Experimental Evidence

Jonathan Levin

May 2006

Dating back to the sixties, researchers have used laboratory experiments to study how people actually might play particular games. Early work focused on simple games such as the prisoners' dilemma and market trading environments. More recent work has broadened the scope of inquiry. These notes briefly survey a few of the many interesting experimental findings.

Experimental work is useful for exploring both human behavior and the performance of institutions. From a behavioral viewpoint, two of the interesting questions for game theory are:

- How do people react when faced with a strategic situation? Do they play dominated strategies? Perform backward induction or reason strategically? Attempt to maximize their payoff?
- How do people learn over time to play games? Does behavior converge toward Nash equilibrium or one of its refinements?

One thing we will see immediately is that people don't just walk into the laboratory and play Nash equilibrium. Sometimes they play dominated strategies. Over time, laboratory play often converges toward equilibrium, but not always. In a sense, this is not surprising. Experimental subjects don't know game theory; they're just given a sheet of paper with numbers. The interest in the experiments comes from finding behavior that is consistent across different games, or that leads to some insight into how people approach problem solving, or how they learn to behave over time.

Beyond these behavioral questions, there is another motivation for experiments, which is to test the performance of different institutions. A basic question here is:

- Do different institutions — auctions, market mechanisms, matching algorithms, bargaining protocols — achieve good outcomes even when players are inexperienced? How do changes in the rules governing these institutions affect outcomes?

1 Bargaining: The Ultimatum Game

One of the most debated results in experimental economics comes from studies of the so-called "ultimatum" game. In this game, one player (the "proposer") goes first and offers a split of a pie of given size. The second player chooses whether to accept or reject. If the second player rejects, both get noting. If he accepts, they split the pie as was proposed. The ultimatum game has a unique subgame perfect equilibrium where the proposer gets (essentially) the whole pie. It also has many other Nash equilibria, where the proposer offers a more generous split, fearing an aggressive offer will be rejected.

The experimental results are in stark contrast to the backward induction solution. On average, proposers offer about 40% of the pie to the second player. Moreover, offers are rejected about 15-20% of the time. Perhaps not surprisingly, lower offers are more likely to be rejected. Table 1 shows data from an experiment by Roth et. al. (1991). The modal proposal is about 500 (out of 1000). Offers are a bit dispersed at the outset, but tend to cluster at the mode over time.

Roth et. al. also ran the same experiment in three other countries — Yugoslavia, Japan and Israel (see table 2). The modal proposal in Yugoslavia looks similar to the U.S., while in Japan and Israel it was closer to 400. Interestingly, countries with lower offers did not have higher rates of disagreement. Instead, somewhat lower offers were accepted in countries where low offers were made.

Discussion of these results often raises two issues:

- First, why do responders reject offers?
- Second, does it seem that proposers are behaving "rationally" given responder behavior?

There are many interpretations for why offers are rejected. One leading explanation is that players punish greed – their behavior exhibits some sort of concern for fair outcomes or reciprocity motive. A striking finding is that these empirical patterns do not seem to be due to low stakes. Hoffman et al. (1998) ran ultimatum experiments in the U.S. with $10 and $100 stakes. While they found somewhat lower offers and more willingness to accept in the $100 treatment, their results are quite far from the backward induction prediction.

