

**Economics 207, Experimental Economics**  
**Vincent P. Crawford, Winter 2002**  
**(partly adapted from notes by Miguel Costa-Gomes)**

Economics 207 introduces the subject matter, methods, and results of experimental economics. It will stress interaction of theory and experiment, seeking to relate questions in the theories of markets, games, and decisions to issues in experimental design and the analysis and interpretation of results. After an initial overview, these themes will be developed by discussing series of related experiments.

### **Introduction and Overview**

- Role of experiments: to fully test economic theories, game-theoretic or not, need control of institutions, information, structure of interactions, often more than possible in the field
  
- Econometrician Guy Orcutt (quoted in Smith (1982)): "the econometrician as being in the same predicament as that of an electrical engineer who has been charged with the task of deducing the laws of electricity by listening to a radio play."
  
- Importance of scientific culture, innovation within culture, part arbitrary but crucial to replicability, progress, publication
  
- Importance of ethics in treatment of subjects: see links at <http://eexcl.ucsd.edu> for guidelines for proper treatment
  
- Importance of ethics in generating and reporting data: role in research delegated to subjects, unbiased design and choice of data to analyze and report (Sir Arthur Evans's model of Knossos, entire murals reconstructed from small parts; more informative than raw data, but less than ideal of summary plus mapping from observation to conclusions)

Advantages of experimental methods: replicability, control

Limitations of experimental methods:

- Student subjects may not be representative of relevant decision makers in the field (can remedy by comparing results across more kinds of subjects, realistic framing, field experiments (more control than "natural experiments"))
- Simplicity of laboratory environments may limit "external validity"—or transfer of results to parallel field environments (valid theories should work in simple settings too, but need to explore boundaries of applicability, transfer to field)
- Technical difficulties in establishing and controlling laboratory environments may also limit effectiveness, e.g. when seeking to elicit information about individual preferences via hypothetical contingent valuation choices

### **Types of Experiments**

- Tests of behavioral hypotheses ("theory falsification"): laboratory environment should satisfy as many of the *structural* assumptions of a particular theory as possible, so that its *behavioral* implications have the best chance (more useful if, when the results falsify the theory, they also suggest alternative explanations of behavior, because all experiments implicitly test the joint hypothesis that the theory is correct *and* the experimenter implemented it correctly; contrast exemplifying experiments popular in psychology)
- Theory Stress Tests: if key *behavioral* assumptions of a theory are not rejected in simple lab settings, useful to explore sensitivity to simplifying assumptions unlikely to be satisfied in the field, explore boundaries of theory's validity
- Searching for Empirical Regularities ("fishing expeditions")

## Terminology

- Session: sequence of periods, games, or other decision tasks involving the same group of subjects on the same day
- Cohort: a group of subjects that participated in a session
- Treatment: a unique configuration of treatment variables (information, experience, incentives, rules)
- Cell: a set of sessions with the same treatment conditions
- Experimental design: a specification of sessions in one or more cells to evaluate the propositions of interest
- Experiment: the collection of sessions in one or more related cells (even if subjects' instructions say otherwise!)
- Within-subjects design: single subject observed in different treatments, so subject serves as his/her own control group
- Between-subjects design: different subjects observed in different treatments

If sessions have repeated decisions, a decision unit is called:

- Trial in individual decision experiments
- Game in game experiments
- Trading period in market experiments

## Design

- Strategies: Control, measure, or assume (e.g. risk preferences)
- Control of environment and procedural regularity ("scripts"): Standardize procedures and report them accurately enough to allow replication, including:

Instructions (clarity here is essential)

Illustrative examples and tests of understanding

Criteria for answering questions

Nature of monetary or other rewards

Presence of "trial" or "practice periods" with no rewards

Number and experience levels of subjects

Procedures for matching subjects and assigning roles

Locations, dates, and durations of sessions

Physical environment, use of assistants, special devices, and computerization

Any intentional or unintentional deception of subjects (John Hey (1998): "There is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not.")

