Over thirty years ago, one of the first reviews of experimental economics began by noting that "[experimental] research is rapidly becoming voluminous, but an overview of it can still be crammed into one article" (Rapoport and Orwant 1962, 1). The task facing contemporary reviewers is orders of magnitude larger, and in the general overview of the literature presented in this chapter, and even in the more focused chapters that follow, a good deal of selection has been required.

Nevertheless, as this multiyear task nears completion, I find myself feeling about experimental economics as a distinct enterprise much as I felt about game theory in the late 1970s. At that time there was already a very considerable body of game theoretic work, but the revolution that has today made game theory an important part of most areas of economic theory was still in its early stages. And in the late '70s it was still—just barely—possible to know, or at least know of, everyone who had done important work in game theory.

Today, as we approach the midpoint of the 1990s, experimental economics looks much the same way. Quite a substantial body of replicable experimental evidence has been gathered on a growing number of topics, in an increasingly cumulative way, as different groups of experimenters build upon one another's work. Experimental evidence appears regularly in the major economics journals, and it has begun to be reflected in the work of economists who do not themselves do experiments—both in research and in teaching. And with the ever growing numbers of economists who conduct experiments as at least a part of their research, it is fast becoming impossible to keep abreast of all the important new work.

In short, it is both an exhilarating time to do experimental economics and an excellent time to take stock of the experimental enterprise.

I begin this task at the beginning, with a brief history. My historical description, in section I of this chapter, concentrates mostly on the pioneering work done in the 1930s, '40s, and '50s, during which time a number of themes emerged that are still important in contemporary experimental economics. I content myself with painting the more recent history only in the broadest strokes, as this more
recent work will provide the substance both of section III of this chapter and of
the other chapters in the handbook.

Section II is also an attempt to set the stage for what follows, but in a different
way. It describes some of the different uses to which experimentation has been
and can be put in economics. I argue that there are few if any areas of economics
in which experimental methods do not have the potential to complement, at least
indirectly, more traditional methods of investigation. And experimentation gives
us a way to attack many important questions that do not yield easily if at all to
other methods.

Section III begins to follow series of experiments, in each of the general areas
covered by the handbook chapters. I take the series of experiments (rather than
single experiments) to be the unit around which the discussion is organized, be-
cause series of experiments allow the full power of the experimental method to be
displayed best.

I. A Brief History of Experimental Economics

In the course of coediting this handbook, it became clear to me that many contem-
porary experimental economists carry around with them different and very partial
accounts of the history of the field. The account that follows began as an attempt
to merge these "folk histories."1

I won't try to pin down the first economic experiment, although I am partial to
Bernoulli (1738) on the St. Petersburg paradox. The Bernoullis (Daniel and
Nicholas) were not content to rely solely on their own intuitions and resorted to
the practice of asking other famous scholars for their opinions on that difficult
choice problem. Allowing for their rather informal report, this is not so different
from the practice of using hypothetical choice problems to generate hypotheses
about individual choice behavior, which has been used to good effect in much
more modern research on individual choice.

But I think that searching for scientific "firsts" is often less illuminating than it
is sometimes thought to be. In connection with the history of an entirely different
subject, I once had occasion to draw the following analogy (Roth and Sotomayor
1990, 170):

Columbus is viewed as the discoverer of America, even though every school
child knows that the Americas were inhabited when he arrived, and that he
was not even the first to have made a round trip, having been preceded by
Vikings and perhaps by others. What is important about Columbus' discov-
ery of America is not that it was the first, but that it was the last. After
Columbus, America was never lost again.

That being the case, I will try to identify the historical context out of which
contemporary experimental economics has grown, by identifying early experi-
ments that have initiated streams of experimental investigation that continue to
the present. For this purpose, I begin in the 1930s. Starting from a low level of
activity, the literature of experimental economics has experienced exponential growth in every decade since, which has yet to level off.\footnote{2}

I will concentrate on three strands of the early experimental literature, each of which have left both substantive and methodological trails in the modern literature.

The first strand concerns experiments designed to test theories of individual choice. I will focus on an experiment reported by Thurstone (1931), concerned with ordinal utility theory, on an influential critique of this experiment by Wallis and Friedman (1942), and on subsequent experiments taking account of this critique by Rousseas and Hart (1951) and Mosteller and Nogee (1951), as well as on the celebrated work of Allais (1953).

The second strand I will concentrate on concerns tests of game-theoretic hypotheses. I will start with the experiment performed by Dresher and Flood in 1950, which formulated the now famous Prisoner's Dilemma game (Flood 1952, 1958), and continue with the work of Kalisch, Milnor, Nash, and Nering (1954), and Schelling (1957). And I will discuss the work of Suppes and Atkinson in the late 1950s, which investigated learning in game environments.

The third strand I will concentrate on concerns early investigations in Industrial Organization. I will focus on the work of Chamberlin (1948) and Siegel and Fouraker (1960).\footnote{3}

One of the methodological themes that can be traced in all three of these strands is how economists have come to rely today primarily on experiments in which subjects' behavior determines how much money they earn.

Finally, each of these strands of experimental economics was profoundly influenced by the publication in 1944 of von Neumann and Morgenstern's \textit{Theory of Games and Economic Behavior}, and I shall try to follow this connection also.

\textbf{A. Early Experiments: 1930-1960}

1. Individual Choice and the Wallis-Friedman Critique

An early formal experiment on individual choice, whose direct descendants in the economics literature are easy to follow, was reported by L. L. Thurstone (1931), who considered the problem of experimentally determining an individual's indifference curves.\footnote{4} Thurstone was concerned with testing the indifference curve representation of preferences and with the practicality of obtaining consistent choice data of the sort needed to estimate indifference curves. To this end he reported an experiment in which each subject was asked to make a large number of hypothetical choices between commodity bundles consisting of hats and coats, hats and shoes, or shoes and coats. (For example, the questions about hats and shoes would involve expressing a preference between a bundle consisting of eight hats and eight shoes and one consisting of six hats and nine shoes, and so on for many such pairs of bundles.) He reported the detailed data for one subject and found that, after estimating from the data the relative trade-offs the subject was prepared to make between hats and shoes and between hats and coats (under
the assumption that the indifference curves were hyperbolic), it was possible to estimate a curve that fit fairly closely the data collected for choices involving shoes and coats. Thurstone concluded that this kind of choice data could be adequately represented by indifference curves and that it was practical to estimate them in this way.

A lengthy and critical review of Thurstone's experiment was given by W. Allen Wallis and Milton Friedman (1942, particularly 177-83). One of their lines of criticism was that the experiment involved ill specified and hypothetical choices. They summarized their position as follows (179, 180):

It is questionable whether a subject in so artificial an experimental situation could know what choices he would make in an economic situation; not knowing, it is almost inevitable that he would, in entire good faith, systematize his answers in such a way as to produce plausible but spurious results.

For a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli. . . . Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react.5

Rousseas and Hart (1951) reported a subsequent experiment on indifference curves designed in reply to Wallis and Friedman and as a follow-up to Thurstone. They constructed what they viewed as a more concrete and realistic choice situation by having subjects choose from different possible breakfast menus, with each potential breakfast consisting of a specified number of eggs and strips of bacon. For added concreteness they specified that "each individual was obliged to eat all of what he chose—i.e. he could not save any part of the offerings for a future time" (p. 291).6 In this experiment individual subjects made only a single choice (repeated subsequently a month later) and also were asked to state their ideal combination of bacon and eggs. While this had the advantage of avoiding the artificiality of having subjects make many choices of the same type, it left Rousseas and Hart with the problem of trying to combine individual choice data collected from multiple individuals. They adopted the approach of seeing whether choices made by individuals with similar ideal combinations could be pieced together to form consistent indifference curves. Although they pronounced themselves satisfied with the results, we will see that the practice of testing theories of individual choice primarily on data from groups of subjects was regarded as questionable by subsequent experimenters.7

To put subsequent experiments in perspective, however, it is important to note that 1944 was the year in which von Neumann and Morgenstern's Theory of Games and Economic Behavior appeared. This presented and brought to wide attention both a more powerful theory of individual choice and a new theory of interactive behavior, and both had a profound influence not only on economic
theory but also on experimental economics. The predictions of expected utility theory gave a new focus to experiments concerned with individual choice, and the predictions of game theory—and its concern with precisely specified "rules of the game"—sparked a new wave of experimental tests of interactive behavior.8

Starting with the individual choice experiments, various aspects of expected utility theory were soon subjected to experimental investigation—see, for example, Preston and Baratta (1948), Mosteller and Nogee (1951), Allais (1953), Edwards (1953a, 1953b), May (1954), Davidson, Suppes, and Siegel (1957), and Davidson and Marschak (1959), to name only a few.9 Of these, the most closely connected to that of Thurstone (1931) is the experiment of Mosteller and Nogee (1951), who essentially sought to test expected utility theory in much the same spirit that Thurstone had examined ordinal utility theory. (Mosteller and Nogee were also well aware of the Wallis-Friedman critique of Thurstone's experiment.)10

Mosteller and Nogee (1951) began their paper thus (371):

The purpose of this paper is to report a laboratory experiment that measured in a restricted manner the value to individuals of additional money income. Although the notion of utility has long been incorporated in the thinking of economic theoreticians in the form of a hypothetical construct, efforts to test the validity of the construct have mostly—and in many cases necessarily—been limited to observations of the behavior of groups of people in situations where utility was but one of many variables.

Their point was that von Neumann-Morgenstern expected utility functions are derived from assumptions about individual choice behavior and that laboratory experimentation provides an opportunity to look at this behavior unconfounded by other considerations. Their general plan of attack had four main steps (372-3):

(a) to have subjects participate in a game with opportunities to take or refuse certain gambles or risks entailing use of real money; (b) from behavior in the game to construct a utility curve for each subject; (c) to make predictions from the utility curves about future individual behavior toward other and more complicated risks; and (d) to test the predictions by examining subsequent behavior toward more complex risks.

The method underlying their construction of the utility curves involved observing whether subjects would accept lotteries with given stakes as the probabilities varied. (They also devoted some attention to arguing that the size of the payoffs could be regarded as significant in terms of alternative employment opportunities available to their subjects.) Their general conclusions (399) were that it was possible to construct subjects’ utility functions experimentally and that the predictions derived from these utility functions "are not so good as might be hoped, but their general direction is correct." And with differences of emphasis, I think that this is a conclusion with which many experimental economists would still agree in the light of much subsequent work.
However, much more is now known about various systematic violations of expected utility theory that can be observed in the lab. Perhaps the most famous of these is the "Allais paradox." (Allais suggested, incidentally, that experiments could be used not only to test the predictions of particular theories of rational choice, but also to define rational behavior.) Allais asked subjects to make two hypothetical choices. The first choice was between alternatives A and B defined (Allais 1953, 527) as

A: Certainty of receiving 100 million (francs)

and

B: Probability .1 of receiving 500 million
   Probability .89 of receiving 100 million
   Probability .01 of receiving zero

and the second choice was between alternatives C and D defined as

C: Probability .11 of earning 100 million
   Probability .89 of earning zero

and

D: Probability .1 of earning 500 million
   Probability .9 of earning zero.

It is not difficult to show that an expected utility maximizer who prefers A to B must also prefer C to D. However, Allais reported that a common pattern of preference was that A was preferred to B but D was preferred to C. Note that although Allais's choices were hypothetical, the phenomenon he reported has subsequently been reproduced with real choices (involving much smaller amounts of money). Camerer discusses these matters in some detail in chapter 8.

It is worth noting that not all of the individual choice experiments motivated by von Neumann-Morgenstern expected utility theory in fact depended in any critical way upon the novel parts of that theory. For example, May (1954) reported that it was possible to elicit intransitive preferences in choices involving no uncertainty. His results thus show a violation of even ordinal utility theory, and his experiment could in principle have been conducted as a test of the earlier theory. However (as has often seemed to be the case since), the further development of the theory may have clarified the role that experiments could play.

2. Game-Theoretic Hypotheses

As mentioned earlier, following von Neumann and Morgenstern (1944), there also began to be considerable attention paid to experiments involving interactive behavior. We turn next to some of these.

In January of 1950, Melvin Dresher and Merrill Flood conducted at the Rand Corporation an experiment which has had an enormous if indirect influence, since it introduced the game that has subsequently come to be known as the prisoner's
The game they studied was the hundred-fold repetition of the matrix game given below, between a fixed pair of subjects who communicated only their choices of row (1 or 2) or column (1 or 2).

\[
\begin{array}{cc}
-1, 2 & 1/2, 1 \\
0, 1/2 & 1, -1 \\
\end{array}
\]

Payoffs were in pennies, with each player receiving the sum, over the one hundred plays of the game, of his payoffs in each play. The unique Nash equilibrium prediction is that the players should choose (2,1)—the second row and the first column—at each of the hundred repetitions. Thus the predicted earnings of the players are 0 for the row player (henceforth "Row") and $0.50 for the column player (henceforth "Column"). Of course this is inefficient, since if the players instead played (1,2) at every period, for example, their earnings would be $0.50 for Row and $1.00 for Column—i.e., they would both earn more. But this is not equilibrium behavior. The fact that equilibrium play is substantially less profitable than cooperative play made Dresher and Flood anticipate—correctly, as it turns out—that this game would present a difficult test of the equilibrium predictions.

The observed payoffs, for a pair of players whose play was reported in detail in Flood (1952, 1958) were $0.40 for Row and $0.65 for Column. This outcome is far from the equilibrium outcome, although it also falls considerably short of perfect cooperation. (As will be discussed in section III.A of this chapter, this observation has since been replicated many times.) Dresher and Flood interpreted this as evidence against the general hypothesis that players tend to choose Nash equilibrium strategies, and in favor of the hypothesis that a cooperative "split the difference" principle would be more powerful in organizing the data from games of this kind.

Despite their own interpretation of the data, Dresher and Flood included in their report of the experiment the following passage, describing an alternative interpretation given by John Nash (Flood 1958, 16):

Dr. Nash makes the following comment (private communication) on this experiment:

"The flaw in this experiment as a test of equilibrium point theory is that the experiment really amounts to having the players play one large multi-move game. One cannot just as well think of the thing as a sequence of independent games as one can in zero-sum cases. There is much too much interaction, which is obvious in the results of the experiment.

"Viewing it as a multimove game a strategy is a complete program of action, including reactions to what the other player had done. In this view it is still true the only real absolute equilibrium point is for [Row] always to play 2, [Column] always 1.

"However, the strategies:

[Row] plays 1 'til [Column] plays 1, then 2 ever after,
[Column] plays 2 'til [Row] plays 2, then 1 ever after,
are very nearly at equilibrium and in a game with an indeterminate stop point or an infinite game with interest on utility it is an equilibrium point.

"Since 100 trials are so long that the Hangman's Paradox cannot possibly be well reasoned through on it, it's fairly clear that one should expect an approximation to this behavior which is most appropriate for indeterminate end games with a little flurry of aggressiveness at the end and perhaps a few sallies, to test the opponent's mettle during the game.

"It is really striking, however, how inefficient [Row] and [Column] were in obtaining the rewards. One would have thought them more rational.

"If this experiment were conducted with various different players rotating the competition and with no information given to a player of what choices the others have been making until the end of all the trials, then the experimental results would have been quite different, for this modification of procedure would remove the interaction between the trials."

Dr. Dresher and I were glad to receive these comments, and to include them here, even though we would not change our interpretation of the experiment along the lines indicated by Dr. Nash.

Despite the limitations of this very exploratory, preliminary experiment, there are many ways in which it foreshadows some of the best of experimental economics. It tests the clear predictions of a general theory, on a difficult test case. And the results allow alternative hypotheses to be developed. When they are as clearly stated as Nash's comments, they suggest further experiments. We will return to some of the more modern of these subsequent experiments in section III.A. And as the quoted passage makes clear, some of the most interesting outcomes of an experiment may be the manner in which its results pit alternative interpretations against each other.

Note that in choosing a difficult test case, Dresher and Flood formulated a game that has since engaged both theorists and experimenters in a number of disciplines, as a large literature has developed around the prisoner's dilemma, which has been used as a metaphor for problems from arms races to the provision of public goods. This too is one of the indirect virtues of experimentation. The design of an experiment to test a particular theory often forces the experimenter to focus on specific aspects of the theory other than those that naturally come to the fore in the theoretical literature. The insights gained from designing an experiment are, as in this case, often of value even apart from the actual conduct of the experiment. Thus there is an interplay, on many levels, between theory and experiment.

In 1952 the Ford Foundation and the University of Michigan sponsored a conference on "The Design of Experiments in Decision Processes," which was held in Santa Monica (in order to accommodate the game theorists and experimenters associated with the Rand Corporation). Some of the experimental papers from this conference appear in Thrall, Coombs, and Davis (1954). The paper by Kalisch, Milnor, Nash, and Nering, which reported a small-scale experiment in-
volving several different $n$-person games, anticipates some issues of experimental
design that have played important roles in the subsequent literature. Some of the
games they looked at were constructed to allow particular theoretical predictions
to be tested on a domain on which the theories in question would make un-
ambiguous predictions. They write as follows:

The negotiation procedures were formalized (e.g., the identities of a player's
opponents were concealed from him and he was allowed to bid, accept, de-
cline, or counter-bid in a very limited number of ways...). The construction
of a theory to deal with an unlimited or very large number of negotiation
possibilities is as yet so difficult that it seems desirable to restrict and se-
verely formalize the negotiation procedure to the point where a meaningful
theory can be constructed. (302)

The choices the players made were not hypothetical, rather the profits they would
take home from the experimental session were proportional to their payoffs in the
experimental games. And (after finding mixed support for various game-theoretic
hypotheses) the authors concluded with a discussion of design features that might
make it easier to interpret future experiments, saying (326):

The same set of players should not be together repeatedly since there is too
much of a tendency to regard a run of plays as a single play of a more
complicated game.

It would be better to play an unsymmetrical game so that there would be no
obviously fair method of arbitrating the game and avoiding competition.

These two bits of advice are very different from one another, but are each repre-
sentative of what have proved to be important aspects of the design of economic
experiments.

The first bit of advice is solidly grounded in theory. If the same players play
a game more than once, their behavior even the first time they play may be differ-
ent than if they were going to play only once, since in the repeated case they
can anticipate that actions in the first period may affect the outcome in future
periods.18

The second bit of advice was grounded not in theory, but in a clearly observed
experimental regularity: in symmetric situations players often agreed on equal
divisions. By suggesting that this is because equal division in symmetric games
is a "fair" method of division and that experimenters should seek to avoid such
situations, the authors seem to have been suggesting that subjects are sometimes
motivated by considerations that the experimenter can only imperfectly control.
In this view, the demands of fairness in situations that seem to subjects to call for
fair behavior may sometimes overwhelm the motivations that the experimenters
are trying to induce (via the monetary rewards), so that the game being played is
different than the one intended by the experimenter.

Another hypothesis about why equal divisions are so often observed in sym-
metric situations was offered by Thomas Schelling. He proposed that in many
situations the problem facing economic agents is predominantly one of coordina-
tion, and that by focusing on outcomes that might be "prominent," some of the costs of coordination failure could be avoided. Schelling (1957) reported an experiment in which he confronted "an unscientific sample of respondents" with a variety of (hypothetical) problems. The following are two examples:

You and your partner (rival) are to be given $100 if you can agree on how to divide it without communicating. Each of you is to write the amount of his claim on a sheet of paper; and if the two claims add to no more than $100, each gets exactly what he claimed. If the two claims exceed $100, neither of you gets anything. (24)

You and your two partners (or rivals) each have one of the letters A, B, and C. Each of you is to write these three letters, A, B, C, in any order. If the order is the same on all three of your lists, you get prizes totaling $6, of which $3 goes to the one whose letter is first on all three lists, $2 to the one whose letter is second, and $1 to the person whose letter is third. If the letters are not in identical order on all three lists, none of you gets anything. (23)

Schelling reports that in the first of these problems, thirty-six out of forty subjects chose $50. Of course, since this yields an equal division, it could have been caused by a desire to be fair, instead of because it is a "prominent" outcome. But it is harder to explain the results of the next problem as a result of anything but the prominence of alphabetical order: 9 out of 12 As, 10 out of 12 Bs, and 14 out of 16 Cs chose the order ABC. This illustrates the power of experiments to test a hypothesis in different ways, the better to distinguish it from alternative hypotheses that might yield similar predictions on some domains.

Schelling's point was that a wide variety of cues could serve to make an outcome prominent and facilitate coordination. His comments were directed primarily at game theorists, the point being that highly abstract models might exclude factors that play an essential role in facilitating coordination. But there is a lesson for experimenters too, which is that details of how experiments are conducted may be of considerable importance, even if they concern features of the environment not addressed by existing theories. Sometimes these details will be worth study in their own right, and sometimes the experimenter will wish to avoid constructing the environment in a way that introduces unwanted influences (e.g., think how the results for the second problem would differ if the players were identified by colors instead of letters). The considerable influence of Schelling's experiments was for many years felt mostly indirectly, through the ways in which various kinds of phenomena were interpreted by subsequent authors. Recently, however, there has been a renewed interest in coordination experiments, motivated in part by macroeconomic questions, and these are discussed by Jack Ochs in chapter 3.

A final set of studies worth mentioning here were not designed to test game-theoretic hypotheses directly, but nevertheless involved careful study of strategic environments. Suppes and Atkinson (1960) reported an extensive series of experiments (involving more than 1,000 subjects) designed to investigate the predic-
tive power of simple learning theories in game situations (see also Atkinson and Suppes 1958, 1959). They started by reporting experimental sessions in which subjects didn't even know that they were playing a game, but were simply asked to make a sequence of choices, after each of which they were told what outcome had resulted, without being told that the outcome was determined in part by the actions of another subject in the experiment, with whom they were in fact playing a game. They then reported sessions in which subjects were increasingly informed about the game. In these sessions, subjects either knew they were playing a game, but did not see the payoff matrix, or else knew both that they were playing a game and how the outcomes were determined through the actions of the players. All of these sessions were conducted with hypothetical payoffs, and a final set of experimental sessions were conducted in which subjects were fully informed about the game and earned actual payoffs based on their performance.

In general, Suppes and Atkinson found that their results corresponded better to the predictions of learning theories than to those of game theory. They took care to emphasize that they found this unsurprising in view of the fact that in most cases their subjects did not know what game they were playing. However, when they reported experimental sessions in which subjects were fully informed about the game and rewarded for the outcomes, their results were somewhat different. In particular, when they considered (in chap. 10) the effects of monetary payoffs, they observed significant effects due to the presence and size of monetary payoffs on subject behavior, which they summarized as follows.

In the present study the subjects tended to approach an optimal game strategy as the monetary reward increased, and it may well be that this result can be directly generalized to more complex reinforcement schedules. However, results of the type described in Chapter 9 (where subjects were shown the payoff matrix and deviated from learning-theory predictions, but not in the direction predicted by game theory) leave the issue open to further analysis. (198)

Because Suppes and Atkinson used different games (zero and nonzero sum, two and three person, and with mixed and pure strategy equilibria) in their different information and payoff conditions, no general conclusion about how the conditions affect the accuracy of the game theoretic predictions can be drawn. (Recall that their experiments were designed primarily to test predictions of their learning theory.) However, their work foreshadows some contemporary experimental work that has developed in conjunction with a growing interest among game theorists in models of learning and adaptation.23

3. Industrial Organization

Turning now to the organization of markets, one early experiment that has exerted a major, if delayed, influence on modern experimentation was reported in 1948 by Edward Hastings Chamberlin. Chamberlin prefaced his article with an explana-
tion of what he thought laboratory experiments might bring to economics, beginning with a description of what he took to be the conventional wisdom on the subject. He wrote as follows:

It is a commonplace that, in its choice of method, economics is limited by the fact that resort cannot be had to the laboratory techniques of the natural sciences. On the one hand, the data of real life are necessarily the product of many influences other than those which it is desired to isolate—a difficulty which the most refined statistical methods can overcome only in small part. On the other hand, the unwanted variables cannot be held constant or eliminated in an economic "laboratory" because the real world of human beings, firms, markets, and governments cannot be reproduced artificially and controlled. The social scientist who would like to study in isolation and under known conditions the effects of particular forces is, for the most part, obliged to conduct his "experiment" by the application of general reasoning to abstract "models." He cannot observe the actual operation of a real model under controlled conditions.

The purpose of this article is to make a very tiny breach in this position: to describe an actual experiment with a "market" under laboratory conditions and to set forth some of the conclusions indicated by it. (1948, 95)

Chamberlin went on to describe the hypothesis motivating his experiment, which was that—contrary to the prevailing orthodoxy—market outcomes would often differ from competitive equilibrium "under conditions (as in real life) in which the actual prices . . . are not subject to 'recontract' (thus perfecting the market), but remain final" (95).

Chamberlin created an experimental market by informing each buyer and seller of his reservation price for a single unit of an indivisible commodity (i.e., for each buyer the price below which he could profitably buy, and for each seller the price above which he could profitably sell), and he reported the transactions that resulted when buyers and sellers were then free to negotiate with one another in a decentralized market. He noted that the reservation prices of the buyers, in aggregate, determined the market's demand curve, while the reservation prices of the sellers determined the supply curve, so that the competitive equilibrium (price and volume) could be established unambiguously and controlled by the experimenter (under only the assumption that buyers and sellers were willing to trade at the reservation prices established for them in this way).

The experiment he reported involved forty-six markets, with slightly varying equilibrium prices. He observed that the number of units transacted was greater than the competitive volume in forty-two of these markets and equal to the competitive volume in the remaining four markets, while the average price was below the competitive price in thirty-nine of these markets and higher in the rest. Chamberlin interpreted the systematic differences he observed between actual transaction prices and volumes and those predicted by the competitive equilibrium as supporting his hypothesis. At the same time, he noted that the results he observed caused him to correct an erroneous assertion he had made in Chamberlin
Figure 1.1. The induced supply and demand curves (top) and the observed prices compared to the equilibrium predictions. Source: Chamberlain 1948.
(1933, 27) that none of the "normally included buyers and sellers" (i.e., those who would transact at equilibrium) could fail to transact even when the market did not achieve equilibrium. In fact, what he observed was that sometimes a buyer, for example, might find that all of those sellers from whom he could afford to buy had already sold their unit to some other buyer, at a price below the equilibrium price. Figure 1.1 records the path of transactions over time in one of his markets. Chamberlin closed by cautioning that his results should be regarded as preliminary and noted that some of the regularities he observed might be sensitive to the shape of the supply and demand curves.

In the years since Chamberlin's experiment, his technique for constructing experimental markets with known supply and demand curves has been widely employed, and we shall return to it in section III of this chapter, and in the chapters by Holt, Kagel, and Sunder. More generally, Chamberlin's experiment illustrates the empirical power that comes from being able to create an environment in which the predictions of a theory (in this case competitive equilibrium) can be precisely known.

Like May's experiment on intransitivities in individual choice, Chamberlin's is an experiment that could have been done before von Neumann and Morgenstern, since it tested hypotheses that predated their work. Also, it should be noted that Chamberlin's experiment employed only hypothetical payoffs.

The end of the decade of the 1950s, and the beginning of the next, was marked by experiments concerning duopoly and oligopoly behavior, in the work of Hoggatt (1959), Sauermann and Selten (1959, 1960), and Siegel and Fouraker (1960) (which won the 1959 Monograph Prize of the American Academy of Arts and Sciences). The work of Siegel and Fouraker was perhaps the most extended experimental economics study reported up until that time.

Siegel and Fouraker (1960) reported a series of experiments in which subjects bargained in pairs until they reached agreement over a price and quantity, which served to determine their profits (each subject was given a payoff table that informed him of his own profits for each possible price and quantity). They designed a series of careful experiments to distinguish among a large variety of hypotheses concerning bilateral monopoly; hypotheses drawn from diverse sources in classical economic theory, game theory, psychology, and from the earlier game theory experiments of Schelling (1957). (They concluded [69] that "consideration of traditional economic forces cannot be depended on to yield an adequate explanation of the prices arrived at in bilateral monopoly bargaining.") One of the notable features of their experiments was the attention they paid to the information available to the subjects about each other's payoffs. They compared the case in which each subject knew only his own payoff table with the case in which one subject knew both payoff tables and the case in which both subjects knew both payoff tables. They found that, as the information increased in this way for the game they considered, the frequency with which subjects chose the Pareto optimal quantity increased, as did the frequency with which they chose a price that gave them equal payoffs.