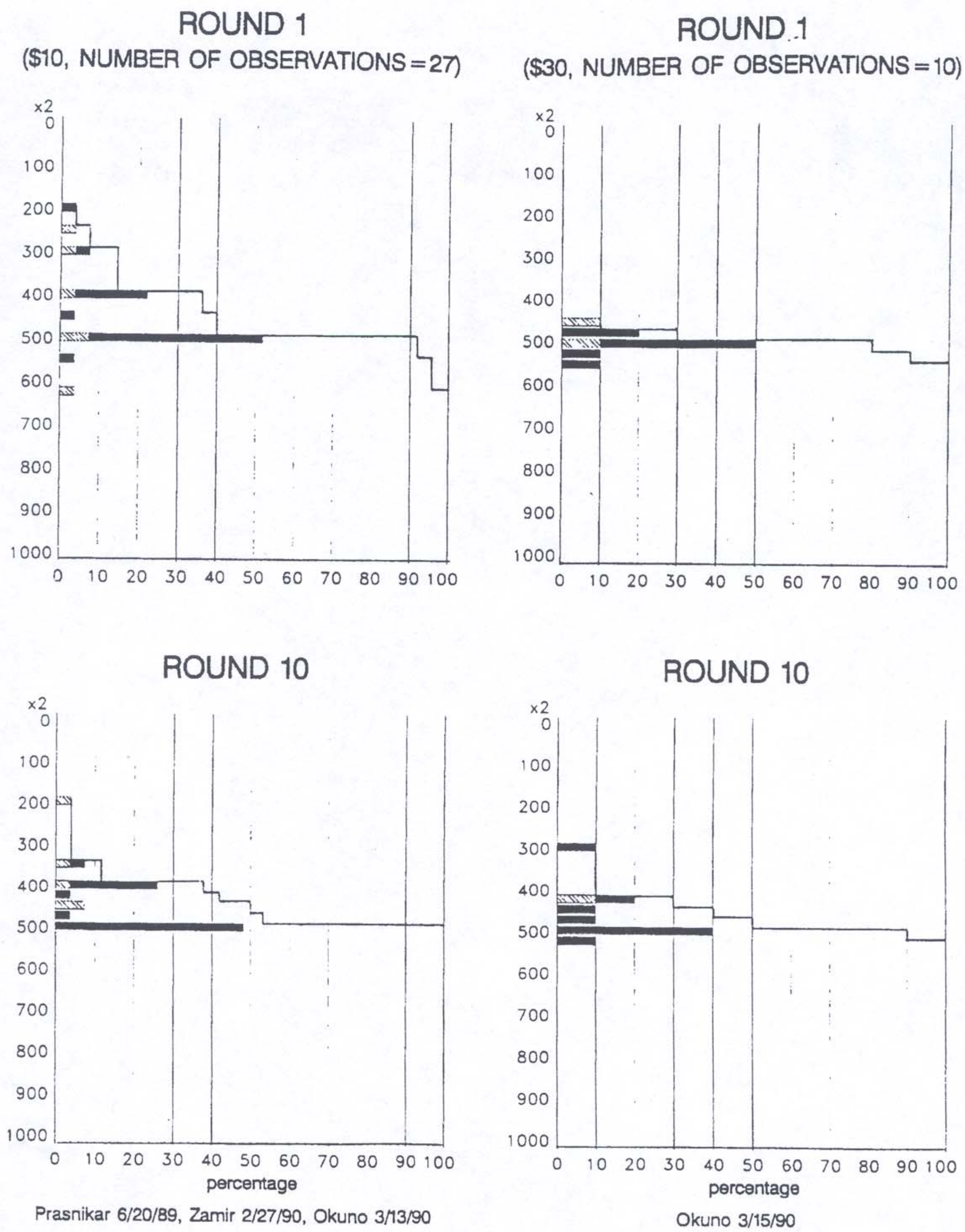


FIGURE 3. DISTRIBUTION OF BARGAINING OFFERS IN THE UNITED STATES (SOLID BARS = ACCEPTED OFFERS; STRIPED BARS = REJECTED OFFERS)