Any procedural irregularities in particular sessions that might affect the interpretation of the results

Pilot sessions (the decision whether a session is a pilot rather than a regular session should be made in advance, and not changed after looking at the data)

- Control of knowledge: public announcements, practice runs, and understanding tests; "public knowledge" as approximation to common knowledge in games; problems in controlling beliefs, especially about stochastic processes, e.g. probability matching experiments

- Control of preferences: Subjects must receive salient rewards that conform to incentives in the relevant theory or application: decisions have a prominent effect on rewards, and rewards are important enough to subjects to dominate subjective costs of decisions, boredom, altruism, etc.; "flat maximum" critique and need for strong *marginal* incentives

Studies have used goods like coffee mugs, extra credit toward grades, "points", and "money" (economists prefer money, which seems to reduce noise in "our" designs)

Controlling for risk preferences with binary lottery procedure

Experiments should be framed as neutrally as possible, and conducted so participants don't perceive any behavior as being correct or expected (unless this is a treatment variable); e.g. "Clever Hans", Binmore et al. *AER*: "You will be doing us a favor if you maximize your money payoffs."

Abstract framing is helpful when experiment concerns effects of "structure" rather than "context", but not strictly necessary (concrete framing can be neutral, though hard to know for sure); e.g. abstract decision or role labels may be preferable to labels that might induce cultural bias: "#" or "^" rather than "help lady cross the street" or "sell crack cocaine"

Subjects should be anonymous to each other and the experimenters, and should not interact face to face (unless this is a treatment variable), to minimize "social" effects on preferences; some argue for double-blind procedures, in which assistants conduct sessions without knowing purpose (can be selected from subjects, and paid fixed amount)

Even if interest is in "social" effects, enhanced control helps

- Control of game:

In game experiments, usually need to give subjects experience and observe the effects of learning; these problems usually solved by repeated-game design in which subjects play same game or series of games repeatedly

Such designs can lead to repeated-game strategies, which blur identification of the game to which subjects respond

Repeated-game strategies can be avoided by repeated random pairing or grouping from large population, or "large"-group interactions with negligible individual influences, so that subjects treat repeated interactions with any given other subject as negligible; e.g.:

- No-repeat matching: subjects never rematched
- No-contagion matching: subjects never rematched, and are never rematched with somebody who will be matched with somebody they will later be matched with, and so forth
- Random matching: subjects rematched randomly and blindly (may be rematched with previous partner, but prior probability is low and they do not know when it occurs)
- Mean-matching: each subject plays every other subject all at once, earning average payoff from all matches

Resulting design is used to test theories of behavior in the "stage game"; free to vary stage game within this structure (can even be a mini-repeated game), so the possibilities are practically unlimited despite restrictions on overall structure

- General considerations:

To maximize external validity, as in "testbedding," design should emphasize closeness to field environment rather than theories; to test theories, vice versa

Importance of clear separation of predictions of alternative theories, in presence of behavioral "noise"; falsification of a theory is more convincing if a reasonable alternative is not falsified, because this increases confidence in design, and rejection of reasonable alternatives strengthens inferences based on failure to reject the maintained hypothesis

Importance to confidence of consistency of results between subjects across sessions or within across treatments

Design should also account and/or control for theoretically irrelevant factors that regularly affect performance, such as experience, group effects, and order effects

*Blocking*, systematic pairing of observations, may be used to neutralize the effects of such nuisance variables. E.g.:

- *Order effects*: When sessions sequence two treatments, A and B, some should be run AB, others BA; otherwise any difference in results for A and B might be an order effect
- *Randomized block*: Sometimes the number of things that can be systematically blocked is unreasonably large and the alternative configurations can be selected randomly
- Use of simple statistical tests in psychology and early economics experiments, versus heavier contemporary use of econometrics in response to complexity of phenomena, as substitute for unattainable full control

## Demonstration experiment 1:

Tacit normal-form "order statistic" coordination games with multiple, Pareto-ranked equilibria (VHBB (1990, 1991))

Run in lab with paper and pencil, instructions in papers linked on subject desktop. Sample median instructions at pp. 908-909 near end of 1991 *QJE* and  $\Gamma$ ,  $\Omega$ ,  $\Phi$  median payoff tables at pp. 890-891, A and B minimum payoff tables at p. 238 in 1990 *AER* (no published instructions, but similar) minimum A:  $\pi(e_i, \underline{e}_i) = a[\min(e_i, \underline{e}_i)] - be_i$ , where  $a > b > 0$   
median  $\Gamma$ :  $\pi(x_i, M) = aM - b[M - x_i]^2 + c$ , where  $a, b > 0$

Questions: Do subjects learn to play equilibria in tacit coordination games? What determines their initial responses? What determines equilibrium selection in the long run?