Two methodological aspects of Siegel and Fouraker's work are especially no-
table. First, they took pains to insure that their subjects interacted anonymously, in order to avoid introducing into their experiment uncontrolled "social" phenomena. This is a subject to which I will return in chapter 4. Second, not only did they follow the increasingly common practice of motivating subjects with monetary payoffs, but they investigated the effect of changing the incentives by changing the size of the payoff differences that resulted from different decisions. That is, they were not content to observe that subjects could make substantial profits from choosing, for example, the Pareto optimal quantity. They also considered how much of a difference it made if the quantity chosen was only a little more or a little less than the Pareto optimum. They noted that in the first of the payoff tables they used this difference was small and conjectured that the variance they observed around the Pareto optimal quantity might be due to the fact that the subjects felt that the potential payoff difference was not worth the hazards of continued bargaining. They say (34):

If this reasoning is correct, then increasing the difference in payoff to each bargainer between contracts at the Pareto optima and contracts at quantities adjacent to the optima should lead to the negotiation of a higher percentage of contracts on the optima.

They then went on to present results obtained from payoff tables that increased the size of these differences and reported that they did indeed observe much less variance around the Pareto optimal quantity.

Siegel and Fouraker used their results to motivate a theory based on the "level of aspiration" of the subjects, which they proposed was the variable effected by the differing amounts of information. They went on to explore this hypothesis in oligopoly models as well (Fouraker and Siegel 1963). Independently, Sauermann and Selten (1959, 1960) formed related hypotheses on the basis of rather different oligopoly experiments.28

I think Siegel and Fouraker's views on the place of experimentation in economics have stood the test of time. They said (1960, 72-3):

Our data have been observations made specifically to meet the purposes of this research. We have not turned to preexisting data. In the specific case of bilateral monopoly, it would be extremely unlikely that appropriate naturalistic data could be collected to test the theoretical models. . . . Although exchanges under bilateral monopoly conditions are common, such . . . descriptions as may be available will not generally be in an appropriate form for testing theoretical models. Following Boulding [unpublished speech], we may say that in science the shift from relying on existing information collected for other purposes to using information collected specifically for research purposes is analogous to primitive man's shift from food collecting to agriculture, and "provides the same kind of stimulus to development and accumulation. It is when a science starts to go out to ask exactly what it wants to know, rather than relying on information collected for other purposes, that it begins to obtain control of its own growth."
We have made our observations under controlled conditions. We have not only collected observations especially for this research, but we have also done so under conditions which make the observations relevant to the research purposes. In using the laboratory rather than the field, we have been able to isolate the phenomena of interest to the research.

We have used the experimental method. That is, we have manipulated certain variables and observed the effects of variations in these upon certain other variables. By so doing, we have demonstrated that the amount of information available to a bargainer and his level of aspiration are significant determinants of the price-quantity contracts which will be reached. We aver that only the experimental method could have demonstrated the influence and importance of these determinants.

Note that, in analyzing their experimental results, Siegel and Fouraker sought to develop game theory in new directions; that is, they sought not merely to test game theoretic predictions, but to develop new theories better able to predict the outcome of games such as those they studied. Schelling (1958; see also Schelling 1960) further proposed that the relationship between game theory and experimentation must be two way in this sense. In a section of his paper entitled "Game Theory and Experimental Research" he wrote (1958, 257):

> some essential part of the study of mixed-motive games is necessarily empirical. This is not to say just that it is an empirical question how people do actually perform... It is a stronger statement: that the principles relevant to successful play, the strategic principles, the propositions of a normative theory, cannot be derived by purely analytic means from a priori considerations.

It is striking to note a number of distinguished game theorists among the earliest experimenters (Nash, Schelling, Selten, and Shubik, for example, set a high standard of distinction by any measure). I have already indicated that I think this is no accident. Rather, game theory brought to economics a kind of theory that lent itself to experimental investigation, and in some cases demanded it. The reason is that it seeks to provide precise models of both individual behavior (in von Neumann-Morgenstern utility functions) and of economic environments. This concern with the "rules of the game," the institutions and mechanisms by which transactions were made, together with precise assumptions about the behavior of individuals and the information available to them, gave rise to theories that could be tested in the laboratory.

Before moving on to more modern experiments, it should be noted that by the end of the 1950s two of the features that have come to distinguish experimental economics were already clearly in evidence. The first of these, just referred to, is the concern for testing theories of great potential generality (such as theories of equilibrium) on specific, controlled environments and the consequent attention to rules of play. The second is the fact that many of the experiments attempted to gain control of subjects' motivations by paying the subjects based on their perfor-
performance so that subjects' performance could be analyzed under the assumption that they were seeking to maximize the utility (or sometimes simply, the expected value) of the money earned. That is, by this time the reaction of experimental economists to the Wallis-Friedman critique of hypothetical choices was already beginning to take shape in a tendency to rely primarily on experiments in which subjects' behavior determined their monetary payoffs.\(^{30}\)

And although the end of the 1950s marks a time that is still quite early in the development of experimental economics, a number of the experiments that were completed well before then have continued to exert a powerful influence on modern research, in ways that will be covered at greater length in both the other sections of this chapter and the other chapters of this volume. Individual choice experiments in the spirit of Allais (1953) have inspired the search for other systematic violations of expected utility theory, and these are surveyed by Camerer in chapter 8. Prisoner's dilemma experiments in the spirit of Dresher and Flood (Flood 1952, 1958) became a small cottage industry by themselves, and also influenced game theory in ways that make their full effect hard to grasp, but they are very close kin to the public goods experiments described by Ledyard in chapter 2. And the basic design of Chamberlin (1948) for inducing individual reservation prices and aggregate supply and demand curves has become one of the most widely used techniques in experimental economics, and plays a role in many of the chapters of this volume, as do the methodological considerations raised by Siegel and Fouraker (1960). Finally, the theoretical work of the early game theorists, especially of von Neumann and Morgenstern and of Nash, have had such profound effects on both modern economic theory and on experimental economics that it is fair to say that their influence is pervasive in every chapter of this volume.

**B. The 1960s to the Present**

The 1960s were a decade of steady growth for experimental economics, and the first reviews of economics experiments began to appear (see Rapoport and Orwant 1962; Cyert and Lave 1965; and Friedman 1969).\(^{31}\) Rapoport and Chammah (1965) compiled a considerable body of work associated with the prisoner's dilemma, and a set of German experiments is reported in Sauermann (1967) (who may have coined the term "experimental economics").\(^{32}\) Well over a hundred experimental economics papers were published in the 1960s.\(^{33}\) By the end of the decade a good deal of thought had begun to be given to questions of experimental methodology as such (see, for example, the description of a computerized laboratory by Hogatt, Esherich, and Wheeler 1969).

An important methodological advance, which illustrates the close connection between economic theory and experimental design, came in the work of Becker, DeGroot, and Marschak (1964). They conducted an experiment to measure individuals' expected utility functions and were concerned with the problem of how to motivate experimental subjects to reveal their "true" reservation prices for lotteries. The solution they hit upon was that each subject was endowed with a
lottery and then asked to name the amount of money for which he would be willing to sell it. Each subject was told that the selling price he named would be compared with a price to be determined randomly. If the randomly determined price (the offer price) was higher than the named selling price, then the experimenter would buy the lottery from the subject for the randomly determined offer price (not for the named selling price); otherwise the subject would keep and play the lottery and earn its random outcome. It is not hard to see that the dominant strategy for a utility maximizer faced with such a mechanism is to state his true selling price, (i.e., the price that makes him indifferent between selling the lottery or keeping it). Because many modern economic theories are based on the assumption that agents are expected utility maximizers, so that it is frequently the case that the predictions of the theory can only be known if the utility functions of the subjects can be accurately estimated, this technique has found wide application even in experiments whose primary purpose is not to estimate utility functions.

Note that if the subjects are not expected utility maximizers, the selling price elicited in this way may not have all the properties that it would for utility maximizers. But to test the predictions of a theory that assumes the subjects are utility maximizers one first needs to know what the theory predicts, and the BDM procedure just described gives experimenters one way of knowing precisely what are the predictions of the theory they are testing, for the subjects they are examining, in the environment they have created. The ability to test the predictions of a theory when they are precisely known in this way is one of the principle attractions of controlled experimentation in the laboratory.

An important experiment from this period that established another way in which individuals may systematically deviate from being perfect subjective expected utility maximizers was Ellsberg (1961), who showed that Bayesian probability assessments are not always descriptive of observed behavior. Camerer will report on some of the many followups to this line of investigation in chapter 8. Another influential experiment from this period, which employed, to quite different effect, the basic design of Chamberlin's (1948) experiment, was reported by Vernon Smith (1962), and this is discussed at greater length in section III.D of this chapter and by Holt in chapter 5. Smith (1992) reports that his first experiment (Smith 1962) was motivated by a desire to see if competitive outcomes could be observed using Chamberlin's basic design but with different pricemaking rules and with repetition using constant parameters. The 1970s brought further growth, including growth of research support. In this latter regard, the fact that the National Science Foundation began providing sustained support to a number of different laboratories had a significant impact on the development of experimental economics. Jim Blackmun and Dan Newlon are the NSF officials who came to be most closely associated with this critical support. The 1970s were also marked by a number of conferences, in (then West) Germany and in the United States, that began to bring different groups of experimenters into contact (see Sauermann 1972, 1978a, 1978b, for collections of papers from the German conferences). Experimenters who played a leading role
in these activities include Heinz Sauermann, Reinhard Selten, and Reinhard Tietz in Germany, and Charles Plott and Vernon Smith in the United States. During this time, the experimental economics literature became increasingly distinct from experimental psychology, although the two literatures retain many points of overlap.37

In the 1980s and early '90s the growth of experimental economics became explosive, and with this growth came the accompanying signs of having become a "mainstream" subject.38 An early experimenter—Maurice Allais—even won the 1988 Nobel Memorial Prize in Economics.39 The vastly increased numbers of experiments and experimenters also instigated a sea change in the way experimental economics is done. For the first time, there began to be a wide variety of areas in which different groups of experimenters began to study the same issues from different points of view. This meant that there began to be series of experiments in which investigators with different hypotheses responded to one another's experiments, critically examining earlier conclusions. It is this process, in which experimental results suggest new experiments and in which different theoretical points of view suggest different experiments to different groups of experimenters, that allows us to begin to look back on experimental economics as a cumulative process. This kind of dialogue is one of the great sources of strength of the experimental method, a sentiment which has been expressed somewhat jocularly by the experimental psychologist Georg von Bekesy:

Another way of dealing with [experimental research] errors is to have friends who are willing to spend the time necessary to carry out a critical examination of the experimental design beforehand and the results after the experiments have been completed. An even better way is to have an enemy. An enemy is willing to devote a vast amount of time and brain power to ferreting out errors both large and small, and this without any compensation. The trouble is that really capable enemies are scarce; most of them are only ordinary. Another trouble with enemies is that they sometimes develop into friends and lose a good deal of their zeal. It was in this way that the writer lost his three best enemies. (1960, 8-9)

It is the development of experimental economics into this kind of cumulative, progressive dialogue that makes this handbook possible today, when not many years ago it would have been premature.

II. The Uses of Experimentation

Experiments can be done for different reasons, and in reading about the series of experiments described in the remainder of this chapter and of this volume, it may help to keep some of them in mind. In the summer of 1985, in a symposium on experimental economics at the Fifth World Congress of the Econometric Society, I suggested that experiments might be loosely classified according to how they were motivated and to whom they were intended to be persuasive, that is, accord-
ing to the dialogues they were part of.\textsuperscript{40} I referred to these different kinds of dialogues as "Speaking to Theorists," "Searching for Facts" (and, closely related, "Searching for Meaning"), and "Whispering in the Ears of Princes." To the extent that these categories have since gained wider currency, it is perhaps because they can be used as a metaphor for what economists do generally.\textsuperscript{41} For the purpose of looking at experiments, they help to focus attention on how different kinds of experiments emphasize different aspects of experimental design. Most series of economics experiments, and even many individual experiments, have elements of more than one of these motivations.

The category "Speaking to Theorists" includes experiments designed to test the predictions of well articulated formal theories, and to observe unpredicted regularities, in a controlled environment that allows these observations to be unambiguously interpreted in relationship to the theory. Such experiments are intended to feed back into the theoretical literature—i.e., they are part of a dialogue between experimenters and theorists.

The category "Searching for Facts" includes experiments studying the effects of variables about which existing theory may have little to say. Often these experiments are motivated by earlier experiments and are designed to isolate the cause of some observed regularity, by varying details of the way the experiments were conducted. Such experiments are part of the dialogue that experimenters carry on with one another. And as these facts begin to accumulate, "Searching for Meaning" becomes possible, as theories of the observed behavior can be proposed and then tested. So this kind of work contributes also to the dialogue between theorists and experimenters.

The category "Whispering in the Ears of Princes" deals with the dialogue between experimenters and policymakers. These experiments might be motivated, for example, by the kind of question raised by regulatory agencies, about the effect of changes in the way some market is organized. Their characteristic feature is that the experimental environment is designed to closely resemble, in certain respects, the naturally occurring environment that is the focus of interest for the policy purposes at hand. This category of experiments has so far not given rise to any extended series of experiments by different investigators but offers the possibility of bringing scientific methods to bear on one of the traditional responsibilities of economists, to formulate advice on questions of policy whose answers lie beyond the reliable scientific knowledge of the profession.\textsuperscript{42}

What this classification should suggest is that experiments may potentially play a role in most of the things that economists do. Since many economists find (or used to find) this a surprising idea, it may be helpful to point out, by way of a loose analogy, that experiments play a role in most of the things that biologists do, and, like economists, biologists have a lot of ground to cover, from molecular biology to evolution, to medicine. Experiments can obviously play a very direct role in testing and refining theories of molecular biology, since the phenomena in question can be brought entirely into the laboratory. But although experiments cannot be conducted on the fossil record, evolutionary biologists nevertheless obtain much of their understanding of selection and evolution from experiments.
in microbiology, genetics, and plant breeding. And while physicians are often called upon to treat diseases that are beyond the reliable scientific knowledge of their profession, clinical trials help them to discern the effects of different drugs even when the mechanism by which they work is still obscure. In the same way, economic experiments may play a role not only in testing and refining theories concerned with individuals or small groups, but also concerning questions about large markets, industrial organization, and macroeconomics.

The remainder of this chapter, and of this volume, aims to show how economists can use the tools they find in the laboratory to make steady, incremental progress on answering questions that might otherwise be intractable. We focus on series of experiments, which together tell us more than any one of them. By "more" I do not mean that these series of experiments necessarily permit us to draw broader conclusions than might have seemed warranted on the basis of one experiment alone. While this will sometimes be so, subsequent experiments sometimes define more narrowly the conditions under which some initially observed phenomenon occurs and sometimes cause the results of an earlier experiment to be entirely reinterpreted and initial conclusions to be rejected. What I hope this volume will illustrate is how series of experiments can be constructed to allow us to draw more reliable conclusions, both about what we know and about what we know we don't know.

This process is particularly informative when different groups of investigators with different theoretical points of view conduct experiments intended to test each others' conclusions. Series of experiments arising in this way may still leave room for experts to differ, but they narrow the room for disagreement and clarify the nature of the remaining disagreement by forcing investigators to refine their hypotheses. And, long after many of the particular experiments so far conducted have receded to no more than historical importance, it is this that is likely to be the chief contribution of controlled experimentation to economics.

III. Some Series of Experiments

My aim in this section is to introduce some topics of contemporary interest to experimenters and in doing so to introduce and provide some context for the specialized chapters of this volume.

Chapter 2, by Ledyard, deals with the problems associated with providing public goods, and the incentives that agents have to "free ride" and enjoy the goods that may be produced by others without contributing to their production. As in the prisoner's dilemma, if no one cooperates to produce the public good, everyone is worse off than if they had cooperated. Ledyard proposes that public goods provision games in which the equilibrium behavior is for no one to contribute to the production of public goods can best be thought of as prisoners' dilemmas, while those games in which the equilibria call for some players to contribute and others
to free ride are like the game of "chicken" (see Table 2.11). I will accordingly start my discussion in section III.A with some prisoner's dilemma experiments, before moving on to series of experiments dealing with more traditional formulations of public goods provision. This is an area in which experimental results have led to new directions of theoretical investigation, including various ways of considering whether players may have motivations not well captured by the payoffs in the game, such as altruism, or other kinds of preferences concerning more than one's own welfare.

Chapter 3, by Ochs, considers problems of coordination. Coordination problems arise in a fundamental way in games with multiple equilibria, since in general if the players fail to coordinate their expectations and actions on a single equilibrium, the outcome is inefficient. The frequency of inefficient outcomes and the conditions which make them more or less likely to occur, are a theme that this area shares with the study of public goods provision, which is one of the very many domains in which coordination problems arise. Schelling's work in the 1950s poses some of the problems still being investigated today. Compared to other chapters in this volume, however, there are relatively few extended series of experiments in this area, and so my introduction in section III.B will be relatively brief and will concentrate on the way in which coordination experiments cast light on the manner in which equilibrium is reached. Data from coordination games suggest that models of learning and adaptive behavior will be important in understanding which (if any) equilibria are reached as players gain experience in a game.

Bargaining is the subject of chapter 4, and, since I am its author, I won't try in this introduction to discuss the same material from a different point of view. Instead, I concentrate in chapter 4 on experiments concerning models of bargaining with highly structured rules for exchanging offers. A very active interchange among experimenters has developed in the last ten years on this subject. I will lay the groundwork for this in section III.C of this chapter by discussing a largely earlier (but somewhat overlapping) series of experiments concerned with unstructured bargaining, in which many of the same issues arise. (And as I was a principal investigator in those earlier experiments, this will also serve to indicate the point of view that I brought to the dialogue among many experimenters, which is the subject of chapter 4.) Like the two chapters that precede it, chapter 4 is concerned with inefficient outcomes, which in bargaining take the form of disagreements and delays. As in the case of public goods problems, theories of other-regarding preferences have been proposed and tested to explain deviations from received theory (with "fairness" playing the role in bargaining that "altruism" plays in the discussion of public goods). And bargaining has some resemblance to the coordination problems of chapter 3, with bargainers needing to coordinate their expectations to reach an agreement.

Chapter 5, by Holt, and chapter 6, by Sunder, both consider series of experiments that arise out of the vibrant tradition whose origins are in Chamberlin's (1948) experiment. Like Chamberlin's experiment, the experiments in these two chapters seek to test theories of exchange that can be formulated in terms of the
aggregate supply and demand curves of the market, and these are induced in the laboratory using Chamberlin's technique of giving each buyer and seller a reservation price for each unit they demand or supply. The two chapters between them study a variety of different forms by which market exchange can be organized, as well as differences due to the kind of commodity being traded. (Sunder considers trading in financial assets, which may have different value to different participants depending on the state of the world, about which different participants may have different information.) One important form of market organization is the double auction market, first experimentally studied by Smith (1962), who observed rapid convergence to competitive equilibrium when the market was repeated several times with stationary parameters. Subsequent investigators have made substantial progress in understanding this phenomenon, and this is one thread of the literature that I consider in this introduction.

Chapter 7, by Kagel, concerns a variety of markets organized as auctions and focuses primarily on tests of game-theoretic hypotheses about the effect of different auction rules and different kinds of commodities. One important distinction among commodities is whether their value to each bidder is known or unknown, and, if unknown, whether or not the value to each bidder is independent of the value to other bidders. In many cases of interest the value of the object being auctioned is unknown to the bidders, but is essentially the same for all of them. (Consider, for example, the auction of the right to drill for oil at a given location—the value of this right is highly dependent on the unknown amount of recoverable oil beneath the surface.) One question that arose in the trade literature, and has been addressed in the experimental literature, was whether there might be a tendency for the winning bidder in such a "common value" auction to be a bidder who overestimated the value. This is the theme I concentrate on in my introduction, as it forms a good bridge between the equilibrium considerations of the preceding chapters and the errors and choice anomalies that are the focus of chapter 8.

Chapter 8 is concerned primarily with experimental observations of individual choice behavior that are at odds with the view of decision makers as idealized rational information processors and (expected utility) maximizers. The work in this chapter, more than any other in the volume, represents an interaction between experimental economists and psychologists, and as such it gives a view not only of the substantive issues involved in various particular debates, but also of how these are played out across a disciplinary boundary that sometimes divides a common subject matter. By way of introduction, I concentrate on a phenomenon called preference reversal, in which subjects can sometimes be observed to choose one of a pair of lotteries when faced with a choice between the two, but to name a higher price for the other lottery when asked to specify a price for which they would be prepared to sell the right to participate in the lotteries. This is a phenomenon that has raised both questions of experimental procedure, and, once its robustness was established, questions about its implications for economic theory. In this respect, the phenomenon of preference reversal is similar to many of the phenomena discussed by Camerer in chapter 8.
Variations on the prisoner's dilemma have been the subject of virtually continuous experimental interest since the 1950 experiment of Dresher and Flood was reported in Flood (1952, 1958). But whereas they deliberately chose an asymmetric form of the game, much of the subsequent literature has focused on the symmetric game, corresponding to the story of the two prisoners formulated by Tucker (1950). He referred to a game we can represent by the following matrix, with $b > a > c > d$.

\[
\begin{pmatrix}
  c & n \\
  n & ( \\
\end{pmatrix}
\]

The "dilemma," of course is that it is a dominant strategy for each prisoner to confess, since $c > d$ and $b > a$, but that both of them would be better off if neither confessed, since $a > c$. So the only equilibrium of this game is the dominant strategy equilibrium at which both prisoners confess and receive the (non-Pareto optimal) payoff of $c$ each. (In much of the literature the strategy "not confess" is called "cooperate," and "confess" is called "defect").

The observation that equilibria could be inefficient did not strike game theorists as odd (always assuming, of course, that the situation facing the players, and their preferences over the outcomes, are accurately modeled by the above matrix); rather it served to emphasize the usefulness of being able to write binding contracts. However to many social scientists this conclusion seemed to represent an error in analysis, their feeling being that when players properly understood the game, they would choose to cooperate with one another and not confess.

A related observation, however, struck (even) game theorists as symptomatic of problems with the notion of equilibrium. If a prisoner's dilemma game is repeated finitely many times, say 100, and if the payoffs to the players are the sum of their payoffs in each game, then it can be seen by backwards induction starting from the last period that no equilibrium of the game yields cooperation at any period. That is, the unique equilibrium behavior involves confessing at every period, even though confessing is no longer a dominant strategy (recall the discussion in section I). Not only did this seem contrary to intuition, it was also disturbing to note that the equilibrium prediction was unchanged no matter how many times the game was repeated. So even as the number of repetitions increases, the finitely repeated game does not approach the infinitely repeated game (or the game played in continuous time) in which cooperation is (also) an equilibrium behavior. For these reasons the finitely repeated game received special note in the game theory literature.

The prisoner's dilemma has motivated literally hundreds of experiments, and so I will not even attempt to review them individually. (Representative examples of early work are Lave 1962, and Rapoport and Chammah 1965.) Typical ex-
experiments concerning the one-period game reported a level of cooperation that responded readily to various kinds of experimental manipulation but that was bounded well away from either zero or 100 percent. A number of experiments were conducted to isolate various factors contributing to the level of cooperation.

However many experiments which were analyzed as one period games were in fact conducted on various kinds of repeated games, using rules that made it difficult to determine precisely what the equilibria were. In a paper about designing prisoner's dilemma experiments Roth and Murnighan (1978) wrote:

It is often contended in the literature that if subjects are not informed of the number of periods to be played, the resulting game yields the same equilibria as the infinite game, since no period is known to be the last. However, this is a considerable oversimplification. Since it is apparent that the game must eventually terminate, subjects must form subjective probabilities greater than zero that a given period might be the last. Although such probabilities have neither been observed nor controlled by experimenters, we shall see that they play a critical role in determining the nature of equilibrium outcomes.

The paper goes on to derive the conditions for equilibrium in the repeated game with a fixed probability $p$ of continuing after each play: cooperation can be achieved at equilibrium only if the probability of continuing is sufficiently large.

A pilot experiment was then conducted, in large part to show that the design was feasible. The payoff matrix was chosen so that cooperation was consistent with equilibrium if and only if $p \geq 1/3$, and subjects played three games, with probabilities of continuing of .1, .5, and .9. (Half the players played in that order, half the players in the opposite order.) The results of the experiment were that significantly more cooperative choices were made in the two higher probability conditions (in which these are equilibrium choices) than in the low probability condition. However, even in the high probability condition, only 36 percent of first period choices were cooperative, while in the low probability condition 19 percent of the first period choices were (nevertheless) cooperative. So the results remain equivocal.

Similarly equivocal results seem to be typical. A recent experiment, whose results help crystallize a lot of what I think has been observed piecemeal in previous experiments, is reported by Selten and Stoecker (1986). In their experiment, subjects played twenty-five "supergames," each of which was a (ten-period) repeated prisoner's dilemma. So this experiment looked at repeated play of the repeated game and thus gave subjects the opportunity to gain experience with the ten-period game.

By far the most common pattern of observed play was initial periods of mutual cooperation (at least four), followed by an initial defection, followed by noncooperation in the remaining periods. After about round 16 almost all of the plays exhibit this pattern in each round. (A round is a play of the supergame; that is,
round 22 is the twenty-second repetition of the ten-period repeated game.) Even more common is the pattern of "end-effect play," which the authors define to be at least four consecutive rounds of mutual cooperation (not necessarily starting from period 1), with no further cooperation following the first defection. (Notice that this pattern includes the previous one.)

The most striking results concerns the progress in the observed (and "intended") period of first defection. Having learned to cooperate, players start to defect earlier and earlier in subsequent supergames—i.e., the cooperation starts to unravel from the end.50

The paper then develops a learning theory model in which each player is represented by the period in which he intends to defect, and updates this, via three probabilities, depending on whether he defects first, simultaneously, or last. Steady state probability distributions are computed for various parameter configurations: It appears that in the typical stable distribution, cooperation either breaks down very early or very late. Monte Carlo simulations based on parameters estimated for each subject based on the first twenty rounds are then made for the pairings in the last five rounds. Like the observed behavior, these predictions have cooperation unravelling from round 20 to round 25.

I think it is fair to summarize these observations as follows: in the initial rounds players learned to cooperate (and consequently exhibited more periods of mutual cooperation starting from the very beginning and breaking down only near the end). In the later rounds, players learned about the dangers of not defecting first, and cooperation began to unravel. There is a sense in which this observed behavior mirrors the game-theoretic observation that the equilibrium recommendation never to cooperate isn't a good one, but that all other patterns of play are unstable. These observations are consistent with many earlier observations of finitely repeated games in which cooperation is observed for some periods, but breaks down near the end. A number of new theories have been motivated by such experimental observations. For example Kreps, Milgrom, Roberts, and Wilson (1982) propose a model in which players may entertain certain slight doubts about the nature of their opponent, who may not after all be the sort of player to whom the backward induction logic applies (either because of limitations on what strategies he may play, or because he may in fact have some inclination to cooperate not captured by the standard payoffs of the game). They show that at equilibrium of such a model in which the uncertainties are of the right sort, there will be cooperation until near the end, because players will have a motive to engage in building reputations as the kind of players who will continue to cooperate when faced with cooperation. Andreoni and Miller (1993) report an experiment motivated by this model and observe both that some players in their subject pool apparently do have some inclination to cooperate not captured by the monetary payoffs and that the observed rates of cooperation are consistent with a model of reputation building on the part of the other players.51

In summary, interest over the course of many experiments has shifted from the one-time game to the repeated game. The contemporary discussion proceeds on both theoretical and experimental lines.
a. Experiments versus Simulations: A Methodological Digression

There was a time when computer simulations, and the kinds of investigations one can do with them, were sometimes confused with experiments involving the observation of real people in controlled environments. Selten and Stoecker's use of both technologies makes the distinction clear. Computer simulations are useful for creating and exploring theoretical models, while experiments are useful for observing behavior.