Figure 1: Roth et. al. (1991, *AER)*

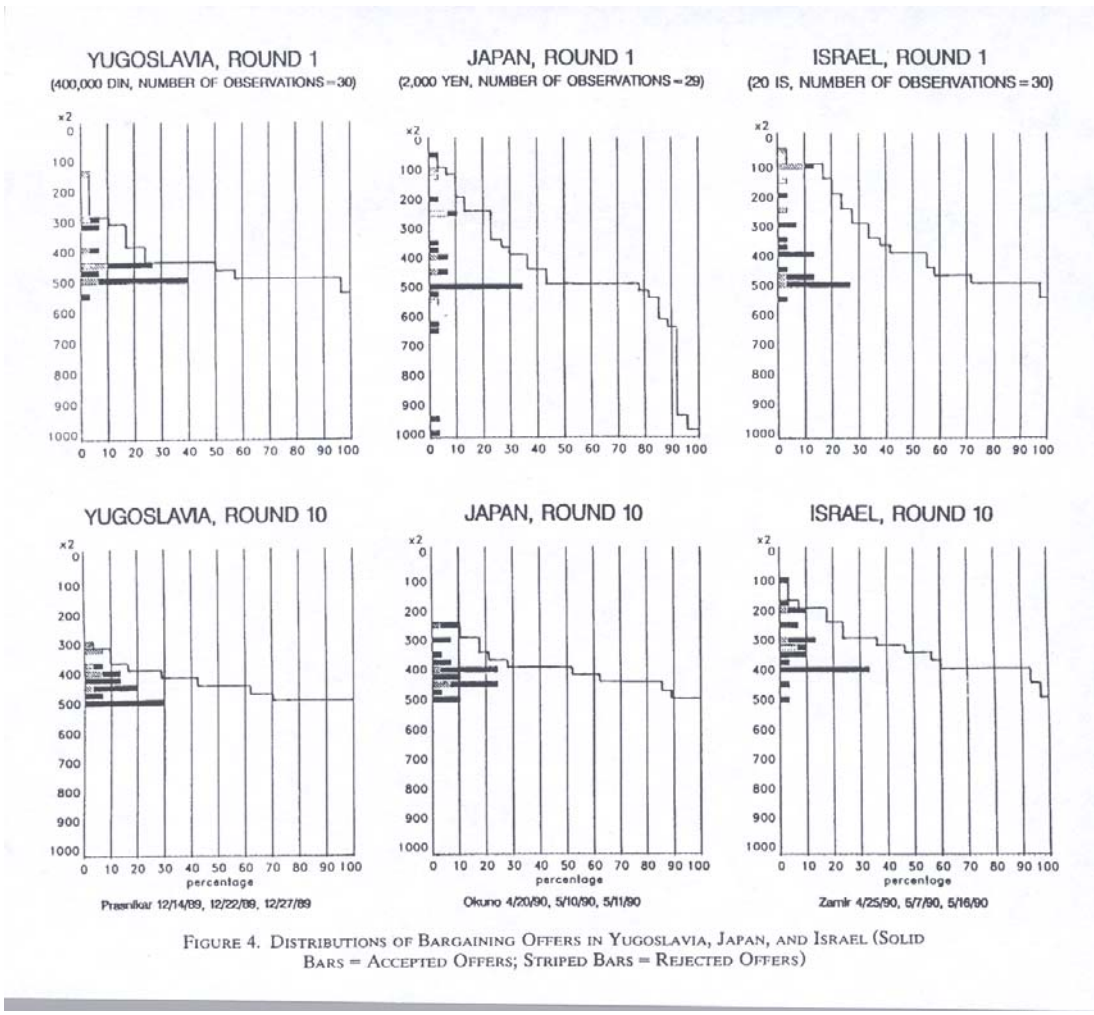
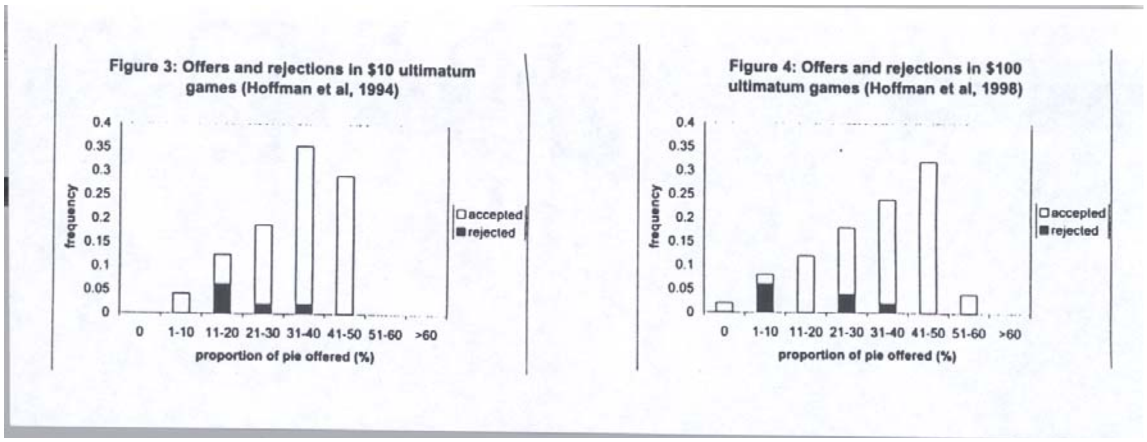


Figure 2: Roth et. al. (1991, *AER*)



Ultimatum Game with High Stakes

As Colin Camerer (2000) points out, a weakness of fairness-based explanations is that they typically don't address the question of where such preferences might come from. One possibility is social conditioning. Camerer discusses an experiment done by anthropologists in primitive cultures in Africa, the Amazon, Papua New Guinea, Indonesia and Mongolia. In most cases, offers were very low (about $1.50 out of $10) and offers were accepted practically every time. This suggests that cultural forces may need to be incorporated to generate a satisfactory theory for the data.