Design: Normal-form complete-information coordination games in which effects of context are minimized, with common set of Pareto-ranked equilibria, varying off-equilibrium payoffs to stress-test traditional theories of equilibrium selection (risk- and payoff-dominance), and large (but finite) strategy spaces to give learning dynamics room to vary widely across treatments

Results: Minimum results at pp. 240, 241, 243, 245, 246, 247 in 1990 *AER*. Median results at pp. 895, 898, 901, 902 in 1991 *QJE*; see also VHBB 1993 *GEB*, Crawford and Broseta 1998 *AER*; little difference in initial responses, but modal responses give weak support for notions like risk-dominance; large differences in subsequent play, with adaptive dynamics driven by strategic uncertainty determining equilibrium selection in the long run

## 2. Competitive Markets

Demonstration experiment: Caltech multiple-unit double-auction market, courtesy of Charles Plott; demonstration experiment at <http://eeps3.caltech.edu/market-demo/>

Questions: What does "perfect competition" require (in 1960 most theorists would have said large numbers of well-informed traders on both sides of the market)? How well do competitive markets aggregate participants' private information? How do institutions affect performance?

Design: Inducing supply and demand, providing incentives, controlling information

Results: Robustly competitive outcomes for double oral auction with small numbers of traders on both sides, better results when traders are *not* informed about others' values, powerful but not unlimited aggregation of private information for some market institutions

History: Edward Chamberlain, (1948), "An Experimental Imperfect Market," *Journal of Political Economy*, 56, 95-108.

Induced Supply and Demand: Each buyer (seller) receives a redemption value (cost) for her or his single unit (look at Table 1 and at Figure 1). Each trader knows her own redemption value (or cost), but not others' costs or values.

*Experiment:* subjects walk around, bargain in pairs, groups. Once a buyer and seller reach a deal, they drop out. The transaction price is recorded on the blackboard (in fact, not always...). Market operated for a single trading period. The competitive equilibria in this market have an equilibrium quantity of 15, and an equilibrium price of 56-58.

## Results

- *Number of Units Traded*: Too much trading (an average of 19 trades), higher than in any of the competitive equilibria of the market in 42 out of 46 experimental sessions, and equal to it in the remaining 4 sessions.

- *Prices*: Average price higher than any equilibrium price in 7 sessions and lower in 39 sessions.

“No tendency for prices to move toward equilibrium during the course of the market.” Why should they? “Information during the market as to the equilibrium price would help establish a trend in that direction, but information as to actual prices may do the opposite, in so far as they are divergent from equilibrium and are falsely interpreted to be near it.”

More history: Vernon Smith, “An Experimental Study of Competitive Market Behavior,” *Journal of Political Economy*, 70, (1962), 111-137.

Induced Supply and Demand: Each buyer (seller) receives a redemption value (cost) for her or his units (1 or more units) (look at Chart 1). Each trader knows own redemption value (cost), but not others' costs or values.

Experiment: Oral Double Auction, buyers and sellers can freely enter limit orders (bids or asks) and accept others' asks or bids. No restrictions on messages. Prices of completed transactions recorded on blackboard. Traders with no more units drop out. Market operated for several periods (3-6), each lasting 5-10 minutes.

Smith (1962) on differences from Chamberlin:

“The design of my experiments differs from that of Chamberlin in several ways. In Chamberlin’s experiment the buyers and sellers simply circulate and engage in bilateral haggling and bargaining until they make a contract or the trading period ends. As contracts are made the transaction price is recorded on the blackboard (in fact, not always...). Each trader’s attention is directed to the one person with whom he is bargaining, whereas in my experiment each trader’s quotation is addressed to the entire trading group one quotation at a time.”

“[Also] Chamberlin’s experiment constitutes a pure exchange market operated for a single trading period. There is, therefore, less opportunity for traders to gain experience and to modify their subsequent behavior in the light of such experience. It is only through some learning mechanism of this kind that I can imagine the possibility of equilibrium being approached in any real market.”

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

## Results

- *Number of Units Traded*: Sometimes subjects trade too little, sometimes too much. Overall quantity close to equilibrium.
- *Prices*: The average price tends to approach the equilibrium price over time. Smith's “coefficient of convergence”—which measures exchange price variation relative to predicted equilibrium exchange price—decreases over time.