Recently, however, there have been a growing number of investigations that combine experimentation and computer simulation. For example, an interesting set of computer simulations that have an unusually experimental flavor are reported in Axelrod (1980a, 1980b, 1984). These have their origin in a pair of computer "tournaments." In the first of these, the author elicited, from fourteen scholars in several disciplines who had written on the prisoner's dilemma, short computer programs encoding a strategy to play the repeated game. Each of the programs played each of the others, as well as a copy of itself and a program that generated random choices, in a 200-play repeated prisoner's dilemma. The strategy with the highest cumulative score was "tit for tat," which starts with cooperation and then echoes the other program's previous move. It and all of the other highest scoring rules were "nice" in the sense that they never defected first. Some programs got into sequences of alternating moves with tit for tat, with one program defecting on the odd numbered moves and cooperating on the even numbered moves and tit for tat doing the opposite, which for the parameters used in the tournament was not nearly as profitable as steady cooperation. This is a pattern you might expect humans would be able to avoid, although it is easy to see how short computer programs could fall into it.

Axelrod (1980b) presented a second round of the tournament, with new entries, in which the game was repeated with a fixed probability of continuation after each round (with p = .99 so that now cooperation is an equilibrium strategy), as discussed above in connection with Roth and Murnighan (1978). Again, tit for tat was the winner. Some simulations of different possible tournaments were presented to show that there are some senses in which this result is robust, but other results were reported to show that this is not an entirely simple matter: "Had only the entries which actually ranked in the top half been present, then TIT FOR TAT would have come in fourth after the ones which actually came in 25th, 16th, and 8th" (402).

These computer tournaments thus suggest that behavior will eventually converge to cooperation. This conclusion is at odds with experimental results such as Selten and Stoecker's. I suspect that the difference in results has a great deal to do with the learning that goes on when experimental subjects are allowed to adapt their strategies as they gain experience with the game and with the behavior of the rest of the subject pool.

Apart from the use of simulation, there are also interesting questions raised by eliciting from the subjects in an experiment a complete strategy—i.e., a rule of action that determines in advance their decision at each of their information sets in the game—instead of having them make only those decisions that arise in the
course of the play of the game, as they arise. Experimental designs using this feature may first have been explored by Selten (1967c). I discuss such designs, in chapter 4, in the methodological digression on the "strategy method."

I turn now from the experiments motivated directly by the prisoner's dilemma to those motivated by the provision of public goods.55

2. The Free Rider Problem in Public Goods Provision

The free rider problem in the provision of public goods was noted in connection with the debate among nineteenth-century economists about whether taxation for public goods should be related to the benefit each agent derived from those goods. The nature of a public good is that once it has been created everyone may use it, and so if each individual is to be taxed in proportion to the profit he derives from the public good, there will be an incentive for individuals to claim that these profits are small, since small contributors will derive the same benefit from the good as if they had been large contributors. The potential for under-contribution to a public good is particularly clear when contributions are voluntary. (American listeners to National Public Radio will immediately recognize the problem.)

The first clear formulation of the free rider problem is generally attributed to an essay written at the end of the last century by the Swedish economist Knut Wicksell, who also anticipated the direction of much subsequent theoretical research by suggesting that the mechanism by which public projects were decided upon would be important. (He suggested that a way to deal with the problem would be to require that proposals for public projects be considered together with proposals to raise the necessary revenue and that the whole package should be subject to [close to] unanimous approval.) For references and an introduction to much of the subsequent theory focusing on the role of the decision mechanism, see Green and Laffont (1979).

Because it is readily apparent that some more-or-less public goods are in fact produced even though they depend on voluntary contributions, the focus of debate shifted both to assessing how serious the free rider problem might be and what circumstances or mechanisms might ameliorate it. So at the same time as a good deal of theoretical progress was being made in "solving" the free rider problem (e.g., Groves and Ledyard 1977), skepticism was being voiced about the importance of the problem and the quality of the empirical evidence in support of it (e.g., Johansen 1977). Since it is difficult to collect field data to determine, for example, how close to the optimum amount of some public good is being supplied, this problem presented a natural opportunity for laboratory experiments. In addition, since some of the mechanisms proposed for solving or ameliorating the free rider problem had no counterpart in existing institutions, some of the questions that presented themselves could not be addressed except by experimentation.
An early public goods experiment, by Bohm (1972), was sponsored by the Swedish Radio-TV broadcasting company. A sample of adult residents of Stockholm was invited to come to the broadcasting company for an interview and asked to state how much (of their interview fee) it would be worth to them to see a half-hour program by two well-known comedians. They were told they would see the program only if the sum of the amounts stated (by their group and others) exceeded the cost of showing it. The experimental variable consisted of five different rules for how they would in fact be charged on the basis of their stated amounts, ranging from the full amount to some percentage of that amount, to a lottery related to the amount, to a small fixed fee, to nothing.

The responses of the different groups of subjects given these different instructions were found not to vary significantly. Bohm argues that the first payment mechanism (everyone pays their stated amount) gives no incentive to overstate willingness to pay, and the last (no actual payment required) gives no incentive to understate willingness to pay, so the similarity of the responses under the two conditions suggests there may not in fact be much of a practical problem in estimating people's demands for a public good. In short, these results suggest that free riding may not be a big problem.

Several other experiments employed what I will loosely call the same general design, of presenting subjects with some public good whose value to them was unknown to the experimenter and of comparing the results of different methods of eliciting their willingness to pay. Sweeney (1973) considered the willingness of subjects to power an electric light by pedaling an exercise bicycle (free "riding" indeed) and found that this responded to whether they perceived themselves as being in a small or large group (a perception he manipulated by controlling the brightness of the light with a rheostat). The public good was if they would all receive credit for participating in the experiment, which depended on how brightly the light remained lit. Scherr and Babb (1975) compared voluntary contributions with those elicited in pricing schemes proposed for public goods by Clarke (1971) and by Loehman and Whinston (1972) and found no significant differences in the amount of public goods (in this case concert tickets and books donated to the library) provided under the three schemes. In general, the experiments using this design support the proposition that at least some public good can be supplied even by voluntary contribution. But it is much more difficult to interpret how much (if any) free riding is being observed, since the true value of the public good to each subject is unknown.

In order not to miss the opportunity to tell a colorful story, let me describe one more experiment of this general type, which was conducted by Schneider and Pommerehne (1981) at the University of Zurich and which would be unlikely, I think, to have been permitted at an American university. The subjects for their experiment were a group of economics students preparing for their comprehensive examinations. Without knowing that they were the object of an experiment, these students were approached by a confederate of the experimenters posing as the representative of a publishing company. She informed them that their professor (who, I surmise, would write the comprehensive exam) was writing a book
on the subject of the exam, which would not be available until after the exam. However, the publishing company was interested in getting feedback on the book, and for this purpose might be willing to make specimen copies available, before the exam. (The authors remark that the students "had a strong incentive to try to obtain the book beforehand" [694-695]) The students were then told they could submit written bids of how much they were willing to pay to get an advance copy, with copies going to the ten highest bidders from both this group and two other groups from which bids had supposedly already been solicited. After these bids were collected the two highest bidders were told that they were among the ten winners. The remaining students were then told that there was another way in which they could obtain the book before the exam: if together with the two other groups they could raise SFr4,200, they would each get a copy. Again, written bids for the now public good were collected, and the heart of the analysis is the comparison of the two bids. The authors note that the second bids were less than the first, but not by much. They conclude (702) that "there is only modest evidence for free riding as compared with the importance attributed to it in the literature."

A different kind of experimental design, in which the public good is an artificial one, makes it possible to employ an experimental strategy of trying to control each subject's value for the good, rather than trying to measure it. The idea is that if the experimenter assigns to each agent a payment depending on the quantity of the public good, then so long as the public good is not one that itself induces strong preferences among the agents, their preferences can be assumed to correspond to their monetary payoffs. In this way the payments of the agents for the public good and the amount of public good provided under a given decision mechanism can be compared not only with the amounts under another decision mechanism, but also with a priori notions about the optimal amount, such as the Lindahl equilibrium.

Smith (1979a, 1979b, 1980) reports three such experiments. In the first, he compared a version of a mechanism proposed by Groves and Ledyard (1977), designed to eliminate incentives to free ride by disentangling the price each agent pays from the price he states, with a procedure in which each agent pays his stated willingness to pay. Both procedures were implemented in an iterative manner that allowed agents to revise their statements in light of those of the others. Smith observed that, under some settings of the experimental parameters determining agents' demands, the Groves-Ledyard mechanism resulted in decisions at the Lindahl equilibrium, while the other mechanism exhibited substantial free riding, sometimes to the point that no public good was produced. A third iterative mechanism was then investigated, which Smith (1979a) called the auction mechanism and which incorporates the features suggested by Wicksell, in that the quantity of the public good and the amount to be contributed by each agent must be unanimously agreed to before the agreement is effective. (In the absence of agreement, no public good is produced.) The theoretical properties of this mechanism are somewhat unclear since it has many Nash equilibria. However under this mecha-
nism too, Smith reports that Lindahl prices and quantities were good predictors for the market parameters considered.

In the light of these results, Smith suggests that the results of Bohm's (1972) experiment might be reinterpreted, since the mechanism he considered to have the most probability of producing free riding (everyone pays their stated amount) resembled the auction mechanism in the sense that if too much free riding took place, no public good would be produced. That is, Smith suggests that the similarity of the bids in all of Bohm's procedures may merely reflect that the situation he considered (inadvertently) gave subjects good incentives not to free ride, because of the fear that no public good would be provided. In this spirit, Smith (1979b) reports an experiment designed to determine which aspects of the auction mechanism may have contributed to its apparent success. He compares the auction mechanism, in which agents propose both a contribution and a quantity of the public good, with a "free rider mechanism" in which each agent simply states his contribution and the quantity of the public good is whatever the summed contributions will buy. (A mechanism intermediate between the two was also considered.) All mechanisms were implemented with a unanimity rule that gave agents a chance to examine (and accept or reject) the outcome before it was implemented. Although the auction mechanism provided an amount of the public good nearer to the Lindahl quantity than the other mechanisms when it reached agreement, its frequency of agreement was sufficiently less than that of the other two mechanisms to make the overall quantity of public good similar under all mechanisms. Smith concludes by noting that under none of the mechanisms was a very strong free rider effect observed and conjectures that this may be due to the rule of unanimous approval. However, as Ledyard emphasizes in chapter 2 (see Table 2.18), subsequent experimenters have reached sharply different conclusions about the benign role of a unanimity rule.

Of course, different theoretical dispositions suggest different regularities in the data. For example, Smith (1980) reports in connection with another experiment that (597) "on average subjects contribute approximately one-half their endowments to the public good and retain one-half for private use." Marwell and Ames (1981), drawing primarily on a series of their own studies that also use a controlled, artificial public good, suggest that this may be an important kind of regularity. Noting that previous studies examined fairly small groups (mostly of fewer than ten individuals), they conducted a study in which both small and large groups could be examined. In a series of studies in which subjects were mostly high school students, subjects were told that they were part of a group and that each member of the group had an endowment of tokens to invest in either a public or private good. The public good had the higher return, but its proceeds were distributed equally to all group members. Over a number of conditions, Marwell and Ames report that on average the percentage of resources invested in the public good was surprisingly regular, in the range of 40 to 60 percent, with some indication of a decrease when the stakes were raised. Among the few exceptions they noted was that a group of first semester economics gradu-
ate students only invested 20 percent in the public good, leading them to suggest that economists may be different from everyone else (and hence the title of their paper).

The remaining experiments I will discuss differ from these previous ones in that they investigate how some of these mechanisms behave when they are used repeatedly, instead of just once. Thus each of these experiments, by Kim and Walker (1984), Isaac, McCue, and Plott (1985), and Banks, Plott, and Porter (1988) give subjects the chance to gain some experience with how the mechanisms work.

Isaac, McCue, and Plott (1985) seek to show that the free rider problem is alive and well, by examining a mechanism already suspected of being favorable to free riding and letting repetition have what effect it would. The mechanism chosen was that of direct contribution: each agent stated his contribution, and the amount of public good the summed contributions would buy was produced. (There was no requirement that the allocation be unanimously approved.) After all agents were informed of how much public good had been produced and had computed their payoff for that period, the process was repeated, with the same individuals and the same demand parameters. The results from a number of trials involving groups of ten subjects were that positive levels of the public good were produced in initial periods, but by around the fifth period these levels declined to near zero. The authors write, "Our results unambiguously demonstrate the existence of the under-provision of public goods and related 'free riding' phenomenon and thereby discredit the claims of those who assert as a general proposition that the phenomenon does not or cannot exist" (51-52). (They also note in reply to Marwell and Ames [who are sociologists] that their experiment included a group of undergraduate sociology students as well as groups of undergraduate economics students, and no differences were found.)

Kim and Walker (1984) report a similarly motivated experiment with similar results, using a much larger (simulated) group size. In their experiment subjects were instructed that they were part of a group of 100, and given a payoff table indicating how much each would be paid as a function of the total contributions made that day to a "common fund." (For example, if the fund received $100 [e.g., from $1 per person], each person would be paid $2.) Each day each subject phoned in his contribution, and had his earnings for the day delivered to him that evening. The results of the experiment, like that of Isaac, McCue, and Plott, were that positive initial contributions sharply diminished in succeeding days, so that substantial free riding was observed.

That results from repeated trials may differ from those in a single trial was confirmed by Banks, Plott, and Porter (1988), who examined both the direct contribution mechanism and Smith's auction mechanism, both with and without the rule of unanimous consent. Although they observed that the auction mechanisms outperformed the direct contribution mechanisms as producers of the public good, they found that the unanimity rule decreased efficiency in the repeated setting. They note that (319) "this result is directly counter to expectations formed from data and conjectures found in the literature." They also found that efficiency
decreased over time, suggesting that more free riding occurs with increased experience with these mechanisms. They conclude, "A more reliable process must be found before we proceed with an application at the practical/political level of analysis."

In summary, the experiments discussed here began with studies of one-shot decisions about various kinds of public goods, in which different decision mechanisms were compared. These experiments often reported little or no free riding. These were followed by experiments in which the public good was artificial and, therefore, more easily controllable. These experiments began to detect some degree of free riding and differences among mechanisms and environments. The most recent experiments introduced repetition and reported results at odds with the experiments preceding them. Since the theoretical properties of these mechanisms under repeated play are not well understood, it would be premature to confidently attribute these results merely to increased experience with the mechanisms. So the experimental results suggest a further theoretical agenda, as well as a continued experimental examination of other mechanisms. In the course of these experiments, the debate has thus shifted from whether or not free riding occurs, to how much and under what conditions it occurs, to what mechanisms and environments may be most vulnerable and most invulnerable to its effects. At this stage there still remains a considerable gap between the experimental results and related questions about the free rider problem in natural environments. (But, as we will see in section III.E in connection with the auction phenomenon of the "winner's curse," such gaps between experimental and field data need not remain unbridgeable.)

B. Coordination

A fundamental problem for players in a game with multiple equilibria (and for theories of equilibrium in such environments) is how to coordinate on a particular equilibrium. This kind of problem shows up clearly in macroeconomic theories of rational expectations, where players' expectations about future events need to become coordinated for equilibrium to be achieved. The theory of rational expectations has thus led to several experimental studies, which Jack Ochs discusses in chapter 3. However, the problems of coordination become apparent even in very much simpler environments.

As Schelling showed in his experiments in the 1950s, players are sometimes able to achieve coordination by focusing on aspects of their environment that are often left out of abstract models. (Recall the experiment in which players were identified by letters of the alphabet and were able to achieve a substantial rate of successful coordination by focusing on alphabetical order.) More recent work has shown that, not only is coordination on a particular equilibrium influenced by features of the environment that are ignored in economic models, but also that features of a game that some theories explicitly assume are irrelevant are not.

For example, a very carefully conducted set of experiments by Cooper,
DeJong, Forsythe, and Ross (1990) considered two-person games in which each player had three strategies. In any play of the game, the players each chose one of their strategies without knowing the choice of the other player. In the games they used to study coordination, there were two strict, pure strategy equilibria, which arose when the players coordinated either on their first pure strategy or the second. (That is, when either player employed one of her first two strategies, the other player maximized her own payoff by choosing the same strategy.) In each game, the third strategy of each player was a strictly dominated strategy—i.e., it gave a strictly lower payoff than one of the first two strategies, for every possible action of the other player. Cooper et al. found that, although the dominated strategy was seldom played, especially after the players had gained a little experience with the game, its presence had a profound effect on which of the two equilibria the players tended to reach after acquiring some experience with the game. Players apparently were influenced in their choice between their first and second strategies by what their payoffs would be in each case in the (unlikely) event that the other player chose his dominated third strategy.

The effect that Cooper et al. noticed was made even more dramatic because the equilibria in their games were Pareto ranked—i.e., both players did better at one of the equilibria than at the other. Nevertheless, sometimes the dominated strategy made it easier for them to coordinate on the less desirable equilibrium. An even clearer example of how coordination can fail to produce socially optimal results comes from an experiment reported by Van Huyck, Battalio, and Beil (1990), who considered a family of pure coordination games, or games in which there is no conflict of interest at all among the players.

Van Huyck et al. considered a game played by groups of around fifteen players, each of whom simultaneously had to choose an integer from 1 to 7. The payoff to any player \(i\) depended on both the integer \(e_i\) chosen by that player and the minimum \(m = \min\{e_j\}\) chosen by any of the players (including player \(i\) himself). In the version of their game which I will speak of here (which was their "condition A"), the payoff to player \(i\) was equal to $0.60 + [0.20(m) - 0.10(e_i)]$. (See the payoff table at the top of Figure 1.2.) Thus all the players had a common interest in a high value of \(m\), but there was a penalty for stating a number higher than the minimum chosen by the other players. In particular, this game has seven strict, pure strategy equilibria, at each of which every player chooses the same integer. However, the equilibrium at which all the players choose 7 (and all earn $1.30) gives a higher payoff to every player than any other equilibrium (or any other outcome of the game), while the equilibrium at which all the players choose 1 gives each player the lowest payoff ($0.70) of any equilibrium. (Of course, the lowest payoff that a player can possibly get is when he chooses 7 and at least one other player chooses 1, in which case he earns only $0.10 for that round.)

Van Huyck et al. observed that when the game was played repeatedly, with the outcome made public after each play of the game, the minimum quickly converged to 1 in each group they examined. For nine groups of subjects who played the game ten times each, a diffuse distribution of choices in the first period
INTRODUCTION

quickly evolved into a very high concentration of 1's in each group. (In no group was the minimum choice higher than 1 in any period after the third.) Thus, despite the common interest of all the players in the equilibrium outcome at which each player chooses a 7, this never occurred. To the contrary, behavior in this game quickly evolved in the direction of the least profitable equilibrium. Because it is clear that players learn from their experience in these games, it is natural to look to models of learning and adaptation to begin to explain the observed behavior.

In this spirit, a theoretical model to help explain these observations was put forward by Crawford (1991), who proposed to adapt game-theoretic models of the kind considered by evolutionary biologists. In evolutionary games (see Maynard Smith 1982), an outcome is considered to be an evolutionarily stable equilibrium if, in addition to being a strategic (i.e., a Nash) equilibrium, it is stable against "invasion" by strategies that are not present in the population at equilibrium, but may be introduced at low levels by mutation. In order to be stable against this kind of invasion, it must be the case that when the equilibrium population of strategies is perturbed by the introduction of a non-equilibrium strategy, the strategies present at the equilibrium must do better in this new population than the strategies introduced by mutation. (If not, the mutation strategies would thrive and increase their presence in the population.)

Crawford observed that all but the minimum equilibrium in Van Huyck et al.'s "minimum" game were in fact quite unstable to this kind of invasion. Consider an equilibrium at which all players choose the same integer greater than 1. If some player were now to choose a smaller number, that player would have the highest payoff in the population. To the extent that high payoff strategies are more likely to "reproduce" themselves in the next period, as in the evolutionary model, this induces a dynamic process that tends towards the lowest number chosen. And if new strategies occasionally appear by some sort of mutation, then the dynamic will lead to the lowest feasible number, as was observed in the experiment.

As Crawford is careful to note, while evolutionary dynamics provide a very suggestive model of the observed behavior in this game, we cannot suppose that the dynamics actually being observed in the experiment are evolutionary in nature. Rather than natural selection from a variable population, we are witnessing some kind of learning and adaptation. In this vein, a very simple model of learning is explored by Roth and Erev (1995) in connection with some of the bargaining games to be discussed in chapter 4, and the contrast between the games studied there and the coordination games considered here is instructive.

Roth and Erev considered a family of adaptive models designed to be consistent with two of the most robust properties observed in the large experimental psychology literature on both human and animal learning, namely that choices that have led to good outcomes in the past are more likely to be repeated in the future and that learning curves tend to be steep initially, and then flatter. The models they considered are all variations on the following basic model.
At time \( t = 1 \) (before any experience has been acquired) each player \( n \) has an initial propensity to play his \( k \)th pure strategy, given by some real number \( q_{nk}(1) \). If player \( n \) plays his \( k \)th pure strategy at time \( t \) and receives a payoff of \( x \), then the propensity to play strategy \( k \) is updated by setting

\[
q_{nk}(t+1) = q_{nk}(t) + x
\]

where player \( n \) just played strategy \( k \) and earned \( x \), while for all other pure

\[
q_{nj}(t+1) = q_{nj}(t).
\]

The probability \( p_{nk}(t) \) that player \( n \) plays his \( k \)th pure strategy at time \( t \) is

\[
p_{nk}(t) = \frac{q_{nk}(t)}{\sum q_{nj}(t)},
\]

where the sum is over all of player \( n \)’s pure strategies \( j \).

So pure strategies that have been played and have met with success tend over time to be played with greater frequency than those that have met with less success, and the learning curve will be steeper in early periods and flatter later (because \( \sum q_{nj}(t) \) is an increasing function of \( t \), so a payoff of \( x \) from playing pure strategy \( k \) at time \( t \) has a bigger effect on \( p_{nk}(t) \) when \( t \) is small than when \( t \) is large).

It turns out that this simple model captures quite well the dynamics observed in certain two-person bargaining games to be discussed in chapter 4. However the modification required for it to capture the results observed by Van Huyck et al. (1990) is illuminating. Instead of having players learn from their own experience only, we can consider the otherwise identical model in which equation (1) is replaced by one in which the strategy that is updated at each period is the strategy that was most successful in the previous period, regardless of which player chose it. That is, in the model with imitation (or common learning), equation (1) is replaced by

\[
q_{nk}(t+1) = q_{nk}(t) + x
\]

The first column of Figure 1.2, from Erev and Roth (in preparation), shows the data of Van Huyck et al. displayed so as to indicate the period-by-period evolution of the probabilities of each number being chosen. (So, we observe that the highest frequency choice in period 1 was a 7, but that the frequency of 1’s—i.e., the probability that a 1 will be chosen by a given player—rises steadily from one period to the next.) Column two shows these probabilities as simulated by the learning rule with equation (1’), and column three shows the simulations by the learning rule with equation (1). In each simulation the initial propensities (i.e., the period 1 propensities) are taken to be those observed by Van Huyck et al., with the learning rule determining the subsequent probabilities of play. (The graphs in the latter two columns each represent the average of ten simulations.)
Looking first at the last column, we see that the learning model in which each player learns only from his own experience fails to capture the rise in choices of "1" from periods 1 to 10. However, the model in which players learn from the common experience of the previous period does a much better job. Thus it appears plausible that some of the role played by natural selection in evolutionary models may be played by imitation in the learning dynamics we are witnessing in this experiment.70

In summary, the coordination problems facing players in games with multiple equilibria provide a clear lens through which to view the process of equilibration. Because it is clear that players learn from their experience in these games, it is
natural to look to models of learning and adaptation to begin to explain what is being observed. More generally, experiments with coordination games have shown that factors that many traditional economic models suggest should be irrelevant (like the presence of strictly dominated strategies) may play a decisive role in determining the outcome of a game, while factors that many traditional models suggest should be of great importance (like Pareto dominance) may fail to overcome other influences on which equilibria will be observed. We will see that some of these issues come to the fore in the study of bargaining games as well, which also have a multiplicity of equilibria.

C. Bargaining Behavior

Theories of bargaining that depend on purely ordinal descriptions of bargainers' preferences tend to predict large sets of outcomes, and for this reason many economists (at least since Edgeworth 1881) have argued that bargaining is fundamentally indeterminate. In the language of cooperative game theory, the difficulty is associated with the fact that the core corresponds to the entire set of individually rational Pareto optimal outcomes. Similarly, in strategic models the problem is that all of this typically large set of outcomes can be achieved as equilibria of the game. Theories of bargaining that seek to make stronger predictions have attempted to distinguish among this multiplicity of equilibria by making use of more detailed information about bargainers' preferences or strategic options.

Since this kind of information is hard to observe in uncontrolled environments, these theories have been notoriously difficult to test with field data. Although there have been some attempts to explain observed bargaining outcomes by inferring what the utility functions of the bargainers would have to have been in order to be consistent with the prediction of some particular theory (i.e., with the prediction that could have been made had these utility functions been observable), such exercises cannot serve to provide any sort of direct test of the theory itself. Similarly, the detailed procedural information required to specify a strategic model of bargaining is mostly unobservable in field environments. Consequently, for tests of such theories it is natural to look to the kind of controlled environment and relatively unlimited access to the bargainers that can be obtained in the laboratory.

Although there has been some convergence between the theoretical literature concerned with strategic and cooperative models of bargaining (see, e.g., Osborne and Rubinstein 1990), the bargaining environments for which their predictions can be most clearly derived are rather different. This section will therefore be concerned with experimental tests of cooperative models, and chapter 4 will take up a more recent experimental literature concerned with strategic models. One of the interesting things to note in comparing the series of experiments discussed here with those in chapter 4 is how much experimental designs are shaped by the hypotheses among which they are intended to distinguish. Another thing to note is that the experiments in chapter 4, which were largely conducted in the latter
part of the 1980s and early 1990s, are conducted by a much more diverse group of experimenters than those covered in this section, which were conducted in the late 1970s and early 1980s by a single group of researchers. In this respect, bargaining experiments reflect the change that has occurred in experimental economics generally in recent years.

One of the best known family of game-theoretic models of bargaining arises from the work of John Nash (1950). Because of the way he specified his assumptions, these models are referred to as "axiomatic," and many specific models other than the one originally proposed by Nash have entered the literature (see Roth 1979).

Nash considered the "pure bargaining problem," in which two bargainers must agree on one alternative from a set \( A \) of feasible alternatives over which they have different preferences. If they fail to reach agreement, some fixed disagreement alternative \( \delta \) results. Nash modeled such a problem by a pair \((S, d)\), where \( S \) is a subset of the plane, and \( d \) a point in \( S \). The set \( S \) represents the feasible expected utility payoffs to the bargainers—i.e., each point \( x = (x_1, x_2) \) in \( S \) corresponds to the expected utility payoffs to players 1 and 2, respectively, from some alternative \( \alpha \) in \( A \), and \( d = (d_1, d_2) \) corresponds to the utility payoffs to the players from the disagreement alternative \( \delta \). The theory of bargaining he proposed and the other theories that have followed in this tradition take as their data the set \((S, d)\) and thus represent the feasible outcomes (solely) in terms of the expected utility functions of the bargainers. So such theories predict that the outcome of bargaining will be determined by the preferences of the bargainers over the set of feasible alternatives, together with their willingness to tolerate risk.

Because of the difficulty of attempting to capture the information contained in bargainers' expected utility functions, there were sometimes claims in the experimental literature that the theory was essentially untestable. To get around the difficulty, the earliest experiments designed to test Nash's theory assumed, for the purpose of making predictions about the outcome, that the utility of each bargainer was equal to his monetary payoff. That is, they assumed that the preferences of all bargainers were identical and risk neutral.