2 Market Games

Arguably, the results in the ultimatum game contradict the standard theory. In contrast, experiments on various forms of market games have tended to be strongly supportive of theoretical predictions. There is a long history of market experiments (this is what Vernon Smith won the Nobel Prize for in 2002). The remarkable finding is that even with relatively few players, and a relatively short number of rounds, market outcomes tend to look like competitive equilibria.

A nice example are Roth et al's (1991) experiments on university students in the United States, Yugoslavia, Israel and Japan. Their game works as follows. In their game, multiple buyers (nine in most sessions) submit an offer to a single seller to buy an indivisible object worth the same amount to each buyer. The seller can either reject the bids, in which case everyone gets zero, or accept the highest bid, in which case the seller gets the bid, and the high bidder gets the difference between the valuation and the bid.

The Roth et. al. game has a unique perfect equilibrium where players bid their value, and the seller accepts. The experimental evidence is strongly in line with this. In Roth's experiments, the seller *always* accepted the high bid and within a few rounds, the high bid converged to the buyers' value. Indeed (see Table 3), the bids ended up highly concentrated in this region. Moreover, the cross-country differences are minimal.

3 Iterated Dominance: The Beauty Contest

The ultimatum game seems to shed some light on preferences, while market games speak to the power of competition. Another interesting class of experiments are aimed at understanding the extent to which subjects use iterated dominance reasoning (i.e. think through what people think about what people think ... people will do).

An elegant example is Nagel's (1995) work on p-beauty contest games. In Nagel's experiment, players were asked to choose a number between 0 and 100, with a prize going to the player whose guess is closest to p times the average guess. When $p < 1$, this game can be solved using iterated strict dominance (first elimate strategies greater than $100p$, then greater than $100p^2$, and so on). Of course, the unique equilibrium has everyone guess zero.

In Nagel's experiment, when $p = \frac{1}{2}$, in the first round of play many players guessed between 10 and 30. When $p = \frac{2}{3}$, players tended to guess in the 20–35 range, although in both cases there were higher guesses as well (Table 4). Thus at the outset, people behave as if they are doing perhaps two rounds of iterated reasoning.

As players playes a few more times, play tended to converge toward the Nash equilibrium (Table 5). Interestingly, play converges faster than best-response dynamics would converge. That is, subjects guess lower than a direct best-response to the previous period's average — perhaps moving two steps at a time. Table 5 also shows that with enough time, players eventually play 0.

Nagel's results suggest that while people don't leap immediately to the iterated dominance solution, they do "think through" the game to some extent. Further evidence on "strategic sophistication" is provided by Costa-Gomes, Crawford and Broseta (2002). They asked subjects to play games that were dominance-solvable or that had a unique pure-strategy equilibrium. Their basic conclusion is that many subjects look as if they are best-responding to what they perceive as a random choice on the part of