## Efficiency measure

$$\text{Efficiency} = \frac{\text{Buyers' Surplus} + \text{Sellers' Surplus}}{\text{Surplus Available}}$$

## Sources of Inefficiency:

- *Type U* - If at the end of the trading period there is a seller,  $i$ , with a unit available for sale whose cost is lower than the value of a unit that a buyer,  $j$ , did not acquire  $c_i^k < v_j^l$ , then there are foregone gains from untraded units;
- *Type V* - When a seller sells a unit whose cost is above the competitive equilibrium price, or when a buyer buys a unit whose corresponding value is below the competitive equilibrium price, then there will be foregone gains from units that could have otherwise been traded. This inefficiency is more likely when transaction prices are volatile.

### Efficiency:

- Type U inefficiency usually much higher than Type V; Type U decreases over time, V decreases too, but more slowly.

### Prices:

- Prices approach equilibrium within and across periods.
- Sequence of price changes is typically negatively autocorrelated, around -0.5.
- Subsequences of prices - The *subsequence of prices of trades initiated by buyers* is generally lower than the *subsequence of prices of trades initiated by sellers*.

### Transaction Order:

- Strong positive rank-order correlation between buyers' valuations and the order in which buyers purchase units, and strong negative rank-order correlation between sellers' costs and the order in which units are sold. That is, traders with higher expected surplus (higher difference between valuations/costs and competitive equilibrium prices) usually trade earlier than traders with lower expected surplus.
- Trading activity is usually concentrated at the beginning of the trading period, when many high-surplus units are traded, and toward the end of the period.

Good modern overview:

\*Charles Plott, "Equilibrium and Equilibration in Multiple Market Systems," paper presented at the Nobel symposium on Behavioral and Experimental Economics, December 4-6, 2001; <http://www.iies.su.se/nobel/papers.htm>

### 3. Extensive-Form Games

Demonstration experiment: Normal-form versus extensive-form framing in 2x2 games (David Cooper and John Van Huyck, "Evidence on the Equivalence of the Strategic and Extensive Form Representation of Games," manuscript, Texas A&M University, September 2001, [econlab10.tamu.edu/JVH\\_gtee/Sim1.pdf](http://econlab10.tamu.edu/JVH_gtee/Sim1.pdf))

Questions: Does extensive-form framing yield systematically different results than normal-form framing (e.g. by making backward induction more salient or by creating asymmetries subjects can use to solve coordination problems)? How?

Key theoretical notions:

- *Subgame-perfectness* (e.g. Ultimatum contracting; generalizations perfect Bayesian or sequential equilibrium), relation to iterated weak dominance in the normal form
- *Forward induction* (e.g. Battle of the Sexes with outside option), relation to strict dominance in the normal form)

Design: Presentation of games in extensive form, designs to elicit "one-shot" responses versus designs that allow learning in repeated play

Results: Some failure to follow backward induction logic, some bias in extensive form toward allowing second mover to influence outcome

Beard and Beil (1994 *Management Science*) start with Rosenthal's (1981 *JET*) extensive-form game

- Game tree gives player A the right to opt out (L) with payoffs  $x$  for A and  $y$  for B; or give player B the move (R) with two choices, l with payoffs 0 for A, 0 for B; or r with payoffs  $z$  for A and  $w$  for B;  $z > x$  and  $w > v$  ( $y >$  or  $< w$ )

	l	r
L	$x$ $y$	$x$ $y$
R	$0$ $v$	$z$ $w$

- The unique subgame-perfect equilibrium is (R,r) (which uniquely survives iterated weak dominance), but A players who think B is not certain to play r are tempted by L; thus the game is a simple test for reliance on other's dominance
- Assuming that utility maximization is close to own-money-payoff maximization (can test for player B in this design), intuitively, (H1) A players should be more willing to play R when  $x$  is lower (R is less risky), (H2)  $w - v$  is higher (B has more incentive to choose r), (H3)  $y$  is lower (B is less likely to resent A's choice of R and choose l), or  $w$  and  $v$  are higher (B is more likely to reciprocate A's choice of R by choosing r)
- Subgame-perfect equilibrium doesn't say any of this, but McKelvey and Palfrey's (Rosenthal's) notion of QRE does
- Beard and Beil used a series of treatments to test these intuitions, holding the probability that B chooses r that makes A indifferent between L and R constant near one (higher than the frequency with which subjects respect dominance, so making A subjects not rely on it) in most treatments