Important aspects of the predictions of the theory obtained in this way were inconsistent with the experimental evidence. This disconfirming evidence, however, was almost uniformly discounted by economists, who felt that the results simply reflected the failure to measure the relevant parameters. Nash's theory, after all, is a theory that predicts that the preferences and risk aversion of the bargainers exercise a decisive influence on the outcome of bargaining (and, furthermore, that these are the only personal attributes that can influence the outcome when bargainers are adequately informed). If the predictions made by Nash's theory under the assumption that bargainers had identical risk neutral preferences were disconfirmed, this merely cast doubt on the assumption. The theory itself had yet to be tested.

It was, therefore, clear that, in order to provide a test of the theory that would withstand the scrutiny of theorists, an experiment would have to either measure or control for the expected utility of the bargainers.
A class of games that control for the bargainers' utilities was introduced in the experiment of Roth and Malouf (1979). In these binary lottery games, each agent $i$ can eventually win only one of two monetary prizes, a large prize $X_i$ or a small prize $\sigma_i$ (with $\lambda_i > \sigma_i$). The players bargain over the distribution of "lottery tickets" that determine the probability of receiving the large prize: for example, an agent $i$ who receives 40 percent of the lottery tickets has a 40 percent chance of receiving $\lambda_i$ and a 60 percent chance of receiving $\sigma_i$. Players who do not reach agreement in the allotted time each receive $\sigma_i$. Since the information about preferences conveyed by an expected utility function is meaningfully represented only up to the arbitrary choice of origin and scale (and since Nash's theory of bargaining is explicitly constructed to be independent of such choices), there is no loss of generality in normalizing each agent's utility so that $u_i(\lambda_i) = 1$ and $u_i(\sigma_i) = 0$. The utility of agent $i$ for any agreement is then precisely equal to his probability of receiving his large prize $\lambda_i$—i.e., equal to the percentage of lottery tickets he has received. Thus in a binary lottery game, the pair $(S, d)$ that determines the prediction of Nash's theory is precisely equal to the set of feasible divisions of the lottery tickets.

Note that no assumptions have been made here about the behavior of the experimental subjects themselves in binary lottery games. (That is, the subjects might not be utility maximizers [see section III.F and chapter 8], or they might have preferences over distributions of payoffs to both players, rather than over their own monetary payoffs [see chapter 4].) What binary lottery games do allow us to know is the utility of utility maximizers who are concerned with their own payoffs. Since this is the kind of data required by Nash's theory, experiments using binary lottery games allow us to use the theory to make precise predictions. It is this which was missing from earlier experiments and from efforts to analyze bargaining data by inferring ex post what the utility of the bargainers might have been.

Under the assumptions of the theory, the set of relevant outcomes—i.e., of expected utility payoffs to the players—of a binary lottery game is insensitive to the magnitudes of $\lambda_i$ and $\sigma_i$ for each agent $i$. Furthermore, the bargainers have what the game theory literature calls "complete" information whether or not they know the value of one another's prizes, since knowing a bargainer's probability of winning his prize is equivalent to knowing his utility. Thus a theory of bargaining under conditions of complete information, which depends only on the utility payoffs to the bargainers, predicts that the outcome of the game will depend neither on the size of the prizes, nor on whether the bargainers know the monetary value of one another's prizes.

The experiment of Roth and Malouf (1979) was designed in part to test this prediction and to determine whether or not changes in the size of the prizes, and if the bargainers knew one another's prizes, influenced the outcome. In this experiment (and in the other binary lottery experiments described in this section) the small prizes of both bargainers were always equal to $0.00$. In this experiment the large prizes of the two bargainers were equal in some games, while in others they were in a ratio of 1 to 3 ($1.25$ to $3.75$). All games were played by bargainers.
seated at separated computer terminals, who could send text messages to each other, but who were prevented from identifying themselves to one another or from otherwise determining who they were bargaining with. Each bargainer played games with different prizes against different opponents in one of two information conditions. In the "full information" condition, each bargainer knew both his own prize and his counterpart's, while bargainers in the "partial information" condition each knew only their own prize value. (In each of these games, under both information conditions the prediction of Nash's theory is that the bargainers would each receive 50% of the lottery tickets.)

The results were that, in the partial information condition and also in those games of the full information condition in which the two bargainers had equal prizes, observed agreements clustered very tightly around the "equal probability" agreement that gives each bargainer 50 percent of the lottery tickets. In the full information condition, in those games in which the bargainers' prizes were unequal, agreements fell between two "focal points": the equal probability agreement and the "equal expected value" agreement (75%,25%) that gives each bargainer the same expected value. The mean agreement in these games fell approximately half way between the equal probability and equal expected value agreements. That is, in these games the bargainer with the lower prize tended to receive a higher share of the lottery tickets. Thus, contrary to the prediction of the theory, the monetary values of the bargainers' prizes were clearly observed to influence the agreements reached when the bargainers knew each other's prizes.75

One difference between the two information conditions of the Roth and Malouf (1979) experiment, which might account for the different outcomes observed, has to do with the messages the players could formulate. The transcripts of the messages show that comparisons of the two bargainers' prizes played a considerable part in the negotiations in the full information condition, in which both players knew each others' prize values. The equal probability (50%,50%) proposal and the (75%,25%) equal expected value proposal occupied prominent places in these negotiations. Although notions of "fairness" were mentioned by both parties, there was clearly a strategic aspect to how these notions of fairness were employed, since the player advancing the (50%,50%) outcome as fair could be reliably counted on to be the player with the larger prize.76 One natural question is whether the different agreements reached in the two conditions might be entirely due to the different messages possible when prizes could be compared, or whether some sociological factors relating to commonly held notions of equity might be an essential ingredient in the effectiveness of the strategic appeals to "fairness."

The experiment of Roth, Malouf, and Murnighan (1981) was therefore designed to see whether arbitrary focal points could be created. It employed binary lottery games with prizes stated in terms of an intermediate commodity, "chips," having monetary value. Each player always knew the number and value of chips in his own prize, but a player's information about his opponent's prize was an experimental variable. The conditions of the previous experiment were essentially replicated with "low information" and "high information" conditions,
and in addition there was an "intermediate information" condition, in which each player knew the number of chips in his opponent's prize, but not their value.

The observed results were that the low and high information conditions replicated the partial and full information conditions of the previous experiment, but the outcomes observed in the intermediate information conditions did not differ significantly from those in the low information condition: the observed agreements tended to give both players equal probabilities, regardless of the size of their prize in chips. Thus, the ability to compare prizes in terms of the artificial commodity, chips, did not affect the outcomes in the same way as did equivalent information about money. This supports the hypothesis that there is a "social" aspect to the focal point phenomenon, that depends on something like the players' shared perceptions of the credibility of any bargaining position.

Given the robustness of the observed (but unpredicted) effect of information about the size of the prizes on the outcome of bargaining, a subsequent experiment (Roth and Murnighan 1982) was designed to separate the observed effect of information into components that could be attributed to the possession of specific information by specific individuals. Each game of that experiment was a binary lottery game in which one player had a $20 prize and the other a $5 prize. In all eight conditions of the experiment, each player knew at least his own prize. The experiment used a 4 (information) x 2 (common knowledge) factorial design. The information conditions were as follows: 1) neither knows his opponent's prize; 2) the $20 player knows both prizes, but the $5 player knows only his own prize; 3) the $5 player knows both prizes, but the $20 player knows only his own prize; and 4) both players know both prizes. The second factor made this information common knowledge for half the bargaining pairs, but not common knowledge for the other half. For example, when the $20 player is the only one who knows both prizes, then the (common) instructions to both players reveal that both players are reading the same instructions, and that, after the instructions are presented, one player will be informed of only his own prize and that the other will be informed of both prizes. In the not-common knowledge condition, the instructions simply state that each player will be informed of his own prize and may or may not be informed of the other prize.

The results of this experiment permitted three principal conclusions. First, the equal expected value agreement becomes a focal point if and only if the player with the smaller prize knows both prizes. When the $5 player knew that the other player's prize was $20, this was reflected not only in his messages and proposals, but also in the mean agreements (when agreement was reached) and in the shape of the distribution of agreements (see Figure 1.3). And the mean agreements reached when neither player knows both prizes and when both players know both prizes replicate the results of Roth and Malouf (1979), both in direction and magnitude.

Second, the frequency of disagreement depends on whether it is common knowledge what information the bargainers possess. The frequency of disagreement in the two not-common knowledge conditions in which the $5 player knew
Figure 1.3. Distribution of payoffs obtained by the player with the $20 prize when agreement was reached (frequency of agreements in terms of the percentage of lottery tickets obtained by the $20 player). Source: Roth and Murnighan 1982.
both prizes is significantly higher than in the other conditions. The highest fre-
quency of disagreement (33%) occurs when the $5 player knows both prizes, the
$20 player does not, but the $5 player doesn't know that the $20 player doesn't
know both prizes. (In this situation the $5 player cannot accurately assess whether
or not the $20 player's [honest] skepticism that his opponent's prize is only $5 is
just a bargaining ploy.)

Third, the regularities among these unpredicted effects of information make it
unlikely that they can be attributed primarily to mistaken or irrational behavior
on the parts of the bargainers. For example, in the four cells in which the bargain-
ers do not know what information their counterpart has, the tradeoffs between
the higher payoffs demanded by the $5 player when he knows both prizes and
the correspondingly increased frequency of disagreements is just what would
be expected at equilibrium, in that the increase in the number of disagreements
just offsets the increased share obtained by the $5 player when agreements
are reached.

A subsequent experiment (Roth and Schoumaker 1983) lends support to the
hypothesis that the effect of information about the cash value of prizes is attribut-
able to its effect on the expectations of the bargainers about what constitutes
a credible bargaining position. Such information may help bargainers (and
theorists) to select from among the multiplicity of equilibria that are found in
bargaining games.

The bargaining experiments discussed above all involved variables which
the theories in question predict will not influence the outcome of bargaining. They
revealed ways in which the theories systematically fail to be descriptive of ob-
served behavior. As such, the experimental results demonstrate serious short-
comings of the theories. However, in order to fully evaluate a theory, we also
need to test the predictions it makes about those variables it predicts are impor-
tant. For theories based on bargainers' expected utilities, risk posture is such a
variable.

The predictions of these theories concerning the risk posture of the bargainers
were developed in a way that lent itself to experimental test in Roth (1979),
Kihlstrom, Roth, and Schmeidler (1981), and Roth and Rothblum (1982).79 (One
indirect virtue of experimentation is that it can provide a discipline to theoretical
work and suggest directions in which theory ought to be explored.) A broad class
of apparently quite different models, including all the standard axiomatic mod-
els,80 yield a common prediction regarding risk aversion. Loosely speaking, they
all predict that risk aversion is disadvantageous in bargaining, except when the
bargaining concerns potential agreements that have a positive probability of
yielding an outcome worse than disagreement.

Three closely related experimental studies exploring the predicted effects of
risk aversion on the outcome of bargaining are reported in Murnighan, Roth, and
Schoumaker (1988). Whereas binary lottery games were employed in the earlier
experiments precisely in order to control out the individual variation due to differ-
ces in risk posture, these studies employed ternary lottery games having three
possible payoffs for each bargainer $i$. These are large and small prizes $\lambda_i$ and $\sigma_i$.
obtained by lottery when agreement is reached, and a disagreement prize \( \delta \) obtained when no agreement is reached in the allotted time. (In binary lottery games, \( \sigma_i = \delta_i \).)

The bargainers' risk postures were first measured by having them make a set of risky choices. (Note that, in contrast to the experiments just discussed, the strategy in this experiment was to measure preferences rather than to control them.) Statistically significant differences in risk aversion were found among the population of participants, even on the relatively modest range of prizes available in these studies (in which typical choices involved choosing between receiving $5 for certain or participating in a lottery with prizes of \( \lambda_i = $16 \) and \( \sigma_i = $4 \)).

Those bargainers with relatively high risk aversion bargained against those with relatively low risk aversion in pairs of games such that the disagreement prizes were larger than the small prizes in one game and smaller in the other. The prediction of game-theoretic models such as Nash's is that agreements reached in the first game should be more favorable to the more risk averse of the two bargainers than agreements reached in the second game.

Let me be precise. The theory actually makes a stronger prediction, but only the weaker form is confirmed by the experiments, and the reasons for this illuminate not only the design and analysis of these experiments, but of many experiments designed to test economic theories. When the prizes of both bargainers are all equal (i.e., \( \lambda_1 = \lambda_2 = \lambda \), \( \sigma_1 = \sigma_2 = \sigma \), and \( \delta_1 = \delta_2 = \delta \)) the theories in question predict that the more risk averse player will get more than 50 percent of the lottery tickets when \( \delta > \sigma \), and less than 50 percent of the lottery tickets when \( \delta < \sigma \). Thus the prediction is not only that the more risk averse player should do better in the first game than he does in the second, but that he should do better than the less risk averse player in the first game, and worse in the second.

Now, as had already been established by the earlier experiments, these axiomatic theories fail to predict the effects of the bargainers' information about one another's prizes. Among the earlier observations was the very high concentration of (50%,50%) agreements in games with equal prizes or in which bargainers know only their own prizes, and a shift in the direction of equal expected values in games with unequal prizes known to both bargainers. The strongest form of the predictions about risk aversion concern games in which the bargainers have equal prizes, and so the first experiment of Murnighan, Roth, and Schoumaker (1988) used such a symmetric game. However a test of the predictions requires data from pairs of agreements between the same subjects, and it was quickly observed that a high percentage of pairs reached (50%,50%) agreements in the game with \( \delta < \sigma \), and ended in disagreement in the game with \( \delta > \sigma \). Although there was a weak effect of risk aversion in the predicted direction, it was not significant. One way to read this, of course, is as a rejection of the prediction, but in view of the relatively small scale of the prizes it was thought that any effect of risk aversion might simply be overpowered by the "focal point" effect already observed in connection with the equal probability agreement. So it was decided to run a subsequent experiment in which the prizes were unequal, in order to give any effect of risk aversion a wider range on which to be observed.81 But, as
had already been noted, this meant that the player with the smaller prize could be expected to receive the higher percentage of lottery tickets, irrespective of the relative risk aversion of the two bargainers. Consequently only the weaker form of the risk aversion prediction could be tested on such a game, and it is this prediction that was ultimately confirmed by the data. That is, the results of these experiments support the predictions of the game-theoretic models that more risk averse bargainers do better when the disagreement prize is high than when it is low.82

But these results also suggest that, in the (relatively modest) range of payoffs studied here, the effects due to risk aversion may be much smaller than some of the effects due to changes in information observed in previous experiments. How much this has to do with the size of the payoffs remains to be determined.83

So one lesson that can be drawn from all this is that it is possible to design experiments to investigate the qualitative predictions of theories that may already be known not to be good point predictors. Because of the relative complexity of economic phenomena compared to the relative simplicity of economic theories intended to account for them, this is frequently the problem facing economic experimenters, one they have in common with econometricians studying field data.

The results also illustrate a frequent and perplexing problem in interpreting experimental work: how should one assess the "size" (or relative importance) of effects observed in the laboratory, particularly when these may be sensitive to the size of the payoffs to the subjects? Since it has so far proved far easier to observe the unpredicted effects of information than the predicted effects of risk aversion on the outcome of bargaining, can we conclude that the former are more important than the latter? I do not think that the available evidence fully justifies this conclusion. The problem is that there is reason to believe that risk aversion is a phenomenon many of whose consequences are easiest to observe when decisions involve very large gambles. In principle this presents no obstacle to experimental investigation (just conduct experiments with very large prizes). But in practice, experimental budgets always make it likely that payoffs in the laboratory will be smaller than those in many situations to which economic theories are naturally applied. Not being able to compare the significance of these unpredicted and predicted effects means that, on the evidence so far available, we cannot deliver a conclusive verdict on the overall health of every aspect of theories of bargaining such as Nash's.

In looking over this whole series of experiments, two other phenomena stand out. First, there was a non-negligible frequency of disagreements. Second, there was a clear "deadline effect." Across all experiments, which varied considerably in the terms and distribution of agreements, the data reveal that a high proportion of agreements were reached in the very final seconds before the deadline. As these observations are closely related to the results observed in the experiments that will be discussed in chapter 4, they will be discussed there.

In summary, the series of experiments discussed here shows several things. First, it disconfirms some important aspects of the received theory of bargaining,
chief among which is what constitutes a "complete" specification of the information available to bargainers. This is particularly notable in light of the fact that much of the theoretical criticism of the complete information assumption is founded on the assumption that bargaining is better modeled by assuming that the bargainers have strictly less than complete information. The experimental evidence suggests that, although bargainers may certainly be expected in general to have less of some kinds of information, the outcome of bargaining is also highly sensitive to information the bargainers may have about each other in addition to what is included in "complete" information about utility functions. These experiments also provide preliminary support for some of the subtle but robust predictions of these bargaining theories about the effect of risk aversion.

D. Market Organization and Competitive Equilibrium

Both chapter 5 by Holt and chapter 6 by Sunder review experiments that primarily involve markets created using the basic design of Chamberlin's 1948 experiment for motivating subjects' supply and demand behavior. That is, each potential buyer of an artificial commodity is told an amount that the experimenters will pay him if he purchases the commodity (so that a buyer's total profit will be this amount minus the price he paid for the purchase), and each seller is told an amount that will be subtracted from the sale price if he sells a unit of the commodity. The assumption that subjects treat these amounts as their reservation prices then permits the experiment to be analyzed in terms of the resulting aggregate supply and demand curves, from which predictions or benchmarks such as competitive equilibria, consumer surplus, etc. can be computed.

However, in most of the markets reviewed in those chapters, the pricemaking mechanism differs from the bilateral negotiations employed by Chamberlin. (And indeed one focus of Holt's chapter is on the effect of different pricemaking mechanisms.) One of the frequently explored pricemaking mechanisms is the "double auction," in which sellers may make offers and buyers may make bids. Sellers and buyers may make new offers and bids whenever they like; the lowest outstanding offer is the "market asked price," and the highest outstanding bid is the "market bid price." If any seller offers the market bid price or if any buyer bids the market asked price, a transaction is consummated between the seller and buyer whose prices coincide.

In the first part of this section I will concentrate on a series of experiments that study the behavior of double auction markets when the market is repeated several times with identical parameters. In the second part of the section I will discuss experiments that compare different forms of market organization. Both of these sets of experiments, and chapters 5 and 6, testify to the continuing importance of Chamberlin's 1948 experimental design. In the final part of this section I will describe some experiments underway at the University of Iowa, which use double auction markets without seeking to control supply and demand behavior, but which use the laboratory environment to allow such behavior to be carefully observed and measured.
1. Repeated Double Auctions with Stationary Parameters

Smith (1962) employed a double auction procedure in which all payoffs were hypothetical, and once all transactions had been made, subjects repeated the process, with the same reservation prices, several times. He describes the differences between his experiment and Chamberlin's as follows (Smith 1962, 114):

The design of my experiments differs from that of Chamberlin in several ways. In Chamberlin's experiment the buyers and sellers simply circulate and engage in bilateral higgling and bargaining until they make a contract or the trading period ends. As contracts are made the transaction price is recorded on the blackboard. . . . Each trader's attention is directed to the one person with whom he is bargaining, whereas in my experiments each trader's quotation is addressed to the entire trading group one quotation at a time. Also Chamberlin's experiment constitutes a pure exchange market operated for a single trading period. There is, therefore, less opportunity for traders to gain experience and to modify their subsequent behavior in the light of such experience. It is only through some learning mechanism of this kind that I can imagine the possibility of equilibrium being approached in any real market.

About this latter difference Smith further notes (115):

One important condition operating in our experimental markets is not likely to prevail in real markets. The experimental conditions of supply and demand are held constant over several successive trading periods in order to give any equilibrating mechanisms an opportunity to establish an equilibrium over time. Real markets are likely to be continually subjected to changing conditions of supply and demand.

Thus the focus of this experiment was on producing conditions under which competitive equilibrium might be observed, and in this respect it succeeded—the repeated double auction markets that Smith reports all show convergence of transaction prices towards the competitive price in the course of only a small number of repetitions of the market. In subsequent experiments Smith reproduced this phenomenon with real monetary payoffs, rather than the hypothetical payoffs employed in Smith (1962).85 These and other experiments confirmed a tendency for transactions to converge towards competitive equilibrium in repeated double auction markets. (The behavior of repeated double auction markets is thus one of the phenomena to have been first observed with hypothetical payments, and later confirmed with real payments, a point that I do not recall seeing mentioned in the debate about the role that experiments with hypothetical payments might continue to play in experimental economics.)

In the ensuing years, Smith and his students and colleagues replicated this result many times, further demonstrating convergence to competitive outcomes in repeated double auctions. However, as the attention of other experimenters was
drawn to the repeated double auction, experiments were designed both to explore conditions under which competitive outcomes would or would not be observed in double auctions, and to explain the convergence that was often observed. Two experiments that are especially worth mentioning are those reported by Holt, Langan, and Villamil (1986), and by Gode and Sunder (1993).

Holt, Langan, and Villamil (1986) considered whether or not convergence to competitive equilibrium in double auctions might be influenced by the parameters determining the supply and demand curves. They considered, for example, a double auction examined by Smith, which he reported (Smith 1982a, 172), "provides the most stringent of all reported tests of the equilibrating tendency in double auction trading." since the competitive equilibrium gave all of the exchange surplus to one side of the market. Holt et al. observed that, while in this dimension the test is indeed stringent, nevertheless the supply and demand is such that (112) "the lack of market power is so severe in this design that even if one buyer unilaterally withholds demand for all four of his units, he has no effect on market price."

They proposed to examine double auction markets that differed from those previously examined in that agents on one side of the market had market power, in the sense that, by foregoing some trades, they could move the competitive price sufficiently in their favor so that they would earn a larger profit on their remaining transactions than if they had made all the trades that would be profitable for them. Holt et al. note that this is a natural place to begin an investigation aimed at finding nonconvergence to competitive equilibrium, since the hypothesis that this kind of market power may lead to noncompetitive outcomes goes back as far as Cournot (1838). Holt (1989, S. 108) further notes that this definition of market power is embodied in the Department of Justice horizontal merger guidelines.

Holt et al. first replicated the results of the double auction discussed by Smith, confirming the tendency of the market to converge to competitive equilibrium (see, e.g., Figure 5.4 for the graph of a similar market). They then conducted seven multiperiod double auctions each involving five buyers and five sellers, using the procedures of previously published experiments, but with parameters that gave market power either to some of the buyers or to some of the sellers. Three of their double auctions used experienced subjects. Contrary to the convergence to competitive equilibrium uniformly observed in earlier experiments (Smith 1982b), four of the seven auctions observed here, including all of those with experienced subjects, failed to converge to the competitive price and converged instead to a price reflecting the distribution of market power (see, e.g., Figure 5.7).

The results of this experiment thus support the hypothesis that the parameters of the market may influence the convergence to competitive equilibrium previously observed in double auction markets, particularly when experienced subjects are involved. And subsequent experiments have also shown that the high efficiency observed in early double auction experiments may also depend on the
market parameters employed (see Banks, Ledyard, and Porter 1989; Van Boening and Wilcox 1993). However, as Holt notes in chapter 5, the significance and implications of such results are still a matter of lively controversy.

One conclusion that can clearly be drawn from the many repeated double auction experiments is that it is possible for competitive equilibrium outcomes to be observed with very few buyers and sellers. This simple fact provides a powerful challenge and stimulus to developing new theories of competitive behavior. One of the most interesting of these is developed by Gode and Sunder (1993), who compare experimentally observed (human) behavior with the simulated behavior of what they call "zero-intelligence traders."

Gode and Sunder report an experiment in which they used Chamberlin's procedure to construct four sets of supply and demand curves that past experience suggested would reliably lead to competitive equilibrium in a repeated double auction, and one that might not. (In this latter case, the reservation prices of some buyers and sellers would prevent them from trading at the equilibrium price, so they had strong incentives to try to transact trades at other prices.) Gode and Sunder compared the behavior of human subjects in six repetitions for each set of parameters with the behavior of a simulated market in which the traders were represented by simple computer programs.

Two kinds of programmed traders were considered, each of which generated random bid or asked prices (depending on whether it was a buyer or a seller). The first kind of programmed trader generated bid or asked prices from a uniform distribution on the integer prices from 1 to 200, a range that defined the feasible range of trading prices. The second kind of programmed trader only was able to generate random bid or asked prices that would not yield a negative profit if accepted; that is, these traders were constrained by their reservation prices (and thus, although still random, could be viewed as having a little intelligence). The first kind of programmed trader was called a ZI-U (zero intelligence, unconstrained) trader, while the second was called a ZI-C (constrained) trader.

Gode and Sunder simulated double auction markets consisting of six buyers and six sellers, all of the same type, i.e., all ZI-U or all ZI-C. The rules of the market were that a transaction occurred whenever any bid or asked crossed (i.e., if some bid price was at least as high as some asked price), with the transaction price equal to whichever of the two prices had been posted earlier. Each transaction was for a single unit.

As expected, the human subjects converged fairly smoothly to the competitive equilibrium in the markets with the first four sets of parameters, with prices near competitive prices, and efficiency (percentage of feasible buyer plus seller profits) near 100 percent after the first few periods. Also as conjectured, the fifth set of parameters was further from the competitive outcome, with a lower efficiency, as some of the "wrong" units were traded away from the equilibrium price.

In contrast, the simulated markets with the ZI-U traders exhibited no trend—transaction prices were distributed with high variance, and efficiency was dimin-
ished by the fact that the maximum feasible number of units (equal to the smaller of the total units sellers were allowed to sell or the buyers were allowed to buy) were always traded, regardless of whether these trades were profitable.

However, the results of the simulated markets with the ZI-C traders are in some important ways much closer to that of the human traders than to the ZI-U traders. First, the transaction prices are much closer to the equilibrium price than those of the ZI-U traders, although not as close to equilibrium as those of the human traders. Second, the efficiency achieved by the ZI traders, while it falls short of the 100% seen in many periods with the human traders, is very high (and in fact higher than the efficiency of the human traders under the fifth parameter set). Third, and to my mind most striking, there is a tendency for the prices of transactions reached later in a given market period to be nearer to the competitive price than are the prices of early transactions. That is, although the ZI-C traders cannot learn and although (in contrast to the human subjects) their transactions in any period therefore look exactly like the transactions in any other period, their transactions within a period exhibit a trend towards competitive prices. This is seen most clearly in their fourth parameter set, shown in Figure 1.4.

Gode and Sunder explain that this trend

... is caused solely by the progressive narrowing of the opportunity sets of ZI-C traders. The left end of the market demand function represents units with higher redemption values. Expected values of the bids generated for these units by ZI-C traders are also higher. Therefore, these units are likely to be traded earlier than units further down the market demand function. As the higher-value units are traded, the upper end of the support of ZI-C bids shifts down. Similarly, as the lower-cost units are sold earlier in a period, the lower end of the support of ZI-C offers moves up.

This means that the feasible range of transactions prices narrows as more units are traded. Though individual units may not be traded in the order in which they appear in the market demand and supply functions, there is a greater probability that the last transaction represents an exchange between the marginal buyer and the marginal seller. (129)

Thus Gode and Sunder have identified a feature of the double auction trading rules that, when combined with very minimal trader intelligence (just enough to know not to buy or sell at a loss), produces high levels of efficiency and exhibits a tendency towards equilibrium prices. And, although the simulated ZI-C traders cannot learn as human subjects do, these results suggest why human subjects are able to learn so quickly to make equilibrium-price transactions when the market is repeated with the same parameters. Because the repeated parameter design makes the supply and demand conditions identical from one period to the next (so that past consumption does not influence future demand, but instead the market "starts over" at each period) the information obtained in the latter part of one period is of very high relevance for the next period.88
Thus, over a long period of time, we see a series of experiments that begin with a demonstration that competitive equilibrium outcomes can be observed in markets even with relatively few buyers and sellers and that eventually move from demonstrations to investigations of the causes and conditions under which this occurs. As can be expected of such a productive series of experiments, the latter experiments raise as many issues as they settle and suggest new hypotheses and further experiments. For example, the results of both Holt et al. (1986) and Gode and Sunder (1993) suggest that further exploration of the sensitivity of double auction outcomes to supply and demand parameters is well justified. And the
results of Gode and Sunder help explain observed convergence results in double auction markets and suggest how some of the properties of these markets may interact with the experimental repetition of a stationary set of parameters. This suggests that the exploration of markets without stationary repetition of parameters may be called for, particularly when it is desirable to compare the performance of different forms of market organization.

