Economic Science:
An Experimental Approach for Teaching and Research

Charles A. Holt

A. Willis Robertson Professor of Political Economy
Department of Economics
University of Virginia
Charlottesville, VA 22904
cah2k@virginia.edu
http://www.people.virginia.edu/~cah2k

January 2003

JEL code: C7, C9

Acknowledgement Footnote: Most of the research discussed in this paper is joint research with the coauthors cited in the text, and I am indebted to them for lots of good ideas and hard work. This work was funded in part by a National Science Foundation Infrastructure Grant (SES 0094800).

Abstract

Economics experiments have been used to study theories and policies that are often difficult to evaluate with data from naturally occurring markets. This paper is a selective summary of my work in the main areas of experimental economics: markets, individual decisions, games, and public choice. The discussion pertains both to scientific discoveries and teaching innovations.

1. Introduction

My doctoral dissertation was a study of the effects of contract provisions on competitive bidding (Holt, 1979). The resulting papers contained the word “auction” in the titles, which seems to have caused journal editors to begin sending me experimental papers with similar sounding titles to be reviewed. I became interested in laboratory methods after observing that subjects’ bidding strategies were approximately linear functions of bidders' private values, as would be predicted by a Nash equilibrium assuming risk neutrality (Vickrey, 1962). Moreover, bids in private-value auctions were biased away from Nash predictions in a systematic pattern of over-bidding (Coppinger, Smith, and Titus, 1980). I wrote Vernon Smith at the time and noted that this bias might be explained by the incorporation of risk aversion into Vickrey's original game-theoretic analysis (Holt, 1980). I then arranged for Smith to discuss his work on auctions in a seminar at the University of Minnesota, where I was teaching. We spend some time discussing how risk aversion might be induced in the laboratory. His enthusiasm was contagious, and I have been using experiments in my teaching and research ever since.

2. Vernon Smith and Market Efficiency

Some of my first experiments were motivated by Smith's (1962) discovery of the surprisingly competitive tendency of "double-auction" markets in which buyers and sellers would bargain through a centralized listing of bids, asks, and trades. Like most other economists, I always began a classroom discussion of the competitive equilibrium (C. E.) model of supply and demand with a list of extreme assumptions. In particular, these included the notions of perfect information and "large" numbers of traders. In contrast, Smith (1982) observed that "Markets economize on information in the sense that strict privacy together with the public messages of the market are sufficient to produce competitive C. E. outcomes."

Smith's key insight was that traders have good information about the going market conditions, i.e. about bid, ask, and agreed-on contract prices. This work was motivated, in part, by Chamberlin's (1948) seminal market experiments in which participants could wander about the classroom and negotiate in dispersed groups. This locational decentralization sometimes produced a range of prices that permitted low-value buyers to make purchases at prices that were below the competitive equilibrium, and that permitted high-cost sellers to get into the action at supra-competitive prices. In this manner, price dispersion can result in a loss of efficiency associated with the extra transactions that would be excluded in a competitive equilibrium with a uniform price. Smith's experiments generated more price uniformity and high efficiency by 1) forcing all bids and asks to be centrally announced, and 2) re-opening the market in a sequence of "periods" or "trading days." In retrospect, it is easy to understand how repetition may promote price uniformity, as a seller who agrees to a low price in one period may refuse to go that low if other sellers are observed to obtain higher prices. Conversely, buyers who pay relatively high prices will be more cautious if many others are seen to have secured lower prices. These arguments also suggest how the provision of good information about the going terms of trade will enhance price uniformity and trading efficiency. A comparison of market performance with and without centralized information was the beginning of an impressive string of studies by Vernon Smith and coauthors on the effects of changes in trading rules on market outcomes. This line of research was recognized in Smith's 2002 Nobel Prize in Economics.

3. Teaching from the Trading Pit

Experimental work is having a major effect on the way economics is taught, as professors try to integrate active-learning exercises into material that is often formal and abstract. Indeed, if I had only one lecture to give to a class, I would begin it with a "pit market" trading experiment (Holt, 1996). Figure 1 shows the results of a classroom pit market experiment reported in Holt (2003). The students were given buyer and seller roles and were led to a crowded trading area (the "pit") in front of the class. There were four buyers with $10 cards, who were told that they could "keep the difference" if they could make a purchase for less than $10. There were also four sellers with $8 cards, who were told that they could keep the difference if they could make a sale at a price above $8. If the market had only consisted of these people, then prices would have been in the $9 range, with four units traded and a profit of about $1 per person. In addition, however, there were also 4 buyers with $4 cards and four sellers with $2 cards. If these people had been isolated from those with higher values and costs, then the result would have been prices in the $3 range, with earnings of about $1 per person. All together, the 4 trading pairs with high cards would earn a total of $8, and the 4 trading pairs with low cards would earn $8, for a total of $16. The total quantity traded would be 8 units, and the prices would range from $3 to $9, yielding a high variability in the aggregate.

