experiments with a Finite Horizon Job Search Model," *Journal of Risk and Uncertainty*, *2*, 301-329.

Davis, Douglas D., and Charles A. Holt (1990) "Equilibrium Cooperation in Three-Person Choice-of-Partner Games," forthcoming in *Games and Economic Behavior*.

——— (1991a) "Equilibrium Cooperation in Two-Stage Games: Experimental Evidence," working paper, Virginia Commonwealth University.

——— (1991b) "Capacity Asymmetries, Market Power, and Mergers in Laboratory Markets with Posted Prices," working paper, Virginia Commonwealth University.

——— (1992) *Experimental Economics*. Princeton: Princeton University Press.

Davis, Douglas D., and Arlington W. Williams (1986) "The Effects of Rent Asymmetries in Posted Offer Markets," *Journal of Economic Behavior and Organization, 7,* 303-316.

——— (1991) "The Hayek Hypothesis in Experimental Auctions: Institutional Effects and Market Power," *Economic Inquiry, 24,* 261-274.

Dyer, Douglas, John Kagel, and Dan Levin (1989) "A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis," *Economic Journal*, *99*, 108–115.

Encyclopedia Britannica, *Macropaedia: Knowledge in Depth*, *27,* (1991) 15th edition. Chicago: University of Chicago Press.

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton (1988) "Replicability, Fairness and Pay in Experiments with Simple Bargaining Games," Working Paper 88-30, Department of Economics, University of Iowa, forthcoming in *Games and Economic Behavior*.

Forsythe, Robert L. and Russell Lundholm (1990) "Information Aggregation in an Experimental Market," *Econometrica*, *58*, 309-347.

Harrison, Glenn W., and Michael McKee (1985) "Monopoly Behavior, Decentralized Regulation, and Contestable Markets: An Experimental Evaluation," *Rand Journal of Economics*, *16*, 51-69.

Hey, John D. (1991) *Experiments in Economics*. Oxford: Basil Blackwell.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith (1991), "Preferences, Property Rights, and Anonymity in Bargaining Games," working paper, University of Arizona.

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

Holt, Charles A., and Fernando Solis-Soberon (1992) "The Calculation of Equilibrium Mixed Strategies in Posted-Offer Auctions," in R. M. Isaac (ed.) *Research in Experimental Economics*, *vol. 5,* Greenwich: JAI Press, 189-229.

Isaac, R. Mark, 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.

Kahneman, Daniel and Amos Tversky (1979) “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica*, *47*, 263–291.

Kagel, John H., Don N. MacDonald, and Raymond C. Battalio (1990) “Tests of ‘Fanning Out’ of Indifference Curves: Results from Animal and Human Experiments,” *American Economic Review*, *80*, 912–921.

Ketcham, Jon, Vernon L. Smith, and Arlington Williams (1984) “A Comparison of Posted-Offer and Double-Auction Pricing Institutions,” *Review of Economic Studies, 51,* 595-614.

Knetsch, Jack L., and J. A. Sinden (1987) “The Persistence of Evaluation Disparities,” *Quarterly Journal of Economics, 102*, 691–695.

Kruse, Jamie L., and Steven R. Elliott (1990) “Strategic Manipulation of Pollution Permit Markets: An Experimental Approach,” working paper, University of Colorado.

Loomes, Graham, and Robert Sugden (1982) “Regret Theory: An Alternative Theory of Rational Choice under Uncertainty,” *Economic Journal*, *92*, 805–824.

Machina, Mark J. (1982) “‘Expected Utility’ Analysis without the Independence Axiom,” *Econometrica*, *50*, 277–323.

McCabe, Kevin A., Stephen J. Rassenti and Vernon L. Smith (1990a) ”Auction Design for Composite Goods: The Natural Gas Industry,” *Journal of Economic Behavior and Organization*, *14,* 127-149.

——— (1990b) “Experimental Research on Deregulated Markets for Natural Gas Pipeline and Electric Power Transmission Networks, forthcoming in ”*Research in Law and Economics*.”

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

Plott, Charles R. (1982) “Industrial Organization Theory and Experimental Economics,” *Journal of Economic Literature*, *20*, 1285-1527.

——— (1986) “Laboratory Experiments in Economics: The Implications of Posted-Price Institutions,” *Science*, *232*, 732–738.

——— (1989) “An Updated Review of Industrial Organization: Applications of Experimental Methods,” in R. Schmalansee and R. D. Willig eds., *Handbook of Industrial Organization*, vol. 2. Amsterdam: North-Holland, 1109-1176.

Plott, Charles R., and Vernon L. Smith (1978) “An Experimental Examination of Two Exchange

Institutions," *Review of Economic Studies*, *45*, 133–153.

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

Roth, Alvin E. and John H. Kagel (eds.) (1992) *Handbook of Experimental Economics*, Princeton, N.J.: Princeton University Press.

Rubinstein, Ariel (1991) "Comments on the Interpretation of Game Theory," *Econometrica*, *59*, 909-924.

Schotter, Andrew and Yale M. Braunstein (1981) "Economic Search: An Experimental Study," *Economic Inquiry*, *19*, 1-25.

Smith, Vernon L. (1962) "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy*, *70*, 111–137.

——— (1964) "The Effect of Market Organization on Competitive Equilibrium," *Quarterly Journal of Economics*, *78*, 181–201.

——— (1982) "Microeconomic Systems as an Experimental Science," *American Economic Review*, *72*, 923–955.

——— (1989) "Theory, Experiment and Economics," *Journal of Economic Perspectives*, *3*(1), 151–169.

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams (1988) "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, *56*, 1119–1151.

Smith, Vernon L, 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.

Starmer, Chris, and Robert Sugden (1991) "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review*, *81*, 971–978.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil (1990) "Tacit Coordination Games, Strategic Uncertainty and Coordination Failure," *American Economic Review*, *80*, 234–248.

Von Neumann, J., and O. Morgenstern (1944) *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

Williams, Arlington W. (1987) “The Formation of Price Forecasts in Experimental Markets,” *The Journal of Money, Credit and Banking*, *19*, 1–18.

Williams, Fred E. (1973) “The Effect of Market Organization on Competitive Equilibrium: The Multi-unit Case,” *Review of Economic Studies, 40*, 97–113.