2. Some Policy-Oriented Comparisons of Market Rules

The next set of experiments I will discuss have been motivated by questions of policy, of the kind raised by government regulatory agencies, typically about the effect of changes in the rules by which some market is organized. These investigations bring scientific methods to bear on one of the traditional nonscientific vocations of economists, which is whispering in the ears of princes who require advice about pressing practical questions about which little is known.

One of the studies I will speak about (Hong and Plott 1982) arose in a matter of concern to the Interstate Commerce Commission; the other (Grether and Plott 1984) in a case before the Federal Trade Commission. Both cases had to do with complex posted price markets, and in both cases an attempt was made to mirror as closely as possible in the laboratory the industrial structure of the market in question.

The ICC case concerned whether or not barge operators should be required by the ICC to post their prices and announce price changes in advance. The existing market allowed rates to be set by individual negotiations between barge operators and their customers, so that the terms of each contract were private information. Plott (1986) reports that the question arose because railroad companies were lobbying to require such price posting. The reasons offered by the railroads were that "public information on prices would make prices more competitive and protect small barge owners from large barge owners, who were allegedly making secret price concessions."

In their introductory comments, Hong and Plott (1982) say the following about their use of laboratory experimentation to illuminate the issues raised by the proposed change (1):

The full consequences of a rate filing policy are unknown. Plausible theoretical arguments can be made on both sides of the policy argument. When existing theory does not yield a definitive answer, one can usually turn to previous experiences with policies, but in this case we are aware of no industrial case study that would provide direct evidence on either side of the controversy.

They go on to note that it would be difficult to draw any compelling policy conclusions regarding the barge industry from previous laboratory experiments concerning posted price markets (in particular those of Williams 1973 and Plott and Smith 1978), since (2)
Any extrapolation from published experimental results to the barge industry itself is open to two potential criticisms, the reasonableness of which this study was designed to assess. First, the barge industry has several prominent economic features that are not incorporated in existing laboratory market studies. Examples include the relative sizes of buyers and sellers, the demand and supply elasticities, and the cyclical nature of demand. Naturally, we can never be certain that all the important features have been included in the present design. If something important has been misspecified or omitted, then the observed behavior of the laboratory market may not extend to the barge industry, and additional appropriately modified experiments can be conducted as checks on our conclusions. The second potential criticism is that the effects of price posting in laboratory studies have only been measured relative to the performance of oral auction markets. Since auction markets differ from the negotiated price markets of the industry, the relevance of the comparison can be questioned.

Hong and Plott proceeded to design their experiment around a laboratory market scaled to resemble, in the features mentioned above, the market for transporting grain along the upper Mississippi River and Illinois Waterway during the Fall of 1970. (This market was chosen because it was believed to be representative of a significant portion of the dry bulk barge traffic in the United States and because adequate data about the market parameters were available.) Aggregate supply and demand functions for the laboratory market were scaled to estimates available for the target market, as was the distribution of large and small firms on each side of the market. The laboratory market was divided into periods representing two weeks of the target market, and the seasonal aspects of the target market were modeled by having demand in the laboratory market scaled to resemble two months of normal demand followed by two months of high demand, followed by two months of normal demand. The experimental design involved running the market under both posted price and negotiated price policies.

In presenting the data from this experiment, Hong and Plott report that (10-1)

The results are easy to summarize. The posted price policy causes higher prices, reduced volume, and efficiency losses. Furthermore, the posted price policy works to the disadvantage of most market participants, especially the small ones, and helps only the large sellers.

They also conclude that the posted price markets react more slowly to the seasonal change in demand than do the negotiated price markets. Plott (1986) reports that this experimental evidence helped to deter the lobbying of the railroads on this matter, and that the price posting policy that they had advocated was not pursued.

We will consider the relationship of these experimental results to policy conclusions after briefly considering the experiment of Grether and Plott, which was motivated by an FTC complaint that also involving posted prices, among other things.
The FTC case involved a complaint by the FTC against the pricing practices of the Ethyl Corporation, E.I. du Pont de Nemours and Company, PPG Industries, Inc., and Nalco Chemical Corporation, the four domestic producers of tetraethyl and tetramethyl lead, the additives in "leaded" gasoline that raise its octane level. The FTC sought to have the producers cease and desist from a number of unusual pricing practices used for these additives, that, according to the FTC theory, had the effect of reducing price competition.

One of these pricing practices was that suppliers agreed to give at least a thirty-day notice of all proposed price increases, and usually such announcements were made with even more than the contractual thirty-day warning. Another practice was that "most favored nation" clauses were commonly included in contracts, by which a buyer was assured that he would receive the best terms being offered by the seller to any customer. (Apparently "meet or release" clauses were sometimes also used, which assured a buyer that the seller would meet the price offered by any other seller, or else release the buyer from any contractual obligation to buy from that seller.) In addition, all prices were quoted in terms of "delivered prices," for goods delivered to the purchaser regardless of his location.

Some of these practices might appear to favor the buyers, but the FTC theory was that, together, they worked to allow the producers to cooperate in raising prices. One way to explain this idea is as follows. If a producer thought that a price increase was desirable, he could announce, with somewhat more than the required thirty-days warning, his intention to raise his price. This wouldn't cause a customer with a "meet or release" clause to start searching for a supplier with a better price, because such a customer is assured that, in any event, he will only be charged the lowest price. (The lowest price is known, since prices are announced, and it is unambiguously defined, since only delivered prices are quoted, so there can be no hidden discounts in transportation costs.) If the other producers agree that this price increase is desirable, they can also announce it; otherwise, it will be rescinded by the initial producer. So a producer faces little cost in exploring the possibility of a price rise, while at the same time has little incentive to explore a price cut, since he will not be able to increase his market share (again, because of advance announcements, and "meet or release" clauses).

In its defense, the industry advanced the competing theory that price levels were determined entirely by the concentrated structure of the industry, and that, in such a concentrated industry, prices were unaffected by the pricing practices described.

Expert testimony by economists was available in support of both positions. The experiment of Grether and Plott was intended to be a possible source of evidence for rebuttal of the industry theory that the indicated pricing policies could not be affecting the price in such a concentrated industry. A scaled-down model of the industry was implemented in the laboratory, with careful attention paid to preserving the relative costs, capacities, and numbers of participants. The experimental design involved a number of multiperiod repeated markets, each of which would be examined both with and without (some or all of) the pricing practices in question. The results were fairly clear—when all the practices were in force, the
observed prices were above those that are observed when none of the practices were employed.89

Looking at these two experiments together, it is apparent that one of the differences between these experiments and those described in the previous sections has to do with the complexity of the economic environment being studied. There is some tension between the goal of designing an experimental market to resemble a particular naturally occurring market and the goal of designing an experiment whose results will be likely to support some fairly general conclusion.

On the other hand, it should also be apparent that these "policy-oriented" experiments have something in common with "theory-testing" experiments, since both involve the testing of hypotheses, whether those hypotheses arise from formal economic theories or from the arguments of lawyers, lobbyists, and expert witnesses. However, in contrast to hypotheses drawn from general economic theories, which are presumably applicable to any market in which the conditions of the theory are met, the hypotheses of interest in this case are explicitly concerned with the target market and not with the experimental market. Therefore, the bearing that the experimental evidence has on the hypotheses is different in these policy-oriented experiments than in the theory-testing experiments.

Hong and Plott aptly describe the role of experimental evidence in policy debates of this kind as serving to *shift the burden of proof*. Speaking of the experiments modeled on the barge industry, they put it this way (16-8):

> From a scientific point of view, we have solid evidence only that price posting markets do not necessarily operate better than negotiated price markets under the parametric conditions we considered. From a policy point of view, this evidence presumably shifts the burden of proof to the price posting advocates, who must now identify the specific features of the barge industry which, if incorporated in the experiment, would reverse the conclusions.

Plott (1986) closes by noting (737-8) that this research "demonstrates that laboratory experimental methods can be used in economics for basic, applied, and policy research. Such a demonstration presents a real challenge to the commonly held belief that economics is not a laboratory science as a matter of principle."

3. Information Aggregation: Markets as Forecasters

While I am on the subject of the range of uses to which economic experiments can be applied, I would be remiss not to at least mention a fascinating set of ongoing experiments at the University of Iowa, one piece of which is reported in Forsythe, Nelson, Neumann, and Wright (1992).90 What Forsythe and his colleagues have done is create and operate a variety of double auction markets in which, over a period of months, participants can buy and sell assets whose ultimate value depends on some future event, which occurs after the close of the market. The extent to which market prices at the close can be used to predict the future event can then be examined, the larger aim being to understand better the ways in which markets can aggregate information.
Forsythe et al. (1992) report the results of a market that opened for trading on June 1, 1988, in which the ultimate value of the assets traded would be determined by the proportion of votes received by each candidate in the election in November of that year for the Presidency of the United States. They describe the market as follows (1143-4):

Aspiring traders were sold portfolios of shares in candidates at $2.50 each, with each basic portfolio consisting of one share in each major candidate in the campaign. The slate of candidates included George Bush, Michael Dukakis, Jesse Jackson, and a candidate labeled "Rest-of-Field." Shares were given value by the dividends paid after the election, with the dividend on each share determined as the candidate's fraction of the popular vote times $2.50. Since Rest-of-Field covered all third-party candidates who earned votes in the election, the vote shares summed to 1 across the four candidates, and the total dividend paid on a basic portfolio of one share in each candidate just matched the fee charged for that portfolio. This investment/payoff rule was adopted ... because it provides a direct translation of market prices into estimates of vote shares,

$$\text{expected vote share} = \frac{\text{price}}{2.50}$$

and thus offers a prediction of not only the election winner, but also the margin of victory.

Traders in this electronic market also maintained cash accounts from which they could finance additional purchases and could log on at any time to post bids or offers for shares in a particular candidate. So the full record of the market allows conclusions to be drawn not only about closing prices, but also about the reaction of the market to particular events during the course of the election campaign, and not only about aggregate results, but also about the behavior of particular traders. The experimenters also conducted regular surveys of the traders and so were able to identify supporters of the different candidates.

Forsythe et al. (1992) conclude, on the basis of over 16,000 transactions observed over the course of the market, that although there is evidence that individual traders exhibited biases related to their political opinions, the market as a whole, as measured by the prices on November 7, was a very good predictor of the November 8 election results and was at least comparable to the major national opinion polls. (In particular, they report that the vote share forecast by the November 7 market prices were 53.2 percent for Bush, 45.2 percent for Dukakis, and 2 percent for any other candidate, while the actual vote shares were 53.2 percent, 45.4 percent, and 1.4 percent respectively, and major opinion polls in that week forecast the Bush vote in a range from 48 to 54 percent.) So although the average trader is not free of biases, the fact that the transaction prices are determined by the traders on the margin reduces the influence such biases play on the market price of each candidate's shares.

Similar political stock markets have since been conducted at Iowa and elsewhere, and, since January of 1993, two financial futures markets have been
initiated at Iowa (the Iowa Earnings Market and the Iowa Economic Indicators Market), open to traders with appropriate electronic access from around the world. \cite{alvin1995} The contracts initially offered on the Iowa Earnings market were based on the earnings per share (EPS) of five companies—three in the airline industry (American, Delta, and United) and two in the computer industry (Apple and IBM). Each company had a separate set of contracts for reported EPS in the first and second quarters of 1993. The Iowa Economic Indicators Market initially offered two types of contracts: inflation rate contracts whose liquidation values are determined by the monthly change in the Consumer Price Index, and currency contracts in the Mexican peso whose liquidation values are determined by the end-of-month exchange rate between the dollar and the peso.

The detailed scrutiny available in experimental markets in contrast to field markets (e.g., concerning the behavior of individual traders) promises to make these markets a unique resource for studying how the market aggregates information across traders. This is a topic that has attracted the attention of experimenters using more traditional laboratory based techniques, and Sunder surveys much of this work in chapter 6.

We turn next to consideration of another class of markets in which we can consider the relationship of experimental and field data and in which aggregation of different information held by different traders is a crucial issue.

\textit{E. Auction Markets and Disequilibrium Behavior}

1. The Winner's Curse

My topic in this section is in some ways the reverse of the discussion of free riding in public goods provision. Instead of discussing a theoretical prediction that seemed difficult to investigate with field data and initially proved difficult to detect experimentally, this section discusses an \textit{un}predicted effect that was initially postulated on the basis of field data, whose existence was debated, and which proved to be easy to observe in the laboratory. Of course, questions remain about how the experimental evidence applies to assessing the importance of this phenomenon in field data. Experience and motivation (of the experimental subjects in comparison with agents in the natural markets), the usual suspects, play a role here too. But in this case some ingenious comparisons between experimental and field data have been suggested, which I think have promise of furthering this part of the debate.

The story begins with a 1971 article by Capen, Clapp, and Campbell, three petroleum engineers employed by the Atlantic Richfield Company. They claimed that oil leases won by competitive bidding yield unexpectedly low rates of return, "year after year," and that this has to do with the fact that the winning bidder is typically the one with the highest estimate of the value of the recoverable oil and that the highest estimate is often an overestimate.

The important feature of this kind of auction is that all the bidders are trying to estimate a common value, in this case the value of the oil in a given tract. So even
if all bidders have unbiased estimates of the true value, one bidder’s estimate would convey valuable information to other bidders: the expected true value given a single estimate is higher than the expected true value given the information that the estimate is the highest of \( n \), where \( n \) is the number of bidders. The hypothesis behind the “winner’s curse” is that winning bidders must frequently have the highest estimate but fail to take this into account.

Now, the idea that bidders persistently make mistakes flies in the face of standard notions of equilibrium, and so this thesis was greeted with skepticism by many economic theorists, particularly as the details of equilibrium behavior in auctions became increasingly well understood (in which regard see particularly Wilson 1977 and Milgrom and Weber 1982). It seemed likely to many that a simpler explanation of why oil company engineers might urge others to lower their bids could be found in cartel theory rather than bidding theory.

Nevertheless, evidence from field data drawn from common value auctions of other kinds was increasingly cited in support of the thesis that this winner’s curse might frequently account for low or negative returns to the winners. But such field data as is available is sufficiently complex and incomplete so as to allow many interpretations. The profitability of an oil field, for example, cannot be known for years after the auction of drilling rights, and so the auction price is only one of many determinants of the rate of return. So the debate continued much as before.

Laboratory experiments provide an opportunity to investigate at least the basic questions associated with whether or not the winner’s curse is a robust phenomenon, and to what features in the auction environment it might respond. As we will see, they also reveal patterns in the data, associated with the presence or absence of a winner’s curse, that suggest directions in which the field data can be further investigated.

Bazerman and Samuelson (1983) reported an experiment designed to see not merely if the winner’s curse could be observed in the laboratory, but to explore how it might be related to the bidders’ uncertainty about the value of the object being auctioned. The basic idea of their experiment was the following: subjects were asked to estimate the number of coins in a jar that in fact contained 800 pennies. (To motivate the subjects to be accurate in this part of the task, a small prize was given for the closest estimate.) Subjects were then asked to bid for the jar, with the understanding that the highest bidder would pay the amount of his bid and receive in return the value of the coins in the jar. Subjects were also asked to write down their 90 percent confidence interval around their estimated value and to bid on other similar objects (e.g., a jar of nickels) also worth $8.

The main results were that a clear winner’s curse was observed in the data, with average winning bids around $10, which is $2 more than the value of the objects being auctioned. This is in contrast to the average estimated value, which at around $5 underestimates the number of coins in the jar. So auctions were mostly being won by bidders with high estimates, and these were overestimates often enough to make the average winning bid higher than the true value. Analysis of
various factors contributing to the level of bids suggested that, when the reported valuations were more uncertain, a winner's curse would start to appear among smaller numbers of bidders. (The amount that the highest value estimate must be discounted is greater when it is the highest of twenty than when it is the highest of four, so it is unsurprising that the winner's curse should be more readily observable among larger numbers of bidders.)

While the results show that the winner's curse is not hard to observe, the subjects in this experiment had no prior experience, and so the results could be attributed to the mistakes of novice bidders. Also, there was a wide range of bidding behavior, so the results could potentially be attributed to the mistakes of just a few bidders. (Bazerman and Samuelson report that the average winning bid is sensitive to [629] "a handful of grossly inflated bids.") One might suppose that in the natural economic environments in which questions about the winner's curse arose, bidders would have some opportunity to learn from their mistakes, and those who did not might be driven from the market by their losses. It is therefore still a reasonable question if the phenomenon observed in this experiment could occur in environments in which experience could be gained and in which bankruptcy could occur.

The experiment of Kagel and Levin (1986) was designed to address these issues and also to control (rather than simply to measure) the uncertainty surrounding the value of the object being auctioned. Their experiment involved auctions in which a value \( x_0 \) was chosen from a known uniform distribution, and each bidder was given a private information signal \( x_i \) drawn from a uniform distribution on \([x_0 - \epsilon, x_0 + \epsilon]\), for known \( \epsilon \) (which was one of the experimental variables, varying from $12 to $30).\(^9\) If the high bid is \( b \), the high bidder earns \( x_0 - b \) and everyone else earns 0. Subjects were given an initial cash endowment, and the opportunity to bid in a series of auctions. Subjects whose losses exhausted their initial endowments were declared bankrupt and were no longer allowed to bid. In addition, after each auction, the subjects were all given substantial feedback about the results: not only was the winning bid announced, but all bids were posted next to the signal that had been received by that bidder, and the true value \( x \) was announced. Thus bidders not only had an opportunity to learn from their own experience, but also from the experience of others. In particular, all bidders had an opportunity to observe the actual earnings of the high bidder. In addition, all subjects in this experiment had some prior experience in experimental auctions.\(^9\)

The main results for this part of the experiment are that bids were observed to be above the (risk neutral) Nash equilibrium bids. Profits were generally positive for groups of three or four bidders (at around 65 percent of the equilibrium profits) and negative for groups of six or seven bidders.\(^9\) Overall, the data are consistent with the conclusion that the winner's curse diminishes with experience, but that changes in the environment (particularly in the number of bidders) require some readjustment during which profits are lower than they are after some additional experience has been accumulated.
Although a winner's curse was clearly observed in this experiment, there is still room to question the relevance of the findings for the kinds of field data that motivated the initial questions. After all, the results do suggest that the phenomenon might eventually disappear as bidders become more experienced. One might suppose that professional bidders for, say, oil companies would have far more experience than can be obtained in a series of laboratory auctions. This may be so, but it should be noted that the argument can also be made the other way: in this experiment, bidders received immediate feedback on the true value of the object and the profit made by the winning bidder. The field data on, say, drilling rights in the Gulf of Mexico, come from bids most of which were made before good information on the value of oil fields ultimately became available. And in many cases, only the winning bidder knows this information in any detail, so, unlike in the experimental environment, it might be that the only bidders to have experience with the winner's curse are its victims. Under this point of view, the bidders in the experimental environment might be thought to have more relevant experience than do bidders in natural environments.

Another line of attack concerns the subject pool itself: maybe the students who were the subjects in this experiment have not been selected for the kind of judgement that successful bidders may possess. Dyer, Kagel, and Levin (1989) address this question in a subsequent experiment, in which the behavior of student subjects was compared with that of construction industry executives and found to be qualitatively similar.96

Of course, there are always going to be differences between laboratory and field environments that make judgments such as these partly matters of taste. However, the second part of the experiment reported by Kagel and Levin (1986) suggests a way to make a direct connection between the experimental and field data. That part of the experiment concerned the effect of introducing public information.

To understand what is at issue here, first note that the equilibrium prediction is that as public information about the value of the object being auctioned is increased, winning bids will rise. The reason is that, at equilibrium, agents must discount their private information about the value, in order to avoid the winner's curse. The more uncertainty there is about the value, the more they must discount their private information. So in a market at equilibrium, additional public information, which reduces uncertainty about the true value, will cause agents (particularly those whose private estimates are low) to discount their private information less, and this should in turn, on average, cause winning bids to rise (as even bidders with high estimates have to bid higher to account for the reduced discounting of other bidders). However the winner's curse occurs when winning bidders overestimate the true value. To the extent that increased public information reduces the uncertainty about the value, it should help bidders with high private signals to correct their overestimates. So, in a market in which the winner's curse is present, additional public information should on average cause winning bids to fall.
Kagel and Levin's experimental results are that in auctions with small numbers of bidders and positive profits, introducing public information (e.g., by announcing the lowest signal value publicly) does cause the winning bids to rise, but in auctions with large numbers of bidders and negative profits the public information causes winning bids to fall.

So, when the effect of public information can be observed, this suggests a test of field data for whether the winner's curse is present. In fact, some data about the effects of information are available for oil auctions, from the work of Mead, Moseidjord, and Sorensen (1983, 1984), who compare differential rates of return between wildcat and drainage tracts. A wildcat tract is one for which no positive drilling data are available, while a drainage tract is one in which hydrocarbons have been located on an adjacent tract. The neighbors of a drainage tract are the companies who lease the adjacent tract(s). They have some private information unavailable to other bidders. There is also a public component to this information. Kagel and Levin argue that (915)

If the information available on drainage leases were purely public, it should, according to Nash equilibrium bidding theory, raise average seller's revenues, hence reducing bidder's profits. . .. If the information were purely private, under Nash equilibrium bidding theory it would increase the rate of return for insiders (neighbors) relative to outsiders (nonneighbors) and reduce the average rate of return for nonneighbors. . .. If the added information on drainage leases contains both public and private information elements, rates of return for neighbors should be greater than for nonneighbors, but with nonneighbor returns definitely less than in the absence of the additional information (both the public and private information components push in this direction for nonneighbors)."

What Mead et al. found were higher rates of return on drainage compared to wildcat leases for both neighbors (88.6% higher) and nonneighbors (56.2% higher). Further, nonneighbors won 43.2 percent of all drainage leases. While the higher rate of return for neighbors compared with nonneighbors can be explained by the presence of insider information (the explanation Mead et al. offer, 1983, 1984), the substantially higher rates of return for nonneighbors remains puzzling within the context of Nash equilibrium bidding theory. However, the higher rate of return for both neighbors and non-neighbors on drainage leases is perfectly consistent with our experimental findings, given the existence of a winner's curse in bidding on wildcat leases. According to this explanation, the additional information available from neighbor tracts served to correct for the overly optimistic estimate of lease value recorded in the average winning bid on wildcat tracts, thereby raising average profits for both neighbors and nonneighbors alike. In this respect the OCS lease data parallel our experimental results with public information in the presence of a winner's curse.
Although this argument may go somewhat beyond the available mathematical theory and although Kagel and Levin are careful to note that there are alternative explanations for why both nonneighbors and neighbors do better on drainage leases, the experimental results establish a qualitative relationship among the data that are associated with the presence of the winner's curse, and this relationship opens new avenues for the investigation of field data. 97

2. Some Other Auction Results

Two papers that offer some interesting comparisons with those just discussed are by Kagel, Harstad, and Levin (1987), and Kagel, Levin, and Harstad (1988).

Kagel, Levin, and Harstad (1988) study second price common value auctions. In a second price auction, the high bidder wins but the price he pays is the second highest bid. Vickery (1961) noted that it is a dominant strategy in private value auctions conducted in this way to bid one's true willingness to pay (and second price auctions are sometimes called "Vickery auctions"). In a common value auction it is still necessary for bidders to discount their private sample in order to arrive at an unbiased estimate of the value (in order to take into account that the winning bidder is the one with the highest of \(n\) samples), but no essentially strategic considerations influence the optimum bid. Thus, in contrast to the first price auctions considered earlier (in which the price is equal to the highest bid and in which there is a strategic incentive to underbid), second price auctions disentangle the issue of evaluating how much an object is worth from strategic questions about how much to bid. Nevertheless, the authors report similar behavior as was observed by Kagel and Levin (1986): positive profits were earned in small groups of bidders, and negative profits in larger groups. Thus these results support the hypothesis that the winner's curse derives primarily from errors in judgement about the value of the object.

Kagel, Harstad, and Levin (1987) study a number of issues concerned with private value auctions, in which each agent knows with certainty the value to him of the object being auctioned, but has only probabilistic information about the value of the object to other agents. So in private value auctions there is no problem of evaluating how much the object is worth; the problem of choosing a bid is all strategic. Nevertheless, the authors observed that in the second price auctions, bidders had a persistent tendency to bid somewhat above their true values and that the bids did not exhibit any tendency to converge to the true values over time. (Recall that it is a dominant strategy to bid true values in such an auction.) Because the winning bidder does not pay his bid, but only the amount of the next highest bid, this tendency to overbid had only a small effect on the (positive) expected payoffs to the bidders.98 The authors conjecture that the overbidding is due to (1299) "the illusion that it improves the probability of winning with no real cost to the bidder. ..."

A striking feature of this result is that it is just the opposite of some previously reported results about second price auctions, which had concluded that bids
tended to be below true values. However upon inquiry Kagel et al. learned that (1286) "in these earlier private value auction experiments subjects were not permitted to bid in excess of their private values."

They go on to remark (1298):

This persistent excess of market price above the dominant strategy price stands in marked contrast to reports of second price sealed bid auctions with independent private values (Coppinger, Smith and Titus, 1980; Cox, Roberson and Smith, 192). Results from those experiments show average market price consistently below the dominant strategy price... . The key institutional feature responsible for these different outcomes is, we believe, that those earlier second-price auction experiments did not permit bidding in excess of private valuations."

Notice what this illustrates about the power of experimental methods. As economists, we have become accustomed to the fact that, because field data is noisy and incomplete, apparently similar data sets may yield different conclusions. With experimental data, however, since the collection of the data is fully under the control of the researchers, we can hope to be able to identify the causes of such differences. In this case, by inquiring of the authors, Kagel et al. were able to learn that the earlier experiments had a restriction that bids in excess of a bidder's private value were not allowed. Once this point had been clarified, the differences between the two data sets also became rather clear.

Kagel, Harstad, and Levin (1987) also consider the effect of public information on bids in first price private value auctions with affiliated values. Here, there is a closer correspondence between the equilibrium predictions and the observed outcomes than there was in the case of common value auctions discussed earlier. This adds some weight to the conclusion of Kagel and Levin (1986) that the contrary information effects they observed for common value auctions were due to the presence of the winner's curse.

a. Controlling Incentives: A Methodological Digression

One topic I should mention in passing concerns the recurring methodological theme that it is difficult to control subjects' preferences. Recall from section LA that, in this regard, Siegel and Fouraker (1960) investigated the effects not only of the total payoffs available to bargainers in their experiments, but also of the effect of increasing the difference in payoffs available at different contracts. A closely related question is raised in a critique by Harrison (1989), who reanalyzes the conclusions reached by Cox, Roberson, and Smith (1982), and Cox, Smith, and Walker (1983, 1988) in a series of experiments concerned with first price private value auctions. In those experiments, subjects' ordinal preferences were taken to be equivalent to their monetary payoffs, and the authors estimated the expected utilities of the bidders, under the assumption that each of the experimental data points that they observe represents a Nash equilibrium that is reached immediately and is constant over time, and that the utility functions are of a certain functional form. Their analysis of the bid data under this assumption led
them to reject the hypothesis that all the bidders have identical and risk neutral preferences. However, they observed that the bidding data conforms well to the equilibrium hypothesis once different, unobserved risk aversion parameters have been estimated for each bidder.