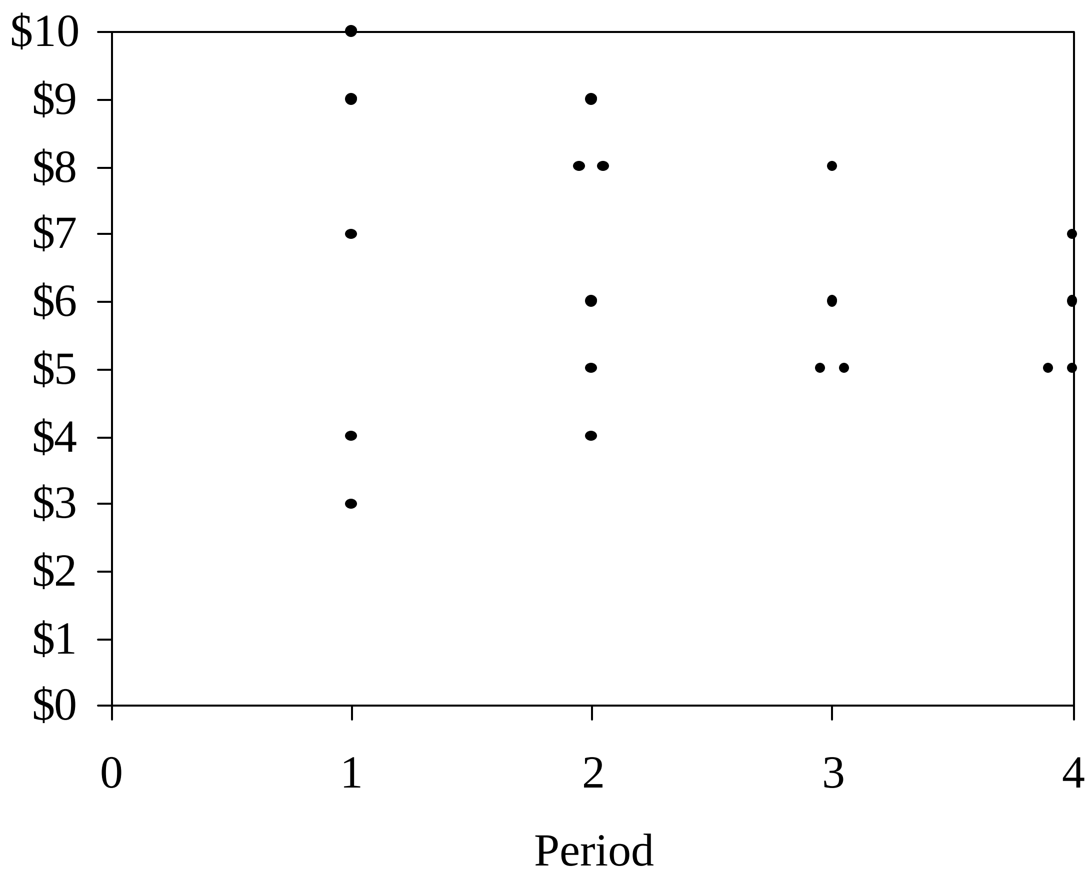

Figure 1. The Contract Price Sequence for the Classroom Experiment

The trading pit prevented the isolation of traders into two groups, since the buyers and sellers facing each other soon dissolved into a fluid mix of changing groups and loud negotiations. When a buyer and a seller agreed on a price, they came together to the recording desk, where the price was checked, called out, and written on the blackboard. The dots on the left side of Figure 1 represent contract prices in the first trading period, where prices ranged from $3 to $10. Price variability declined in subsequent periods, and there were four units traded in the $5-$7 range in the final period. With prices in this range, the high-cost ($8) sellers are excluded, as are the low-value ($4) buyers. Thus it was the four sellers with costs of $2 trading with the four buyers who had high values of $10. Each buyer-seller pair had an $8 difference ($10 - $2) between the buyer value and the seller cost, so the earnings total (for the four traded units) was $32, which is exactly double what would have been observed if the high-cost sellers had traded with the high-value buyers, and if the low-cost sellers had traded with the low value buyers. This exercise illustrates how the pressures of market trading promote a price uniformity that enhances the wealth created by the market.

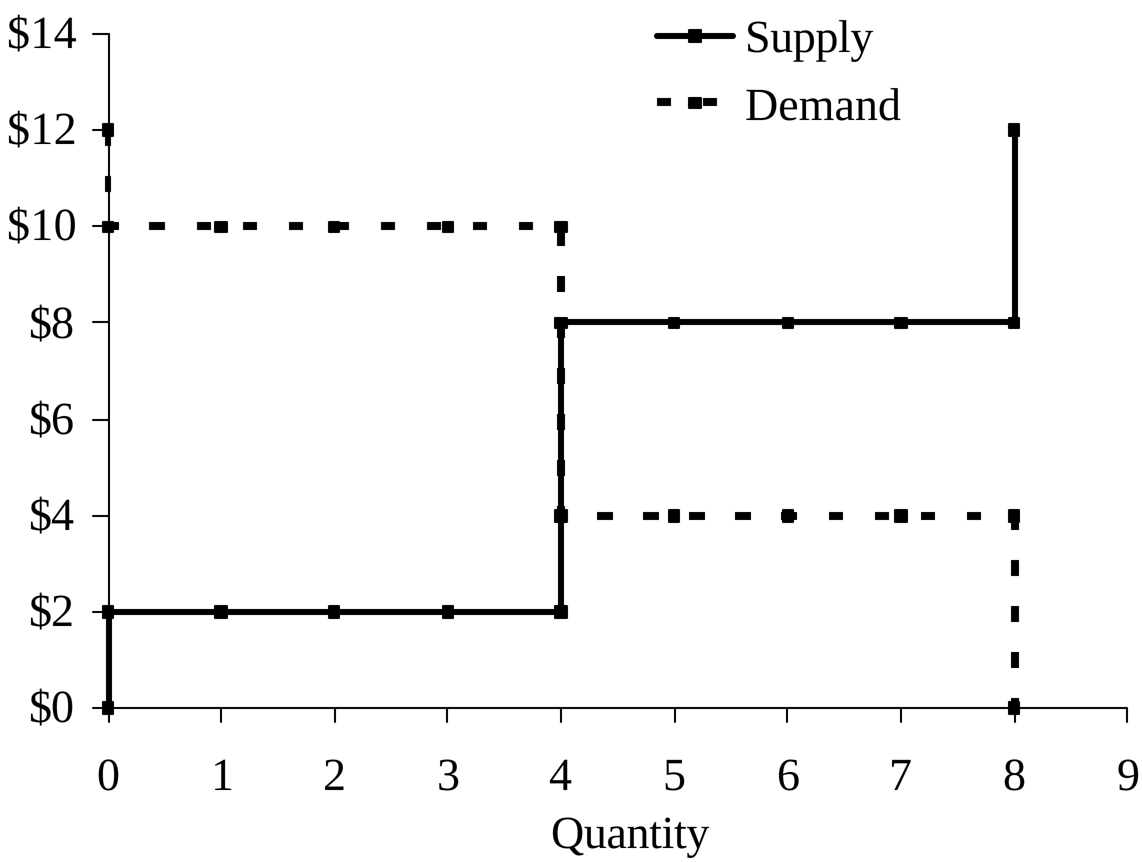

Figure 2. The Structure of the Pit Market Classroom Experiment

The most important secret involved in teaching with classroom experiments is to resist that professorial impulse to show off your knowledge by rushing to the black board and explaining why economic theory was accurate. I have found that it is best to follow the pit trading exercise with a series of carefully thought-out questions that lead students into a self-discovery of the notions of supply and demand in this context. For example, you might call attention to the first-period prices that were above $8 and ask: if those high prices had persisted, would there have been more willing buyers or more willing sellers. Then you might ask what sellers would do if there were lots of willing sellers and only a few interested buyers at those high prices. These questions lead to the notion that prices would fall until there is more of a balance, which leads students to the type of balanced supply and demand that is inherent in a competitive equilibrium. The final step is to lead them to the construction of the supply and demand graph shown in Figure 2. See Holt (1999) for a more detailed discussion of how to use experiments in class. That issue of the *Southern Economic Journal* also contains a collection of other useful classroom experiments on voting, the law of one price, and macroeconomics.

One of my Virginia colleagues, Bill Johnson, once described the pit market trading exercise that he does as being “the gift that keeps on giving,” since it provides a clear example that students can remember and re-consider later in the semester. For example, the graph in Figure 2 illustrates the notions of consumer and producer surplus, which sum to the $32 earnings amount obtained in the final periods of trading. A tighter band of prices in these rounds prevented the production of inefficient units with costs that were higher than the values for buyers at the margin. The resulting surplus of $32 is double that which would have been achieved if the inefficient units had been traded in a market with wide price variability. This discussion is at the heart of Adam Smith’s insight that the pursuit of private gain may generate wealth, resulting in an outcome that is beneficial from a social point of view. But notice that the pursuit of private gain would not have been good enough if the market had been segmented or decentralized, and experiments have provided key insights into the design of efficient trading mechanisms.

4. Market Power

The most widely used trading institution in experimental work is the “double auction,” which has a little more structure than the give-and-take (sometimes push-and-shove) of the pit market. The buyer side in a double auction is somewhat like an ascending-price (English) auction in which each bidder is free to top the highest current bid as it rises in a succession of upward jumps. The word “double” comes from the fact that sellers are operating in reverse, with a downward movement in asking prices as sellers undercut each other’s offers. The bid-ask spread narrows as bids rise and asking prices fall, and a contract is made when these meet, i.e. when a seller accepts a buyer's bid or a buyer accepts a seller's ask.

In his widely-cited survey of industrial organization experiments, Charles Plott (1982, p. 1486) notes that with double auctions, "...the overwhelming result is that these markets converge to the competitive equilibrium even with very few traders." My first reaction to this result was one of disbelief. What about demand elasticity, concentration indices, market power, and all of those things that we talk about in the industrial organization courses? Notice that the market setup in Figure 2 involved giving each seller a single unit to sell in each period, as was the case in most of Chamberlin's and Smith's early experiments. Clearly, there is no market power in the sense that it cannot be profitable for a seller to restrict quantity in order to raise price when there would be no units remaining to sell at the higher price.