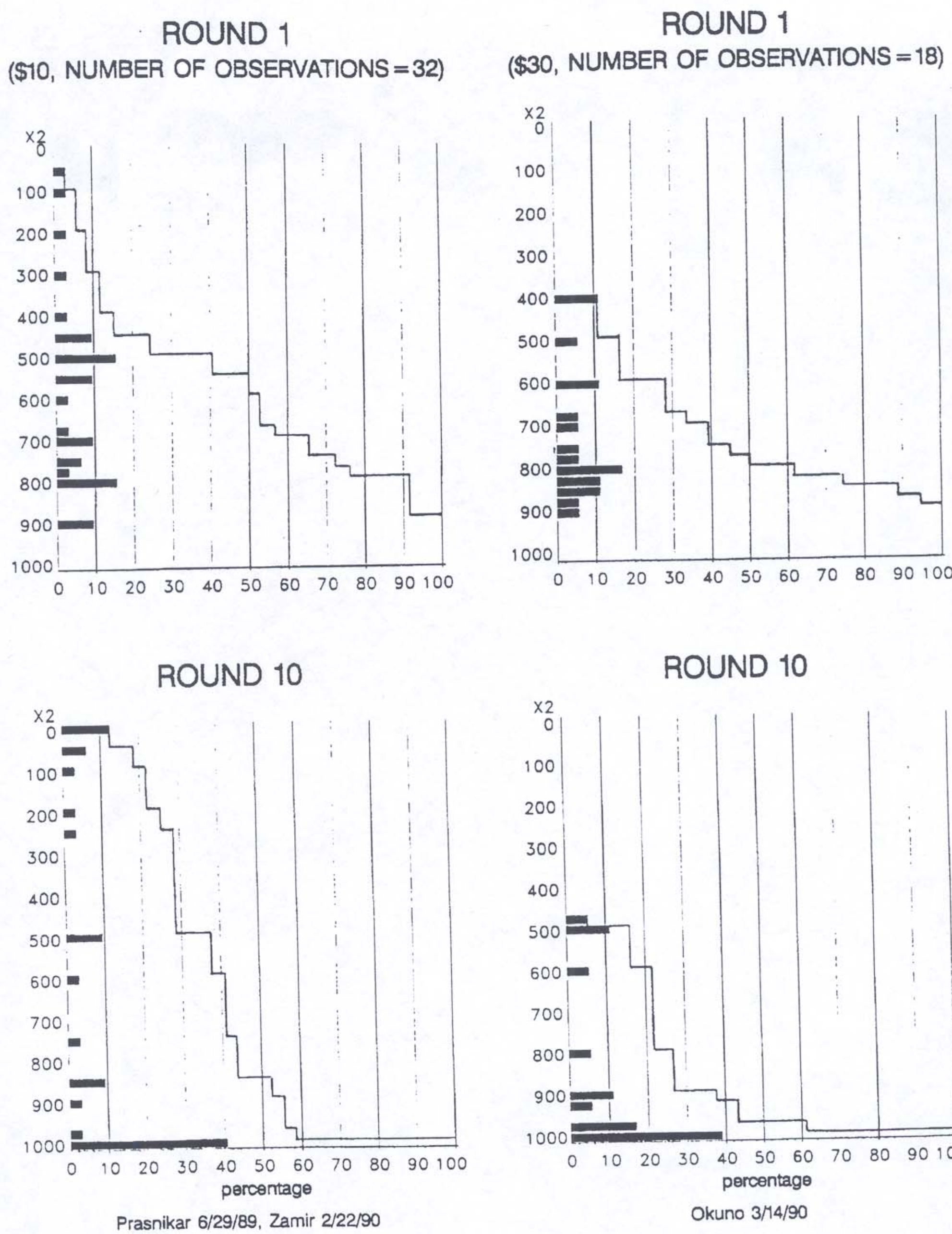


FIGURE 1. DISTRIBUTION OF MARKET OFFERS IN THE UNITED STATES

Figure 3: Roth et. al (1991, *AER)*

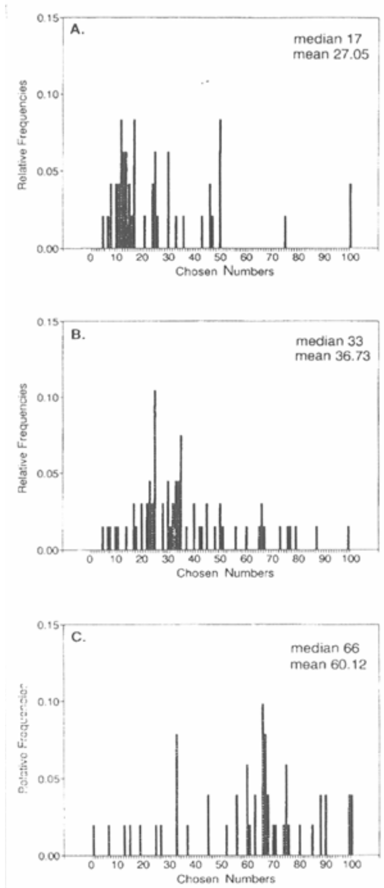


FIGURE 1. CHOICES IN THE FIRST PERIOD: A) SESSIONS 1–3 ($p = {}^1/_2$); B) SESSIONS 4–7 ($p = {}^2/_3$); C) SESSIONS 8–10 ($p = {}^4/_3$)

Figure 4: Nagel (1995, *AER*)

