Laboratory Experimentation in Economics: A Methodological Overview

Alvin E. Roth

The Economic Journal, Vol. 98, No. 393 (Dec., 1988), 974-1031.

Stable URL:
http://links.jstor.org/sici?sici=0013-0133%28198812%2998%3A393%3C974%3ALEIEAM%3E2.0.CO%3B2-8

The Economic Journal is currently published by Royal Economic Society.

The Economic Journal, 98 (*December* 1988) 974–1031
Printed in Great Britain

LABORATORY EXPERIMENTATION IN ECONOMICS: A METHODOLOGICAL OVERVIEW*

Alvin E. Roth

Informal experimentation in economics goes back at least as far as Bernoulli (1738) (in connection with the Petersburg game), and formal reports of laboratory experiments as such have appeared for some time now (see e.g. Allais, 1953; Chamberlin, 1948; Flood, 1958; Friedman, 1963; May, 1954; Sauermann and Selten, 1959; Siegel and Fouraker, 1960; Smith, 1962; Stone, 1958). There are even survey articles on the subject which precede this present one by more than twenty years.[1] Nevertheless, it is probably only in the last ten or fifteen years that laboratory experimentation can be clearly seen to have truly begun its now steady and sustained transformation from an occasional curiosity into a regular means for investigating many kinds of economic phenomena.

Incidentally, when I speak of 'laboratory' experiments, I am not speaking of the location where experiments are conducted, which we will see may be in a casino or an Indian village as well as at a university. Rather I am speaking of experiments in which the economic environment is very fully under the control of the experimenter, who also has relatively unimpeded access to the experimental subjects. This distinguishes laboratory experiments from 'field' experiments, in which relatively few aspects of the environment can be controlled, and in which only limited access to most of the economic agents may be available.[2] It is precisely this control of the environment, and access to the agents (sufficient to observe and measure attributes that are not controlled) that give laboratory experiments their power.

The volume and variety of contemporary experimental work is now sufficiently large that, in a survey of (even) this size, I would be unable to represent it all adequately. Instead, I will aim for the more modest,

* I have enjoyed many helpful discussions about this paper with Jack Ochs and Emilie Roth, both of whom have cheerfully read various preliminary drafts, and suggested many improvements. I have also received useful comments and references to the literature from John Kagel and Graham Loomes, and from John Hey and Andrew Oswald. Any errors of omission, commission, or confusion which have nevertheless made their way into the paper are of course my own responsibility. This work has been partially supported by grants from the National Science Foundation, the Office of Naval Research, and the Alfred P. Sloan Foundation.

[1] E.g. Cyert and Lave (1965) extolled the promise of controlled experimentation (and excoriated theorists) in a review of experiments bearing on oligopolistic collusion. Friedman (1969) gave a thoughtful review (still well worth reading today) of several experiments, including his own, concerned with oligopolistic competition. Interestingly, his paper appeared in a special 'Symposium on Experimental Economics' that also included reports of experiments by Hogatt (1969), MacCrimmon and Toda (1969), Sherman (1969), Carlson and O'Keefe (1969), and Cummings and Harnett (1969), and which is probably as good an indicator as any that experimentation had by that time already begun to draw sustained attention.

[2] A discussion of field experiments is well beyond the scope of this survey, as they raise some very different methodological issues (cf. Ferber and Hirsch, 1982 or Hausman and Wise, 1985).

methodological goal of illustrating how economists can use the tools they find in the laboratory to make steady, incremental progress on answering questions that would otherwise be intractable.

For this reason my survey will focus on *sets* of experiments which together tell us more than any one of them. By 'more' I do not mean that these sets of experiments permit us to draw *broader* conclusions than might have seemed warranted on the basis of one experiment alone. While this will sometimes be so, it can also be that subsequent experiments serve to define more narrowly the conditions under which some phenomenon initially observed can be expected to occur, or even that subsequent experiments will cause the results of an initial experiment to be entirely reinterpreted. What I want to 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.

The first four sections of this paper recount several series of experiments, chosen to illustrate both some of the different areas of inquiry to which laboratory experimentation has been applied in economics, and some of the important methodological themes around which debates among experimenters sometimes revolve. Each of the areas I have chosen is sufficiently complex to allow room for experts to differ in their assessment of the nature and causes of some of the principle phenomena encountered. In each of the series of experiments I will describe, initial experiments serve to focus the investigation, and subsequent experiments help to identify potentially critical variables. When we get to the present in each of these series, there is still room for experts to differ, but the room for disagreement has been considerably narrowed, and the nature of remaining disagreements has been clarified. This is, I think, typical of what we can expect of experiments.

Section I concerns several series of experiments investigating two-person bargaining. Among the methodological themes which these experiments will introduce is the importance, and the difficulty, both of creating an experimental environment in which theories being tested give unambiguous predictions, and of controlling or measuring the preferences of experimental subjects.

Section II concerns experiments investigating both the free rider problem and the prisoner's dilemma. The early experiments concerned with each problem concentrated on single decisions, and the later experiments concentrated on repeated decisions, shifting the debate to the role of experience in each case.

Section III concentrates on a controversy that arose over certain kinds of conjectured auction behaviour. While the controversy arose out of differing interpretations of field data, experimental methods proved useful in studying the nature of the phenomenon, and have suggested that the problem can be studied in field data by observing certain qualitative relationships identified in the laboratory.

Section IV concerns experiments involving individual choice behaviour, and focuses on one of many anomalies, from the point of view of expected utility theory, that have been experimentally observed by both economists and psychologists. Even more than the other sections, the experiments discussed

here illustrate the extent to which the interpretation of experimental results is influenced by the theoretical predispositions of the investigators.

Section V attempts to tie together some of the methodological themes that arise in the course of the paper, by means of some questions and answers, and Section VI concludes.

Let me close this introduction by apologising to the reader for the fact that I have found it necessary to include a reasonable amount of detail in my description of at least some of the experiments described here. While the details are obviously important in all kinds of work (and good surveys nevertheless manage, mercifully, to leave them out), I think a survey of experiments would be critically incomplete if it failed to at least suggest how very much the details of experimental design are at the heart of how experimental results are interpreted, and sometimes contested. The reader unfamiliar with experimentation may derive what comfort he can from knowing that my experimental colleagues will criticise me for the many crucial details of each experiment (as well as for the many important experiments) that I have omitted.

I. BARGAINING BEHAVIOUR

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 (i.e. in the tradition of non-cooperative game theory) the problem is that all of this typically large set of outcomes can be achieved as equilibria of the game. Consequently, theories of bargaining that seek to make stronger predictions have attempted to make 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. While 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 an increasing convergence between the theoretical literature concerned with strategic and cooperative models of bargaining (see e.g. Sutton, 1986), the bargaining environments for which their predictions can

be most clearly derived are rather different. Section I.1 will therefore be concerned with experimental tests of cooperative models, and section I.2 will take up a more recent experimental literature concerned with strategic models.

I.1 *Experimental tests of axiomatic models.*[3]

Perhaps 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 δ results. Nash modelled 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 α in A, and $d = (d_1,d_2)$ corresponds to the utility payoffs to the players from the disagreement alternative δ. That is, 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 such a theory was essentially untestable.[4] To get around the difficulty, the earliest experiments [5] 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 game theorists, 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

[3] Some of the descriptive material in this section is adapted from the generally more detailed presentation in Roth (1987*a*).

[4] For example, Morley and Stephenson (1977) state 'these theories... do not have any obvious behavioral implications' (p. 86).

[5] Which are reviewed in Roth and Malouf (1979).

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.[6]

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 λ_i or a small prize σ_i (with $\lambda_i > \sigma_i$). The players bargain over the distribution of 'lottery tickets' that determine the probability of receiving the large prize: e.g. an agent i who receives 40% of the lottery tickets has a 40% chance of receiving λ_i and a 60% chance of receiving σ_i. Players who do not reach agreement in the allotted time each receive σ_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 normalising 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 the amount λ_i, i.e. equal to the percentage of lottery tickets he has received. Thus in a binary lottery game, the pair (S,d) which determines the prediction of Nash's theory is precisely equal to the set of feasible divisions of the lottery tickets.[7]

Note that the set of feasible utility payoffs to the players of a binary lottery game is insensitive to the magnitudes of λ_i and σ_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, that 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 determine whether or not changes in the size of the prizes, and whether the bargainers knew one anothers' prizes, influenced the outcome. All

[6] 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.

[7] Note that no assumptions have been made here about the behaviour of the experimental subjects themselves in binary lottery games. (That is, the subjects might not be utility maximisers (see Section IV), or they might have preferences over distributions of payoffs to both players, rather than over their own monetary payoffs (see Sections I.2 and I.3). What binary lottery games do allow us to know is the utility of utility maximisers 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 analyse bargaining data by inferring *ex post* what the utility of the bargainers might have been.

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 with whom they were bargaining. 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% of the lottery tickets. In the full information condition, in those games in which the bargainers' prizes were unequal, agreements tended to cluster around two 'focal points': the equal probability agreement, and the 'equal expected value' agreement 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.

Subsequent experiments (Roth *et al.*, 1981; Roth and Malouf, 1982; Roth and Murnighan, 1982; Roth and Schoumaker, 1983) were conducted to verify that these results were not artifacts of the particular experimental procedures, and to explore further the unpredicted effects of information in bargaining that the data from Roth and Malouf (1979) so clearly revealed. Because I have recently discussed most of these experiments in Roth (1987*a*), I will not describe them again here. But for the purposes of this paper, three further points are worth making.

The first is that 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).

The second point is methodological. All these experiments make critical use of the control that laboratory environments allow. For example, in Roth and Murnighan (1982), one of the eight cells of the experiment involved a binary lottery game, with prizes $\lambda_1 = \$20$, $\lambda_2 = \$5$, and $\sigma_1 = \sigma_2 = 0$, in which both players were aware that the \$5 player knew the size of both players' prizes and the \$20 player knew only his own prizes. Another cell of the experiment

involved different players playing the same game, again with the $5 player knowing both players' prizes and the $20 player knowing only his own, except in this cell each player was uninformed about what the other player knew. (In both cells, players were free to make any statements they wished about what their prizes were, and about what they knew). This experimental manipulation, which did not alter what the bargainers knew about each other's prizes, but only what they knew about what the *other* knew, was observed to have a significant effect on the frequency of disagreement. (The highest frequency of disagreement in the experiment, 33%, was observed in the second of the two cells described above, in contrast, for example, to a 19% frequency of disagreement in the first of those cells.) It is hard even to 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.[8] Laboratory experiments give us what has at least so far proven to be the only direct way to investigate them.

Note finally that the regularities among these unpredicted effects of information make it unlikely that they can be attributed primarily to mistaken or irrational behaviour on the parts of the bargainers. For example, in the four cells of the above experiment in which the bargainers 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 what would be expected at equilibrium, in that the increase in the number of disagreements almost exactly offsets the increased share obtained by the $5 player when agreements are reached.

These experiments 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 observed behaviour. As such, the experimental results demonstrate serious shortcomings of the theories. However, in order to evaluate a theory fully, we also need to test the predictions it makes about those variables it predicts *are* important. 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 *et al.* (1981), and Roth and Rothblum (1982). A broad class of apparently quite different models, including all the standard axiomatic models, 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 *et al.*

[8] For example, just such distinctions are made in the theoretical literature on reputations, such as that growing out of Selten's (1978) famous example of the 'chain store paradox'. See for instance the papers by Kreps and Wilson (1982) and Milgrom and Roberts (1982), or the general review by Wilson (1985).

(1988). Whereas binary lottery games were employed in the earlier experiments precisely in order to control out the individual variation due to differences in risk posture, these studies employed *ternary* lottery games having three possible payoffs for each bargainer i. These are large and small prizes λ_i and σ_i obtained by lottery when agreement is reached, and a disagreement prize δ_i obtained when no agreement is reached in the allotted time. (In the 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 λ_i = \$16 and σ_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 favourable 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% of the lottery tickets when $\delta > \sigma$, and less than 50% 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 anothers' 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 *et al.* (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.[9] 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. 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.[10]

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

[9] That is, it was thought that the high concentration of agreements around a 'focal point' such as (50%, 50%) might reflect forces at work which 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.

[10] 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 (1988), Radner and Schotter (1987), and Cox *et al.* (1985*a*) also observe only small differences between certain observations made with and without binary lottery games. (Cox *et al.* take the position that, if one is confident that all experimental observations conform to some theoretical prediction stated in terms of expected utility functions, then the risk posture of bargainers can be estimated from experimental results without any need to directly measure or control for differences in subjects' risk posture. In the context of an auction experiment they further argue that, in this way, the results of binary lottery games can be used to detect violations of expected utility theory (on which see Section IV), by using observed results to estimate what agents' utilities would have to have been to make these results conform to the theoretical prediction, and comparing these estimated utilities to the risk neutrality that utility maximisers exhibit on binary lottery games.) Baiman and Lewis (1988) report in passing a direct test to see if subjects are indifferent between equivalent compound and simple binary lotteries (as utility maximisers would be), and based on somewhat ambiguous evidence they tentatively conclude that the binary lottery procedure does induce risk neutral behaviour in their subjects. (Their paper also contains an exceptionally clear exposition of their methodological concerns.)