Soon after Smith's visit to Minnesota, I began working with two graduate students, Anne Villamil and Loren Langan, on designing a market structure in which some sellers had large numbers of units that were only marginally profitable at the competitive price. The withholding of these units, therefore, would involve very little loss of earnings, but the resulting leftward shift in supply might raise the price received for the seller's other, low-cost units. The market demand that we came up with was highly inelastic, so that a small quantity restriction would be predicted to have a large effect on price. This demand inelasticity also had the effect of making the buyers' values very high for the units predicted to sell in a competitive equilibrium. Thus it would be quite costly for buyers to withhold purchases in an effort to resist small-to-medium price increases. There were 5 buyers and 5 sellers in all, and the trading was done with a double-auction format. We were encouraged when the very first research session that we conducted resulted in prices that were about ten percent above the competitive prediction. But prices locked tightly onto the competitive prediction in the next session, run on a different day with a new group of 10 traders in buyer and seller roles. This pattern was continued, with about half of the sessions generating small but significant price deviations in the direction predicted by market power, and with a tight convergence to the competitive equilibrium price in the other half of the sessions. Even the markets that produced supra-competitive prices yielded a high level of efficiency, with little if any quantity restriction. And the tight convergence in the others sessions was impressive, given that it occurred in spite of our best efforts to provide traders on one side of the market with a strong market-power advantage (Holt, Langan, and Villamil, 1986).

The Holt, Langan, and Villamil (1986) experiment was replicated by Davis and Williams (1991), who also looked at the effects of changing from the double auction to a market in which sellers post prices simultaneously on a take-it-or-leave-it basis. This "posted-offer auction" produced strong and reliable price increases in the Holt, Langan, and Villamil seller-market-power design. As a result, I began working with Doug Davis, who was nearby at Virginia Commonwealth University. To clarify the effects of market power, we came up with an experimental design that held the aggregate supply-and-demand structure and the trading institution (posted-offer) constant. Market power was created by transferring units from some sellers to others, creating a couple of sellers with shares of market capacity that were high enough to make unilateral quantity restrictions profitable. The creation of market power in this manner resulted in very large price increments over competitive predictions, which disappeared when the capacity concentration was reversed and market power was eliminated without changing the aggregate supply-and-demand structure (Davis and Holt, 1994). Recent work by Rassenti, Smith, and Wilson (2000) has proceeded to explore the effects of market power in the generation and transmission of electric power.

One of the most interesting series of laboratory markets, from the point of view of the experimenter, was one that Doug Davis and I ran with collusion. In particular, we let sellers discuss prices while buyers were taken out of the room on the pretext of making some role assignments. The sellers could talk freely until the buyers were about to return, at which time sellers would go back to the computers that they were using to enter their prices at the start of each round. This collusion was typically successful in the sense that prices ended up being close to the joint-profit-maximization level. In contrast, prices converged to competitive levels in a control treatment without seller discussions.

The collusion experiments were motivated by a belief that secret discounts were strongly pro-competitive, a viewpoint that I encountered while working at the Federal Trade Commission for several summers in the mid-1980's. Once we had established that the supply-and-demand structure in use yielded near-monopoly prices with collusion and near-competitive prices without it, we began a series of sessions that involved collusion with secret discounts. As before, sellers entered their posted prices after their discussions and after the buyers had returned. Then buyers began shopping, but instead of making purchases at the posted prices, any buyer could request a discount from a particular seller. Sellers could offer price reductions or not, with the confidence that any reduction was not observed by other buyers and sellers. Thus discounts were secret, selective, and made in a sequential manner. The sellers would typically establish a common posted price with some quantity restriction in the early periods, but discounts from these posted prices were common. Deep discounts resulted in sales imbalances across sellers, and those with lower sales in one period were quick to discount in the next, despite discussions of the importance of sticking with agreed-on prices. Sellers responded to these discounts by lowering the common list price. In some cases, the price was fixed and common to all conspirators, but it was *fixed at an essentially competitive level*, with discounts driving average transactions to competitive levels. One group became so frustrated that they refused to speak with each other during the discussion periods while buyers were absent (Davis and Holt, 1998).

In conclusion, this line of experimentation led me to understand that double auction markets are quite competitive, not always, but surprisingly so. A key aspect of these markets is the symmetric treatment of buyers and sellers, with all bids, asks, and acceptances being announced and observed. In contrast, when sellers post prices on a take-it-or-leave-it basis, seller concentration can interfere with efficient, competitive outcomes. Even with posted seller prices, however, the ability of sellers to offer secret discounts from those prices may effectively counter the effects of power and collusion. I can now teach the notions of supply and demand with more confidence, and I spend a lot more time on the importance of market institutions in industrial organization classes.

5. Voting and Public Choice

Many people view voting outcomes in democratic meetings as somehow reflecting a group preference that would be otherwise difficult to discover. A more cynical view can be found in the literature on public choice. I became interested in this topic after meeting Charles Plott, who, like Smith, has served as President of the

Southern Economic Association. Plott has been influential in bringing experimental methods to the study of voting and public choice issues. For example, Levine and Plott (1977) had been members of a flying club that was to meet and decide how to spend a large sum of money on a collection of airplanes to be used by the membership. After being appointed to serve on the Agenda Committee, they distributed a survey of members' preferences to assist in structuring the discussion at the meeting. The survey results were used to design an agenda that the authors believed would yield a fleet of new aircraft that they preferred. The president of the club had different preferences and repeatedly tried to deviate from the agenda during the meeting, but was ruled out of order in each case. The authors were asked to resign from the club after an account of the agenda strategy was published in the *Virginia Law Review*.

This field experiment was followed with a series of laboratory experiments in which Plott and Levine (1978) developed a theory of voting behavior in such situations. In order to explain this, consider a simple agenda in which two alternatives, A and B, are pitted against each other in the first-stage vote, with the winner being pitted against C, the incumbent, in the final vote. In this sense, you can think of the initial vote as being between {A, C} and {B, C}, i.e. between which second-stage vote to hold. A commonly mentioned notion in the political science literature is that of "sincere voting," which in this context would mean voting in the first stage for the option, A or B, that is most preferred (without consideration of whether it will win in the second stage). Of course, not everyone will vote sincerely, and the Plott-Levine theory was designed to allow different types of behavior. The main idea was that some people would vote for the set that contains their most-preferred option, some would vote against the set with their least-preferred option, and some would vote for the set that had the highest average payoff. The average-payoff approach may be appropriate if one believes that each option in the set selected will have an equal probability of winning in the final stage, a kind of naïve-expectations approach. Each of these behavioral rules has some intuitive appeal, and Plott and Levine observed some votes consistent with each of these three heuristics.

Given my background in economic theory, I was curious about what rational behavior would be in such cases, especially when expectations about the second-stage vote satisfy some notion of rational expectations. With only two options in the final stage, it is easy to imagine that people (in the final stage) will vote for the one that they prefer, which yields a unique final-stage outcome with an odd number of voters. Then a strategic voter might vote for the set, {A, C} or {B, C} that is expected to produce the best outcome for them in the final stage. For example, suppose that the person prefers A to B to C, but the others' preferences are such that C would beat A in the final-stage vote, and that B would beat C. Then even though the voter's top choice is A, it would be best to vote for {B, C} over {A, C} in the first stage in order to avoid the worst outcome. In this manner, it is easy to see how strategic voting may differ from sincere voting. The relevance of this observation is that you have to know whether people will vote naïvely or strategically if you are going to use the agenda to manipulate the outcome.