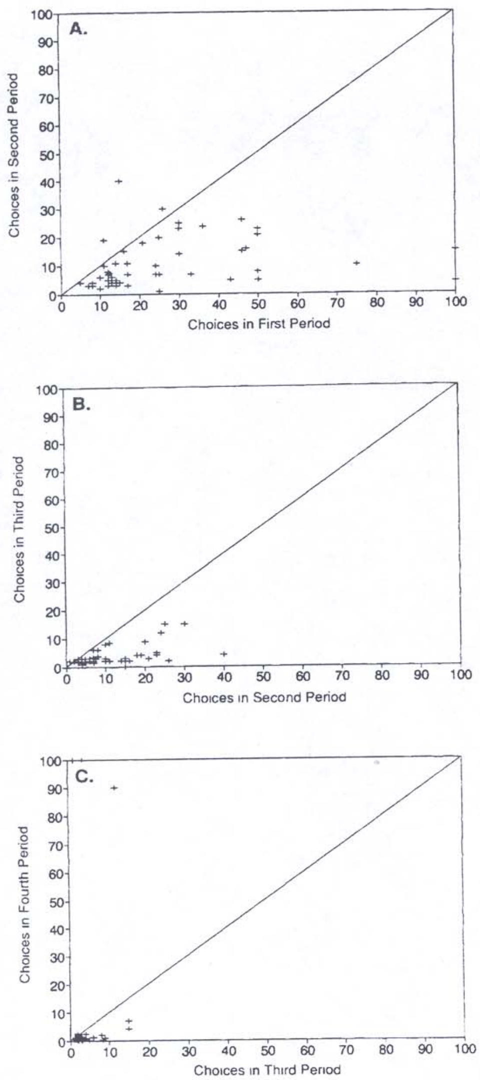


FIGURE 3. OBSERVATIONS OVER TIME FOR SESSIONS 1–3 ($p = 1/2$): A) TRANSITION FROM FIRST TO SECOND PERIOD; B) TRANSITION FROM SECOND TO THIRD PERIOD; C) TRANSITION FROM THIRD TO FOURTH PERIOD

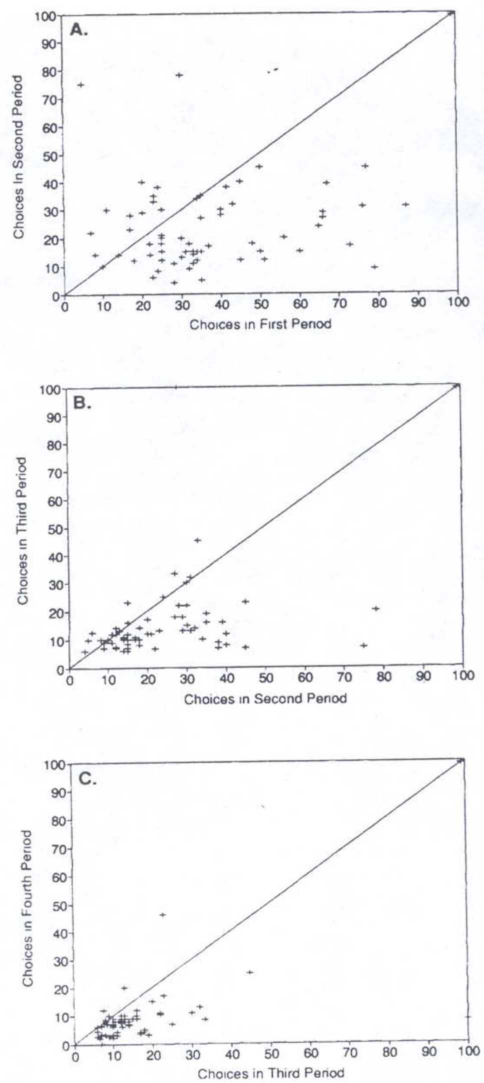


FIGURE 4. OBSERVATIONS OVER TIME FOR SESSIONS 4–7 ($p = 2/3$): A) TRANSITION FROM FIRST TO SECOND PERIOD; B) TRANSITION FROM SECOND TO THIRD PERIOD; C) TRANSITION FROM THIRD TO FOURTH PERIOD

Figure 5: Nagel (1995, *AER*)

opponents, or perhaps a random choice from among undominated strategies.