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 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. While in principle this presents no obstacle to experimental investigation (just conduct experiments with very large prizes), in practice, experimental budgets always make it likely that payoffs in the laboratory will be smaller than those in some 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 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'. (In all the experiments discussed above, there was a fixed time limit, typically from 9 to 12 minutes, by which time any agreements must be concluded. Three minutes before the deadline, a clock came on the screen). 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 (see Roth *et al.*, 1988). The methodological point I want to make about these two regularities is that the experiments reported above, which were designed to permit the effects of particular manipulations to be observed, do not cast much direct light on either of these phenomena, other than to indicate that they are easy to observe.[11] Convincing evidence that particular variables are of importance in producing deadline effects, for example, must come from experiments in which the concentration of agreements near the deadline can be shown to respond to changes in those variables.

In closing, I should remark that while I have emphasised in this section the way theoretical predictions can be tested experimentally, experimental results also suggest theoretical directions. The clearest and most often replicated result discussed above is that information about the underlying commodities influences the outcome of bargaining in a way precluded by theories stated only in terms of bargainers' utilities. This suggests we should examine bargaining theories in which information about commodities can play a role, and there have been a number of recent attempts in this direction, as well as renewed interest in earlier theories that can be reinterpreted in light of the evidence (See Roth (1979) for some of these earlier theories which allow comparisons between bargainers. For one of the most recent, see Roemer (1988).) These theories too yield testable predictions, and experiments have begun to help distinguish among them (see e.g. Roth and Malouf, 1982).

[11] Some other experiments cast some light on factors that may contribute to disagreements. Malouf and Roth (1981) look at variations in the set of feasible outcomes, and Coursey (1982) compares single play bargaining under loose time constraints with repeated bargaining under tight time constraints.

I.2 *Experimental Tests of Strategic Models*

Recently a good deal of attention has been given to models of bargaining in which two bargainers, 1 and 2, alternate making offers over how to divide some amount k (of money). Time is divided into periods, and in even numbered periods t (starting at an initial period $t = 0$) player 1 may propose to player 2 any division $(x, k-x)$. If player 2 accepts this proposal the game ends and player 1 receives a utility of $(\delta_1)^t x$ and 2 receives a utility $(\delta_2)^t(k-x)$, where δ_i is a number between 0 and 1 reflecting player i's cost of delay. (That is, a payoff of y dollars to player i at period t gives him the same utility as a payoff of $\delta_i y$ dollars one period earlier, at period $t-1$.) If player 2 does not accept the offer, and if t is not the final period of the game, then the game proceeds to period $t+1$, and the roles of the two players are reversed. If an offer made in the last period of the game is refused, then the game ends with each player receiving 0. A game with a maximum number of periods T will be called a T-period game.[12]

A *subgame perfect equilibrium* can be computed by working backward from the last period. An offer made in period T is an ultimatum, and so at such an equilibrium player i (who will receive 0 if he rejects the offer) will accept any non-negative offer. So at a subgame perfect equilibrium, player j, who gets to make the proposal in period T, will receive 100% (if payoffs are continuously divisible) of the amount k to be divided, if the game continues to period T. Consequently at period $T-1$ player j will refuse any offer of less than $(\delta_j)k$ but accept any offer of more, so that at equilibrium player i receives the share $k-(\delta_j)k$ if the game goes to period $T-1$, and so forth. Working back to period 0, we can compute the equilibrium division: i.e. the amount that the theory predicts player 1 should offer to player 2 at period 0, and player 2 should accept. (When payoffs are continuous this equilibrium division is unique.)

Recent experimental studies of this kind of bargaining have reported markedly different results. Their authors have drawn quite different conclusions about the predictive value of perfect equilibrium models of bargaining, and about the role that experience, limited foresight, or bargainers' beliefs about fairness might play in explaining their observations. (Questions of fairness arise because in some of these experiments many observed agreements give both bargainers 50% of the available money.) In reviewing these experiments here, my aim is to show how, even in the earliest stages of a programme of experimental research, when there is room for substantial disagreement about what is being observed, early experiments suggest later ones, subsequent experimental results suggest reinterpretations of earlier ones, and the process of experimentation offers the prospect of steadily (if somewhat slowly) narrowing the areas of potential disagreement.

[12] Much of the recent theoretical work using this kind of model follows the treatment by Rubinstein (1982) of the infinite horizon case. An exploration of various aspects of the finite horizon case is given by Stahl (1972). This literature considers the cost of delay in more general form than only the discounting discussed here. An experiment motivated by this literature which considers a fixed cost per period of bargaining is Rapoport *et al.* (1988). An experiment with a similar cost structure, motivated by earlier theoretical attempts to deal with costs of delay, is Contini (1968).

In each of the following experiments, the predictions tested involved only the ordinal utilities of the bargainers, not their risk posture. Following standard practice in the experimental literature when only ordinal utilities are of concern, the utility of the bargainers was assumed to be measured by the amount of money they receive.

Guth *et al.* (1982) examined one-period ('ultimatum') bargaining games. Player 1 could propose dividing a fixed sum of k Deutsche Marks any way he chose, by filling out a form saying 'I demand DM x'. Player 2 could either accept, in which case player 1 received x and player 2 got $k-x$, or he could reject, in which case each player received 0 for that game.

The perfect equilibrium prediction for such games is that player 1 will ask for and get (essentially) 100% of k. However the average demand that players 1 were observed to make was for under 70%, both for players playing the game for the first time and for those repeating the game a week later. About 20% of offers were rejected. The authors conclude that '...subjects often rely on what they consider a fair or justified result. Furthermore, the ultimatum aspect cannot be completely exploited since subjects do not hesitate to punish if their opponent asks for "too much".'

Binmore *et al.* (1985) write: 'The work of Guth *et al.* seems to preclude a predictive role for game theory insofar as bargaining behaviour is concerned. Our purpose in this note is to report briefly on an experiment that shows that this conclusion is unwarranted...'[13] Their experiment studied a 2-period bargaining game, in which player 1 makes a proposal of the form $(x, 100-x)$ to divide 100 pence. If player 2 accepts, this is the result. Otherwise 2 makes a proposal $(x', 25-x')$ to divide 25 pence. If player 1 accepts, this is the result, otherwise each player receives 0. Thus in this game $\delta_1 = \delta_2 = 0{\cdot}25$, and (since proposals are constrained to be an integer number of pence) at any subgame perfect equilibrium player 1 makes an opening demand in the range 74–6 pence, and player 2 accepts any opening demand of 74 pence or less. Subjects played a single game, after which player 2 was invited to play the game again, as player 1. In fact there was no player 2 in this second game, so only the opening demand was observed.

The modal first demand in the first game was 50 pence, and 15% of the first offers were rejected. In the second game (in which only first demands were observed), there was a mode around a first demand near 75 pence. There was thus a clear shift between the two distributions of first demands, in the direction of the equilibrium demand. The authors conclude 'Our suspicion is that the one-stage ultimatum game is a rather special case, from which it is dangerous to draw general conclusions. In the ultimatum game, the first player might be dissuaded from making an opening demand at, or close to, the "optimum" level, because his opponent would then incur a negligible cost in making an

[13] They add: 'This does not mean that our results are inconsistent with those of Guth *et al.* Under similar conditions, we obtain similar results. Moreover our full results would seem to refute the more obvious rationalizations of the behavior observed by Guth *et al.* as "optimising with complex motivations". Instead, our results indicate that this behaviour is not stable in the sense that it can be easily displaced by simple optimizing behavior, once small changes are made in the playing conditions.'

"irrational" rejection. In the two-stage game, these considerations are postponed to the second stage, and so their impact is attenuated.'

Guth and Tietz (1987) responded with an experiment examining two two-stage games with discount factors of 0·9 and 0·1 respectively. So the subgame perfect equilibrium predictions (in percentage terms) for the two cases are (10%, 90%) and (90%, 10%) respectively. They say 'Our hypothesis is that the consistency of experimental observations and game theoretic predictions observed by Binmore *et al.* as well as by Fouraker and Siegel is solely due to the moderate relation of equilibrium payoffs which makes the game theoretic solution socially more acceptable.' Subjects played one of the two games twice, each with a randomly chosen other bargainer. Subjects who played the first game as player 1 played the second game as player 2. One difference from the sequential bargaining games discussed above was that disagreement automatically resulted if player 2 rejected an offer from player 1 but made a counterproposal that would give him less than player 1 had offered him.[14]

In the first game, the average first demand in games with a discount factor of 0·1 was 76%, and in the second game 67%. For games with a discount factor of 0·9, the average first demand in the first game was 70%, and in the second game 59%. (Recall that when the discount factor is 0·9, the equilibrium first demand is only 10%.) The authors conclude 'Our main result is that contrary to Binmore, Shaked and Sutton "gamesmenship" is clearly rejected, i.e., the game theoretic solution has nearly no predictive power.'

Neelin *et al.* (1988) also responded to Binmore *et al.* (1985). They reported two experiments involving 2-period, 3-period, and 5-period bargaining games. Neelin *et al.* observe that the data for all their (2, 3, and 5 period) games are near the perfect equilibrium prediction for 2 period games. They conclude '...the strong regularity of the behaviour we observed is one of the most noteworthy aspects of our results and lends power to our rejection of both the Stahl/Rubinstein theory and the equal-split model'.[15]

Following most of this exchange, Ochs and Roth (1988) noted that the prior analyses had focused on the accuracy of the perfect equilibrium as a point predictor, i.e. on whether the observed outcomes were distributed around the perfect equilibrium division or around some other division of the available money. Their experiment was designed to test the predictive accuracy of some of the *qualitative* predictions of the perfect equilibrium in sequential bargaining, and was designed to detect whether changes in the parameters of the game influence the observed outcomes in the predicted direction, even in the case that there might be a systematic error in the point predictions (recall the discussion of risk aversion in Section I.1). To this end the experiment was

[14] Note that this rule makes the games more like ultimatum games, since some demands of player 1 (e.g. demands of less than 90% in games with discount factor of 0·1) can only be rejected at the cost of disagreement.

[15] In a reply, Binmore *et al.* (1988*a*) decline to attribute the same significance to these results, and conjecture that the various differences described among these experiments may be due to the various differences in experimental procedures employed, which, they suggest, may make the theoretical model more appropriate for their experiment than for some of the others. (A related view is expressed by Harrison and McCabe (1988).)

implemented in a way that allowed the discount factors of the two bargainers to be varied independently.[16] In order to compare games like those considered in the earlier experiments, the experimental design allowed comparisons between different combinations of discount factors for games of fixed length, as well as between games of different length for given discount factors. The eight cells of the experiment compare two and three period games using all four combinations of discount factors (δ_1,δ_2), with δ_i equal to 0·4 or 0·6.

Overall, although the data reveal some striking regularities, the perfect equilibrium predictions do poorly both as point predictions and in predicting qualitative differences between cells, such as mean first period offers. (The authors write '...we can just barely reject the null hypothesis that, as a predictor of the direction of differences in pairwise comparisons of means, the theory does no better than coin flipping.') While parts of the data appear to be consistent with similar observations made in the earlier experiments, larger experimental designs allow more comparisons to be made, so that observations which, piecewise, appear contradictory, emerge as part of a larger picture.[17] Perhaps the most interesting part of this picture, for our present purposes, concerns what happens when first period offers are rejected, both in this experiment and, as it turns out, in the previous experiments.

Briefly, approximately 15% of first offers met with rejection (including those in games with experienced subjects who had played ten games against different opponents), and of these well over half were followed by counterproposals in which player 2 demanded *less* cash than he had been offered. That is, a significant number of players 2 were rejecting small shares of the relatively large gains available in the first period in favour of large shares of the much smaller gains available in the second period. Since, after player 1 has made a proposal, player 2 is faced with an individual choice problem, we can conclude by revealed preference that these player 2's utility is *not* measured by their monetary payoff, but must include some non-monetary component. When the data of the previous experiments were reanalysed with this in mind, it turned out that this pattern of rejections and counterproposals was strikingly similar in all of these experiments.[18]

[16] Each of the earlier experiments was designed to correspond to the case that the players have equal discount factors, i.e. $\delta_1 = \delta_2 = \delta$, with the costliness of delay implemented by making the amount of money being divided in period $t+1$ equal to δ times the amount available at period t. Since half the cells of the experimental design of Ochs and Roth require different discount rates for the two bargainers, the discounting could not be implemented in this way. Instead, in each period, the commodity to be divided consisted of 100 'chips'. In period 1 of each game, each chip was worth \$0.30 to each bargainer. In period 2, each chip was worth δ_1(\$0.30) to player 1 and δ_2(\$0.30) to player 2, and in period 3 of the three period games each chip was worth $(\delta_1)^2$ (\$0.30) and $(\delta_2)^2$ (\$0.30) respectively. That is, the rate at which subjects were paid for each of the 100 chips that they might receive depended on their discount rate and the period in which agreement was reached.