Eckel and Holt (1989) designed a theory to test the Plott and Levine theory by using a preference profile for which the predicted outcome with strategic voting was different from the outcome predicted from any of the three behavioral types in the Plott/Levine theory and from any mix of these types. There were nine students serving as committee members in each session. Students' preferences were induced by telling them how much money they would earn for each of the three possible committee decisions. Students selected a chair for the meeting and then were free to discuss the issues prior to voting in the manner prescribed by the agenda. In some sessions, there was a student "monitor" who looked at each person's earnings calculation sheet and made a public report of each person's (ordinal) preferences. To our surprise, the voting outcome was *never* strategic in the first meeting, even when they were given information about each other's preferences. This provided strong support for the Levine/Plott theory that used a mix of naïve voting heuristics. But after several meetings with the same agenda, the voting outcome would switch to the one predicted by strategic behavior. Subjects were, therefore, unable to perform backward-induction calculations in this situation, and they had to learn from experience.

Voting experiments are useful in teaching, where a richer context than the usual (A, B, C terminology) is appropriate to raise student interest. Holt and Anderson's (1999) paper in this journal provides a convenient setup involving playing cards that are related to preferences for public spending on "schools," "roads," etc. The agendas provided can be used to illustrate voting cycles, agenda effects, and the difference between strategic and sincere (naï ve) voting. Another useful classroom voting experiment is Hewett et al. (2003), which is used to discuss the Tiebout hypothesis. Here people are given playing cards that determine preferences for one of four public goods: Hearts, Spades, Diamonds, or Clubs. The number on the card determines the intensity of preference, and cards are distributed randomly, with each person getting 2-3 cards. People are then divided into several "towns," where they meet, elect a mayor, and decide on a type and level of one (and only one) of the four possible public goods. High levels result in high taxes, according to an announced formula. After all towns have made their decisions, the results are announced, and people are free to move to a town with a tax and public good choice that is more to their liking. The newly configured towns then meet and vote again, which may change the results. Nevertheless, almost all people are made better off by the moving process, and this can lead naturally to a discussion of the Tiebout hypothesis and related notions of efficient public goods provision. Finally, the preferences of each person, as determined by the cards, determine a median-voter prediction for the level of each public good, and discussion of the "median voter theorem" takes almost no time after people have observed outcomes that are often close to this prediction. This is a fun experiment to run outdoors on a nice day, and the instructions for it are available on my web page:

http://www.people.virginia.edu/~cah2k

along with papers describing many other "hand-run" experiments.

I have also written a web-based voting program that lets the instructor supply names for each of the options under consideration by a committee. The voting institutions that can be used include a two-stage agenda and a simple plurality vote (with or without a runoff if no option receives a majority). In addition, it is possible to include a non-binding opinion poll prior to the vote, or to include a randomly determined cost of voting that is not incurred if the person decides not to vote. Finally, there is an "approval voting" option, in which a voter may vote "approve" for one or more options, and the one with the most approval votes from the committee as a whole is selected. (The Economic

Science Association uses approval voting to select its "Section Heads.") This voting program is freely available on the internet; the instructor would go to:

http://veconlab.econ.virginia.edu/admin.htm

and select the Public Choice Menu. The students would later log on from: **http://veconlab.econ.virginia.edu/login.htm** and select the Public Choice Menu. The other public choice programs include: a common-pool resource game, a public goods (voluntary-contributions) game, rent seeking, and the "volunteer's dilemma." In addition, there are over 20 other programs at this site for those who wish to use web-based experiments in the classroom; these include Monopoly, Cournot, Bertrand, Double Auction, Bargaining, Auctions, Signaling, Trust and Reciprocity, Statistical Discrimination, a flexible Matrix Game (e.g. Prisoner's Dilemma), and the Traveler's Dilemma to be discussed in the next section.

6. Games: Nash and Beyond

Game theory is sometimes defended as being a "normative" theory about how rational people should play against rational opponents. This theory is often able to provide precise predictions about outcomes in mathematical models of strategic situations, and these predictions have been used to evaluate aspects of public policies on antitrust, tort law, etc. These policy implications may be seriously flawed if the theory provides poor predictions. In this sense, the way game theory is used (outside of pure mathematics) is only appropriate if it is a theory with *positive*, predictive value. It is easy to be skeptical, since even a perfectly rational person may not want to follow the prescriptions of game theory if other players might not be rational. There is clearly an important role for experimental work in this area, and in fact, experiments with real cash payoffs were being run by economists and mathematicians at the RAND Corporation even prior to John Nash's (1950) seminal paper that led to the notion of a Nash equilibrium. One of these experiments was devised on the same day that two researchers heard about Nash's notion of a noncooperative equilibrium (see Al Roth's introduction in Kagel and Roth, 1995). The payoffs for that experiment were later used to devise the well-known story of the Prisoner's Dilemma, a game that has been used extensively in experimental studies ever since. In the Prisoner's Dilemma, the optimal choice for each player is to defect, regardless of what the other person decides to do. The discussion in this section is largely based on a different game, the Traveler's Dilemma, in which the optimal decision *does* depend on what the other player is expected to do. The Traveler's Dilemma is similar to a Prisoner's Dilemma in the sense that the unique game-theoretic prediction is an outcome that is much worse for both players than an outcome that could be reached by cooperative behavior.

To get a feel for the Traveler's Dilemma, you may wish to play an on-line demonstration of this game. There is no required registration or password for this demonstration; just go to the "admin.htm" site provided in the previous section, click on **Guide to Experimenters**, and then on the **On-line Demonstration for a Traveler's Dilemma**. Alternatively, you can go straight to the demo by typing:

http://veconlab.econ.virginia.edu/tddemo.htm

It will take no more than a few minutes to play this game, which will give you a clear picture of the strategic setting and how the Veconlab software works. You will be playing against a sequence of University of Virginia Law School students who took a class in Behavioral Game Theory that I taught last fall. The students were taking notes with laptop computers, about half of which had wireless access to the internet. Therefore, I divided them up into 10 pairs or "teams," each with a wireless-connected laptop computer. The teams were randomly matched in a sequence of one-shot games. When you log in, you will temporarily replace one of these teams in the database, and you will meet the same sequence of other teams that the person you replaced met. After each of your decisions is entered and confirmed, you will find out the decision of other team with whom you are matched in that period. After you finish, you can compare your earnings for the 5 periods with those of the law student team who met the same sequence of other decisions that you encountered. (There is also a Traveler's Dilemma game on the Veconlab web site that you can use to let people in your class play against each other. In the setup, you specify the number of participants and the number of rounds for each of two treatments. The treatments can have different penalty-reward rates and ranges of permitted claims.)

The Traveler's Dilemma game, due to Basu (1995), is motivated by the story of two travelers who go on a tropical vacation and purchase identical items, which are lost on the return trip. The airline representative asks each person to go into a separate room to fill out a claim, with the understanding that claims must be at least 80 and no greater than 200 (here I am using numbers from the experiment to be discussed). Any pair of claims in this range will be fully reimbursed if they are equal. But if one claim is higher than the other, then the airline representative will infer that the higher one is inflated, and will reimburse both people at the lower of the two claims. In addition, the lower claimant will receive a small reward (5), and the higher claimant will incur a small penalty (5) deducted from the reimbursement (that equals the minimum of the two claims). For example, claims of 120 and 130 will result in payments of 125 to the low claimant and 115 to the high claimant. Obviously, each person has an incentive to "undercut" the other, so no common claim above 80 can constitute a Nash equilibrium. In fact, 80 is the unique Nash equilibrium in pure or mixed strategies in this game, and it provides payoffs of 80 that are much lower than the 200 amount that would result from cooperation. If you already played the on-line demonstration game, you will have noticed that the others' decisions were generally nowhere near the Nash prediction, and that there was no clear movement in that direction.