Harrison concludes that the control of the bidders’ ordinal preferences via their monetary payoffs in these experiments was insufficient to reach even this conclusion. His point, like that of Siegal and Fouraker (1960), is that it is not the total payoff to the bidders that is relevant, but the difference in payoffs that bidders get corresponding to different bids they might make. Harrison's key observation is that when one examines the expected payoffs to the bidders, the bids that appear to be significantly different from the risk neutral equilibrium bids differ only by pennies in expected payoff.100

The methodological thrust of the argument is that the whole point of paying subjects in experiments is to gain control of their incentives, that is, to create an environment in which their incentives are known. But if the observed bids frequently differ from the equilibrium bid by only pennies of expected income, other (uncontrolled) incentives that the bidders might have may be stronger than the effective monetary incentive.

I think that there is an even more general point involved here, which is that the difficulty of indirectly inferring an unobserved or uncontrolled variable may be as great in experimental data as in other kinds of data.101 This will be discussed at greater length in chapter 7. But this "marginal payoff critique" is of special importance in investigations of equilibrium phenomena, precisely because many equilibrium phenomena are predicted to happen on the margin, where agents may be more or less indifferent between a number of choices, regardless of the total payoffs available.102

**F. Individual Choice Behavior**

Almost simultaneously with the rise of expected utility theory to pride of place among economists’ models of individual choice behavior, early experiments began to establish that there are at least some situations in which a substantial percentage of experimental subjects can be observed to exhibit systematic patterns of choice that violate predictions of the theory. The best known of these is due to Allais (1953), who, as discussed in section LA, observed that certain kinds of risky choices could not be squared with utility theory. Around the same time, May (1954) observed that intransitive choices could be systematically elicited over multidimensional alternatives that did not even involve risk.

These observations did not materially impede the adoption of utility maximization as the primary vehicle for modeling individuals in economic theory. To the extent that utility theory is in part viewed as a prescriptive theory of rational choice, this is unsurprising, since it is unclear how experimental evidence of this kind can, or should, be incorporated into a theory of "ideally rational" behavior. But even when utility theory is viewed as an approximately descriptive theory of actual choice behavior, this is not too surprising, since the nature of the regular
violations of the theory were still unclear, no powerful alternative theories had
been proposed, and there was ample room to question the importance for eco-
nomic applications of the reported violations.

In the intervening years, the nature of these and many other reliable
"anomalies" in choice behavior have started to be much more thoroughly ex-
plored in experiments by both psychologists and economists, and in the last few
years these have prompted the proposal of several interesting alternative theories
of choice. There still remains ample room to question the importance of these
anomalies for economics, but of necessity these questions must now be more
pointed and specific and, hence, seem more likely to be answerable.

1. Preference Reversals

The anomaly I will consider in detail here is the discovery that it is possible to
construct pairs of lotteries with the property that many people, when asked at
what price they would be willing to sell (or buy) the lotteries, put a higher price
on one, but when asked to choose which they would prefer to participate in,
choose the other.

Investigation of this phenomenon, called "preference reversal," had its roots in
a paper by Slovic and Lichtenstein (1968) that considered how different ways of
assessing lotteries were differently influenced by the lotteries’ prizes and prob-
abilities. On the set of (hypothetical) lotteries they examined, how much subjects
were willing to pay to play a given lottery was correlated more highly with the
amount of the potential loss than with any other dimension, while the stated "at-
tractiveness" of the lottery correlated most highly with the probability of winning.
They argued that this difference was evidence that subjects considered different
kinds of information when asked to choose between lotteries than when asked to
price them. They conjectured that being asked to bid (an amount of money) for
the right to participate in a lottery caused subjects to concentrate on the monetary
values of the prizes, in a way that choosing between lotteries did not.

This motivated a subsequent study (Lichtenstein and Slovic 1971), in which
preference reversals were first reported. In that paper they wrote (47):

The notion that the information describing a gamble is processed differently
for bids than for choices suggested that it might be possible to construct a
pair of gambles such that S[subject]s would choose one of them but bid more
for the other. For example, consider the pair consisting of Bet P (.99 to win
$4 and .01 to lose $1) and Bet $ (.33 to win $16 and .67 to lose $2). Bet P
has a much better probability of winning but Bet $ offers more to win. If
choices tend to be determined by probabilities, while bids are most influ-
enced by payoffs, one might expect that S[subject]s would choose Bet P over
Bet $, but bid more for Bet $.

To test this conjecture, three experiments were performed. In the first, subjects
were presented with matched pairs of P and $ bets with positive expected values,
and asked to pick the bet they would prefer. Later, subjects were presented with
the bets singly, and asked to name the minimum price for which they would be willing to sell each bet rather than play it. Subjects were told that all lotteries were hypothetical and would not actually be played or sold.

The results were that, although subjects preferred the P bets to the $ bets only about half the time, they put a higher price on the $ bet far more often. In fact, 73 percent of the subjects were observed to always make the predicted reversal (p48): "For every pair in which the P bet was chosen, the $ bet later received a higher bid." In contrast the unpredicted reversal (choosing the $ bet but putting a higher price on the P bet) was much less frequent, and only 17 percent of the subjects ever made this kind of reversal.

The second experiment was much like the first except that, instead of being asked at what price they would be willing to sell each bet, subjects were asked at what price they would be willing to buy it. The prices thus elicited were lower than the corresponding selling prices in the first experiment, and this decrease in price was substantially more pronounced for the $ bets than for the P bets. This decreased the number of predicted reversals and increased the number of unpredicted reversals.

In the third experiment, which was intended (51) "to maximize motivation and minimize indifference and carelessness," transactions were actually carried out. All outcomes were stated in "points," which would be converted into cash at the end of the experiment. (However, subjects were not informed of the rates at which they would be paid.) The data again yielded a high proportion of predicted reversals and a low proportion of unpredicted reversals. One feature of this third experiment worth mentioning is that care was taken to motivate the subjects to reveal their "true" selling prices for each lottery. The technique employed was proposed for this purpose by Becker, DeGroot, and Marschak (1964). (Recall the description of the BDM technique in section I.B.)

Based on these three experiments, the authors concluded that the preference reversal effect is robust, that it is inconsistent with not only utility theory but with "every existing theory of decision making," and that it gives strong support to the view that subjects process information differently in making choices and in stating prices. They favor the view that subjects employ what has come to be called an "anchoring and adjustment" heuristic in stating prices, in which they first "anchor" on the amount of money to be won and then "adjust" their price to reflect that a win is not certain. In this view, preference reversals arise because subjects fail to adjust sufficiently. (For an account of other decision heuristics considered in the psychology literature, see Kahneman, Slovic, and Tversky 1982.)

A similar experiment by Lindman (1971) found qualitatively similar results over a set of hypothetical lotteries that included some with negative expected values. Shortly thereafter, Lichtenstein and Slovic (1973) sought to replicate the basic results using potentially significant amounts of money and a different subject pool. (In the previous studies, subjects had been college students.) In the new experiment, subjects were volunteer participants in a Las Vegas casino. Lichtenstein and Slovic describe the environment as follows (17):
The game was located in the balcony of the Four Queens Casino.... The game was operated by a professional dealer. . . . The S[ubject]s were volunteers who understood that the game was part of a research project. Only 1 S[ubject] could play the game at a time. Anyone could play the game, and the player could stop playing at any time (the dealer politely discouraged those who wanted to play for just a few minutes; a single complete game took 1-4 hr) . . . At the start of the game, S was asked to choose the value of his chips. Each chip could represent 50, 100, 250, $1, or $5, and the value chosen remained unchanged throughout the game. The player was asked to buy 250 chips; if, during the game, more chips were needed, the dealer sold him more. At the end of the game (or whenever the player quit), the player's chips were exchanged for money.

In the choice part of the experiment, each subject was faced with four bets at a time, all with the same absolute expected value, two positive and two negative. Subjects were instructed to choose one of the positive and one of the negative expected value bets, and these were played with the aid of a roulette wheel. In the pricing part of the experiment, subjects were presented with the lotteries, one at a time, and told to state a price such that either "I will pay the dealer — chips to get rid of this bet" or "The dealer must pay me — chips to buy this bet." The Becker, DeGroot, and Marschak (1964) procedure was used to determine transaction prices, with the dealer's offer being determined by the roulette wheel, so it was a dominant strategy for utility maximizers to state their true reservation price. Again, predicted reversals were frequent and unpredicted reversals rare. The authors conclude that (20)

The widespread belief that decision makers can behave optimally when it is worthwhile for them to do so gains no support from this study. The source of the observed information-processing bias appears to be cognitive, not motivational.

These results, which all appeared in the psychology literature, were viewed with suspicion by many economists. This is well expressed in the report of a subsequent experiment by Grether and Plott (1979), who were concerned that the earlier experiments (and also Slovic 1975) either did not use real payoffs or did not control for income effects. (That is, in the course of choosing between real lotteries the subjects become richer, which might change their preferences sufficiently to produce the reported reversals.) They also expressed concerns related to the fact that most of the experimental subjects were psychology undergraduates (629) "one would be hesitant to generalize from such very special populations") and that the experimenter were psychologists (629) ("Subjects nearly always speculate about the purposes of experiments and psychologists have the reputation for deceiving subjects"). They therefore proposed experiments to address these questions, designed, in their words (623), "to discredit the psychologists' works as applied to economics."
They employed the same gambles as in the third experiment of Lichtenstein and Slovic (1971), using subjects recruited from economics and political science classes. In the first experiment, subjects were divided into two groups. Subjects in the first group were paid a flat rate of $7 for participating and made only hypothetical choices, while subjects in the second group were told they had a credit of $7, with their final payment being the sum of the initial $7 and any gains or losses they might get from the lotteries. They were told that, at the end of the experiment, one of their decisions would be chosen at random to be actually played. (The authors remark that this procedure, rather than one in which all lotteries are played, should reduce any income effect.) Finally, the design of the experiment counterbalanced the two tasks so that subjects first chose between lottery pairs, then priced lotteries, then chose between the remaining lottery pairs. Prices were elicited as selling prices using the Becker, DeGroot, and Marschak (1964) procedure. (In a second experiment, all mention of "selling" was suppressed, in case this should be a reason why subjects might overstate their reservation prices.)

The chief result was that preference reversals persisted. There were observable differences between the data from hypothetical and from real lotteries, with a higher percentage of reversals arising from the real lotteries. The propensity to reverse was the same for lottery choices made before the pricing task as for those made after it. As before, prices for $ bets were generally higher than those for P bets, and higher than their expected values, so the data remains consistent with the hypothesis that pricing decisions are reached by "anchoring" and (insufficient) "adjustment."

These results did not settle the matter. Two subsequent studies, by Pommerehne, Schneider, and Zweifel (1982) and Reilly (1982), were motivated by concern that the experiment of Grether and Plott had not been effective in giving the subjects substantial motivation, because the amounts involved were not large. Pommerehne et al. conducted an experiment with higher payoffs and reported a frequency of reversals that is still substantial, but lower than that observed by Grether and Plott. Reilly's experiment provides a within experiment comparison that supports the conclusion that increased payoffs do reduce the rate of reversals. But in his experiment also, substantial percentages of reversals were observed. Thus this series of experiments supports the notion that preference reversals are not simply an artifact of certain narrow experimental procedures.

Reilly's results suggest that the rate of reversals does decrease as financial motivation increases (at least for some range of payoffs, since Grether and Plott report the reverse effect in moving from hypothetical lotteries to small payoffs), so it is reasonable to ask whether the rate of reversals might decline to insignificance if the subjects were sufficiently well motivated. This kind of question remains after many experimental studies. However, the following experiment of Berg, Dickhaut, and O'Brien (1985) shows that such questions can sometimes be addressed by means other than simple extrapolation.

Briefly, their experiment was designed to assess the effect of making subjects
pay for every preference reversal they stated, by running them around a "money pump." Using a pricing task in which subjects were required to state, for each lottery, a single price at which they would be willing either to buy or sell it, they extracted a fine from subjects who stated preference reversals in a first set of choices by first selling them the high price lottery (the $ bet) for the indicated price, then trading it for the low price lottery (the preferred P bet) and then buying back that lottery at its (lower) price. (Note that at this point these transactions were not voluntary: subjects had been told that they would be obliged to honor their stated preferences and prices, for either buying or selling.) Comparing those conditions of their experiment that do not extract this fine with those that do, they found no significant differences in the number of reversals, but a significant decrease in the dollar value of the reversals (i.e., the difference in prices between the two bets). As subjects gained more experience, the dollar value of the reversals declined, but reversals did not disappear. Thus the evidence suggests that subjects tried to eliminate reversals but were unable fully to do so.

This lends some indirect support, I think, to the view among psychologists that preference reversal may reflect a "cognitive illusion" in the pricing task, similar in some ways to familiar optical illusions. By analogy, consider an experiment where the paired comparison task is to estimate the length of two horizontal lines each of which is "framed" with sideways Vs facing either out or in to make them look longer or shorter, respectively. The "pricing" part of the experiment is to look at lots of horizontal lines, framed one way or the other, in random order, and estimate their length in inches. Even after you know that outward Vs make the lines look longer, it might remain hard to estimate them in inches, and increasing your motivation would not be expected to solve the problem.

\textit{a. Alternative Theoretical Directions}

But this is not the only way to view the evidence, and here the different theoretical points of view of psychology and economics suggest different directions in which it might be fruitful to proceed. This is a subject I will only introduce here, as it is elaborated on in chapter 8.

Loosely speaking, much of the work by psychologists on this and related subjects has been motivated by the point of view that people make choices in a manner analogous to interrogating a data base, and that how questions are asked therefore makes a difference in what answers will be obtained. In contrast, economists (who are generally interested in choice behavior at a somewhat different level of detail, and are therefore typically more willing to sacrifice some accuracy for some generality) have viewed choice behavior as reflecting underlying, already existing, and reasonably stable preferences. The assumptions about such preferences embodied in standard expected utility theory are of course not the only ones imaginable, and one way of seeking to capture the kinds of behavior discussed here is by relaxing such assumptions, while preserving the idea that at some useful level of approximation agents do indeed have preferences. A number of such theories have now been proposed (for a good introductory survey, see Machina 1987).
Loomes and Sugden (1983) discuss how preference reversals are consistent with a theory of choice, called "regret theory," which they earlier proposed in a 1982 paper, in which choices between risky alternatives reflect not only some underlying "choiceless" utility, but also comparisons ("regret" or "rejoicing") with what might have been. These comparisons depend on the subsequent realization of the underlying random events. Different comparisons are involved in choosing between two bets than are involved in choosing between each of them and a selling price, and the previous experiments allowed subjects to make some of these comparisons. That is, a subject might choose the P bet over the $ bet in part because of the regret he would feel if he chose the $ bet and the random device (e.g., roulette wheel) subsequently produced a number that meant a loss in the $ bet but would have meant a win in the P bet. But the same subject, with appropriately specified regret function, might still set a higher price on the $ bet, because the pricing task involves different comparisons between the random outcome of each lottery and the selling price.

Another hypothesis to account for preference reversals has been suggested independently by Holt (1986) and Kami and Safra (1987). This is that individuals may possess preferences that violate the "independence" assumption of expected utility theory, but not necessarily the transitivity assumption. (Following Machina [1982] a number of choice theories without independence have been proposed.) Independence is the assumption that says an outcome A is weakly preferred to B if and only if a lottery between A and C is weakly preferred to a lottery with the same probabilities between B and C, for any outcome C. It is this assumption that makes the utility of a lottery a linear function of the probabilities, so that compound lotteries may be decomposed in the standard way. And it is this that implies that the price a utility maximizer will state in the Becker, DeGroot, and Marschak (1964) elicitation procedure can be interpreted as his reservation price: that is, it implies he is indifferent between a lottery A and a selling price $p if and only if $p is the price that maximizes the utility of the compound lottery between A and prices greater than $p that he faces after stating a price.106 To emphasize that preference reversals may be compatible with transitive preferences over lotteries, Kami and Safra (1987) couch their discussion in terms of such a family of generalizations of utility theory proposed by Quiggin (1982) and Yaari (1987).

Holt (1986) further notes that the procedure of paying subjects for only one of their decisions, randomly chosen after all decisions are made, which was employed by Grether and Plott (1979) to control for income effects, only can be interpreted as having that effect if the independence assumption is satisfied. That is, the assumption is that the optimal choice in each decision evaluated separately is also the optimal choice when each decision is evaluated as part of the compound lottery consisting of the whole experiment; but without independence this may not be the case. And Kami and Safra note that the direct elicitation of preferences may present difficulties if preferences do not satisfy independence. So there is ample room for further experiments exploring these hypotheses.107

It should be emphasized that the Becker, DeGroot, and Marschak (1964) elici-
tation procedure allows us to predict what prices utility maximizers would state. That non-utility-maximizers may have incentives to respond differently is in no way a criticism of the experimental designs that incorporate this procedure. On the contrary, the virtue of those experimental designs is that they allow us to test predictions made in terms of utility theory, by permitting unambiguous predictions about what utility maximizers would do. In the absence of such a design, we would be unable to conclude that the observed phenomenon constituted a violation of the theory. This is a point worth repeating: one of the major virtues of laboratory experiments well designed to test theoretical predictions is that they allow us to make observations in theoretically unambiguous circumstances. To appreciate the power of this, consider the difficulty presented by any attempt to obtain unambiguous field observations concerning preference reversals. (In this connection, see the interesting experiment of Bohm [1994], which reports a lack of preference reversal observations in an experiment in which subjects purchased a used car. See also Bohm and Lind, forthcoming.)

b. Market Behavior

The very difficulty of making such observations in the field raises again the question of what is the importance of such phenomena for economics. I think it is fair to say that quite a broad range of opinions have been expressed on this point, with some economists taking the view that choice anomalies have not yet been shown to occur in typical economic environments such as markets.

To give a brief account of how that discussion has begun to be pursued by experimental means, it will be helpful to consider not only preference reversals, but the related phenomenon that stated buying prices have been observed to be substantially below stated selling prices (more than can be accounted for by income effects) in a number of studies of hypothetical choice (recall experiment 2 of Lichtenstein and Slovic 1971). Knetsch and Sinden (1984) review these results from hypothetical choice experiments and report an experiment showing this disparity between buying and selling prices persists for real transactions.

In a reply to Knetsch and Sinden, Coursey, Hovis, and Schulze (1987) propose to test what I will call the market hypothesis, which is that agents in a market environment will behave like utility maximizers: that is, experimental subjects in a market will receive feedback and experience of a kind that will extinguish such anomalies as the buying and selling price disparity. The market environment in their experiment is a second price auction, so that it will be a dominant strategy for utility maximizers to state their true reservation prices. (Buying and selling auctions were conducted separately: what is being bought and sold is the right not to taste, a "bitter... non-toxic... very unpleasant" substance called SOA.) In addition, the auction result would only be considered final if it was unanimously agreed to; otherwise another trial would be conducted to determine which (four out of eight) subjects would taste the SOA. The authors report that, although initial trials yielded the familiar disparity between buying and selling prices, and although the auction results continued to show some continued disparity, this diminished over auction periods, and by the final period the remaining gap be-
between the two prices was no longer statistically significant. (Most of the move-
ment came in subjects' declining prices for agreeing to taste the SOA.) They
conclude that "the divergence obtained in early trials of the experiment. . . may
result mainly from lack of a market experience."110

In their rejoinder, Knetsch and Sinden (1987) decline to attribute the same
significance to the diminution of the buying and selling price disparity in the
above experiment. Apart from critiquing aspects of the experiment (they are not
persuaded that tastes of SOA are a typical economic commodity), they also cite
some stylized facts about market behavior that they think may reflect choice
anomalies similar to this price disparity.

Kahneman, Knetsch, and Thaler (1990) present some evidence in support of
this view, from a market experiment in which half the subjects have been en-
dowed with a small consumer good (e.g. mugs, pens). In these experiments
subjects who wished to buy or sell a good (with their own money) were free
to do so, and the authors report that substantially fewer trades were transacted
than would be expected in the absence of an endowment effect, i.e. without an
unexplained tendency of subjects to prefer what they already have.111 This exper-
iment thus lends support to the view that observed anomalies in choice behavior
do not vanish simply because the environment in which they might arise is a
market.

2. Other Choice Phenomena

I have concentrated on preference reversals here because they have been the sub-
ject of a long series of experimental investigations, from different points of view,
which serve well to illustrate some ways in which experimental investigations
may proceed. There are other individual choice "anomalies" with equal (and
equally contested) claims to importance. For a discussion of some of these, see
Thaler (1987), who particularly concentrates on what in his view is the impor-
tance of these phenomena for economics. As Camerer discusses in chapter 8,
these other choice phenomena offer other possibilities to design experiments by
which various generalizations of expected utility theory may be tested. Battalio,
Kagel, and Jiranyakul (1990) succinctly summaize the state of theoretical affairs
by stating (46): "None of the alternatives to expected utility theory considered
here consistently organize the data, so we have a long way to go before having a
complete descriptive model of choice under uncertainty." In fact, even laboratory
animals have been observed to exhibit some choice anomalies of this kind (see
Battalio, Kagel, and McDonald 1985; Kagel 1987), so experiments on these mat-
ters need not be confined to humans.

Before leaving the subject of individual choice behavior, I will mention one
more experiment, that differs from those so far discussed in the scale of rewards
that were offered. Binswanger (1980) reports the results of an experiment carried
out among village farmers in areas of India, where (397) "the average physical
wealth of the households . . . is very low by international standards." Villagers
from a sample with substantial variations in wealth were repeatedly given the
opportunity to choose among a set of gambles that could be ranked in order of riskiness (395-396). ("To overcome moral problems confronting low income people involved in gambling, the gambling was limited so that the worst possible outcome was a zero gain.") First, relatively small gambles were offered, with the prizes eventually increasing to levels (405) "commensurate with monthly wage rates or small agricultural investments." Subjects considered their choices for several days.

The chief results were that, at very low payoff levels, there was a wide distribution of observed levels of risk aversion, but at higher payoff levels there was much less variance, with most responses concentrated in an "intermediate to moderate" level of risk aversion. Furthermore, subjects' risk aversion at these high payoff levels did not appear to be significantly influenced by their wealth. Binswanger also observed that these results obtained from actual gambles varied in important ways from the answers initially obtained from hypothetical questions about high stakes gambles. The hypothetical results showed both many more severely risk averse choices, and more risk neutral or risk preferring choices, than did the comparable data from actual choices.

In summary, in the series of experiments reported here the focus of debate was initially on if certain kinds of anomalous choice behavior were artifacts of the experimental procedure, and in particular if they would persist in nonhypothetical choice situations. In the case of preference reversals, the phenomenon survived both the change from hypothetical to real choices and increases in the payoff level. (However, questions about the reliability of hypothetical choices are not always resolved in this way, as is shown by the results of Binswanger.) The debate has now shifted to the underlying causes of the phenomena and if market environments will moderate the observed effects. Some of the various theories that have been advanced as possible explanations of certain kinds of choice behavior not only suggest new experiments, but new directions for pursuing traditional kinds of economic theory (see, e.g., Crawford 1990). It is clear that the degree of success such theories achieve in organizing and explaining phenomena in domains other than individual choice behavior itself will be important. At the same time, the question of whether or not individual choices in market and other economic environments are systematically different from what can be observed in various unstructured environments (either because certain kinds of choices do not arise in markets or because markets provide a certain kind of feedback) will undoubtedly require proponents of different points of view to sharpen their hypotheses about market phenomena.

3. Why Haven't These Demonstrated Anomalies Swept Away Utility Theory?

As the number of replicable violations of utility theory, and of even more basic models of rational choice, has grown, a question we frequently hear from some of our psychologist colleagues, and one that we can reasonably ask ourselves, is "what accounts for economists' reluctance to abandon the rationality model, despite considerable contradictory evidence?"
In a handbook of this sort, this question deserves to be addressed at at least two levels. The first, which I shall attempt in this section, concerns the implications of these "anomalous" experimental results for economic theory, and the second, which I'll address in the next section, concerns the implications for experimental economics per se.

Regarding economic theory, I'll argue that

1. There are quite defensible reasons for a reluctance to abandon theories of rationality in favor of nonrational theories (and, in view of this, some kinds of evidence and theories are likely to be more successful in attacking these defenses than others), and

2. There is (nevertheless) a considerable and growing attempt by economic theorists to respond to experimental evidence by moving away from an overdependence on idealized models of hyper-rationality.

Regarding the first point, to the extent that economists view expected utility maximization (merely) as a useful approximation of human behavior and not as a precisely true description in all cases, then to attack the central role it plays in economic theory, it isn't sufficient to show counterexamples, even many counterexamples.

Notice in this regard that even the simpler model of expected value maximization (i.e., risk neutral expected utility maximization for monetary rewards) remains a useful, and much used, approximation, although economists must universally agree that it does not provide a precise description of human behavior. But the reason utility theory at least partially replaced expected value theory as economists' "typical" approximation was because economists became convinced that the phenomena that couldn't be explained by the approximation that individuals maximize expected value were of central importance, and justified a model with unobservable parameters of risk aversion. (The phenomena that could only be captured this way include whole industries, such as the insurance industry.) I very much doubt that expected utility theory would have made such inroads into economics only on the strength of anomalies (from the expected value point of view) like the St. Petersburg paradox (for which Bernoulli proposed a kind of expected utility theory as a resolution.)

In this respect, I think that economists are likely to find most persuasive those attacks on rational models of choice that produce (more) examples of how not fully rational phenomena (such as framing effects) are reflected and exploited in important economic activities (e.g., in marketing and advertising and retailing) in ways that cannot be accounted for by rational models of individual choice.

There is also increasing attention being paid in the economics literature to alternatives to expected utility maximization as a model of individual behavior. One approach is to consider theories that are "plug compatible" with utility theory, in the sense that they are meant to replace it but to do the same job, and fit into strategic and market theories in the same way. Some of these are reviewed by Camerer in chapter 8 and are typically generalizations of expected utility theory, which relax some of its assumptions and which thus yield theories with more unobserved personal parameters (much as the replacement of expected value with
expected utility introduced personal parameters of risk aversion). Many of these are motivated by particular experimental anomalies (the Allais paradox has played an unusually large motivating role in this connection). By introducing additional parameters they are able to provide a better fit to the motivating data. But it remains an open question whether any of these theories organizes all of the data in a way that is clearly superior to utility theory, particularly when the costs (in diminished predictive power) of adding unobserved parameters are counted in.

Another approach involves the exploration of nonrational or boundedly rational foundations for economic phenomena involving collective choice—e.g., for theories of equilibrium and market behavior. The evolutionary and adaptive learning models I have discussed are good examples of these. The object of the theoretical work on these models is to understand the extent to which equilibrium phenomena, of the kind which economists have traditionally motivated with models of very rational economic agents, might also arise from more bounded sorts of rationality. To the extent that the reason economists don't dispense with hyperrational models of choice is that their primary interest lies in strategic and market phenomena (rather than in individual choice per se) this second approach may be even more important than the first. If important economic phenomena can be explained without reliance on rational models (or if the uses to which economists put rational models can be shown not to depend on them), the apparent gap between economists and psychologists in their models of individual behavior may diminish in importance.113

In summary, there exists a substantial body of experimental evidence that shows that individuals are not ideally rational utility maximizers; on the contrary, there are a growing number of systematic violations of utility theory that can be robustly demonstrated and reliably replicated in the laboratory. Nevertheless, even a brief review of the contemporary economic literature reveals that economists remain by and large quite content to model individuals as utility maximizers. What can account for this?

I've argued here that, to the extent that utility maximization is viewed as a useful approximation of behavior, it can't easily be displaced by counterexamples, since approximations always admit counterexamples. The experimental evidence is nevertheless very valuable—it is of the utmost importance to know where approximations break down. But in order to inspire theorists to replace a model that is regarded as a useful approximation, it is probably necessary to show that it is not useful for the purposes that it is being used. Some of the debate among experimenters about whether or not choice anomalies persist in market environments is addressed to this issue, and I expect that this debate will continue and become more specific about the types of market environments in which particular choice anomalies may or may not persist. And the usefulness of an approximation depends on what other approximations are available, and the further development of boundedly rational models of strategic and market behavior (also a response to the experimental evidence) may reduce the weight placed on rational models of individual choice by economists primarily interested in studying collective behavior.
Of course even an approximation that is adequate for many purposes may sometimes fail to provide the precise control desired in experimental design. We turn next to consider some of the implications of experimental results on individual choice for the design of economic experiments generally.