[17] In this regard, the paper notes '...if we had looked only at Cell 1 our conclusions might have been similar to those of Binmore *et al.*, since the data for that cell looks as if after one or two periods of experience, the players settle down to perfect equilibrium proposals... And if we had looked only at Cells 1 and 5, our conclusions might have been similar to those of Neelin *et al.*, since in those two cells both the two and three period games yield observations near the two period predictions... And if we had looked only at cells 5 and 6, we might have concluded, like Guth and Teitz, that the phenomena observed here was closely related to the relatively extreme equilibrium predictions in those cells.'

[18] When the necessary data from these earlier experiments were not contained in published accounts, they

Ochs and Roth (1988) go on to argue that this and other patterns in the data can plausibly be explained if the unobserved and uncontrolled components of utility in these experiments have to do with subjects' perceptions of 'fairness', which involve comparing their share of the available wealth to that of the other bargainer. They note that in most cases agents propose divisions that give them more than half of the proceeds, and say '... we do not conclude that players "try to be fair". It is enough to suppose that they try to estimate the utilities of the player they are bargaining with, and ... at least some agents incorporate distributional considerations in their utility functions.' That is, if agents' preferences are such that they will refuse 'insultingly low' offers, then this must be taken into account in making offers.

Note that uncontrolled elements in the bargainers' utility in these experiments suggests that none of them can be easily interpreted as tests of perfect equilibrium *per se*, since to compute a perfect equilibrium we need to know the preferences of the players (and so do they).[19] But the uniformity with which 'disadvantageous counterproposals' have appeared in the experiments to date, in contrast to their otherwise quite varied results, suggests that bargaining may be an activity that systematically gives bargainers motivations distinct from simple income maximisation. One natural direction in which to continue this (still young) series of experiments is to attempt to observe directly or manipulate these so far uncontrolled motivating factors.

I.3 *Bargaining as a Complex 'Social' Phenomenon*

To the extent that these other motivations may reflect some element of bargainers' perceptions of 'fairness', this may help explain why the results of these various experiments may have been more sensitive than might have been expected to details of the experimental environment. Studies of fairness (based on survey questions) suggest that peoples' ideas about what is 'fair' may be both clear and very labile, subject to dramatic change in response to how the issue is presented. (This is particularly clear in the study reported by Yaari and Bar-Hillel (1984). See also the related work of Kahneman *et al.* (1986*a*; *b*) and of Bazerman (1985) and Farber and Bazerman (1986).[20]) So variations in experimental procedures may have had unanticipated effects.

---

were readily available from the working papers circulated by the authors. That this is a good experimental practice cannot be overemphasised, since the easy availability of data permits just these sorts of comparisons.

[19] However Ochs and Roth (1988) do report consistency across subgames, which could be interpreted as indirect evidence supporting the subgame perfectness hypothesis with respect to the unobserved preferences.

[20] In the psychology literature, some ideas concerned with beliefs about fairness have been explored under the name of 'equity theory'. The interpretation of various experimental results in this literature is that the outcome of various kinds of interactions can best be understood as reflecting the common beliefs among the participants about what constitutes a fair outcome. I think that one of the contributions of the experimental bargaining results is that they reveal that subjects may possess multiple, different notions of fairness, and employ them selectively, for strategic purposes. Thus, for example, the agreements reported in the binary lottery games of Roth and Murnighan (1982) had modes at both the equal probability and the equal expected value agreements, and the transcripts of the bargaining reveal that notions of 'fairness' were employed by both bargainers in arguing for their claims. But the bargainer who saw the equal expected value agreement as 'fair' was invariably the one with the smaller ($5) prize, while the player with the larger ($20) prize was the champion of the fairness of equal division of the lottery tickets.

Just such effects were revealed in a rather different bargaining environment in experiments conducted by Hoffman and Spitzer (1982; 1985). In one of the experiments in their 1982 paper pairs of subjects were asked to agree on how to divide up to $14, in face to face negotiations. However if no agreement was reached one of them (the 'controller', chosen just before negotiations began by the toss of a coin) could simply choose an outcome that would give him up to $12 and the other bargainer nothing. When bargainers negotiated with each other twice under these conditions, twelve out of twelve agreed to split the proceeds equally, so that the controller settled for a smaller cash payoff than he could have obtained unilaterally. In their 1985 paper, Hoffman and Spitzer report that in similar experiments in which the position of 'controller' was allocated to the winner of a simple game, and in which the instructions to the participants gave the controller 'moral authority' to claim his prize unilaterally, the frequency of equal splits was reduced.[21]

Now, we should hesitate to draw too close a parallel between these differences in outcomes caused by differences in experimental environments and those that may have occurred in other experiments, because there is reason to believe that face to face negotiations for relatively small prizes may be a rather special situation. In particular, face to face interactions call into play all of the social training we are endowed with, and may make it unusually difficult to control preferences. (Ask yourself if you would agree to be very rude to the next stranger you meet at a party if I offer to pay you $5.) There is now a small body of comparable experiments that make it possible to begin to assess how results of face to face negotiations may differ from results obtained in more anonymous laboratory negotiations. Such comparisons will allow us to assess some of the potential sensitivity of reported results to experimental procedures.

First, it seems that the frequency of disagreement is sharply less in face to face negotiations than in anonymous negotiations concerning comparable prizes. Aside from the general results of the experiments cited above to this effect (which involve comparisons which must be treated cautiously, since they involve other differences than whether bargaining was face to face), a reduced frequency of disagreement in face to face bargaining is reported by Radner and Schotter (1987), who report a 'within experiment' comparison involving just this variable.[22] Regarding the frequency of equal divisions when one of the bargainers has an 'outside option' worth more than half the total proceeds, Binmore *et al.* (1988*b*) report an experiment roughly parallel to that of Hoffman and Spitzer (1982) but involving anonymous negotiations, and find a preponderance of proposals that give the player with the outside option as much as he could have got unilaterally. And in another pair of closely parallel experiments Roth and Malouf (1982) report fewer equal splits in anonymous

[21] For some related results which look at other variations in experimental procedures, see Harrison and McKee (1985). For a subsequent experiment on a somewhat more complicated environment (related to the one explored by Plott (1983)), see Harrison *et al.* (1987).

[22] Radner and Schotter's experiment concerns the efficiency of bargaining between a buyer and seller neither of whom know the other's reservation price. For some related experiments concerning bargaining with incomplete information, see Hogatt *et al.* (1978) and Forsythe *et al.* (1987).

bargaining than are observed by Nydegger and Owen (1975) in face to face bargaining. Again, 'between experiment' comparisons have to be treated with caution. But they give us a strong indication that the difference between face to face and anonymous interaction that has been reliably reported in multi-person negotiations (see e.g. Murnighan, 1985) is also liable to account for significant differences among experiments of the kinds considered here.

As I have indicated, my own suspicion is that these differences in results are due to the fact that bargainers' incentives in the more social, face to face environment may correspond less closely to their monetary payoffs than when bargaining is anonymous. However this is not the only possible explanation. Face to face bargaining also allows many more channels of communication (e.g. tone of voice and facial expression) than does anonymous bargaining in which subjects may be restricted to written messages, and perhaps the differences are due to this. The difference between these two hypotheses could be important, for example in judging the claim that some of the face to face evidence should be seen as supporting policy implications of the Coase Theorem (Coase, 1960), which rest on the presumed efficiency of bargaining.[23]

In conclusion, the evidence to date suggests that bargaining is a complex social phenomenon, so that special care must be taken in designing and interpreting bargaining experiments. In addition to refining theories of bargaining stated in terms of abstract preferences, the evidence suggests that it will be worth investigating what (non-monetary) motivations may be engendered by bargaining itself. The extent to which such non-monetary motivations would remain important if the bargaining concerned much larger payoffs is of course also an open question, but I see no obvious reason to jump to the conclusion that some of the very consistent patterns of behaviour discussed in this section would disappear as the stakes become large, particularly when they become large for both bargainers. Also, there is evidence that the 'social' aspects of bargaining involve more than just the incentives of the players, but also their expectations and their ability to coordinate with one another.[24] Finally, this conclusion need not be 'bad news' for the prospect of developing more descriptive theories of bargaining: the very regularity of the behaviour, and the fact that bargainers often appear to be able

[23] If the absence of disagreement is due to the lack of control of subjects' incentives, we might suspect it would diminish as the monetary consequences become larger. But if it is due to having many channels of communication, then we might investigate whether similarly effective channels might be available to bargainers in situations of the sort to which the Coase Theorem is typically thought to apply. Thus the experimental results suggest some different directions to investigate the efficiency assumption of the Coase Theorem than those which have surfaced from other considerations (see e.g. Farrell, 1987).

[24] Roth *et al.* (1981) show that strategically equivalent games can produce different results, and argue that this may be due to the existence of social conventions which contribute to determining the credibility of different bargaining positions, and which (thus) help bargainers coordinate their expectations. (Roth and Schoumaker (1983) report an experiment which manipulates the expectations of the bargainers, and provides some indirect support for this view.) To the extent that bargaining involves elements of coordination, a better understanding of the problems of pure coordination will likely be of help in developing theories of bargaining. In this regard I have been glad to notice a number of recent experiments concerned with coordination, such as Cooper *et al.* (1987), Ochs (1988), and Van Huyck *et al.* (1987).

to anticipate it and turn it to strategic use, suggest that theorists should also be able to do so.

II. FREE RIDERS, AND PERPLEXED PRISONERS

This section is in part a story of a dog that didn't bark, at least not initially, or at least not as loudly as might have been expected. It concerns two closely related problems, both of which have captured the imaginations of theorists and experimenters. Both the free rider problem and the prisoner's dilemma concern the potential difficulties in achieving gains from cooperation when those gains must be allocated in a fixed way. For both problems, there were theoretical formulations predicting that virtually none of such potential gains could be achieved, and different opinions about how real this predicted predicament was. Early experiments concerned with both problems did not observe the phenomenon, at least not in its severest form, and later experiments did, at least to a degree.

Curiously, the experimental literatures concerned with the two problems seldom refer to each other. Although my discussion of both literatures will be necessarily brief, I hope that some of the parallels between them will become apparent.

II.1 *The Free Rider Problem*

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 recognise 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.

The experiment of 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 programme by two well known comedians. They were told they would see the programme 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.[25] 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 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 pedalling 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 whether they would all receive credit for participating in the experiment, which depended on how brightly the light remained lit. Scherr and Babb (1975) compare 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

[25] However he notes that a sixth group of subjects who were asked in a purely hypothetical way how much such a programme would be worth to them gave significantly different responses from the other five groups. He says (p. 125) '... this result may be seen as still another reason to doubt the usefulness of responses to hypothetical questions...'

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 colourful story (in the midst of so many dry recitals), 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.[26] 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'.) 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 others from which bids had 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 SFr.4,200, 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.[27] The authors note that the second bids were less than the first, but not by much.[28] They conclude (p. 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

[26] Experiments with human subjects in the United States are now regulated by State and Federal laws which 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.

[27] 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 (p. 696) '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...)

[28] There are some complexities in the data, since 10 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.

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 which itself induces strong preferences among the agents, their preferences can be equated with their monetary payoffs.[29] 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 Lindhal equilibrium.

Smith (1979*a*, *b*; 1980) reports three such experiments.[30] 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 which 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 Lindhal 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 (1979*a*) 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 mechanism too, Lindhal prices and quantities were observed to be good predictors for the market parameters considered.

In the light of these results, Smith suggests that the results of Bohm's (1971) 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 (1979*b*) 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

[29] 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.

[30] Another interesting experiment using this general design is that of Ferejohn *et al.* (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 purchase collectively.

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.

While the variable of primary concern in these experiments was the decision mechanism, in the design of an experiment of this complexity it is clear that many essentially arbitrary choices have to be made, and are held constant over the course of the experiment. Smith (1980) reports a subsequent examination of the Auction mechanism with different functions assigning subjects' values to the public good (and allowing for income effects). In these markets, a substantial amount of free riding was observed. So these results suggest it may be fruitful to study the way decision mechanisms interact with demand parameters for public goods.

Of course, different theoretical dispositions suggest different regularities in the data. For example Smith (1980) reports in connection with his experiment that (p. 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. They noted that previous studies examined fairly small groups (mostly of fewer than ten individuals), and 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%, with some indication of a decrease when the stakes were raised.[31] Among the few exceptions they noted was that a group of first semester economics graduate students only invested 20% in the public good, leading them to suggest that economists may be different from everyone else (and hence the title of their paper).[32]

[31] One feature of the procedures in this study which differed from the studies so far discussed is that subjects knew they would be required to explain their decisions to the experimenter (cf p. 297).

[32] 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),

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 *et al.* (1985) and Banks *et al.* (1988) give subjects the chance to gain some experience with how the mechanisms work.

Isaac *et al.* seek to show that the free rider problem is alive and well, by examining a mechanism already suspected of being favourable 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.' (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.[33] The results of the experiment, like that of Isaac *et al.*, 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 *et al.* (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

---

I suspect that the authors do not take this result as seriously as some of the other 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. See sections III and V on this point.