Figure 3 shows the results of the Capra et al. (1999) research experiment with groups of ten subjects who were randomly matched in a sequence of ten periods. At the start of each period, all subjects would choose claims on the range from 80 to 200 pennies. The penalty/reward rate was changed from one session to another, but none of these changes altered the unique Nash equilibrium, which remained at the lowest permitted claim. With a high penalty/reward parameter of 80 cents, the claims averaged about 120 in the first round and fell to near-Nash levels in the final rounds, as shown by the thick solid line at the bottom of the figure. The data for the 50-cent treatment show a similar pattern. In contrast, the data for the 5 and 10-cent treatments started at about a dollar above the Nash prediction and actually rose slightly, moving *away* from the Nash prediction. The data for the intermediate treatments (20 and 25) are not shown, but they stayed in the middle range ($1.00 to $0.50) below the dashed lines and above the solid lines, with some more variation and crossing over.

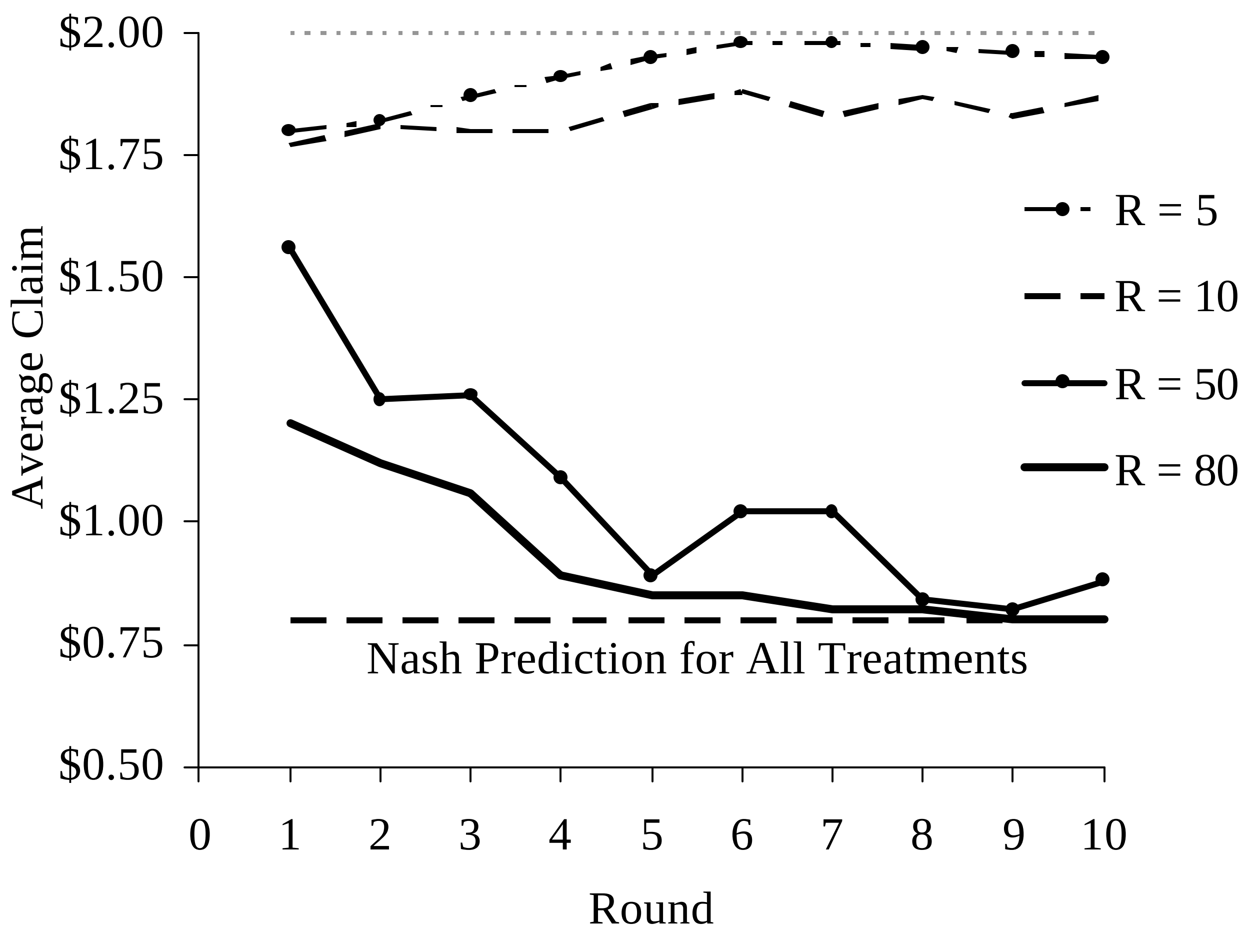

Figure 3. Traveler's Dilemma Data Averages (Capra et al. 1999)

I used to believe that the choices of subjects in experimental games would eventually converge to Nash predictions with enough repetition in a random-matching protocol, at least as long as issues of relative payoffs and fairness did not intervene. The Traveler's Dilemma data seem to contradict this belief, since average claims are diverging from Nash predictions for relatively low values of the penalty-reward parameter. This is my *favorite* game, and it comes to mind every time that I see a theoretical paper in which some mathematical refinement or learning model is proved to converge to a Nash equilibrium.

The data in Figure 3 do show patterns that are consistent with a type of intuition; it is more likely that people will take the risk of making a high claim if the penalty for being high is not very large. The trouble with standard game theory is that the predictions depend on the *sign* of the payoff difference, not on the *magnitude*. But payoff magnitudes have a strong effect on the outcomes of the Traveler's Dilemma and on other games like the minimum-effort coordination game (Anderson, Goeree, and Holt, 2001; Goeree and Holt, 1999b, 2001). The challenge for game theory is to generalize the notion of a Nash equilibrium so that it is sensitive to payoff magnitudes, with the goal of having a single theory that explains the data patterns that converge to Nash predictions and those that do not. My work with various coauthors has made me a lot more optimistic that this will be possible. For example, Capra et al. (1999) and Goeree and Holt (1999a) report how Traveler's Dilemma data averages are consistent with predictions of the ***quantal response equilibrium*** (McKelvey and Palfrey, 1995). Goeree and I have developed related models of learning (for games with repeated interactions) and introspection (for games played only once); see Goeree and Holt (2000b, 2000d, 2001) and Capra et al. (2002).

The key element in all of these models is the notion of probabilistic choice, i.e. that choices do not respond perfectly to payoff differences, but rather, that correct decisions are more likely when payoff differences are large (Luce, 1959). Payoff magnitudes matter in these models. Formally, the models introduce a (possibly small) amount of randomness. The error terms represent the aspects of the strategic situation that are not explicitly modeled, i.e. a collection of residual effects due to heterogeneity, omitted variables, calculation and recording errors, and random preference shocks. These models predict probability distributions of decisions, and hence, are ideal for estimation using experimental data that typically show some degree of unpredictability.

With a continuum of feasible choices, as in the Traveler's Dilemma, the quantal-response equilibria are density functions. An existence proof, therefore, involves finding a fixed point in a function space, see the Appendix of Anderson, Goeree, and Holt (2002) in this journal. This appendix also uses a proof-by-contradiction approach to derive symmetry, uniqueness, and comparative-statics results. A typical comparative-statics result, for example, is that an increase in the penalty-reward rate in the Traveler's Dilemma game will cause a decrease in claims in the sense of first-degree stochastic dominance. We have applied this approach to the analysis of behavior in coordination games (Anderson, Goeree, and Holt, 2001), rent-seeking (Anderson, Goeree, and Holt, 1998a), bargaining (Goeree and Holt, 2000a), auctions (Goeree, Holt, and Palfrey, 2002), voting (Goeree and Holt, 2000c), matrix games (Goeree, Holt, and Palfrey, 2003), and public goods games (Anderson, Goeree, and Holt, 1998b; Goeree, Holt, and Laury, 2002).