4. Experimental Control of Individual Preferences

Laboratory experiments make available a high degree of control of the environment, which makes possible two strategies for experimental design when the incentives facing participants in an experiment must be known: the experimenter can try to control the preferences of the subjects, or try to measure them. And because the theories being tested in an economic experiment are often phrased in terms of utility maximizing economic agents and also because the techniques of measurement and control often posit utility maximizing subjects, experimenters have to be alert to deviations from utility maximizing behavior on the part of experimental subjects.

In this section I will briefly discuss, from this point of view, some of the techniques for control and measurement of preferences that I have already mentioned in this chapter. But before discussing particular features of experimental design, there is one result of the experimental investigations of individual choice that has quite general implications. This is that the choice an individual makes is sometimes sensitive to the way it is presented, or "framed," in the sense that even theoretically equivalent choices may elicit different responses when presented differently (as in the case of preference reversals when comparisons are made by choosing between two lotteries or by stating a reservation price for each). This reinforces the general conclusion that the most reliable comparisons will be "within experiment" comparisons, in which the effect of a single variable can be assessed within an otherwise constant environment and "frame." For this reason also, economic experiments can rarely be interpreted to yield "constants" (e.g., the percentage of altruists in the population or the percentage of disagreements in bargaining), since the precise values observed in any experiment may be sensitive in unexpected ways to details of the experimental environment.

We turn next to specific concerns, related to utility maximization, which arise in connection with particular experimental procedures.

a. The BDM Procedure for Measuring Reservation Prices

As already noted, the Becker, DeGroot, and Marschak (1964) procedure has the property that it gives utility maximizers the incentive to reveal their true reservation price for an object, that is, the price at which they would just be indifferent between selling it and not selling it (or buying it and not buying it). The chief role it plays in experimental designs, therefore, is to allow the experimenter to determine unambiguously what are the predictions of theories that assume (among other things) that the subjects are utility maximizers and that their choices depend on their reservation prices. The substantial use (and usefulness) of the technique arises from the fact that so much of economic theory is of precisely this sort.
However, what if the subjects are not behaving as utility maximizers? This is a question that could in principle arise even when the predictions of a theory that assumes utility maximization are supported by the observed behavior (since many different theories may yield the same predictions in any given case). In practice, the question is raised most often when the results of an experiment violate the predictions of the theory, as was the case in the preference reversal literature. We saw there that when subjects' violation of utility theory is of the right sort (i.e., violation of the independence axiom) then the BDM procedure loses its "truth revealing" properties. Indeed, when subjects are not utility maximizers, the whole idea that they have a single reservation price, or that it is an indicator of their choice behavior, may be on shaky ground—this is after all the phenomena that preference reversals explore. Consequently different designs may be required to test some of the generalized utility theories for which the information obtained by the BDM procedure is not germane.

b. Controlling Preferences with Monetary Payoffs

The same issues arise with experimental designs intended to control subjects' preferences, starting with the simplest designs in which subjects are paid in money and the predictions of the theories being tested are formulated as if subjects were interested only in maximizing their own income. As I discussed in the context of both public goods and bargaining, there is abundant experimental evidence that these designs may sometimes fail to control subjects' preferences, because subjects may in fact also be concerned with the payoffs of other subjects, Chapters 2 and 4 both discuss experiments designed to test such hypotheses, that is, experiments in which it is not assumed that subjects' preferences are successfully controlled and can be equated with their payoffs in this way. Note, however, that to design an experiment that allows one to contrast a more complex theory of preferences with a theory based on simple income maximization, it is nevertheless necessary to know what the income maximization theory predicts, so that it remains necessary to control for the predictions of the simpler theory even when more complex theories of behavior are being examined.

Even when the experimenter only wishes to control subjects' reservation prices, as in the Chamberlin design for implementing markets with given supply and demand curves, there remains a question of whether subjects will in fact buy and sell at the prices the experimenter wishes. For example, Gode and Sunder (1993) express a good deal of skepticism that the Chamberlin procedure for inducing supply and demand behavior in fact delivers the intended experimental control. In motivating their study of programmed traders (discussed in section III.D), they say:

It is not possible to control the trading behavior of individuals. Human traders differ in their expectations, attitudes toward risk, preferences for money versus enjoyment of trading as a game, and many other respects. The problem of separating the joint effects of these variations, unobservable to
the researcher, can be mitigated by studying market outcomes with participants who follow specified rules of behavior. We therefore replaced human traders by computer programs. (Gode and Sunder 1993, 120)

As their programmed traders show, the nature of the individual behavior of the traders cannot be inferred from the fact that for some market structures and parameters the results may reliably approach equilibrium, and so the extent to which the Chamberlin procedure controls the preferences of the traders is a subject yet to be fully explored.

c. Controlling for Unobserved Risk Preferences with Binary Lottery Payoffs

One technique for controlling preferences that has not only been very widely used, but has also been the object of a good deal of experimental study designed to assess its effect, is the binary lottery procedure of Roth and Malouf (1979) for controlling risk preferences. The primary use for this technique has been for introducing a medium in which utility maximizers will be risk neutral, regardless of their (unobserved) natural risk postures. However, the technique also allows artificial intermediate commodities to be created in which utility maximizers can be induced to have arbitrary risk postures, and a number of experiments have begun to employ binary lotteries this way also, following the work of Berg, Daley, Dickhaut, and O'Brien (1986). They introduced a set of procedures by which subjects could be introduced to the notion that their payoffs would be computed in "points," and that their probability of winning the binary lottery would be a function of the points they received. Since the expected utility of a utility maximizing subject is precisely equal to his probability of winning the binary lottery, if the function $f$ converts points into probability, then the subject's utility function for points should be precisely the function $f$. Since $f$ can be chosen by the experimenter, the idea is that any utility function can be induced. This extended notion of binary lottery control of preferences thus gives considerable scope for tests of whether the procedure in fact induces actual experimental subjects to behave in accordance with the predictions for utility maximizers.

The most comprehensive test to date of the binary lottery procedure, both in its risk neutral and in its arbitrary utility function forms, is by Prasnikar (1993). She presented subjects with a large number of individual decision tasks in the form of choices between lotteries, after first attempting to endow each subject with one of five expected utility functions (whose functional forms were risk neutral, constant relative risk averse, constant relative risk preferring, and constant absolute risk averse and risk preferring). She could then test both if the aggregate responses deviated detectably from the precise coefficient of risk aversion she had attempted to induce in each group of subjects, and also if and how each individual subject might deviate from the choices that would be made by an ideal utility maximizer. An ingenious feature of Prasnikar's experiment was that after completing all choices each subject was tested to see if he understood how to calculate the probability of winning the binary lottery resulting from a compound lottery, with
prizes in points, using the induced utility function he had been endowed with. (This information had not been included in the instructions, which had simply presented, in tabular form, the probability of winning the binary lottery corresponding to each number of points.) This questionnaire thus allowed those subjects who had a good understanding of the binary lottery procedure to be distinguished from those who did not.

Prasnikar's main results were as follows. First, there were no substantial differences between subjects who knew how to quantitatively decompose compound lotteries and those who did not, indicating that even those subjects who knew how to do so apparently did not make their choices by conducting precise arithmetic computations. Second, the coefficients of risk aversion estimated from the observed data for each group of subjects (i.e., for all subjects endowed with the same utility function) did not differ significantly from those that had been induced. However, the data were quite noisy; individuals frequently deviated from the choices that would have been made by ideal utility maximizers. Prasnikar then sought to detect whether there were any systematic components to the deviations from the choices predicted for ideal utility maximizers. She found that the choices of the subjects who did not understand how to decompose compound lotteries were somewhat correlated with their natural risk aversion (which had been tested before each subject was introduced to the binary lotteries). However, no such correlation was detected among those subjects who had understood how the binary lotteries worked.

Prasnikar concludes that "the gross features of risk preference can be reliably implemented, albeit with a non-negligible amount of error." In particular, her results suggest that whether the amount of control of subjects' choice behavior achieved by the binary lottery technique is adequate for particular experimental purposes depends on both what those purposes are, and how carefully the binary lotteries are introduced to the subjects. If the design requires that each individual invariably perform as an ideal utility maximizer, it is unlikely that this technique delivers the desired control. However if the design requires that the aggregate risk posture of the subjects should be controlled, or that the unobserved natural risk aversion that subjects bring to the experiment should be neutralized, then the technique, carefully implemented, performs largely as expected.

Similar conclusions are reached by Rietz (1993), who tests, in a very different way, the ability of the binary lottery procedure to produce risk neutral behavior. Unlike Prasnikar (1993), who conducts a direct test of the induced individual choice behavior, Rietz explores the effect of the binary lottery procedure in an indirect manner, by considering its effect on bidding behavior in auctions. The equilibrium predictions for the first price auctions he considers are that risk averse players should bid higher than risk neutral players. Rietz therefore compares auctions with monetary payoffs to those with binary lottery payoffs to see if the attempt to induce risk neutral behavior via the binary lottery design will cause observed bids to decline. Of course, this is a test of the joint hypothesis that the players are utility maximizers (in the sense required for the binary lottery proce-
due to control their risk aversion) and that they are making equilibrium bids. As Kagel discusses in chapter 7, there is good reason to believe that players may have a tendency to bid higher than the equilibrium predictions (for reasons other than risk aversion) so that even a completely successful attempt to render the players risk neutral would not be expected to cause them to reach equilibrium. However, the power of Rietz’s test comes from the ability to see whether the introduction of binary lottery payoffs has a substantial effect on the players’ behavior in a situation in which a change in their risk aversion is predicted to have such an effect.114 Rietz concludes that

... [binary lottery] procedures can perform well in complex market environments such as sealed-bid auctions. However, they should be implemented carefully. The implementation should fit the environment, and the procedures should be simple enough for subjects to understand procedural implications completely. (212).

Thus the results of both Prasnikar (1993) and Rietz (1993) indicate that binary lotteries are not a magic wand, which can be waved over subjects to change their behavior. But carefully used, it appears that binary lottery payoffs can give experimenters a substantial measure of control over subject behavior, by neutralizing the unobserved natural risk postures that subjects bring with them into the laboratory.

Note again that, because the binary lottery procedure controls for the risk aversion of (ideal) utility maximizers, it allows experimental environments to be created in which the predictions of theories which depend on expected utility maximization can be known. This is what makes the technique so useful. However, in interpreting the results of an experiment which uses binary lottery payoffs, it is helpful to know the extent to which the procedure successfully controls the unobserved risk posture of experimental subjects who are not ideal utility maximizers. The evidence to date suggests that here too a substantial, although far from perfect degree of control can be achieved by careful application of binary lottery procedures.

(1) Preferences and Probabilities: A Historical Digression on Binary Lotteries and Related Experimental Designs

If I can end this chapter as I began it, with a foray into the history of thought, I recently became aware of an interesting precursor to the introduction of the binary lottery technique into experimental economics by the paper of Roth and Malouf (1979), in a 1961 paper by Cedric A. B. Smith. Smith’s paper was on the foundations of subjective expected utility theory, and he was interested in exploring the ways in which subjects’ assessments of probabilities could be separated from their preferences over outcomes. That is, if one individual can offer a bet to another who is willing to accept it, the problem is how to distinguish between the case in which the individuals have different subjective probabilities about the outcome, and the case in which they have different risk preferences, that is, differ-
ent expected utility functions on the payoffs. Although Smith's paper was purely theoretical, his approach to this problem actually took the form of an experimental design (albeit not a completely practical one) as follows (1961, 13):

To avoid these difficulties [differences in risk aversion between individuals] it is helpful to use the following device, adapted from Savage (1954). Instead of presenting cash to Bob and Charles, the Umpire takes 1 kilogram of bees-wax (of negligible value) and hides within it at random a very small but valuable diamond. He divides the wax into two parts, presenting one to each player, and instructs them to use it for stakes. After all bets have been settled, the wax is melted down and whoever has the diamond keeps it.

Effectively this means that if, say, Bob gives Charles \( y \) grams of wax, he increases Charles's chance of winning the diamond by \( y/1000 \).

. . . Hence using the beeswax or "probability currency" the acceptability of a bet depends only on the odds . . . and not on the stake.

Smith goes on to outline how, "by using this currency, Bob and Charles can be together; and from these bets we can construct a system of 'personal odds,' i.e. sets of lower and upper odds . . . for pairs of propositions . . ." (14).

Note that Smith's proposed use for his two outcome [diamond/no diamond] probability currency was just the opposite of the use for which the binary lottery technique was eventually introduced into experimentation in Roth and Malouf (1979). Whereas Smith was interested in establishing a method of determining probability assessments, the standard use of binary lotteries in the experimental literature has been for determining the expected utility to experimental subjects of risky outcomes with known probabilities. In this respect, it is interesting to compare binary lottery designs with the Becker, DeGroot, and Marschak (1964) procedure for determining subjects' values for objects, independently of their probability assessments. Their paper, which introduced the technique by reporting an actual experiment, was immediately incorporated into the experimental economics literature. In contrast, although Smith's paper was referred to in discussions of the axiomatization of subjective expected utility, his probability currency idea was before its time and appears not to have entered the discussions among economists until binary lotteries were already a standard procedure among experimenters. (Although I have been able to check the citation indexes going back only to 1967, it appears that Smith's 1961 paper was never cited in an economics journal before 1988, when it was referred to in a paper by Page [1988], concerned with a broad class of mechanisms for eliciting truthful probability revelation.115)

It is interesting to speculate why Smith's binary lottery idea, and certainly its practical implications, were lost for eighteen years and had to be independently rediscovered, while its contemporary, the BDM procedure, was quickly adopted as a practical feature of experimental design. It would be discouraging (but possibly not unrealistic) to conclude that this reflects the impermeability of the scientific literature, so that a practical idea that is first floated in a theoretical discussion in one discipline can take years to find its way to its natural users in another. More
optimistically, the reason may be instead that there are other, simpler ways for experimenters to control probabilities, since probabilities can often be made objective by conducting lotteries in the presence of subjects, with verifiable probabilities (e.g., by spinning a roulette wheel or a bingo cage). In contrast, preferences are more inherently individual and subjective, and therefore experimenters have more need of techniques to control or measure them. The fact that when binary lottery designs finally did become a standard technique it was to control preferences rather than probabilities provides some support for this more optimistic view, that Smith's binary lottery proposal may have gotten temporarily lost merely because he proposed to use binary lotteries to measure subjects' probability assessments and that in many cases experimenters preferred to deal with probability assessments more directly.

This brings me to my final point, which is that there are both costs and benefits associated with adopting possibly cumbersome elements of experimental design in order to better control the experimental environment.

d. To Control or Not to Control? Costs and Benefits

As long as utility based hypotheses are being tested, it is desirable to control for the preferences of utility maximizers in the design of experiments. (Failure to do so might mean that there would be room to dispute what the standard theory was supposed to predict, so that it would be unclear ex post whether the observed behavior supported the theory.) But there are costs to controlling preferences, from the direct cost of making subjects' decisions have nontrivial monetary consequences, to the indirect cost of adding complexity to an experiment (if only because each additional feature of the design must be explained to subjects and has the potential of being misunderstood so that additional complexity can translate into additional variance). And, as the discussion of the previous sections indicates, some of the benefit from particular experimental technologies may be reduced if their effect on subject behavior is unclear, or high variance. So in designing any particular experiment, it is reasonable to try to balance these costs and benefits.

Consider, for example, the binary lottery procedure for controlling risk preferences. The benefits of the control it offers are greatest when the theories being tested are sensitive to the risk posture of the subjects or when risk posture is being offered as a primary explanatory variable of observed behavior, and it is for this kind of experiment that the technique has become fairly standard. In other experiments, in which the theory being tested is relatively insensitive to the unobserved risk aversion of the players and in which this risk aversion does not play a primary role in the hypotheses being offered to explain the data, it is probably a reasonable first approximation to assume that players are risk neutral in monetary payoffs, particularly when the range of feasible payoffs is relatively small. And in experiments in which risk aversion is thought to play no role, nothing is gained by controlling for it.

Much the same could be said for the BDM procedure, and many of the same arguments apply as well to the general question of whether (and on what scale)
monetary payoffs should be the primary means by which an experimenter attempts to control the ordinal preferences of subjects. It is important, I think, to avoid establishing rigid orthodoxies on questions of methodology.

Notes

1. I first circulated a draft of this history at the Handbook Conference in 1990, and it has been substantially revised since then in response to the many comments I received. A slightly shortened version of my account of the period 1930-1960 appears in the *Journal of the History of Economic Thought*, under the title "On the Early History of Experimental Economics" (Roth 1993).

2. In a prescient speculation about experiments (and perhaps with mostly field experiments in mind) Wesley C. Mitchell in his 1924 presidential address to the American Economic Association said:

   The work of experimenting in the social sciences requires a technique different from that of the natural sciences. The experimenter must rely far more upon statistical considerations and precautions. The ideal of a single crucial experiment cannot be followed. The experiments must be repeated upon numerous individuals or groups; the varieties of reactions to the stimuli must be recorded and analyzed; the representative character of the samples must be known before generalizations can be established. . . . But whatever approaches are made toward controlling the conditions under which groups act will be eagerly seized upon and developed with results which we cannot yet foresee.

   In collecting and analyzing such experimental data as they can obtain, the quantitative workers will find their finest, but most exacting opportunities for developing statistical technique—opportunities even finer than are offered by the recurrent phenomena of business cycles. It is conceivable that the tentative experimenting of the present may develop into the most absorbing activity of economists in the future. (1925, 9)

3. Of course there is something artificial about dividing up the work in this way, and there are other ways in which it could be connected. For example, Siegel and Fouraker's work is in the game-theoretic tradition as well and influenced not only subsequent experiments in industrial organization but also in bargaining.

4. Thurstone (1931, 139) remarks:

   "The formulation of this problem is due to numerous conversations about psychophysics with my friend Professor Henry Schultz of the University of Chicago. It was at his suggestion that experimental methods were applied to this problem in economic theory. According to Professor Schultz, it has probably never before been subjected to experimental study."

5. We will see in what follows that, while this line of criticism is by no means uncontroversial, the question of actual versus hypothetical choices has become one of the fault lines that have come to distinguish experiments published in economics journals from those published in psychology journals. Of course, laboratory animals in psychology experiments face very well motivated choices, and Wallis and Friedman expressed some optimism about economic experiments using animal subjects as well and cite Wolfe (1936) and Cowles (1937) as interesting examples. And in fact a modern body of economic experiments has been developed using laboratory animals; see, for example, Kagel (1987) and Kagel, Battalio, and Green (1994).

6. However, this stipulation may have been addressed more at the first of Wallis and Friedman's criticisms (concreteness) than at the second (real payoffs). Although Rousseas and Hart's description of their experiment is somewhat ambiguous on this score (a situation that would
not be seen in a contemporary report of an economic experiment) it appears to me that the choices were still hypothetical and that no breakfasts were in fact cooked and consumed in response to the choices made (although for additional concreteness it was nevertheless specified that all eggs would be scrambled). However MacCrimmon and Toda (1969, 435) read Rousseas and Hart's account differently and conclude that their subjects were indeed required to eat their most preferred choice. MacCrimmon and Toda conducted an experiment themselves in which subjects did eat their choices; see the next note also.

7. See for example MacCrimmon and Toda (1969) who follow up on Thurstone and Rousseas and Hart with a similarly designed experiment that addresses the previous criticisms by using well-motivated individual choice data. In one part of MacCrimmon and Toda's experiment, the choices were among bundles involving combinations of cash and French pastries, "with the stipulation that the pastries had to be consumed in the laboratory, before the subject received any other payoff (441). (Subjects made multiple choices, and the determination of which bundle they would actually receive was made by the Becker, DeGroot, Marschak (1964) elicitation technique, which will be discussed shortly, in the discussion of the 1960s.) MacCrimmon and Toda argue that this procedure squarely addresses the Wallis-Friedman critique of Thurstone's experiment.

8. Expected utility theory had its predecessors in the work of Bernoulli and Ramsey, and there were predecessors of parts of game theory as well (see Weintraub 1992), but recall my earlier comments about scientific "firsts."

9. I make no attempt to include a full list, particularly since many early utility theorists employed an informal, but nevertheless, revealing style of casual experimentation, casually reported. For example, Markowitz (1952) gives a qualitative account of the responses "of my middle-income acquaintances" to hypothetical questions about lotteries involving gains and losses. More formal reports of experiments involving hypothetical choices still play an important role in this literature.

10. They report (372) that "plans for this experiment grew directly out of discussions with [Milton] Friedman and [L. J.] Savage at the time they were writing their [1948] paper. W. Allen Wallis also contributed to the discussions." The Friedman Savage (1948) paper contains a conceptual experiment along these lines, and I think the two papers taken together make a fine illustration of the great distance between the first conception of an experiment and its careful implementation and realization.

11. "Rationality can [also] be defined experimentally by observing the actions of people who can be regarded as acting in a rational manner" (Allais 1953, 504).

12. The famous story of the two prisoners each of whose dominant strategy is to confess, even though both do better if neither confesses, is due to Tucker (1950). Straffin (1980) recounts how Tucker came across the game on Dresher's blackboard and composed the story that has given the game its name. Apparently Howard Raiffa independently conducted experiments with a prisoner's dilemma game in 1950, but did not publish them (see Raiffa 1992).

13. Note that, if the game were played only once, it would in fact be a dominant strategy for Row to play row 2, and for Column to play column 1. In the repeated game these are no longer dominant strategies. That (nevertheless) no other actions occur at any period of the equilibrium of the repeated game follows by backward induction from the now familiar observation that on the last play of the game no player can do better than to play his one period equilibrium strategy, and so for the purpose of calculating the equilibrium we can now treat the game as a ninety-nine period repeated game and repeat the argument.

14. Limitations of which the very first paragraph of Flood (1958) makes clear the investigators were aware.

15. But note that Flood's 1952 report led quickly to follow-up experiments. Two that were funded by the Air Force and conducted at Ohio State University were reported by Scodel, Minas, Ratoosh, and Lipetz (1959) and Minas, Scodel, Marlowe, and Rawson (1960). Like Dresher and Flood's experiment, these used monetary payoffs to avoid hypothetical choices. However, some of the phenomena the authors observed made them question if very small payoffs were significantly different from hypothetical payoffs.
16. See, for example, Poundstone (1992) for a popular biography of von Neumann and his times, which focuses on the prisoner's dilemma. Poundstone devotes a good deal of attention to the early prisoner's dilemma experiments.

17. Specifically Coombs and Beardslee (1954), Estes (1954), Flood (1954a, 1954b), Hoffman, Festinger, and Lawrence (1954), and Kalisch, Milnor, Nash, and Nering (1954). And see Simon (1956) for some reinterpretation of the results of Estes. Oskar Morgenstern gave a talk at that conference, later published as Morgenstern (1954), in which he applauded the appearance of “strictly planned experiments” and anticipated a large future role for economic experiments of various kinds.

18. Recall Nash's comments about the prisoner's dilemma experiment.

19. Schelling's paper and these examples were also reprinted in his influential 1960 book *The Strategy of Conflict*.

20. See Stone (1958) for a related bargaining experiment with more complicated sets of feasible agreements.

21. Of course, it might still be that the equal division in the first problem is prominent because it is fair, so Schelling's prominence hypothesis does not necessarily contradict the fairness hypothesis implicit in the advice of Nash and his colleagues. In fact, hypotheses about fairness play a lively role in the contemporary exchange among experimenters, and we will see in section III and in chapter 4 how experiments have served to advance and to focus the debate.

22. This is not to say that there were not contemporary follow-ups to his experiments: see, for example, Willis and Joseph (1959).

23. Ledyard touches on models of learning in chapter 2 when he discusses the effects of repetition and experience in public goods experiments, and I will discuss in chapter 4 how simple models of learning track the data of repeated play in certain kinds of bargaining experiments. In section III.B of this chapter I discuss learning models of behavior in coordination games.

24. He cautions that there were some accounting differences between different trials (cf. footnote 2, page 98), as well as some differences between how completed transactions were reported to the participants in the different markets.

25. Chamberlin notes that in fact he observed this happen “ten to twelve times out of the forty-six trials” (98).

26. The “moving equilibrium” line in Chamberlin's figure graphs the price determined after each transaction by the intersection of the aggregate supply and demand curves of the agents still in the market.

27. Especially when viewed in the light of their immediate follow-ups to that work, in collaboration with Donald Harnett and Martin Shubik, in Fouraker, Shubik, and Siegel (1961), Fouraker, Siegel, and Harnett (1962), and Fouraker and Siegel (1963).

28. Interestingly, the aspiration hypothesis has attracted different amounts of attention on different sides of the ocean. American experimenters and theorists have subsequently come to regard aspirations as, at most, an intermediate variable, rather than as a primary explanatory variable. Our German counterparts have been more inclined to regard aspirations as a primary explanatory variable (although the two sides do not divide up quite so neatly: see, e.g., the edited volume on “Aspiration Levels in Bargaining and Economic Decision Making” [Tietz 1983]). For other early thoughts on aspirations and expectations, see Simon (1959). The work of Sauermann and Selten in these early papers, and subsequently, has other things in common with the work of Simon, such as their common interests in decision making process (see, e.g., Cyert, Simon, and Trow [1956] for a nonlaboratory study).

29. And although these directions involved "non standard" game-theoretic considerations such as aspirations, Siegel and Fouraker saw both the origins of their experiments and their outcome as squarely in the game-theoretic tradition. Speaking of von Neumann and Morgenstern's *Theory of Games and Economic Behavior*, Fouraker and Siegel say (1963, 6): "The reinforcement of economic theory with the mathematics and general methodology of that magnificent work has provided the impetus for a broad front of new research; we hope this book is a proper element of that movement."
30. It has today become extremely rare to see an experiment that does not use real payments published in an economics journal. However, the relative efficacy (and cost efficiency) of the two kinds of experiments remains a subject of lively debate (particularly when the real payments may be small, or not very sensitive to players' behavior), and the debate is fueled by the fact that a number of individual choice phenomena, which were first identified with (inexpensive) experiments using hypothetical rewards, have subsequently been robustly reproduced with real payments. (See, e.g., Thaler [1987], who after reviewing a number of studies in which the difference between real and hypothetical payments did not yield important differences in results notes that [120]: "Asking purely hypothetical questions is inexpensive, fast, and convenient. This means that many more experiments can be run with much larger samples than is possible in a monetary-incentives methodology.") This debate will show up in several chapters, including chapters 7 and 8 by Kagel and Camerer.


32. The experimental papers included in Sauermann (1967) are Sauermann and Selten (1967a,1967b), Selten (1967a,1967b,1967c), Tietz (1967), and Becker (1967). The volume also includes English summaries of these papers and a bibliography of experimental literature compiled by Volker Haselbarth.

33. See, for example, the extensive bibliographies for this period in Sauermann (1967) and Shubik (1975). Some of the other notable experimenters and experiments from this period are Becker, DeGroot, and Marschak (1963a,1963b,1964); Bower (1965); Contini (1968); Dolbear, Lave, Bowman, Lieberman, Prescott, Rueter, and Sherman (1968); Ellsberg (1961); Friedman (1963); Lave (1962); Lieberman (1960); Maschler (1965); Rapoport and Cole (1968); Shubik (1962); Smith (1962, 1964); and Yaari (1965). Rapoport, Guyer, and Gordon (1976, 423) present a graph of the "number of articles, books, memoranda, etc., published from 1952 to 1971, on various aspects of game experiments," which shows a fairly steady rise from about thirty papers published in 1960 to between ninety and a hundred papers in each of 1967, 1968, and 1969, many by social psychologists.