[33] Since in fact there were only 5 subjects, payoffs were based on calculating the total contributions to the fund as if each subject represented 20, with some modifications designed to conceal from the subjects how few of them there were.

setting. They note that (p. 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 attribute these results confidently merely to increased experience with the mechanisms.[34] 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 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 the various related questions about the free rider problem in natural environments. (But, as we will see in Section III.1 in connection with a different problem, such gaps between experimental and field data need not remain unbridgeable.)

II.2 *The Prisoner's Dilemma*

The now famous story which gives the prisoner's dilemma its name was apparently first told by A. W. Tucker (1950).[35] He referred to a game we can represent by the following matrix, with $b > a > c > d$.

	confess	not confess
confess	(c, c)	(b, d)
not confess	(d, b)	(a, a)

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 each receives the (non Pareto optimal) payoff of c.

The observation that equilibria could be inefficient did not strike game-theorists as odd (always assuming, of course, that the situation facing the

[34] But see Isaac *et al.* (1984) who observe some related results in a design that helps separate experience from repetition among a fixed group.

[35] See Straffin (1980) who arranged for Tucker's 1950 note to be published, and recounts a bit of its history.

players, and their preferences over the outcomes, are accurately modelled by the above matrix), rather it served to emphasise 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.[36] 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 an equilibrium behaviour. 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 from economics and psychology are Lave (1962) and Rapoport and Chammah (1965).) Typical experiments concerning the one-period game reported a level of cooperation which responded readily to various kinds of experimental manipulation but which was bounded well away from either zero or one hundred percent. A number of experiments were conducted to isolate various factors[37] contributing to the level of cooperation.

However, many experiments which were analysed 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 papers goes on to derive the conditions for equilibrium in the repeated game with a fixed probability p of continuing after each play: cooperation can

[36] The argument is that any strategy which calls for cooperation on the last period is dominated, and so we can reduce the problem to the 99 period game, and so on. Note that unremitting non-cooperation remains an equilibrium strategy in the repeated game even though it is no longer a dominant strategy.

[37] From payoffs and number of trials to personality differences: see e.g. Lave (1965) and Terhune (1968).

be achieved at equilibrium only if the probability of continuing is sufficiently large.[38]

A pilot experiment was then conducted, in large part to show that the design was feasible.[39] The payoff matrix was chosen so that cooperation was consistent with equilibrium if and only if $p \geqslant \frac{1}{3}$, and subjects played three games, with probabilities of continuing of 0·1, 0·5, and 0·9. (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% of first-period choices were cooperative. So the results remain equivocal.

Similarly equivocal results seem to be typical. A recent experiment, whose results help crystallise a lot of what I think has (in retrospect) been observed piecemeal in previous experiments, is reported by Selten and Stoecker (1986). In their experiment, subjects played 25 'supergames', each of which was a (ten-period) repeated prisoner's dilemma.[40] 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 4), 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, i.e. round 22 is the 22nd 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 4 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 result 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.[41]

The paper then develops a learning theory model in which each player is

[38] Cooperation can be achieved by some equilibrium if and only if $p \geqslant (b-a)/(b-c)$, and it can be achieved 'easily', i.e. 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 $p \geqslant (b-a)/(a-d)$ also.

[39] 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 discussed in Section I.1) were not introduced until Roth and Malouf (1979).

[40] The payoffs were $b = 1{\cdot}45$ German marks, $a = 0{\cdot}6$, $c = 0{\cdot}1$, and $d = 0{\cdot}5$. 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...])

[41] The authors caution, however (p. 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'.

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 20 rounds are then made for the pairings in the last 5 rounds. Like the observed behaviour, these predictions have cooperation unravelling from round 20 to round 25.

I think it is fair to summarise 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 behaviour mirrors the game-theoretic observation that the equilibrium recommendation is not 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 *et al.* (1982) propose a model in which players may entertain certain slight doubts about the nature of their opponent, and observe that at equilibrium there will be cooperation until near the end.[42]

In summary, as in the case of experiments concerned with the free rider problem, interest over the course of many experiments has shifted from the one-time game to the repeated game. The theory for the repeated case is further developed for the prisoner's dilemma than for the free rider problem, and the contemporary discussion proceeds on both theoretical and experimental lines.

III.3 *Experiments versus Simulations: A Methodological Digression*

Let me digress to note that one still encounters in some quarters a distressing tendency to confuse computer simulations, and the kinds of investigations one can do with them, 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 behaviour.

Perhaps the distinction can be made most clearly by considering an interesting set of computer simulations that have an unusually experimental flavour, and are reported in Axelrod (1980*a*,*b*; 1984). These have their origin in a pair of computer 'tournaments'. In the first of these, the author invited

[42] 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).

fourteen scholars in several disciplines who had written on the prisoner's dilemma to submit 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 which 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.[43] 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 number moves and tit for tat doing the opposite, which for the parameters used in the tournament was not nearly as profitable as steady cooperation.[44] 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 (1980*b*) 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 $p = 0{\cdot}99$, 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: (p. 402) '...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'.

These computer tournaments thus suggest that behaviour 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 difference between computer simulations and actual experiments. While the computer simulations which produce this result were conducted with an element of experimental flavour that is missing from conventional computer simulations (in that tournament entries were solicited from others),[45] experiments with human subjects introduce a certain amount of open-ended complexity in the form of human behaviour, that is absent from a tournament in which individuals are represented by short (or even moderately long) computer programs.

III. THE WINNER'S CURSE, AND OTHER AUCTION PHENOMENA

III.1 *The Winner's Curse*

My topic in this section is in some ways the reverse of the one in section II.1. Instead of discussing a theoretical prediction that seemed difficult to investigate

[43] It turns out that there were two 'kingmakers', i.e. two programs which largely determined how the other programs did.

[44] 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.

[45] One virtue of this, which it shares with experimental work, is to open up the process to ideas that might not have occurred to a single investigator designing a conventional computer simulation.

with field data and initially proved difficult to detect experimentally, this section discusses an *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, 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 returns on 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 scepticism by many economic theorists, particularly as the details of equilibrium behaviour 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 drilling rights auction, and so the auction price is only one of many determinants of the rate of return. So the debate continued much as before. (For some of the flavour of the argument, see Cox and Isaac (1984), Brown (1986) and Cox and Isaac (1986).)

Laboratory experiments provide an opportunity to investigate at least the basic questions associated with whether 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.[46] Subjects were also asked to write down their 90% 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 two dollars 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 20 than when it is the highest of 4, 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 behaviour, 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 (p. 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 whether 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 ϵ (which was one of the experimental variables, varying from $12 to $30).[47] If the high bid is b, the

[46] Not the coins themselves (to control for 'penny aversion').

[47] Thus private signals are 'positively affiliated' in the sense of Milgrom and Weber (1982).

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 feedback about the results. In some of the auctions 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_0 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 had some prior experience in experimental auctions.[48]

The main results for this part of the experiment are that bids were observed to be below the (risk neutral) Nash equilibrium bids. Profits were generally positive for groups of 3 or 4 bidders (at around 65 % of the equilibrium profits) and negative for groups of 6 or 7 bidders.[49] 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 which 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 *et al.* (1988) address this question in a subsequent experiment, in which the behaviour of student

[48] For a similar experiment with previously inexperienced subjects, see Kagel *et al.* (1988), who report similar results.

[49] 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.

subjects was compared with that of construction industry executives, and found to be qualitatively similar.[50]

Of course, there are always going to be differences between laboratory and field environments that make judgements such as these largely 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 to discount their private information less, and should on average cause winning bids to *rise*. 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 is available for oil auctions from the work of Mead *et al.* (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 neighbour(s) 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 (p. 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

[50] 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 (1984; 1987) found that collusion among bidders was pervasive.)

> the rate of return for insiders (neighbors) relative to outsiders (non-neighbors) *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 high 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 nonneighbors 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.

While this argument may go somewhat beyond the available mathematical theory, and while Kagel and Levin are careful to note that there are alternative explanations for why both nonneighbours and neighbours 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.[51]

III.2 *Some Other Auction Results*

There is quite a large and distinguished experimental literature concerned with different auction rules in various economic environments. One of the striking and by now well known results from that literature is that in certain kinds of markets it is possible to observe trades converge to competitive equilibrium, in repeated markets with relatively few traders, often in relatively few periods, as traders gain experience, through repetition, with the parameters of the market. This is particularly so in repeated *double auction* markets, in which both buyers and sellers are free to make and revise bids and offers. For surveys of these results, see Plott (1982) or Smith (1982). (For an account of how some of the earlier conclusions fare in the light of subsequent experiments, in the kind of process that is the subject of this survey, see section 3 of my earlier survey, Roth (1987*b*).)

[51] For a new analysis of the field data, see Hendricks *et al.* (1987).

Before going on, note how these results, like the results of the experiments discussed at greater length here, change the ground of debate, even while not ending it. While there remains room to debate what dimensions of the experimental environment foster competitive outcomes, that competitive outcomes can be observed can be reliably demonstrated even in class. So whether markets are competitive is a question about real markets, not only about those with infinitely many agents. And in evaluating theories of competition, therefore, the experimental results point us in the direction of theories like that of Ostroy (1980), which do not depend on large numbers of agents.

Two recent papers which offer some interesting comparisons with those in Section III.1 are those by Kagel, Harstad and Levin (1987), and Kagel, Levin and Harstad (1987).

Kagel, Levin and Harstad (1987) 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 such auctions to bid one's true willingness to pay (and such auctions are sometimes called 'Vickery auctions'). 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 behaviour 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.[52] The authors conjecture that the overbidding is due to '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 (p. 1286).

[52] The authors report that the probability of losing money based on the observed amount of overbidding averaged only 0·06.

> ...in these earlier private value auction experiments subjects were *not permitted* to bid in excess of their private values. (emphasis in original)

They go on to remark (p. 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, 1982). 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.[53]

Notice what this illustrates about the power of experimental methods. As economists, we have become accustomed to the fact that, because field data are 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 an inadvertently unreported procedure of the earlier experiments had been 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 information on bids in first price private value auctions. 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.

III.3 *Some Methodological Remarks*

One topic I should mention in passing concerns the recurring methodological theme that it is difficult to control subjects' preferences. An aspect of this problem is raised in a penetrating critique by Harrison (1987), who reanalyses the conclusions reached by Cox *et al.* (1982), and Cox *et al.* (1983*a*, *b*; 1985*b*; 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

[53] The 'we believe' is due to the fact that the values here are not *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, 1985 [*b*]).' And in a subsequent experiment, Kagel and Levin (1988) 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.

utilities of the bidders, under the assumption that the experimental data represent a Nash equilibrium. 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 conform well to the equilibrium hypothesis once different 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 this conclusion. His point 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 which appear to be significantly different from the risk neutral equilibrium bids differ only by pennies in expected payoff.[54]

The methodological thrust of Harrison's argument is that the whole point of paying subjects in experiments is to gain control of their incentives, i.e. 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. Harrison concludes that 'It is the purpose of an experimental design to make the monetary payoffs salient enough to be used as a surrogate for subjects' utilities. If we cannot conclude that this is the case for observed deviations from predicted behavior, then we are unable to reject the theory generating those predictions'.

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 is as great in experimental data as in other kinds of data. A similar critique of quite a different kind of experiment is made by Brown and Rosenthal (1987), who reanalyse the data which led O'Neill (1987) to conclude that his experimental subjects had played the minimax mixed strategy in a zero-sum game. Brown and Rosenthal conclude that this is not the case, by disaggregating the data by pairs of players. The part of their analysis that is germane here is that they note that near equilibrium the payoff to the players is very 'flat' as a function of the strategies, and consequently the observation that the average payoff was near the equilibrium payoff cannot be interpreted as strong evidence that the strategies were near the equilibrium strategies.[55]

I think that these critiques show that the issue of control in experimental design is a subtle one, in that a degree of control sufficient to support some conclusions may be insufficient to support others. Since many economic phenomena are predicted to happen on the margin, where agents may be more

[54] 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.

[55] They say (p. 7) 'As it turns out, the structure of O'Neill's game is such that a wide range of non-minimax behaviour can also lead to a winning percentage for the row player that is close to [the equilibrium value of] ·400.'

or less indifferent between a number of choices, the issues raised here will need to be of concern to a wide range of experimenters, both in choosing experimental designs and interpreting experimental results.[56]

IV. PREFERENCE REVERSALS AND OTHER INDIVIDUAL CHOICES

Almost simultaneously with the rise of expected utility theory to pride of place among economists' models of individual choice behaviour, 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 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 multi-dimensional alternatives that did not even involve risk.

These observations did not materially impede the adoption of utility maximisation as the primary vehicle for modelling 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' behaviour. But even when utility theory is viewed as a descriptive theory of actual choice behaviour, 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 of the reported violations for students of economic phenomena.

In the intervening years, the nature of these and many other reliable 'anomalies' in choice behaviour have started to be much more thoroughly explored 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.