One reaction that economists sometimes have to the introduction of noise is that this will just provide a bell-shaped distribution of decisions around the theoretical prediction in the absence of noise. This is not necessarily true. In the Traveler's Dilemma, for example, suppose that each person expects the other to choose the Nash claim of 80. Then a little noise in a player's beliefs about the other person's decision will move the player's own distribution of choices upward, either a little or a lot depending on the size of the penalty-reward parameter. But then an upward movement in one's own choices, if anticipated by the other person, will move the other's choices upward even more. In this manner, a kind of "snowball" effect might result in a predicted choice distribution that is clustered near the upper end of the set of feasible claims, far from the Nash prediction. This is the intuition for why a little noise (due to numerous, un-modeled effects) may result in a large movement in predicted behavior when there is strategic interaction.

When theorists do inject noise into economic models, they often take it out in the limit, in a process of "purification." This is clearly inappropriate in situations where noise is needed to explain sharp deviations from Nash predictions. If the noise represents un-modeled factors and heterogeneity, then there is no reason to expect it to diminish over time to any great degree. In contrast, our approach is to parameterize the degree of randomness by introducing a noise parameter, e.g. a logit error. Then we typically try to prove theoretical results for any value of the error parameter. In empirical work, the model is solved for the fixed-point distribution of decisions for a specific value of the error parameter, and this solution is used to calculate the likelihood as a function of the decisions observed in the experiment. Then iterative methods are used to estimate the error parameter that maximizes the likelihood function (e.g. Capra et al, 1999, 2002; Goeree, Holt, and Palfrey, 2002, 2003).

7. Incentives and Individual Behavior; Did Hollywood Get It Right?

The last category of experiments that I wish to discuss involves individual decisions, in situations involving risk and uncertainty. There are many situations in which people seem to use heuristics and shortcuts when making such decisions. This is a natural and intuitive point of view, since the calculations involved in making purely rational decisions may be considerable. In evaluating the behavioral relevance of a particular bias or heuristic, it is important to examine the experimental procedures, and in particular, the *incentives*. After all, economic theories are largely focused on the incentives provided by money and the goods and services that it buys. This is not to say that other factors are not important; the approach taken in most economics experiments is to hold the psychological setting constant and vary the economic incentives.

As a graduate student at Carnegie-Mellon in the seventies, I remember being told about a type of clearly irrational behavior called "probability matching." These experiments, which had been conducted by psychologists since before World War II, were typically done by letting a subject guess which of two light bulbs would light up next. For example, the subject might be seated on one side of a vertical piece of plywood, with the experimenter on the other side. The subject would make a decision by pressing a lever on the left or right, and then the experimenter would press a key that would light up a light bulb on one side or the other. The sequence of choices of which bulb to illuminate would typically be predetermined by some random device, with the probability of one of the sides being set to something like 0.75. If one wishes to maximize the number of correct guesses, the optimal decision is to figure out which side is more likely and then predict that side every time thereafter. Thus the proportion of guesses of the more likely event should approach one. An alternative mode of behavior is known as "probability matching," i.e. having the choice proportions for each event match the observed frequencies. For example, if the more likely event was for the right-side light to illuminate about three-fourths of the time, then *probability matching* would involve guessing that side with probability 0.75. Probability matching has been observed in numerous psychology experiments.

A psychologist named Sidney Siegel began running experiments in the 1960's that focused on the old (even at that time) issue of probability matching. In one treatment, he paid a 5-cent reward for each correct guess, and he deducted a 5-cent penalty for each incorrect guess. In a second treatment, he simply told subjects to "do your best." The results of one of these experiments (Siegel, Siegel, and Andrews, 1964) are shown in Figure 4. The dashed line represents the "no-pay" treatment, and it is apparent that the proportion of guesses associated with the more likely event converges to 0.75 after about a hundred trials (the data are averaged over subjects and over 20-trial blocks). In contrast, the choice proportion for the more likely event goes above 0.9 in the "pay/loss" treatment. Here we see the strong effect of economic incentives.

Siegel's findings have been largely ignored by psychologists who continued to run these experiments, both with people and with hungry animals pressing levers that had different probabilities of producing a food pellet. For example, the results of this literature were recently summarized in an experimental psychology journal:

> "…human subjects do not behave optimally. Instead they match the proportion of reinforcement associated with each alternative….This behavior is perplexing given that non-humans are quite adept at optimal behavior in this situation." (Fantino, 1998)

So the issue is why the animals tend to do better than the people! I would bet that the difference is not due to the better reasoning abilities of animals, but to the fact that all animal experiments are done with incentives, i.e. food pellets. You cannot just tell an animal to *do your best*!

A second and related issue is the level of incentives. Experimental economists have been criticized for using relatively low levels of money payments, although there are some notable exceptions. In a paper cited by the 2002 Nobel Prize Committee, Kahneman and Tversky (1979, p. 265) state:

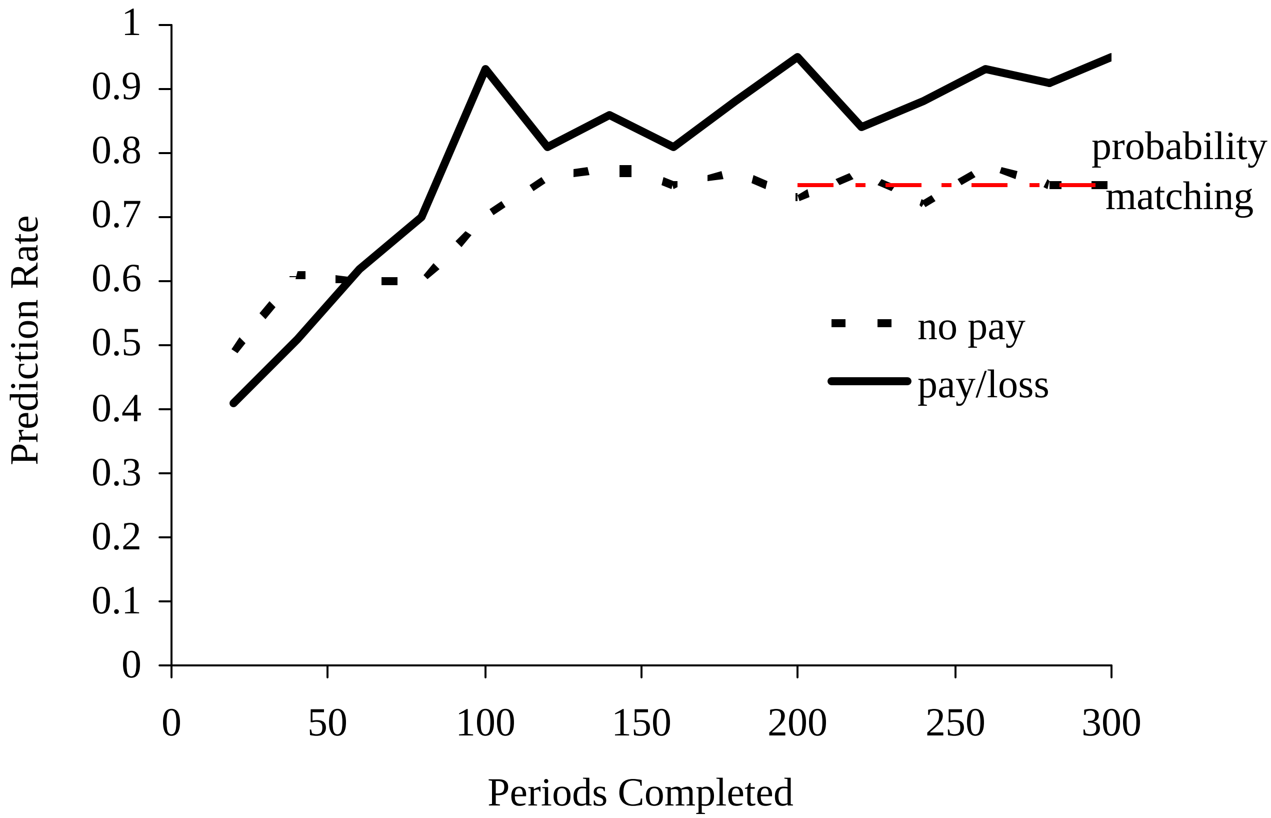

Figure 4. Probability Matching Experiment (Siegel, Siegel, and Andrews, 1964)

> "Experimental studies typically involve contrived gambles for small stakes, and a large number of repetitions of very similar problems. These features of laboratory gambling complicate the interpretation of the results and restrict their generality. By default, the method of hypothetical choices emerges as the simplest procedure by which a large number of theoretical questions can be investigated. The use of this assumption relies on the assumption that people often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reason to disguise their true preferences."

I agree that it is especially interesting to study behavior in high-risk settings, since many important economic decisions are made under high-payoff conditions. The work of Siegel and others, however, suggests that there may be a real danger of using questions with hypothetical incentives, even if people have no incentive to deceive the person administering the questionnaire. The proposition that behavior might be dramatically different when high hypothetical stakes become real is echoed in the film *Indecent Proposal*:

> John (a.k.a. Robert): Suppose I were to offer you one million dollars for one night with your wife.
>
> David: I'd assume you were kidding.
>
> John: Let's pretend I'm not. What would you say?
>
> Diana (a.k.a. Demi): He'd tell you to go to hell.
>
> John: I didn't hear him.
>
> David: I'd tell you to go to hell.
>
> John: That's just a reflex answer because you view it as hypothetical. But let's say there were real money behind it. I'm not kidding. A million dollars. Now, the night would come and go, but the money could last a lifetime. Think of it – a million dollars. A lifetime of security for one night. And don't answer right away. But consider it – seriously.

In the film, John's proposal was ultimately accepted, which is the Hollywood answer to the incentives question. On a more scientific note, incentive effects are an issue that can be investigated with experimental techniques.

Some psychologists and experimental economists have argued that incentives may not matter in many cases. For example, Tversky and Kahneman (1992, p. 315) note:

> "In the present study we did not pay subjects on the basis of their choices because in our experience with choice between prospects of the type used in the present study, we did not find much difference between subjects who were paid a flat fee and subjects whose payoffs were contingent on their

> decisions. ... Although some studies found differences between paid and unpaid subjects in choice between simple prospects, these differences were not large enough to change any significant qualitative conclusions."

Comments like these raise some important questions. It could be the case that people can guess pretty well what they would do for choices of gambles involving several dollars, but like David and Diana in the film, they behave differently when the stakes are high and real.

Susan Laury and I recently investigated some of these issues in the context of a series of choices between two risky gambles. Payoffs for the low-payoff condition are shown in Table 1. The risky choice involved a probability *p* of \$3.85 and 1-*p* of \$0.10. The safe choice involved a probability *p* of \$2.00 and 1-*p* of \$1.60. Subjects in the Laury and Holt (2002) study were given a menu of ten choices between these two lotteries, with the probability of higher payoff ranging from 1/10, 2/10, … 1. They were told that one of these 10 choices would be selected at random, *ex post*, to be used to determine the person's cash earnings. For low values of *p*, essentially everybody chose the safe lottery, and for high values of *p* essentially everybody chooses the "risky" lottery. A risk neutral person will make four safe choices and switch to the risky lottery as soon as *p* = 0.5. Thus the number of safe choices can be used to infer risk aversion, with 4 indicating risk neutrality.

Table 1. A Paired Lottery Choice

Safe Choice	**Risky Choice**
\$2.00 with probability ***p***	\$3.85 with probability ***p***
\$1.60 with probability 1-***p***	\$0.10 with probability 1-***p***

It is straightforward to show that 5 safe choices indicates a small amount of relative risk aversion (about 0.3), and 6 safe choices indicates a fairly large amount (about 0.5) that corresponds to a "square-root" utility function with significant curvature.

The average number of safe choices turned out to be about 5, indicating small but significant risk aversion. When all payoffs were multiplied by a factor of 20, so that the highest payoff was $77, the average number of safe choices increased to about 6, as shown in the first row of Table 2. Further increases in risk aversion were observed as the payoffs were scaled up by factors of 50 and 90 (generating a high payoff of over $300).

Table 2. Average Numbers of Safe Choices: Differing Payment Conditions
(Source: Holt and Laury, 2002)

Payoffs	1x	20x	50x	90x
Full cash for 1 of 10 decisions	5.2	6.0	6.8	7.2
Hypothetical	-	4.9	5.1	5.3

To summarize the results thusfar, it is clear that increases in incentives resulted in dramatic increases in risk aversion. Next consider the effects of offering no incentives, and scaling up the hypothetical payoffs. Everybody began with a real-payoff choice under the low-payoff condition (1x payoffs). Then we asked them to think carefully about what they would do with scaled-up payoffs (20x in some sessions, and 50x or 90x in other sessions), but with the understanding that we were not going to pay actual earnings for these choices. These hypothetical choices were collected and used to determine earnings (not paid) prior to presenting them with the real-payoff menus that were used to get the high-real-payoff results discussed in the previous paragraph. The results of scaling up the hypothetical payoffs are shown in the bottom row of Table 2. If you compare the numbers in this row, 4.9, 5.1, and 5.3, it seems that scaling up payoffs has no clear effect on risk aversion, and that the choices look essentially the same as the average, 5.2, for the low-real-payoff condition. If we had only done the study with low real payoffs of several dollars and successively higher hypothetical-payoff choices, then we might have been tempted to reach the *incorrect* conclusion that it is not necessary to pay in cash. But this conclusion would have been reached by changing two things at the same time, the nature of payoffs (real or hypothetical) and the scale, and then attributing the effects to one of those changes. If you compare high real payoffs with high hypothetical payoffs in each column of the table, however, it is clear that the incentive condition (real or hypothetical) matters a lot, holding scale constant. Hollywood got it right.

8. Experimental and Behavioral Economics: A Bag of Tricks or a New Set of Glasses

Seminar presentations of laboratory results are often focused on the unexpected results, i.e. those that run counter to economic theory. This can leave the audience with the impression that advances in experimental and behavioral economics have produced little more than a "bag of tricks" or anomalies. The subtext is that preferences are unstable and complex, and that formal economic theory might be replaced by a collection of context-specific insights.