34. The argument is the same as the argument of Vickery (1961) that it is a dominant strategy to bid your true value in a second price auction. (Smith [1979a, footnote 1] recounts how he heard this kind of procedure described by Jacob Marschak in 1953.) Becker, DeGroot, and Marschak (1964) report that they used this technique to repeatedly estimate the utility functions of two experimental subjects. They concluded that their subjects' responses were not consistent with utility maximization, although their behavior became more consistent with repeated exposure to the problem.

35. Smith (1992) writes of his early involvement in experimental economics and notes that his first exposure came as a graduate student at Harvard in 1952, when he attended a class taught by Chamberlin in which Chamberlin's 1948 experiment was conducted as a classroom exercise. (As such he may be the first "second generation" experimental economist. He reports that among his contemporaries Jim Friedman also learned of experiments as a graduate student, from Martin Shubik at Yale. Perhaps more common at that time was the experience of Reinhard Selten, whose experimental work Smith reports was inspired by the 1954 experiment of Kalish, Milnor, Nash, and Nering. I suspect that certainly until very recently, and perhaps in many cases still, economists have often ventured into experimentation without having had any direct contact with experimenters as part of their formal education.) Smith notes that he was not active in experiments for a period from the late 1960s through the early 1970s and that his collaboration with Charles Plott played an important role in his reemergence as an active experimenter. Their first paper, Plott and Smith (1978), made an important contribution to the experimental industrial organization literature and will be discussed at greater length by Holt in chapter 5.

36. It was in this period that the editors of the present volume each published their first experimental work.
37. A notable contributor to experimental economics from the psychology side of the divide is Amnon Rapoport, a collection of whose papers, from the 1960s through 1990, are contained in Rapoport (1990). In his introduction he characterizes with regret the separateness of the two literatures, saying in part (ix):

The history of experimentation in psychology is rich and old. It would have been quite natural and highly desirable for psychologists to extend their scope of research and assume a major role in the study of economic decision behavior. Psychology professes to be the general study of human behavior. Most psychologists are trained to regard their discipline as an observational science; they do not have to overcome the conditioning of many economists who think of economics as an *a priori* science. Psychologists' knowledge of experimental techniques is comprehensive, and their experience in conducting experiments, analyzing data, and discovering empirical regularities exceeds that of most economists. However, with the exception of research on individual choice behavior—where psychologists like Tversky, Kahneman, and Slovic have played a major role—psychologists have not contributed in any significant way to the growing research in experimental economics.

More recently there have been indications that this divide is being closed, and important experimental contributions by psychologists to the investigation of economic phenomena have been made in recent years. Many of these, made by students and associates of Rapoport (such as Bornstein, Budescu, Erev, Suleiman, Weg, and Zwick), by Robyn Dawes and his colleagues, and by Keith Murnighan and his colleagues, will be encountered in the chapters of this volume. As Ledyard points out in chapter 2, the interaction between disciplines has been particularly productive in experimental studies of the provision of public goods, with contributions not only in the economics and psychology literatures, but also in sociology and political science. (For a collection of papers representing the growing use of experiments in political science, see Palfrey [1991]. In this connection, see also McKelvey and Ordeshook [1990].)

38. For example, in the middle of the decade the *Journal of Economic Literature* initiated a separate bibliographic category for "Experimental Economic Methods," and by the end of the decade experiments were beginning to become standard fare in introductory economics courses (see, e.g., the excellent graduate textbook by David Kreps [1990], which pays particular attention to bargaining experiments, but also to experiments concerning individual choice and market behavior). Among undergraduate texts, the instructor's manual of Varian's intermediate microeconomics text (Varian 1990) contains a section written by Glenn Harrison (87-116) on running classroom market experiments (and the new text by Stiglitz comes with software for that purpose). (The use of experiments in the classroom as a teaching tool is showing signs of becoming a subject in its own right: see, e.g., the special issue of the *Journal of Economic Education* devoted to Classroom Experimental Economics, containing articles by Williams and Walker [1993], Bell [1993], Williams [1993], DeYoung [1993], Leuthold [1993], and Fels [1993]. There is also a small newsletter devoted to the subject, called Classroom Expernomics, edited and published by Greg Delemeneer at Marietta College in Ohio and John Neral at Frostburg State University in Maryland.) We have even started to see texts devoted to experimental economics—see Hey (1991), Davis and Holt (1993), and Friedman and Sunder (1994). Finally, we have seen the establishment of laboratories that plant the seeds of experimental economics in new communities of economists (e.g., in England, France, the Netherlands, Spain, and Japan).

39. Although Allais was cited for his work in general equilibrium theory, the statement from the Royal Swedish Academy of Sciences referred to his experimental work by noting (1989,3) that "he is perhaps best known for his studies of risk theory and the so-called Allais paradox." In his summary of Allais' career contributions on the occasion of the Nobel Prize, Grandmont (1989, 23-4) writes: "Another of Allais' outstanding contributions—this time one which is well known—concerns the theory of decision-making among risky prospects. . . . Allais' attitude towards the [expected utility hypothesis] has been characteristic of his constant view that theory should be confronted with facts. At a meeting that he organized on this topic in
1952, Allais conducted a series of experiments in order to test the empirical relevance of the hypothesis, the results of which were partially reported in [Allais, 1953]. These experiments showed that actual behavior violated systematically the expected utility hypothesis. This fact, which is known to economists as the Allais paradox, has generated an increasing amount of research work, both empirical and theoretical, especially in recent years. This work is discussed by Camerer in chapter 8.

40. A somewhat reorganized version of that talk appeared as Roth (1987a). While the contents of the present chapter are substantially different from that earlier survey, I have felt free to borrow from it where appropriate and also from my subsequent survey of different experimental topics in Roth (1988).

41. They are used in just this way in Michael Bruno's 1986 presidential address to the Econometric Society, which appears as Bruno (1989) (see particularly 300-1).

42. In this respect, these kinds of laboratory experiments are close kin to field experiments of the kind surveyed by Ferber and Hirsch (1982) or Hausman and Wise (1985), as well as to the analysis of complex special situations, such as the cigarette economies that arose in POW camps (see, e.g., Radford 1945).

43. I will try to use the singular term "experiment" to cover an entire experimental design, which may consist of many observations, each one a potentially complex interaction among many participants. I have elsewhere (Roth 1994) argued against the practice in some experimental economic literature (now much less common than it once was) of regarding each trial as a separate experiment. When combined with a tendency to report only "successful" experiments, this practice raises the risk of introducing a misleading element into experimental reports, similar to that facing econometricians who conduct many regressions and report only those that appear significant (cf. Learner 1983).

44. "Chicken" takes its name from the story of two drivers (presumably adolescent, presumably male) who play a game of nerves by driving their cars directly at one another at high speed. Each player's preferred outcome is that the other will "chicken out" at the last minute and swerve to avoid a collision, while the player who did not swerve is rewarded with a (good) reputation for reckless disregard for his own safety. To swerve earns a less favorable reputation, but the worst outcome occurs when neither player swerves, and the game ends in a fatal collision. So the game has two equilibria, in each of which exactly one of the two players swerves. (What makes the game exciting is the problem of equilibrium selection.)

45. From payoffs and number of trials to personality differences: see, for example, Lave (1965) and Terhune (1968).

46. The nonnegligible observed rates of cooperation have caused many investigators to entertain hypotheses concerning the altruism of the participants. However, a recent experiment by Shafir and Tversky (1992) suggests that the reasons why cooperation is so often observed in the one period game may be complex. They compare rates of cooperation in a conventional one period prisoner's dilemma with rates of cooperation observed in a one period game in which a subject is first told what the other player has done. Somewhat surprisingly they observe lower rates of cooperation in the modified game than in the conventional prisoners' dilemma, not only when subjects are told that the other player defected, but also when they are told that the other player cooperated. Shafir and Tversky therefore attribute some of the cooperation observed in the conventional prisoner's dilemma not to altruism (which should cause subjects who know that the other player has cooperated to also cooperate), but to the difficulty of evaluating whether to cooperate or defect when the consequent outcome of the game will depend on the still uncertain action of the other player, in comparison to the simpler problem faced by subjects in their modified game). A related experiment with a very different design has led Andreoni (1993) to somewhat similar conclusions about the causes of the nonnegligible rates of contribution observed in public goods provision games whose equilibria call for no contributions.

47. Cooperation can be achieved by some equilibrium if and only if \( \frac{(b - a)(b - c)}{(b - a)(b - d)} \), and it can be achieved "easily," by the "tit for tat" strategy of first cooperating and then doing whatever the other player did in the previous period, if and only if \( p \geq \frac{(b - a)(a - d)}{(b - a)(a - b)} \) also.
48. Subjects played against a programmed strategy (without knowing what it was). In fact the programmed opponent always played the "tit for tat" strategy. And the players' incentives were only loosely controlled. Note that, since the equilibrium calculations depend on expected values, it would have been necessary to control for expected utility, not just ordinal utility, in order to do a proper job of controlling the equilibrium predictions. The experimental tools for doing that (via binary lottery games, as will be discussed in sections III.C and III.F) were not introduced until the paper of Roth and Malouf (1979).

49. The payoffs were $b = 1.45$ German marks, $a = 0.6$, $c = 0.1$, and $d = -0.5$. The choices were phrased as setting a high price (cooperation) or a low price. (For early discussions of the prisoner's dilemma as a model for cooperation among oligopolists, see Shubik [1955]. An early experiment on collusion among several oligopolists that refers to the prisoner's dilemma as such a model is Dolbear, Lave, Bowman, Lieberman, Prescott, Rueter, and Sherman [1968]. [I have always suspected that so many authors may indicate a predisposition to collusion.])

50. The authors caution, however (54), "Even if it is very clear from the data that there is a tendency of the end-effect to shift to earlier periods, it is not clear whether in a much longer sequence of supergames this trend would continue until finally cooperation is completely eliminated."

51. Other theoretical attempts have been directed at changing the notion of equilibrium entirely (see, e.g., Rosenthal 1980) or at studying closely related problems (see, e.g., Selten's [1978] chain store paradox and the papers by Kreps and Wilson [1982] and Milgrom and Roberts [1982]). An experimental study motivated in turn by this literature is reported in Camerer and Weigelt (1988), who interpreted their results as supporting the predictions of models in which a small amount of incomplete information is introduced. Subsequent experiments, by Neral and Ochs (1992) and by Jung, Kagel, and Levin (1994), suggest however that this interpretation may have been premature. Orbell and Dawes (1993) point out that allowing players to exit from games if they wish also produces changes in both predicted and observed behavior. In the repeated prisoner's dilemma, this can increase efficiency by making it possible for players to exit rather than retaliate.

52. It turns out that there were two "kingmakers," that is, two programs that largely determined how the other programs did.

53. Of course, the results are also sensitive to the payoff matrix, which in this tournament had payoffs of $b = 5$, $a = 3$, $c = 1$, and $d = 0$, so that this kind of alternation gives up a half point each period in comparison to steady cooperation.

54. Another way of combining simulation with experimentation is explored by Roth and Erev (1994), who simulate via simple learning rules how the behavior of game players evolves as they gain experience. When the initial behavior in the simulations is estimated from the initial behavior observed experimentally, Roth and Erev report that, for a variety of games, the behavior predicted by the simulations for experienced players resembles the experimentally observed behavior. I discuss this at greater length in chapter 4 and some related work in section III.B of this chapter.

55. Although public goods provision is a subject older and larger than the prisoner's dilemma, the role of game theory in the foundations of experimental economics is reflected in the much longer history of prisoner's dilemma experiments than other kinds of public goods experiments. That is not to say by any means that the division between the two is clear. For a particularly interesting set of experiments in which prisoner's dilemma considerations are inextricably entwined with some of the larger issues of public goods, see for example, Bornstein and Ben-Yossef (1993) and Bornstein and Hurwitz (1993), which consider prisoner's dilemmas played by teams, in which the issue of free riding enters not only between teams but within teams. Bornstein, Erev, and Goren (1994) study learning in repeated play of these games, using the learning model of Roth and Erev (1994).

56. However, he notes that a sixth group of subjects who were asked in a purely hypothetical way how much such a program would be worth to them gave significantly different responses from the other five groups. He says (125): "This result may be seen as still another reason to doubt the usefulness of responses to hypothetical questions. . .."
57. Experiments with human subjects in the United States are now regulated by state and federal laws that require that universities maintain review boards to determine in advance that experiments do not violate certain guidelines. These laws were passed in response to some hair-raising abuses, with notable contributions from both psychologists and biomedical researchers.

58. However, the experiment did not end here. The students were told that they had failed to reach the required sum, so only the two original high bidders would get the book, although the offer would remain "open for a few days should the students still want to try to bring the money together." The (now surely desperate?) students were then presented with a third scheme, in which they were told essentially that any bids they submitted would be sufficient. These bids provided a third comparison, and, while they were significantly less than the previous two bids, they were significantly greater than the minimum bid required to be included among those who would (supposedly) receive the books before the exam. (Unfortunately we do not learn how the students did on the exam, or if their bids were good predictors of their grades.)

59. There are some complexities in the data, since ten students bid zero in the first auction, but contributed positive amounts when the book was offered as a public good. The authors consider the possibility that this was a result of coalition formation in the auction.

60. However, Palfrey and Rosenthal (1987) speculate that in a number of these experiments the monetary payoffs cannot simply be taken as equivalent to the utility of the agents, because there may be an unobserved "altruistic" component to agents' preferences. They go on to study the effect that this could have in a strategic situation in which being able to estimate how much others will contribute is important.

61. Another interesting experiment using this general design is that of Ferejohn, Forsythe, and Noll (1979). They examined a public goods provision mechanism abstracted from one used by station managers in the (American) Public Broadcasting Service to decide on what shows to collectively purchase.

62. One feature of the procedures in this study that differed from the studies so far discussed is that subjects knew they would be required to explain their decisions to the experimenter. However, they attribute this to selection rather than training, noting that few of the economics graduate students "could specifically identify the theory on which this study was based." In view of the fact that there were other obvious differences between the subject pools (e.g., graduate students versus high school students), I suspect that the authors do not take this result as seriously as some of the others they report. However, the point that different subject pools may behave differently is always a matter of potential concern, and one which can be addressed empirically.

63. Since in fact there were only five subjects, payoffs were based on calculating the total contributions to the fund as if each subject represented twenty, with some modifications designed to conceal from the subjects how few of them there were. In chapter 2 John Ledyard takes strong exception to designs that involve any deception of the subjects; however, his view is not universally shared.

64. But see Isaac, Walker, and Thomas (1984), who observe some related results in a design that helps separate experience from repetition among a fixed group.

65. In seven of these groups, only the minimum choice was publicly reported, while in two groups the whole distribution of choices was made public after each period of play. This result, known as the "Law of Effect," has been observed in a very wide variety of environments at least since Thorndike (1898).

66. This observation is known as the "Power Law of Practice" and dates back at least to Blackburn (1936).

67. This basic model was proposed as an approximation of evolutionary dynamics by Harley (1981).

68. In the same way, the role of mutation in evolutionary models may be played by experimentation and error in models of learning. For some recent theoretical work in this direction, see Fudenberg and Kreps (1988) for a model of persistent experimentation in extensive form games.
71. For a more detailed survey of this material, see Roth (1987b).
72. For example, Morley and Stephenson (1977) state that "these theories ... do not have any obvious behavioral implications" (86).
73. These experiments are reviewed in Roth and Malouf (1979).
74. Of course, it is commonplace in interpreting field data in economics that one is often obliged to accept or reject joint hypotheses, which cannot be separated from one another. Much of the power of experimental methods comes from the fact that they often allow us to test such hypotheses separately.
75. The appearance from the results of Roth and Malouf (1979) that Nash's theory was a good point predictor of agreements in the partial information condition did not survive examination of a wider class of games. The robust feature of those results, rather, was that when players did not know one another's monetary value for agreements, there was a tendency to reach agreements that gave each bargainer an equal share of the commodity being divided, whether this was lottery tickets or some more usual medium (see Roth and Malouf 1982).
76. For a further exploration of self-serving assessments of fairness, see Babcock, Loewenstein, Issacharoff, and Camerer (forthcoming).
77. The means in each game do reflect a small tendency for the player with the smaller number of chips in his prize to receive a higher percentage of the lottery tickets; however, this effect is approximately an order of magnitude smaller than the difference between the prizes of the players with low and high monetary values in either the full information condition of the previous experiment or the high information condition of this one.
78. It is hard to even imagine any field data that might allow the effect of this kind of information difference to be observed. But, in view of the importance of notions such as "common knowledge" in contemporary game theory, differences of precisely this kind are increasingly encountered in the theoretical literature. Laboratory experiments give us a way to investigate them.
80. Including those of Nash (1950), Kalai and Smorodinsky (1975), and Perles and Maschler (1982).
81. That is, it was thought that the high concentration of agreements around a focal point such as (50%,50%) might reflect forces at work that made it unprofitable for bargainers to try to achieve small deviations from equal division, but that, once the bargaining had shifted away from such a compelling focal point (into a region in which previous experiments had shown agreements would have greater variance), the influence of risk aversion on the precise terms of agreement might be greater.
82. Note that this is an experimental design that depended critically on the theoretical demonstration, in Roth and Rothblum (1982), that there are situations in which theories of bargaining like Nash's predict that risk aversion will be advantageous to a bargainer. Prior to that demonstration it had not been clear how to design an experiment that would separate the predicted (disadvantageous) effects of risk aversion from the possible effects of other personal attributes that might be correlated with risk aversion. If risk aversion were predicted to be disadvantageous in all the bargaining situations under examination and if an experiment were conducted in which it was observed that more risk averse bargainers do worse than less risk averse bargainers in these situations, then it might still be the case that risk aversion was correlated with, say, a lack of aggressiveness and that it is aggressiveness that accounted for the results.
83. When prizes are small, the relatively small effect of risk aversion observed here suggests that it may not be critical to always control for unobserved effects of risk aversion by employing binary lottery games in an experimental design, particularly when risk aversion is not a primary cause of concern for the phenomenon being investigated. Roth and Malouf (1982) report experiments similar to those of Roth and Malouf (1979), but in which players are paid...
in money rather than in lottery tickets, and observe that the qualitative effects of information
are very similar. Harrison and McCabe (1992), Radner and Schotter (1989), Cox, Smith, and
Walker (1985), and Walker, Smith, and Cox (1990) also observe only small differences
between certain observations made with and without binary lottery games. (However, Rietz,
[1993] reports a very careful replication and extension of the experiment reported in Cox,
Smith, and Walker and Walker, Smith, and Cox, and concludes that their results, which
concern an environment in which risk aversion may play a role, are primarily due to artifacts
of their experimental procedure. This will be discussed further in section III.F, when we
discuss tests of the effects of binary lottery designs on subject behavior.)

84. In experiments in which multiple units of a commodity, or of different commodities, are
available, subjects can be given separate reservation prices for each unit (e.g., to induce
increasing or decreasing costs of "production" among sellers).

85. Smith (1962) acknowledged that using hypothetical payoffs amounted to swimming against
the prevailing methodological tide among experimental economists, by noting before begin-
ning (in a footnote to the title) that "the next phase is to include experimentation with mone-
tary payoffs and more complicated experimental designs to which passing references are
made here and there in the present report."

86. Except for 5ct "commissions" that were paid for each transaction in that experiment, in order
to encourage players to make trades at the margin.

87. Van Boening and Wilcox (1993) summarize their conclusions as follows (i):

Thirty years of experiments show that the double auction trading institution achieves
nearly maximal efficiency across a wide variety of market structures. Accordingly the
contemporary consensus is that the performance of the institution is virtually indepen-
dent of market structure and the strategies of agents. We show that this is not so: (1)
Large avoidable costs can undermine both the efficiency and stability of the double
auction; (2) Certain sellers who must produce for full efficiency in such markets earn
negative profits, on average, when they do produce; and (3) Price dynamics in such
markets are described well by recent theory meant to explain price dynamics of ordinary
marginal cost double auctions. We conclude that double auctions are neither as robust,
nor well understood, as currently thought; and that the performance of the double auc-
tion does depend on market structure, and through it on rationality and strategy as well.

88. For example, if prices start high in the first period and trend downward, then everyone learns
from the last transactions in the period that there are sellers willing to trade at below the
average price for the period. Because all parameters are the same in the next period, buyers
can start the next period by trying to buy at those prices, etc. This can be seen in the transac-
tion paths of many repeated double auction experiments, in which, for example, prices of the
first transactions and more complicated experimental designs to which passing references are
made here and there in the present report."

89. Plott (1986) reports that this experimental evidence was ultimately not used in the court
testimony. The government won the case, but was reversed on appeal. For some subsequent
theoretical work that supports the general conclusions of this experiment, see Holt and

90. See also Forsythe et al. (1991a, 1991b).

91. These markets fall under the regulatory purview of the Commodity Futures Trading Com-
mission, which has issued a "no action" letter, stating that no adverse regulatory action will
be taken so long as the market maintains a limit on the size of the initial capital investment
in each account ($500), does not engage in paid advertising, and remains non-profit.

92. Not the coins themselves (to control for "penny aversion").

93. Thus private signals are "positively affiliated" in the sense of Milgrom and Weber (1982).

94. For a similar experiment with previously inexperienced subjects, see Kagel, Levin, Battalio,
and Meyer (1989), who report similar results.

95. It is a little difficult separating group size from experience and selection in these results,
since although group size was one of the design variables, some of the small groups are the
result of bankruptcies by overbidders in early periods.
The authors remark: "We believe that the executives have learned a set of situation specific rules of thumb which permit them to avoid the winner's curse in the field but which could not be applied in the lab." (It is, of course, also possible that the bidding environment encountered in the field is not well represented by the one created for the experiment. For example, in a field study of machine tool auctions, Graham and Marshall [1987] and Graham, Marshall, and Richard [1990] found that collusion among bidders was pervasive.)

For another analysis of the field data, see Hendricks, Porter, and Boudreau (1987).

The authors report that the probability of losing money based on the observed amount of overbidding averaged only 0.06.

The "we believe" is due to the fact that the values here aren't independent, but rather affiliated. However, they note (footnote 22):

> It is unlikely that positive affiliation is responsible for these differences. We have conducted one second-price experiment with independent private values which showed average market prices in excess of the predicted dominant strategy price. Further, recently published nondiscriminatory, multiple unit sealed bid auctions with independent private values, where the dominant strategy is to bid one's value, show a substantial (and persistent) percentage of all bids in excess of private values. (Cox, Smith, and Walker 1985b)

And in a subsequent experiment, Kagel and Levin (1993) replicate the overbidding in second price single object auctions with independent private values, as part of a very interesting experiment designed to investigate the qualitative differences predicted to hold for first, second, and third price auctions.

Deviations in expected payoffs from the equilibrium payoff will differ less than the deviation of the bids from the predicted bid, since the former is the latter times the probability that the bid will be the winning bid. So, particularly for low bids, which have low probability of winning, substantial changes in the bid can have very small consequences for the payoff.

In this connection, the assumptions involved in introducing unobserved parameters always deserve careful, skeptical examination. For example, in the papers of Cox et al., which Harrison critiques, the fundamental assumption is that all observations are at equilibrium. This certainly flies in the face of much of the auction data, which suggests that a great deal of learning takes place in early periods of play. Indeed, learning is one of the ubiquitous phenomena observed in experiments, as will be seen in nearly all of the chapters of this volume.

Note the relationship to other phenomena, such as the observation of Kagel, Harstad, and Levin (1987) of overbidding in second price private value auctions. There (and also in the "disadvantageous counteroffers" observed in bargaining experiments and discussed in chapter 4), actions that would be "irrational" if expected monetary income could be equated with utility result in only small expected monetary losses, and so cannot be regarded as strong evidence of irrational behavior. They may simply be evidence that small monetary differences may be insufficient to override nonmonetary elements in subjects' utility or that negative feedback that occurs with only small probability may be insufficient to correct misconceptions. See Fudenberg and Levine (1993) for a reexamination of several experiments along these lines.

For the pairs of negative expected value bets, the prediction is (since subjects are predicted to focus in the pricing task on the size of the potential loss) that they will be willing to pay more to avoid playing the $ bet, with its large potential loss, than they are to avoid playing the P bet. The "predicted reversal" thus occurs when the bets are priced in this way, but the $bet is chosen over the P bet. For these bets also, the predicted reversals outnumbered the unpredicted reversals. The authors note that, by including negative expected value bets in the design, they are able to rule out one alternative hypothesis for the results in the positive expected value case, namely that subjects price the $ bets in such a way as to increase the likelihood that they will retain them (either out of a strategic impulse to state high selling prices or out of a preference for playing out gambles). Such a strategy for negative expected
value gambles would involve stating a less negative price, and this would have diminished the number of predicted reversals.  

104. In their reply, Grether and Plott (1982) note that Pommerehne et al. did not replicate the earlier experiment so that it is premature to attribute the lower rate of reversals to the higher payoffs, since the experiments differed in other ways as well.  

105. In a comment on these experiments, Slovic and Lichtenstein (1983) urge economists to view such reversals not as an isolated phenomenon, but as part of a family of choice anomalies that may arise from information processing considerations.  

106. As noted above, the argument for the Becker, DeGroot, and Marschak procedure is the same as that for second price auctions, and so violations of independence have the same implications there when outcomes are uncertain. See Karni and Safra (1985).  

107. An initial experiment in this direction is that of Loomes, Starmer, and Sugden (1989).  

108. Note the parallel to the use of binary lottery games discussed in section III.C.  

109. A similarly motivated experiment concerned with anomalies in the perception of probabilities is reported in Camerer (1987).  

110. An initial experiment in this direction is that of Loomes, Starmer, and Sugden (1989).  

111. Interestingly, Marshall, Knetsch, and Sinden (1986) report that individuals exhibit a much smaller buying/selling price disparity when they are asked to act as an agent for someone else than when they are acting on their own behalf.  

112. In Brookshire and Coursey (1987) go on to compare different methods of eliciting values for public goods and report a similar decrease in price discrepancies elicited from a repeated market-like elicitation procedure as compared to data elicited in a hypothetical survey.  

113. The gap is not really between the disciplines of psychology and economics, but between what levels of approximation seem to be useful for addressing what kinds of problems. The cognitive psychologist John Anderson has written persuasively of the need to look at optimizing models: he argues that if you want to know how people solve problems, you can often get good predictions by assuming they find optimal solutions. For example, he says (1990, 22):  

So far, I have discussed three levels of analysis: a biological level, which is real but almost inaccessible to cognitive theorizing, the approximate but essential implementation level, and the real and accessible algorithmic level. . . . The rational level of analysis offers a different cut at human behavior. . . . It is not "psychologically real", in the sense that it does not assert that any specific computation is occurring in the human head. Rather, it is an attempt to do an analysis of the criteria that these computations must achieve. . . . This turns out to be an important level at which to develop psychological theory. . . . This level of analysis . . . can tell us a lot about human behavior and the mechanisms of the mind. The function of this book is to demonstrate the usefulness of the rational level of analysis.  

114. An interesting feature of Rietz's analysis is that he compares experimental procedures that he finds successfully use binary lotteries to control risk aversion with the procedures that an earlier experiment (reported in Walker, Smith, and Cox 1990) had reported did not have much effect. Rietz attributes the difference to the fact that Walker et al. had first trained their subjects with ordinary monetary payoffs, and then simply switched to binary lottery payoffs, with what he regards as insufficient explanation. He concludes that "hysteresis resulting from switching between monetary payoffs and lottery procedures . . . hinders success"(199).  

115. Such mechanisms, which go by the name of proper scoring rules, form a literature that goes back at least as far as Brier (1950), on weather forecasters (see also Savage 1971) and has attracted the attention of experimenters testing theories that require them to assess sub-
jects' subjective probability estimates. Some experiments of this kind are discussed by Camerer in chapter 8.

116. Of course, not all probabilities can be made objective—e.g., players' subjective assessments of how other players will react, and it is for this reason that some experiments employ proper scoring rules and other devices to elicit probability estimates. And not all preferences are entirely subjective—it is for this reason that a considerable measure of control can be obtained by paying subjects in money.

Bibliography


INTRODUCTION


--------. 1987a. Laboratory experimentation in economics. In Advances in economic theory