IV.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

[56] Note the relationship to other phenomena we have discussed, such as the observation of 'disadvantageous counteroffers' by Ochs and Roth (1988) (see Section I.1), or the observation of Kagel, Harstad, and Levin (1987) of overbidding in second price private value auctions (see Section III.2). In both cases, actions that would be 'irrational' if expected monetary income could be equated with utility result in only small expected monetary losses, and so in neither case can the observations be regarded as strong evidence of irrational behaviour. They may simply be evidence that small monetary differences may be insufficient to override non-monetary elements in subjects' utility, or that negative feedback that occurs with only small probability may be insufficient to correct misconceptions. Similarly, in a set of experiments designed to investigate how subjects shop for low prices when search is costly, Hey (1982; 1987) concluded that they employed non-optimal 'rules of thumb' that, because of the flatness of the payoff function, performed almost as well as optimal search.

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 probabilities. 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 'attractiveness' 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 their 1971 study (Lichtenstein and Slovic, 1971), in which preference reversals were first reported. In that paper they wrote (p. 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[ubjects] would choose one of them but bid more for the other. For example, consider the pair consisting of Bet *P* (0·99 to win \$4 and 0·01 to lose \$1) and Bet \$ (0·33 to win \$16 and 0·67 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 influenced by payoffs, one might expect that S[ubject]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, while 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% of the subjects were observed to always make the predicted reversal: (p. 48) '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% 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 (p. 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 *et al.* (1964). Each subject was told that the selling price he named would be compared with a price to be determined randomly, by spinning a roulette wheel. 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), and otherwise the subject would keep the lottery. It is not hard to see that the dominant strategy for a utility maximiser 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.[57]

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 favour 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 *et al.* (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 (p. 17):

[57] The argument, which Lichtenstein and Slovic presented to their subjects, is the same as the now familiar argument of Vickery (1961), that it is a dominant strategy to bid your true value in a second-price auction. The basic idea seems to have been independently developed a number of times: Smith (1979*a*, footnote 1) recounts how he heard this kind of procedure described by Jacob Marschak in 1953. Interestingly, Becker *et al.* (1964) report that they used this technique to estimate the utility functions of two experimental subjects repeatedly. They concluded that their subjects' responses were not consistent with utility maximisation, although their behaviour became more consistent with repeated exposure to the problem.

> 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 5¢, 10¢, 25¢, $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 *et al.* 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 maximisers to state their true reservation price. Again, predicted reversals were frequent and unpredicted reversals rare.[58] The authors conclude that (p. 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 ('... one would be hesitant to generalize from such very special populations') and that the experimenters were psychologists ('Subjects nearly

[58] 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.

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, '...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 *et al.* second price auction 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 remain 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 *et al.* (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.[59] 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.[60]

[59] 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.

[60] 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.

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 in economics. However the following experiment of Berg *et al.* (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 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 honour 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 to do so fully.

This lends some indirect support, I think, to the view among psychologists that this phenomenon 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 'V's (facing either out or in) to make them look longer or shorter. 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 'V's 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.

IV.2 *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. 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 behaviour at a somewhat different level of detail,

and are therefore typically more willing to sacrifice some accuracy for some generality) have viewed choice behaviour 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 behaviour 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', 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 realisation 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.

Now, one of the nice things about experiments is that they can help to distinguish between alternative hypotheses. Consider the kinds of experiments we might contemplate to distinguish between the competing hypotheses that preference reversals arise from information processing or regret-theoretic causes. Since the predictions derived from regret theory appear to be sensitive to whether lotteries share comparable random events (so that the 'might have beens' can be compared), this suggests that choices between lotteries in which these comparisons cannot be made might yield different results. Similarly, since the regret theory hypothesis is stated in terms of the same kinds of comparisons both for the choice between lotteries and for the setting of a price, it suggests that reversals are only one of a species of intransitivities. Since neither of these effects is obviously predicted by the information-processing hypothesis as it has so far been presented, appropriate experimental results will help to distinguish between these competing hypotheses, and to suggest refinements and elaborations of each.[61]

Notice I am not predicting that we will soon see some 'critical experiment' that will cause one of these hypotheses to fall by the wayside. Even in the laboratory, things are rarely so simple, since terse theories about complex phenomena permit a variety of interpretations and elaborations. Indeed, one

[61] In this connection, see Loomes *et al.* (1988*a*).

of the main uses of experiments is to prompt the refinement of the theories used to explain them. And of course, experimental results intended to distinguish between two such hypotheses as these may yield ambiguous results, particularly if the phenomenon turns out to be due to some third cause.

Just such a third hypothesis to account for preference reversals has been suggested independently by Holt (1986) and Karni 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 lottery 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 which implies that the price a utility maximiser will state in the Becker *et al.* elicitation procedure can be interpreted as his reservation price: i.e. it implies he is indifferent between a lottery A and a selling price p if and only if p is the price that maximises the utility of the compound lottery between A and prices greater than p that he faces after stating a price.[62] To emphasise that preference reversals may be compatible with transitive preferences over lotteries, Karni and Safra (1987) couch their discussion in terms of such a family of generalisations of utility theory proposed by Quiggan (1982) and Yaari (1987).

Holt 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 Karni 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.[63]

It should be emphasised that the Becker *et al.* elicitation procedure allows us to predict what prices *utility maximisers* would state. That non-utility-maximisers may have incentives to respond differently is in no way a criticism of the experimental designs which 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 maximisers would do. In the absence of such a design, we would be unable to conclude that the observed phenomenon constituted a violation

[62] As noted above, the argument for the Becker *et al.* procedure is the same as that for second-price auctions, and so violations of independence have the same implications there: see Karni and Safra (1985).

[63] An initial experiment in this direction is that of Loomes *et al.* (1988*b*).

of the theory.[64] 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, see if you can think of any *field* observations that might possibly allow us to observe preference reversals.

IV.3 *Market Behaviour*

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 (i.e. more than can be accounted for by income effects) stated selling prices 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 *et al.* (1987) propose to test what I will call the *market hypothesis*, which is that agents in a market environment will behave like utility maximisers, i.e. 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.[65] The market environment in their experiment is a second price auction, so that it will be a dominant strategy for utility maximisers to state their true reservation prices.[66] (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 while the auction results continued to show some continued disparity, this diminished over auction periods, and by the final period the remaining gap between the two prices was no longer statistically significant. (Most of the movement came in subjects' declining prices for

[64] Note the parallel to the use of binary lottery games discussed in Section I.1.

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

[66] Kagel and Levin (1988) note that Coursey *et al.* rely on the second-price auction results reported by Coppinger *et al.* (1980) and Cox *et al.* (1982) in interpreting their own results. As these now seem to be erroneous (recall the discussion in section III.2), Kagel and Levin suggest that the results of Coursey *et al.* may also require reinterpretation. In particular, while the observed overbidding is in general small, Kagel and Levin observe that this may be consequential in an experiment in which bids are taken to equal the true reservation prices, and they note that the direction of the anomalous behaviour in second price auctions is such that it could provide an alternative explanation for Coursey *et al.*'s results.

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'.[67]

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 stylised facts about market behaviour that they think may reflect choice anomalies similar to this price disparity. Knetsch *et al.* (1987) present some evidence in support of this view, from a market experiment in which half the subjects have been endowed 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.[68]

Another experiment, concerned with both preference reversals and buying and selling price disparities in a market setting, is reported by Knez and Smith (1987). They propose to test a version of what I will call the *strong market hypothesis*, which is that markets equilibrate as if agents were utility maximisers even if the agents do not themselves behave as if they were utility maximisers. They state this point of view as follows (p. 132):

> The efficiency and social significance of markets does not depend on the validity of any particular theory of individual demand... the empirical validity or falsity of efficient markets theory is a proposition that is entirely distinct from the empirical validity or falsity of theories of individual demand in markets.

Knez and Smith observed several repeated markets in which subjects were first asked to state hypothetical preferences and buying or selling prices for a P bet and a $ bet with expected values of $3·85, and were then allowed to trade. Trading was conducted in two separate double auction markets, first a market for P bets, then for $ bets. In each market each buyer was given an endowment of $5·50, and was allowed to buy no more than a single lottery. Each seller was given an endowment of one lottery (and was also paid $1·65). After both markets had been conducted, buyers and sellers were again asked to state hypothetical preferences and reservation prices, and all initial endowments were replenished. The whole process was repeated several times, with each subject remaining always a buyer or always a seller.

Although Knez and Smith observe numerous and persistent price disparities and preference reversals in the hypothetical responses (p. 142, 'It seems clear that there is a hard core of 35 to 38% reversals that continue to be exhibited

[67] In a related experiment, 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.

[68] Interestingly, Marshall *et al.* (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.

by the reported preferences and values...') they argue that very few of the prices at which lotteries change hands are inconsistent with utility theory, in the sense that they do not lie outside the interval of possible payoffs of the lotteries. In fact, most transactions occur sufficiently close to the expected value to be consistent with plausible risk preference or risk aversion. On the basis of this evidence they conclude (p. 152) 'The results of this study call into question the interpretation, reliability, and robustness of preference reversal phenomena in the joint context of repetitive responses and market trading'.

It should be clear by now why it is unlikely that this conclusion will meet with universal acceptance. Rather it provides another piece of the continuing debate among experimenters interested in these issues. Since their experiment is too recent to have drawn a reply as of this writing, let me play the Devil's advocate. For example, one might speculate that the moderate trading prices observed by Knez and Smith have more to do with the specific market parameters they have chosen than with general properties of markets. Since preference reversals seem to be associated with the overpricing of the $ bets (which in this case had a prize of $16), anomalous trades might be easier to observe in a market in which buyers had cash endowments larger than $5·50, and in which $ bets were in shorter supply (e.g. by having fewer sellers than buyers). Or, since preference reversals involve choices between P and $ bets, it might be easier to observe them in some market in which both lotteries could be traded simultaneously. My object here is not to try to design or to guess the outcome of the 'next' experiment, but to reemphasise that experiments provide a method for continuing the argument, in a way that can steadily narrow the grounds for potential disagreement.

IV.4 *Other Phenomena*

I have concentrated on preference reversals here because they have been the subject 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 phenomena with equal (and equally contested) claims to importance. For a discussion of some of these, see Kahneman and Tversky (1979; 1984), or Thaler (1987) (who particularly concentrates on what in his view is the importance of these phenomena for economics). These other choice phenomena offer other possibilities to design experiments by which various generalisations of expected utility theory may be tested. In this regard see e.g. Camerer (1988), Loomes (1988), Loomes and Sugden (1987), or Battalio *et al.* (1988) (who succinctly summarise the present state of theoretical affairs by stating: '... none of the rival formulations considered consistently organises choices, indicating that 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 behaviours of this kind (see Battalio *et al.* (1985), Kagel (1987)), so experiments on these matters need not be confined to humans.

Before leaving the subject of individual choice behaviour, I will mention one more experiment, which 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 '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 which could be ranked in order of riskiness. ('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 '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.[69] 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 whether certain kinds of anomalous choice behaviour were artifacts of the experimental procedure, and in particular whether they would persist in non-hypothetical 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 whether market environments will moderate the observed effects. Some of the various theories which have been advanced as possible explanations of certain kinds of choice behaviour not only suggest new experiments, but new directions for pursuing traditional kinds of economic theory (see e.g. Crawford (1988)). It is clear that the degree of success such theories achieve in organising and explaining phenomena in domains other than individual choice behaviour itself will be important. At the same time, the question of whether 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.

[69] In Binswanger (1981) and Quizon *et al.* (1984), this is interpreted as reflecting utility functions defined in terms of changes in wealth, rather than in terms of net wealth (cf. Markowitz, 1952; Kahneman and Tversky, 1979), or else as reflecting some deviation from expected utility maximisation such as those involving failures of the independence assumption.

V. SOME COMMON METHODOLOGICAL QUESTIONS

I will try to organise briefly some of the methodological themes encountered in this survey into a set of commonly asked questions, together with my answers and evasions (which may not always reflect a consensus among experimenters).

Are experiments involving hypothetical choices reliable? This question is one of the fault lines dividing some economists and psychologists. There are examples which allow one to answer in either direction. For example, preference reversals were initially identified from data on hypothetical choices, and proved to be robust to better motivated choices. On the other hand, the study of Binswanger (1980) showed that large differences in responses will sometimes be observed.[70] Of course, Binswanger also observed important differences between gambles for small stakes and for large stakes, and so the question is really related to questions about the effects of the scale of payoffs.[71] As Thaler (1987) notes (p. 120), 'Asking purely hypothetical questions is inexpensive, fast, and convenient'. Since many experiments have binding budget constraints, this is not a frivolous point. Personally, I am most comfortable with responses to hypothetical choices when they are presented together with some evidence that the answers are not too dissimilar from those obtained from actual choices. To the extent that the size of actual payoffs may be critical for some phenomena, I am prepared to entertain experimental designs in which some parts of the design use hypothetical choices, while other parts use the money thus saved to make a comparison with more substantial monetary payoffs than would otherwise have been possible.