4 Adaptive Behavior in Coordination Games

A striking finding in the beauty contest is that behavior seems to converge in a natural adaptive way toward the unique equilibrium. To the extent that behavior over time can be described as a process of learning and adaptation, a natural question is what will happen in games where there is no "obvious" outcome toward which behavior should converge. One clever experiment along these lines was done by van Huyck, Battalio and Cook (1998).

In their game, players in groups of seven chose numbers between 1 and 14. Payoffs depended on own choice and the median and were symmetric. Best responses are shown below. The game has two equilibrium, 3 and 12, where the higher equilibrium is Pareto-preferred.

	Median Choice													
	1	2	**3**	4	5	6	7	8	9	10	11	**12**	13	14
Best Response	2	3	**3**	3	4	5	6	9	10	11	12	**12**	12	13

This game is called the "continental divide" game because on either side of 7, the best-responses "flow downhill" toward an equilibrium.

The experiment called for players to play for fifteen rounds. Median choices for several runs are shown in Table 6.

Behavior seems to adjust roughly in the direction of best-responses, though perhaps at a slower or faster rate than the simplest best-response dynamics. Nevertheless, an adaptive learning story looks like a reasonable one to explain what is going one. A particularly interesting aspect of these results is that the long-run play depends a lot on the initial behavior. If the median starts out below seven, play tends to converge toward the low equilibrium. If the median starts out above seven, play tends toward the higher equilibrium. This is true despite the fact that the high equilibrium is Pareto-preferred.

References

[1] Costa-Gomes, Miguel, Vincent Crawford and Bruno Bosetta (2002) "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica.*

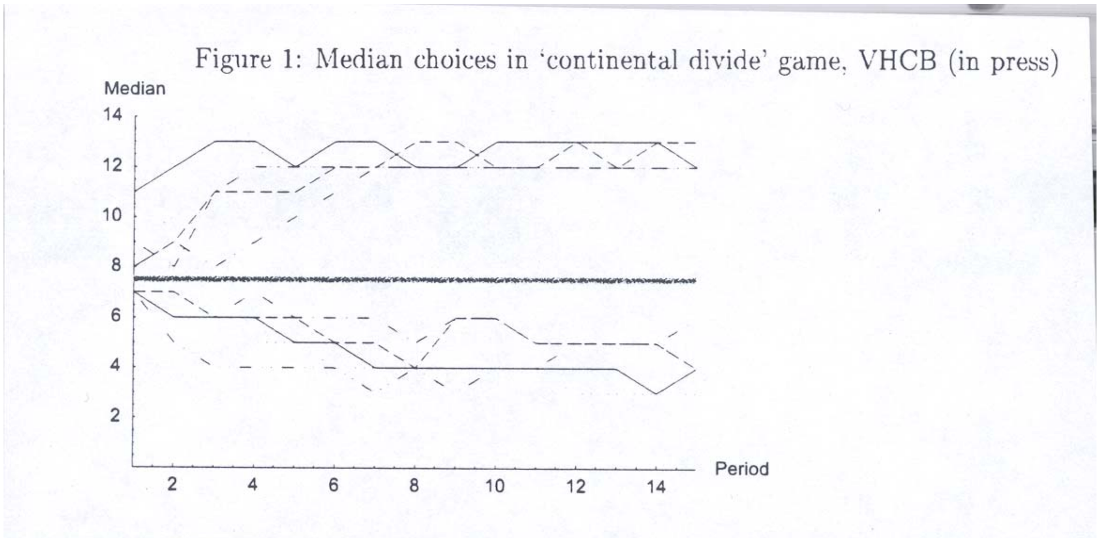


Figure 6: Adaptive Behavior in a Coordination Game

[2] Craword, Vincent (1997) "Theory and Experiment in the Analysis of Strategic Interaction," in *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress of the Econometric Society, ed. D. Kreps and K. Wallis.

[3] Camerer, Colin (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press.

[4] Nagel, Rosemarie (1995) "Unraveling in Guessing Games: An Experimental Study," *Amer. Econ. Rev.,* 85, 1313–1326.

[5] Roth, Alvin, Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991) "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study," *Amer. Econ. Rev.,* 81, 1068–1095.

[6] van Huyck, John, Raymond Battalio and Richard Beil (1991) "Strategic Uncertainty, Equilibrium Selection and Coordination Failure in Average Opinion Games," *Quarterly J. Econ.*, 106, 885–910.