- "Real-time play": didn't use strategy method

### Treatments

Treatment	Player A plays <i>L</i>	Player A Chooses <i>R</i>	
		Player <i>B</i> plays <i>l</i>	Player <i>B</i> plays <i>r</i>
1	(9.75, 3.00)	(3.00, 4.75)	(10.00, 5.00)
2	(9.00, 3.00)	(3.00, 4.75)	(10.00, 5.00)
3	(7.00, 3.00)	(3.00, 4.75)	(10.00, 5.00)
4	(9.75, 3.00)	(3.00, 3.00)	(10.00, 5.00)
5	(9.75, 6.00)	(3.00, 4.75)	(10.00, 5.00)
6	(9.75, 5.00)	(5.00, 9.75)	(10.00, 10.00)
7*	(58.50, 18.00)	(18.00, 28.50)	(60.00, 30.00)

\*subjects in this treatment received the stated payoffs with probability 1/6, but otherwise only a participation fee

Probability of nonmaximizing play of *l* by *B* that a risk-neutral payoff-maximizing *A* subject needs for *L* to be optimal:

Treatment	Prob. of Nonmaximizing
1	3.57%
2	14.29%
3	42.86%
4	3.57%
5	3.57%
6	5.00%
7	3.57%

(H1): Treatments 1, 2, and 3 (\$9.75, \$9.00, \$7.00)

(H2): Treatments 1 and 4 (\$0.25, \$2.00)

(H3): Treatments 1 and 5 (*B*'s payoff from *A*'s secure choice *L* goes from \$3 to \$6)

## Results

- 97.8% of B subjects made choices that maximized own money rewards, suggesting almost all were self-interested
- Despite predictability of most subjects' decisions, A subjects opted out in surprisingly large numbers
- Intuitions all upheld qualitatively: Rate of opting out varied across treatments in a coherent manner, suggesting that payoffs had a significant, intuitive effect on willingness to rely on the self-interested behavior of others

Treatment	# of pairs	A chose	A chose R		% secure by A
			B chose l	B chose r	
1	35	23	2	10	65.7%
2	31	20	0	11	64.5%
3	25	5	0	20	20.0%
4	32	15	0	17	46.9%
5	21	18	0	3	85.7%
6	26	8	0	18	30.7%
7*	30	20	0	10	66.7%

Could explain with QRE-type model, maybe with social utility effects that depend on others' choices as in Rabin 1993 *AER*

Alternative explanations:

- Psychologists have argued that people value control of the kind an A subject can get by playing L. Beard and Beil's subject survey provides evidence for this.
- Experience as a B player was associated with significantly greater willingness to rely on other's maximization

Schotter, Weigelt, and Wilson (1994 *GEB*) reconsidered these issues using strategy method; also considered differences in responses to "same" game presented in extensive and normal form, sometimes varying description (sequential versus simultaneous) independently of play; tested for reliance on up to three rounds of iterated dominance, and forward induction

Results: subjects seldom rely on more than one round of dominance, framing as game tree is more important than whether actual decisions are sequential

#### **4. Normal-Form Games** (see attached notes)

Demonstration experiments: MouseLab matrix games with iterated dominance and unique pure-strategy equilibria (Costa-Gomes, Crawford, and Broseta (2001)), two-person guessing games (Costa-Gomes and Crawford, manuscript, instructions, and script for guessing experiments (2002))

Questions: How does iterated dominance affect behavior in normal-form games, with or without opportunities for learning from experience? How do subjects deviate from equilibrium? What decision rules best describe their behavior?

Design: Presentation of games in normal form, designs to elicit initial responses versus designs that allow learning, using MouseLab to track searches for hidden payoffs

Results: Subjects typically respect 1-3 rounds of iterated dominance, but no more; they tend to play equilibrium in simple games, but deviate systematically in more complex games; much of their behavior (information search as well as decisions) is well described by simple boundedly rational but strategic decision rules like *Naïve* and *L2*

## **5. Unstructured bargaining (see attached notes)**

No demonstration experiment

Questions: What determines outcomes of unstructured bargaining in settings like those studied in cooperative game theory? How well do standard bargaining theories (structured/noncooperative or unstructured/cooperative) describe observed bargaining outcomes?

Design: control of bargaining institutions and information, use of binary lottery procedure and private information to create invariances that can be used to test the theory, use of monitored communication via computer to mimic "no rules" bargaining with a deadline, modern implementation via NetMeeting software as in Moreno and Wooders (1998))