Do college students behave like real people? The answer to this one must be clear to all the college professors among my readers: you can never be sure. But for most purposes there is little evidence that they do not, since results initially obtained with college students seem to be robust more often than not (e.g. Lichtenstein and Slovic's casino gamblers, or Dyer *et al.*'s construction executives). That is not to say that special groups of subjects, such as commodities traders or meteorologists, might not have some properties not found among college students, or in the general population. I certainly do not want to discourage anyone from conducting experiments with different subject populations. In my own opinion, there has been a shortage of carefully controlled cross cultural experiments, which might reveal unanticipated (or anticipated) differences or similarities in various kinds of economic behaviour among subjects in different societies.

Do economic experiments have to be done in computer controlled laboratories? No. Certainly not. Many are, because such a laboratory facilitates many kinds of control (e.g. always presenting subjects with certain information in the same way, or keeping them anonymous), and record

[70] There are several studies that have made explicit comparisons between hypothetical and real choices, and they typically report some differences, but often not qualitatively important ones (cf. Battalio *et al.*, 1988; Grether and Plott, 1979; Jamal and Sunder, 1988; Marshall *et al.*, 1986).

[71] See also Kroll *et al.* (1988) in this regard. They report an experiment in which subjects' behaviour (regarding the diversification of investments) changed significantly in response to an increase in payoffs by a factor of ten.

keeping (e.g. recording the time of each transaction). However many of the experiments discussed here were conducted without such a laboratory, and most of the rest could have been. If you are planning to do a lot of experiments, however, the investment in time and energy to establish such a laboratory may be worth it. But no one should let the start up costs of a computerised lab be a deterrent to conducting experiments, and for many purposes it will be most sensible at least to begin experimental work with simpler arrangements.

Is it possible to 'lie' with experiments? Yes, of course it is. I suppose we will know that economic experimentation has really come of age when we start to have outright fraud to contend with, of the kind that sometimes afflicts other areas of science, in which researchers desperate to establish priority or show progress on their grant proposals, or simply to 'prove' a treasured theory fabricate data out of whole cloth.[72] In the meantime, I think what this question most often means is 'If I believe theory A and you believe theory B, cannot I do an experiment whose results will seem to support A while you do one whose results seem to support B? Here too, I think the answer must be, in a limited way, yes. The reason is that the designer of an experiment may have to make many particular decisions, and choose the level of many parameters that may affect the phenomena under study. If there is reason to believe that the resulting observations depend in important ways on some of these choices, variations may be incorporated into the experimental design, or into subsequent experiments, to see if this is the case. But there are often too many essentially arbitrary choices for this approach to be applied to all but a few of them. (It is for this reason that the strongest conclusions from an experimental study come from *within experiment* comparisons, which report the effect of the change of a single variable, while holding all others constant.)

I think the major pitfall to be aware of here is that, since decisions made in the design of the experiment cannot be regarded as random samples from the space of possible design choices, there is room for an experimenter's prior beliefs about the likely outcome of the experiment to influence the outcome, through these design decisions.[73] Sometimes this is explicit, deliberate, and desirable, as when the experimenter wishes to demonstrate that some phenomenon is possible, rather than general. The danger is of inadvertently reading experimental evidence as supporting an overly general conclusion based on observations made in special cases. Sometimes this will be unavoidable, when the relevant variables simply have not yet been identified. However it may be possible to address some aspects of this problem, since experiments are often preceded by pilot studies, designed to determine inexpensively and quickly if a full scale experiment will yield useful data. Final design decisions may be made on the basis of these pilot studies, which therefore may include parameters and design choices not included in the experiment. I think that reporting such pilot experiments in a little more detail than is presently customary in the

[72] If space allowed, I would report a series of computerised cross cultural experiments involving large numbers of identical triplets, one raised in a market economy, one in a command economy, and one by wolves.

[73] The same could of course be said for econometrics.

experimental literature may be a good practice, particularly insofar as it may sometimes help to alert other investigators to alternative hypotheses for the results presented.[74] (However I should note in this regard that experimenters would often like to report more detail than the guardians of scarce journal space allow.) In the meantime, I think that experimentation is well served by sceptical readers, and particularly by experimenters with different theoretical predispositions.

VI. CONCLUDING REMARKS

This essay has focused on experimental methods, and how experiments can be used to further an investigation, and often a debate.[75] Lest this emphasis on debate leave the reader with the mistaken impression that we never get anywhere with all this, let me recall a few of the things clearly learned from experiments: The outcome of bargaining can be systematically effected by information other than that which goes into determining bargainers' expected utility; The winner's curse is a real phenomenon; Competitive equilibria can be observed in small repeated markets; Reliable violations of expected utility theory can be observed in well motivated individual choice behaviour.

Each of these is now a well established, often reproduced experimental result, but was once a matter capable of generating considerable controversy, at least in some circles. I think the results I have just stated are no longer themselves grounds for much controversy, although their significance may still leave room for debate. But experiments give us directions to further this debate too, and to continue to narrow the grounds for disagreement. This is so even in areas in which some of the experimental results remain equivocal: for example, the regularities observed in repeated play of the prisoner's dilemma have prompted new kinds of theoretical models, just as the results on public information and the winner's curse suggest new kinds of analyses of field data.

What I have tried to make clear is that experimentation is more than the sum of the facts that have so far been established. Like econometrics, and like deductive theory, it is a method of investigation. For this reason, I have concentrated in this survey on methodological themes, broadly construed, by which I mean not merely how experiments are conducted, but how it is decided *what* experiments to conduct, and how results are interpreted, and sometimes challenged. In doing so, I hope I have presented a clearer picture of the experimental endeavour, and how it complements more traditional kinds of economic research, both theoretical and empirical.

University of Pittsburgh

Date of receipt of final typescript: July 1988

[74] For an example, see Conlisk (1987), who makes something of the same point in the context of individual choice behaviour.

[75] There are uses of experiments I have not even touched on here, such as for stimulating particular target markets (see Plott (1987) or section 4 of Roth (1987*b*) or for constructing compelling classroom demonstrations. I do not mean to slight them, I just did not have enough room to speak about them here. Similarly, I have not concentrated specifically on the development of experimental methodology as such (e.g. the development of further procedures for implementing binary lottery games (recall Section I.1, and cf. Berg *et al.* (1986)) which is an as yet small part of the literature that will grow in importance as experiments concentrate on phenomena that are increasingly difficult to measure or control.

References

Allais, Maurice (1953). 'Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école americane.' *Econometrica*, vol. 21, pp. 503–46.

Axelrod, Robert (1980*a*). 'Effective choice in the iterated prisoner's dilemma.' *Journal of Conflict Resolution*, vol. 24, pp. 3–25.

—— (1980*b*). 'More effective choice in the prisoner's dilemma.' *Journal of Conflict Resolution*, vol. 24, pp. 379–403.

—— (1984). *The Evolution of Cooperation*, New York: Basic Books.

Baiman, Stanley and Lewis, Barry L. (1988). 'An experiment testing the behavioral equivalence of strategically equivalent employment contracts.' *Journal of Accounting Research*, forthcoming.

Banks, Jeffrey S., Plott, Charles R. and Porter, David P. (1988). 'An experimental analysis of unanimity in public goods provision mechanisms.' *Review of Economic Studies*, vol. 55, pp. 301–22.

Battalio, Raymond C., Kagel, John H. and Komain Jiranyakul (1988). 'Testing between alternative models of choice under uncertainty: some initial results.' Mimeo.

—— —— and McDonald, Don N. (1985). 'Animals' choices over uncertain outcomes: some initial experimental results.' *American Economic Review*, vol. 75, pp. 597–613.

Bazerman, Max H. (1985). 'Norms of distributive justice in interest arbitration.' *Industrial and Labor Relations*, vol, 38, pp. 558–70.

—— and Samuelson, William F. (1983). 'I won the auction but don't want the prize.' *Journal of Conflict Resolution*, vol. 27, pp. 618–34.

Becker, Gordon M., DeGroot, Morris M., and Marschak Jacob (1964). 'Measuring utility by a single-response sequential method.' *Behavioral Science*, vol. 9, pp. 226–32.

Berg, Joyce E., Daley, L. A., Dickhaut, John W. and O'Brien, John R. (1986). 'Controlling preferences for lotteries on units of experimental exchange.' *Quarterly Journal of Economics*, pp. 281–306.

—— Dickhaut, John W. and O'Brien John R. (1985). 'Preference reversal and arbitrage.' in (V. Smith, *ed.*,) *Research in Experimental Economics*, vol. 3, pp. 31–72. Greenwich: JAI Press.

Bernoulli, Daniel (1738). 'Specimen theoriae novae de mensura sortis.' *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. 5, pp. 175–92. (English translation in *Econometrica*, vol. 22, 1954, pp. 23–36.)

Binmore, Ken, Shaked, Avner and Sutton, John (1985). 'Testing noncooperative bargaining theory: a preliminary study.' *American Economic Review*, vol. 75, pp. 1178–80.

—— —— and —— (1988*a*). 'Testing noncooperative bargaining theory: a reply.' *American Economic Review*, forthcoming.

—— —— and —— (1988*b*). 'An outside option experiment.' Mimeo, London School of Economics.

Binswanger, Hans P. (1980). 'Attitudes toward risk: experimental measurement in rural India.' *American Journal of Agricultural Economics*, vol. 62, pp. 395–407.

—— (1981). 'Attitudes toward risk: theoretical implications of an experiment in rural India.' Economic Journal, vol. 91, pp. 867–90.

Bohm, Peter (1972). 'Estimating demand for public goods: an experiment.' *European Economic Review*, vol. 3, pp. 111–30.

Brookshire, David S. and Coursey, Don L. (1987). 'Measuring the value of a public good: an empirical comparison of elicitation procedures.' *American Economic Review*, vol. 77, pp. 554–66.

Brown, James N. and Rosenthal, Robert W. (1987). 'Testing the minimax hypothesis: a re-examination of O'Neill's game experiment.' Department of Economics, SUNY Stony Brook.

Brown, Keith C. (1986). 'In search of the winner's curse: comment.' *Economic Inquiry*, vol. 24, pp. 513–15.

Camerer, Colin F. (1987). 'Do biases in probability judgment matter in markets? Experimental evidence.' *American Economic Review*, vol. 77, pp. 981–97.

—— (1988). 'An experimental test of several generalized utility theories.' *Journal of Risk and Uncertainty*, forthcoming.

—— and Weigelt Keith (1988). 'Experimental tests of a sequential equilibrium reputation model.' *Econometrica*, vol. 56, pp. 1–36.

Capen, E. C., Clapp, R. V. and Campbell, W. M. (1971). 'Competitive bidding in high-risk situations.' *Journal of Petroleum Technology*, vol. 23, pp. 641–53.

Carlson, John A. and O'Keefe, Terrence B. (1969). 'Buffer stocks and reaction coefficients: an experiment with decision making under risk.' *Review of Economic Studies*, vol. 36, pp. 467–84.

Chamberlin, Edward H. (1984). 'An experimental imperfect market.' *Journal of Political Economy*, vol. 56 (2), pp. 95–108.

Clarke, Edward H. (1971). 'Multipart pricing of public goods.' *Public Choice*, vol. 11, pp. 17–33.

Coase, A. (1960) 'The problem of social cost.' *Journal of Law and Economics*, vol. 1, pp. 1–44.

Conlisk, John (1987). 'Verifying the betweenness axiom with questionnaire evidence, or not: take your pick.' *Economics Letters*, vol. 25, pp. 319–22.