There are certainly lots of surprises and unexplained data patterns in the experiments that I have run and studied. I am still puzzled and fascinated by many of these anomalies, but the impression I am left with is that there is often an underlying consistency that suggests a modification, not abandonment, of economic theory. For example, all of the ten "intuitive contradictions" in Goeree and Holt (2001) *are* intuitive, just as the tendency for claims to rise well above Nash levels in the Traveler's Dilemma when the penalty for being high is not so great. The Nash equilibrium can be (and has been) generalized to allow for behavior that is sensitive to the magnitudes of payoff differences in a probabilistic manner, and the result is a single theory that explains data patterns that converge to Nash predictions in some treatments and those that diverge sharply in other treatments. Moreover, these theories are simple enough to be explained to students using simple spreadsheet iterations (see Chapter 12 in Holt, 2003). The result is a new game theory that is useful "for playing games, not just for doing theory" (Goeree and Holt, 1999a, p. 10567).

In reviewing the results of various experiments, it is important to examine the procedures and the economic incentives. For example, irrational probability matching behavior in individual choice experiments is greatly diminished when money payments are made. The seemingly paradoxical tendency for animal subjects to perform better than humans in these tasks is due to the fact that real incentives are always used with animals, since you cannot just tell them to "do your best."

Finally, it is reassuring to know that the model of supply and demand is alive and well in appropriate institutional settings. And running experiments in class and in research settings helps both the professor and the students understand how the strategic situation looks from the bottom up, which is important in learning and developing new theoretical perspectives. In this manner, the increased use of experimental methods has provided economists with a new set of glasses with which to view the world and re-evaluate old issues and puzzles.

References

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt. 1998a. Rent seeking with bounded rationality: An analysis of the all-pay auction. *Journal of Political Economy* 106: 828-853.

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt. 1998b. A theoretical analysis of altruism and decision error in public goods games. *Journal of Public Economics* 70: 297-323.

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt. 2001. Minimum-effort coordination games: Stochastic Potential and Logit Equilibrium. *Games and Economic Behavior* 34: 177-199.

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt. 2002. The logit equilibrium: A unified perspective on intuitive behavioral anomalies in games with rank-based payoffs. *Southern Economic Journal* 68: 21-47.

Basu, Kaushik. 1994. The traveler's dilemma: Paradoxes of rationality in game theory. *American Economic Review* 84: 391-395.

Capra, C. Monica, Jacob K. Goeree, Rosario Gomez, and Charles A. Holt. 1999. Anomalous behavior in a traveler's dilemma? *American Economic Review* 89: 678-690.

Capra, C. Monica, Jacob K. Goeree, Rosario Gomez, and Charles A. Holt. 2002. Learning and noisy equilibrium behavior in an experimental study of imperfect price competition. *International Economic Review* 43: 613-636.

Chamberlin, Edward C. 1948. An experimental imperfect market. *Journal of Political Economy* 56: 95-108.

Coppinger, V.M., V. L. Smith, and J. A. Titus. 1980. Incentives and behavior in English, Dutch, and sealed-bid auctions. *Economic Inquiry* 18: 1-22.

Davis, Douglas D. and Charles A. Holt. 1994. Market power and mergers in markets with posted prices. *RAND Journal of Economics* 25: 467-487.

Davis, Douglas D. and Charles A. Holt. 1998. Conspiracies and secret price discounts. *Economic Journal* 108: 736-756.

Davis, Douglas D. and Arlington W. Williams. 1991. The Hayek hypothesis in experimental auctions: Institutional effects and market power. *Economic Inquiry* 29: 261-274.

Eckel, Catherine C. and Charles A. Holt. 1989. Strategic voting behavior in agenda-controlled committee experiments. *American Economic Review* 79: 763-773.

Fantino, Edmund. 1998. Behavior analysis and decision making. *Journal of the Experimental Analysis of Behavior* 69: 355-364.

Goeree, Jacob K. and Charles A. Holt. 1999a. Stochastic game theory: For playing games, not just for doing theory. *Proceedings of the National Academy of Sciences* 96: 10564-10567.

Goeree, Jacob K. and Charles A. Holt. 1999b. An experimental study of costly coordination. Unpublished paper, University of Virginia.

Goeree, Jacob K. and Charles A. Holt. 2000a. Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review* 44: 1079-1089.

Goeree, Jacob K. and Charles A. Holt. 2000b. Models of noisy introspection. *Games and Economic Behavior*. In press.

Goeree, Jacob K. and Charles A. Holt. 2000c. An explanation of anomalous behavior in binary-choice games: entry, voting, public goods, and the volunteers' dilemma. Unpublished paper, University of Virginia.

Goeree, Jacob K. and Charles A. Holt. 2000d. Stochastic learning equilibrium. Unpublished paper, University of Virginia.

Goeree, Jacob K. and Charles A. Holt. 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 91: 1402-1422.

Goeree, Jacob K., Charles A. Holt, and Susan K. Laury. 2002. Altruism and noisy behavior in one-shot public goods experiments. *Journal of Public Economics* 83: 257-278.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. 2002. Quantal response equilibrium and overbidding in private value auctions. *Journal of Economic Theory* 104: 247-272.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. 2003. Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*. In press.

Hewett, R., C. Holt, G. Kosmopoulou, C. Kymn, C. Long, S. Mousavi, and S. Sarangi. 2002. A classroom exercise: Voting by ballots and feet. Unpublished paper, University of Virginia.

Holt, Charles A. 1979. Bidding for contracts. In *Bayesian analysis in economic theory and time series analysis, the 1977 Savage dissertation award theses*. New York: North Holland.

Holt, Charles A. 1980. Competitive bidding for contracts under alternative auction procedures. *Journal of Political Economy* 88: 433-445.

Holt, Charles A. 1996. Classroom games: Trading in a pit market. *Journal of Econmic Perspectives* 10: 193-203.

Holt, Charles A. 1999. Teaching economics with classroom experiments: A symposium. *Southern Economic Journal* 65: 603-610.

Holt, Charles A. 2003. Webgames and Strategy: Recipes for interactive learning. Unpublished paper, University of Virginia.

Holt, Charles A. and Lisa R. Anderson. 1999. Agendas and strategic voting. *Southern Economic Journal* 65: 622-629.

Holt, Charles A., Loren Langan, and Anne P. Villamil. 1986. Market power in oral double auctions. *Economic Inquiry* 24: 107-123.

Holt, Charles A., and Susan K. Laury. 2002. Risk aversion and incentive effects. *American Economic Review* 92: 1644-1655.

Kahneman, D. and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263-291.

Kagel, John and Alvin Roth. 1995. *Handbook of experimental economics*. Princeton: Princeton University Press.

Levine, M. E. and Charles R. Plott. 1977. Agenda influence and its implications. Virginia *Law Review* 63: 561-604.

Luce, Duncan R. (1959) *Individual Choice Behavior*. New York: Wiley.

McKelvey, Richard D. and Thomas R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* 10: 6-38.

Nash, John F. 1950. Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences*, U.S.A. 36: 48-49.

Plott, Charles R. 1982. Industrial organization theory and experimental economics. *Journal of Economic Literature* 20: 1485-1527.

Plott, Charles R. and M. E. Levine. 1978. A model of agenda influence on committee decisions. *American Economic Review* 68: 146-160.

Rassenti, Stephen J., Vernon L. Smith, and Bart Wilson. 2000. Market power in electricity networks. Unpublished paper, George Mason University.

Siegel, S., A. Siegel, and J. Andrews. 1964. *Choice, Strategy, and Utility*. New York: McGraw Hill.

Smith, Vernon L. 1962. An experimental study of competitive market behavior. *Journal of Political Economy* 70: 111-137.

Smith, Vernon L. 1982. Markets as economizers of information: Experimental examination of the "Hayek hypothesis". *Economic Inquiry* 20: 165-179.

Tversky, A. and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297-323.

Vickrey, William. 1962. Counterspeculation and competitive sealed tenders. *Journal of Finance* 16: 8-37.