Contini, Bruno (1968). 'The value of time in bargaining negotiations: some experimental evidence.' *American Economic Review*, vol. 58, pp. 374–93.
Cooper, Russell, DeJong, Douglas V., Forsythe, Robert and Ross, Thomas W. (1987). 'Selection criteria in coordination games: some experimental results.' Working paper no. 87–20, Dept. of Economics, University of Iowa.
Coppinger, Vicki M., Smith, Vernon L. and Titus, Jon A. (1980). 'Incentives and behavior in English, Dutch and sealed-bid auctions.' *Economic Inquiry*, vol. 18, pp. 1–22.
Coursey, Don L. (1982). 'Bilateral bargaining, Pareto optimality, and the empirical frequency of impasse.' *Journal of Economic Behavior and Organization*, vol. 3, pp. 243–59.
—— Hovis, John L. and Schulze, William D. (1987). 'The disparity between willingness to accept and willingness to pay measures of value.' *Quarterly Journal of Economic*, vol. 102, pp. 679–90.
Cox, James C. and Isaac, R. Mark (1984). 'In search of the winner's curse.' *Economic Inquiry*, vol. 22, pp. 579–92.
—— and —— (1986). 'In search of the winner's curse: reply.' *Economic Inquiry*, vol. 24, pp. 517–20.
—— Roberson, Bruce and Smith, Vernon L. (1982). 'Theory and behavior of single object auctions.' (In V. L. Smith, ed.), *Research in Experimental Economics*, Volume 2, Greenwich: JAI Press, pp. 1–43.
—— Smith, Vernon L. and Walker, James M. (1983*a*). 'Tests of a heterogeneous bidders theory of first price auctions.' *Economics Letters*, vol. 12, pp. 207–12.
—— —— and —— (1983*b*). 'A test that discriminates between two models of the Dutch-first auction non-isomorphism.' *Journal of Economic Behavior and Organization*, vol. 4, pp. 205–19.
—— —— and —— (1985*a*). 'Experimental development of sealed-bid auction theory; calibrating controls for risk aversion.' *American Economic Review Papers and Proceedings*, vol. 75, pp. 160–5.
—— —— and —— (1985*b*). 'Expected revenue in discriminative and uniform price sealed-bid auctions.' (in V. L. Smith, ed.), *Research in Experimental Economics*, Volume 3, Greenwich, JAI Press, pp. 183–232.
—— —— and —— (1988). 'Theory and individual behavior of first-price auctions.' *Journal of Risk and Uncertainty*, vol. 1, pp. 61–99.
Crawford, Vincent P. (1988). 'Equilibrium without independence.' *Journal of Economic Theory*, forthcoming.
Cummings, L. L. and Harnett, D. L. (1969). 'Bargaining behaviour in a symmetric bargaining triad: the impact of risk-taking propensity, information, communication and terminal bid.' *Review of Economic Studies*, vol. 36, pp. 485–501.
Cyert, Richard M. and Lave, Lester B. (1965). 'Collusion, conflit et science economique.' *Economie Appliquee*, vol. 18, pp. 385–406.
Dolbear, F. T., Lave, L. B., Bowman, G., Lieberman, A., Prescott, E., Rueter, F. and Sherman, R. (1968). 'Collusion in oligopoly: an experiment on the effect of numbers and information.' *Quarterly Journal of Economics*, vol. 82, pp. 240–59.
Dyer, Douglas, Kagel, John H. and Levin, Dan (1988). 'A comparison of naive and experienced bidders in common value offer auctions: a laboratory analysis.' Economic Journal, forthcoming.
Edgeworth, F. Y. (1881). *Mathematical Psychics*, London: Kegan Paul.
Farber, Henry S. and Bazerman Max H. (1986). 'The general basis of arbitrator behavior: an empirical analysis of conventional and final offer arbitration.' *Econometrica*, vol. 54, pp. 1503–28.
Farrell, Joseph (1987). 'Information and the Coase theorem.' *Journal of Economic Perspectives*, vol. 1, pp. 113–29.
Ferber, Robert, and Hirsch, Werner Z. (1982). *Social Experimentation and Economic Policy*, (Cambridge Surveys of Economic Literature), Cambridge: Cambridge University Press.
Ferejohn, John, Forsythe, Robert and Noll, Roger (1979). 'An experimental analysis of decision making procedures for discrete public goods: a case study of a problem in institutional design.' In (V. L. Smith, ed.), *Research in Experimental Economics*: Volume 1, Greenwich: JAI Press, pp. 1–58.
Flood, Merrill M. (1958). 'Some experimental games.' *Management Science*, vol. 5, pp. 5–26.
Forsythe, Robert, Kennan, John and Sopher, Barry (1987). 'An experimental analysis of bargaining and strikes with one sided private information.' Working paper no. 87–4, Department of Economics, University of Iowa.
Friedman, James W. (1963). 'Individual behaviour in oligopolistic markets: an experimental study.' *Yale Economic Essays*, vol. 3, pp. 359–417.
—— (1969). 'On experimental research in oligopoly.' *Review of Economic Studies*, vol. 36, pp. 399–415.
Graham, Daniel A. and Marshall, Robert C. (1984). 'Bidder coalitions at auctions.' Duke University Department of Economics, mimeo.
—— and —— (1987). 'Collusive bidder behavior at single object second price and English auctions.' *Journal of Political Economy*, vol. 95, pp. 1217–37.
Green, Jerry R. and Laffont, Jean-Jacques (1979). *Incentives in Public Decision-Making*, Amsterdam: North-Holland.
Grether, David M. and Plott, Charles R. (1979). 'Economic theory of choice and the preference reversal phenomenon.' *American Economic Review*, vol. 69, pp. 623–38.

—— and —— (1982). 'Economic theory of choice and the preference reversal phenomenon: reply.' *American Economic Review*, vol. 72, p. 575.

Groves, Theodore and Ledyard, John (1977). 'Optimal allocation of public goods: a solution to the "free rider" problem.' *Econometrica*, vol. 45, pp. 783–809.

Guth, Werner, Schmittberger, R. and Schwarz, B. (1982). 'An experimental analysis of ultimatum bargaining.' *Journal of Economic Behavior and Organization*, vol. 3, pp. 367–88.

—— and Tietz, Rienhard (1987). 'Ultimatum bargaining for a shrinking cake – An experimental analysis.' mimeo.

Harrison, Glenn W. (1987). 'Theory and misbehavior of first-price auctions.' *American Economic Review*, forthcoming.

—— and McCabe, Kevin A. (1988). 'Testing bargaining theory in experiments.' Mimeo, University of Western Ontario.

—— and McKee, Michael (1985). 'Experimental evaluation of the Coase theorem.' *Journal of Law and Economics*, vol. 28, pp. 653–70.

—— Hoffman, Elizabeth, Rutstrom, E. E. and Spitzer, Matthew L. (1987). 'Coasian solutions to the externality problem in experimental markets.' Economic Journal, vol. 97, pp. 388–402.

Hausman, Jerry A., and Wise, David A. (eds.) (1985). *Social Experimentation*, National Bureau of Economic Research.

Hendricks, Kenneth, Porter, Robert H. and Boudreau, Bryan (1987). 'Information, returns, and bidding behavior in OCS auctions: 1954–1969.' *Journal of Industrial Economics*, vol. 35, pp. 517–42.

Hey, John D. (1982). 'Search for rules for search.' *Journal of Economic Behavior and Organization*, vol. 3, pp. 65–81.

—— (1987). 'Still searching.' *Journal of Economic Behavior and Organization*, vol. 8, pp. 137–44.

Hoffman, E. and Spitzer, M. L. (1982). 'The Coase theorem: some experimental tests.' *Journal of Law and Economics*, vol. 25, pp. 73–98.

—— and —— (1985). 'Entitlements, rights, and fairness: an experimental examination of subjects' concepts of distributive justice.' *Journal of Legal Studies*, vol. 14 no. 2, pp. 259–97.

Hoggatt, Austin C. (1969). 'Response of paid student subjects to differential behavior of robots in bifurcated duopoly games.' *Review of Economic Studies*, vol. 36, pp. 417–32.

—— Selten, Reinhard, Crockett, David, Gill, Shlomo and Moore, Jeff (1978). 'Bargaining experiments with incomplete information.' In (H. Sauermann, ed.), *Bargaining Behavior*, Tubingen: J. C. B. Mohr.

Holt, Charles A. (1986). 'Preference reversals and the independence axiom.' *American Economic Review*, vol. 76, pp. 508–15.

Isaac, R. Mark, McCue, Kenneth F. and Plott, Charles R. (1985). 'Public goods provision in an experimental environment.' *Journal of Public Economics*, vol. 26, pp. 51–74.

—— Walker, James M. and Thomas, Susan H. (1984). 'Divergent evidence on free riding: an experimental examination of possible explanations.' *Public Choice*, vol. 43, pp. 113–49.

Jamal, Karim and Sunder, Shyam (1988). 'Money vs. gaming: effects of salient monetary payments in double oral auctions.' Working Paper, University of Minnesota.

Johansen, Lief (1977). 'The theory of public goods: misplaced emphasis?' *Journal of Public Economics*, vol. 7, pp. 147–52.

Kagel, John H. (1987). 'Economics according to the rats (and pigeons too): what have we learned and what can we hope to learn?' in (A. E. Roth, ed.), *Laboratory Experimentation in Economics: Six Points of View*, Cambridge: Cambridge University Press, pp. 155–92.

—— Harstad, Ronald M. and Levin Dan (1987). 'Information impact and allocation rules in auctions with affiliated Private Values: A Laboratory Study.' *Econometrica*, 55, pp. 1275–1304.

—— and Levin Dan (1986). 'The winner's curse and public information in common value auctions.' *American Economic Review*, vol. 76, pp. 894–920.

—— and —— (1988), 'Independent private value auctions: bidder behavior in first, second and third-price auctions with varying numbers of bidders.' Mimeo.

—— —— Battalio, Raymond C. and Meyer, Donald J. (1988). 'First-price common value auctions: bidder behavior and the "winner's curse"', *Economic Inquiry*, forthcoming.

—— —— and Harstad, Ronald M. (1987). 'Judgment, evaluation and information processing in second-price common value auctions.' Mimeo.

Kahneman, Daniel, Knetsch, Jack L. and Thaler, Richard H. (1986*a*). 'Fairness and the assumptions of economics.' *Journal of Business*, vol. 59, no. 4 pt 2, pp. S285–S300.

—— —— and —— (1986*b*). 'Fairness as a constraint on profit seeking: entitlements in the market.' *American Economic Review*, vol. 76, pp. 728–41.

—— Slovic, Paul and Tversky, Amos ed. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, Cambridge University Press.

—— and Tversky, Amos (1979). 'Prospect theory: an analysis of decision under risk.' *Econometrica*, vol. 47, pp. 263–91.

—— and —— (1984). 'Choices, values, and frames.' *American Psychologist*, vol. 39, pp. 341–50.
Karni, Edi, and Safra, Zvi (1985). 'Vickrey auctions in the theory of expected utility with rank dependent probability.' *Economics Letters*, vol. 20, pp. 15–8.
—— and —— (1987). '"Preference reversal" and the observability of preferences by experimental methods.' *Econometrica*, vol. 55, pp. 675–85.
Kihlstrom, Richard, Roth, Alvin E. and Schmeidler, David (1981). 'Risk aversion and solutions to Nash's bargaining problem.' In *Game Theory and Mathematical Economics* (ed. O. Moeschlin and D. Pallaschke). Amsterdam: North-Holland, pp. 65–71.
Kim, Oliver and Walker, Mark (1984). 'The free rider problem: experimental evidence.' *Public Choice*, vol. 43, pp. 3–24.
Knetsch, Jack L. and Sinden, J. A. (1984). 'Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value.' *Quarterly Journal of Economics*, 99, pp. 507–21.
—— and —— (1987). 'The persistence of evaluation disparities.' *Quarterly Journal of Economics*, vol. 102, pp. 691–5.
—— Thaler, Richard and Kahneman, Daniel (1987). 'Experimental tests of the endowment effect and the Coase theorem.' Mimeo.
Knez Marc and Smith, Vernon L. (1987). 'Hypothetical valuations and preference reversals in the context of asset trading.' In *Laboratory Experimentation in Economics: Six Points of View* (ed. A. E. Roth.) Cambridge: Cambridge University Press, pp. 131–54.
Kreps, David M. and Wilson, Robert B. (1982). 'Reputation and imperfect information.' *Journal of Economic Theory*, vol. 27, pp. 253–79.
—— Milgrom, Paul, Roberts, John and Wilson, Robert (1982). 'Rational cooperation in the finitely repeated prisoners' dilemma.' *Journal of Economic Theory*, vol. 27, pp. 245–52.
Kroll, Yoram, Levy, Haim and Rapoport, Amnon (1988). 'Experimental tests of the separation theorem and the capital asset pricing model.' *American Economic Review*, vol. 78, pp. 500–19.
Lave, Lester B. (1962). 'An empirical approach to the prisoners' dilemma game,' *Quarterly Journal of Economics*, vol. 76, pp. 424–36.
—— (1965). 'Factors affecting co-operation in the prisoner's dilemma.' *Behavioral Science*, vol. 10, pp. 26–38.
Lichtenstein, Sarah and Slovic, Paul (1971). 'Reversal of preferences between bids and choices in gambling decisions.' *Journal of Experimental Psychology*, vol. 89, pp. 46–55.
—— and —— (1973). 'Response-induced reversals of preference in gambling: an extended replication in Las Vegas.' *Journal of Experimental Psychology*, vol. 101, pp. 16–20.
Lindman, Harold R. (1971). 'Inconsistent preferences among gambles.' *Journal of Experimental Psychology*, vol. 89, pp. 390–7.
Loehman, Edna and Andrew Whinston (1972). 'A new theory of pricing and decision-making for public investment.' *The Bell Journal of Economics and Management Science*, vol. 2, pp. 606–25.
Loomes, Graham (1988). 'When actions speak louder than prospects.' *American Economic Review*, vol. 78, pp. 463–70.
—— Starmer, Chris and Sugden, Robert (1988*a*). 'Preference reversal: information-processing effect or rational non-transitive choice?' Mimeo, presented at the Royal Economic Society Conference, March.
—— —— and —— (1988*b*). 'Observing violations of transitivity by experimental methods.' Mimeo, presented at the Foundations of Utility and Risk IV Conference, Budapest, June.
—— and Sugden, Robert (1982). 'Regret theory: an alternative theory of rational choice under uncertainty.' ECONOMIC JOURNAL, vol. 92, pp. 805–24.
—— and —— (1983). 'A rationale for preference reversal.' *American Economic Review*, vol. 73, pp. 428–32.
—— and —— (1987). 'Testing for regret and disappointment in choice under uncertainty.' ECONOMIC JOURNAL, Conference Papers, vol. 97, pp. 118–29.
MacCrimmon, K. R. and Toda, M. (1969). 'The experimental determination of indifference curves.' *Review of Economic Studies*, vol. 36, pp. 433–51.
Machina, Mark J. (1982). '"Expected utility" analysis without the independence axiom.' *Econometrica*, vol. 50, pp. 277–323.
—— (1987). 'Choice under uncertainty: problems solved and unsolved.' *Economic Perspectives*, vol. 1, pp. 121–54.
Malouf, Michael W. K. and Roth, Alvin E. (1981). 'Disagreement in bargaining: an experimental study.' *Journal of Conflict Resolution*, vol. 25, pp. 329–48.
Markowitz, Harry (1952). 'The utility of wealth.' *Journal of Political Economy*, vol. 60, pp. 151–8.
Marshall, James D., Knetsch, Jack L. and Sinden, J. A. (1986). 'Agents' evaluations and the disparity in measures of economic loss.' *Journal of Economic Behavior and Organization*, vol. 7, pp. 115–27.
Marwell, Gerald and Ames, Ruth E. (1981). 'Economists free ride, Does anyone else? Experiments on the provision of public goods, IV.' *Journal of Public Economics*, vol. 15, pp. 295–310.

May, Kenneth O. (1954). 'Intransitivity, utility, and the aggregation of preference patterns.' *Econometrica*, vol. 22, pp. 1–13.

Mead, Walter J., Moseidjord, Asbjorn and Sorensen, Philip E. (1983). 'The rate of return earned by leases under cash bonus bidding in OCS oil and gas leases.' *Energy Journal*, vol. 4, pp. 37–52.

—— —— and—— (1984). 'Competitive bidding under asymmetrical information: behavior and performance in Gulf of Mexico drainage lease sales, 1959–1969.' *Review of Economics and Statistics*, vol. 66, pp. 505–8.

Milgrom, Paul R. and Roberts, John (1982). 'Predation, reputation, and entry deterrence.' *Journal of Economic Theory*, vol. 27, pp. 280–312.

Milgrom, Paul R. and Weber, Robert J. (1982). 'A theory of auctions and competitive bidding.' *Econometrica*, vol. 50, pp. 1089–122.

Morley, I. and Stephenson, G. (1977). *The Social Psychology of Bargaining*, London: Allen and Unwin.

Murnighan, J. Keith (1985). 'Coalitions in decision-making groups: organizational analogs.' *Organizational Behavior and Human Decision Processes*, vol. 35, pp. 1–26.

—— Roth, Alvin E. and Schoumaker, Francoise (1988). 'Risk aversion in bargaining: an experimental study.' *Journal of Risk and Uncertainty*, vol. 1, pp. 101–24.

Nash, John (1950). 'The bargaining problem.' *Econometrica*, vol. 28, pp. 155–62.

Neelin, Janet, Sonnenschein, Hugo and Spiegel, Matthew (1988). 'A further test of noncooperative bargaining theory.' *American Economic Review*, forthcoming.

Nydegger, Rudy V. and Owen, Guillermo (1975). 'Two person bargaining: an experimental test of the Nash axioms.' *International Journal of Game Theory*, vol. 3, pp. 239–349.

Ochs, Jack (1988). 'The coordination problem in decentralized markets: some experiments.' Mimeo, Department of Economics, University of Pittsburgh.

—— and Roth, Alvin E. (1988). 'An experimental study of sequential bargaining.' *American Economic Review*, forthcoming.

O'Neill, Barry (1987). 'Nonmetric test of the minimax theory of two-person zerosum games.' *Proceedings of the National Academy of Sciences*, vol. 84, pp. 2106–9.

Ostroy, Joseph M. (1980). 'The no-surplus condition as a characterization of perfectly competitive equilibrium.' *Journal of Economic Theory*, vol. 22, pp. 65–91.

Palfrey, Thomas R. and Rosenthal, Howard (1987). 'Private incentives in social dilemmas: the effects of incomplete information and altruism.' Social Science Working Paper 659, California Institute of Technology.

Plott, Charles R. (1982). 'Industrial organization theory and experimental economics.' *Journal of Economic Literature*, vol. 20, pp. 1485–527.

—— (1983). 'Externalities and corrective policies in experimental markets.' Economic Journal, vol. 93, pp. 106–27.

—— (1987). 'Dimensions of parallelism: some policy applications of experimental methods.' In (A. E. Roth, ed.) *Laboratory Experimentation in Economics: Six Points of View*, Cambridge: Cambridge University Press, pp. 193–219.

Pommerehne, Werner W., Schneider, Friedrich and Zweifel, Peter (1982). 'Economic theory of choice and the preference reversal phenomenon: a reexamination.' *American Economic Review*, vol. 72, pp. 569–74.

Quizon, Jaime B., Binswanger, Hans P. and Machina, Mark J. (1984). 'Attitudes toward risk: further remarks.' Economic Journal, vol. 94, pp. 144–8.

Quiggan, J. (1982). 'A theory of anticipated utility.' *Journal of Economic Behavior and Organization*, vol. 3, pp. 225–43.

Radner, Roy and Schotter, Andrew (1987). 'The sealed bid mechanism: an experimental study.' Research report no. 87-41, Department of Economics, NYU.

Rapoport, Anatol and Chammah, Albert M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*, Ann Arbor: University of Michigan Press.

Rapoport, Amnon, Weg, Eythan and Felsenthal, Dan S. (1988). 'Effects of fixed costs in two-person sequential bargaining.' Mimeo, Department of Psychology, University of North Carolina.

Reilly, Robert J. (1982). 'Preference reversal: further evidence and some suggested modifications in experimental design.' *American Economic Review*, vol. 72, pp. 576–84.

Roemer, John E. (1988). 'Axiomatic bargaining theory on economic environments.' *Journal of Economic Theory*, forthcoming.

Rosenthal, Robert (1980). 'New equilibria for noncooperative two-person games.' *Journal of Mathematical Sociology*, vol. 7, pp. 15–26.

Roth, Alvin E. (1979). *Axiomatic Models of Bargaining*, Lecture Notes in Economics and Mathematical Systems no. 170, Springer Verlag.

—— (1987*a*). 'Bargaining phenomena and bargaining theory.' In (A. E. Roth, ed.) *Laboratory Experimentation in Economics: Six Points of View*. Cambridge: Cambridge University Press, pp. 14–41.

—— (1987*b*). 'Laboratory experimentation in economics.' In (Truman Bewley, ed.) *Advances in Economic Theory, Fifth World Congress*. Cambridge: Cambridge University Press, pp. 269–99. (Preprinted in *Economics and Philosophy*, vol. 2, 1986, pp. 245–73.)

—— and Malouf, Michael W. K. (1979). 'Game-theoretic models and the role of information in bargaining'. *Psychological Review*, vol. 86, pp. 574–94.

—— and —— (1982). 'Scale changes and shared information in bargaining: an experimental study,' *Mathematical Social Sciences*, vol. 3, pp. 157–77.

—— —— and Murnighan, J. Keith (1981). 'Sociological versus strategic factors in bargaining.' *Journal of Economic Behavior and Organization*, vol. 2, pp. 153–77.

—— and Murnighan, J. Keith (1978). 'Equilibrium behavior and repeated play of the prisoner's dilemma.' *Journal of Mathematical Psychology*, vol. 17, pp. 189–98.

—— and —— (1982). 'The role of information in bargaining: an experimental study.' *Econometrica*, 50, pp. 1123–42.

—— —— and Schoumaker, Francoise (1988). 'The deadline effect in bargaining: some experimental evidence.' *American Economic Review*, forthcoming.

—— and Rothblum, Uriel G. (1982). 'Risk aversion and Nash's solution for bargaining games with risky outcomes.' *Econometrica*, vol. 50, pp. 639–47.

—— and Schoumaker, Francoise (1983). 'Expectations and reputations in bargaining: an experimental study.' *American Economic Review*, vol. 73, pp. 362–72.

Rubinstein, Ariel (1982). 'Perfect equilibrium in a bargaining model,' *Econometrica*, vol. 50, pp. 97–109.

Sauermann, Heinz and Selten, Reinhard (1959). 'Ein Oligolpolexperiment.' *Zeitschrift fur die Gesamte Staatswissenschaft*, vol. 115, pp. 427–71.

Scherr, Bruce A. and Babb, Emerson M. (1975). 'Pricing public goods: an experiment with two proposed pricing systems.' *Public Choice*, vol. 23, pp. 35–48.

Schneider, Friedrich and Pommerehne, Werner W. (1981). 'Free riding and collective action: an experiment in public microeconomics.' *Quarterly Journal of Economics*, vol. 116, pp. 689–704.

Selten, Reinhard (1978). 'The chain-store paradox.' *Theory and Decision*, vol. 9, pp. 127–59.

—— and Stoecker, Rolf (1986). 'End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach.' *Journal of Economic Behavior and Organization*, vol. 7, pp. 47–70.

Sherman, Roger (1969). 'Risk attitude and cost variability in a capacity choice experiment.' *Review of Economic Studies*, vol. 36, pp. 453–66.

Shubik, Martin (1955). *Strategy and Market Structure*, New York: Wiley.

Siegel, Sidney and Fouraker, Lawrence E. (1960). *Bargaining and Group Decision Making: Experiments in Bilateral Monopoly*, New York: McGraw-Hill.

Slovic, Paul (1975). 'Choice between equally valued alternatives.' *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, pp. 280–7.

—— and Lichtenstein, Sarah (1968). 'Relative importance of probabilities and payoffs in risk taking.' *Journal of Experimental Psychology Monograph Supplement*, vol. 78, no. 3 (Part 2), pp. 1–18.

—— and —— (1983). 'Preference reversals: a broader perspective.' *American Economic Review*, vol. 73, pp. 596–605.

Smith, Vernon L. (1962). 'An experimental study of competitive market behavior.' *Journal of Political Economy*, vol. 70, pp. 111–37.

—— (1979*a*). 'Incentive compatible experimental processes for the provision of public goods.' In (V. L. Smith, ed.) *Research in Experimental Economics*, vol. 1. Greenwich: JAI Press, pp. 59–168.

—— (1979*b*). 'An experimental comparison of three public good decision mechanisms.' *Scandinavian Journal of Economics*, vol. 81, pp. 198–215.

—— (1980). 'Experiments with a decentralized mechanism for public good decisions.' *American Economic Review*, vol. 70, pp. 584–99.

—— (1982). 'Microeconomic systems as an experimental science.' *American Economic Review*, vol. 72, pp. 923–55.

Stahl, Ingolf (1972). *Bargaining Theory*. Economic Research Institute, Stockholm.

Stone, Jeremy J. (1958). 'An experiment in bargaining games.' *Econometrica*, vol. 26, pp. 286–97.

Straffin, Philip D., Jr. (1980). 'The prisoner's dilemma.' *UMAP Journal*, vol. 1, pp. 102–3.

Sutton, John (1986). 'Non-cooperative bargaining theory: An introduction.' *Review of Economic Studies*, vol. 53, pp. 709–24.

Sweeney, John W., Jr. (1973). 'An experimental investigation of the free-rider problem.' *Social Science Research*, vol. 2, pp. 277–92.

Terhune, K. W. (1968). 'Motives, situations, and interpersonal conflict within prisoner's dilemma.' *Journal of Personality and Social Psychology Monograph Supplement*, vol. 8, pp. 1–24.

Thaler, Richard (1980). 'Toward a positive theory of consumer choice.' *Journal of Economic Behavior and Organization*, vol. 1, pp. 39–60.

—— (1987). 'The psychology of choice and the assumptions of economics.' In (A. E. Roth, ed.) *Laboratory Experimentation in Economics: Six Points of View*. Cambridge: Cambridge University Press, pp. 99–130.

Tucker, A. W. (1950). 'A two-person dilemma.' Mimeo, Stanford University. Published under the heading 'On jargon: the prisoner's dilemma.' *UMAP Journal*, vol. 1, 1980, p. 101.

Van Huyck, John B., Battalio, Raymond C. and Beil, Richard O. (1987). 'Keynesian coordination games, strategic uncertainty, and coordination failure.' Mimeo, Department of Economics, Texas A&M University.

Vickrey, W. (1961). 'Counterspeculation, auctions, and competitive sealed tenders.' *Journal of Finance*, vol. 16, pp. 8–27.

Wilson, Robert (1977). 'A bidding model of perfect competition.' *Review of Economic Studies*, vol. 4, pp. 511–8.

—— (1985). 'Reputations in games and markets.' In (A. E. Roth, ed.) *Game-Theoretic Models of Bargaining*. Cambridge: Cambridge University Press, pp. 27–62.

Yaari, Menahem E. (1987). 'The dual theory of choice under risk.' *Econometrica*, vol. 55, pp. 95–115.

—— and Bar-Hillel, M. (1984). 'On dividing justly.' *Social Choice and Welfare*, vol. 1, pp. 1–24.