

ExpEcon Methods: Incentivized Experiments: A History

ECON 8877

P.J. Healy

Updated 21. September 2023

Roth's History

Introduction to Kagel & Roth's (1995) *Handbook of Experimental Economics*

Three threads:

1. Individual choice, from 1931
2. Game theory, from 1950
3. Markets/IO, from 1948 & 1960

Briefly: Ledyard's Public Goods survey

Individual Choice



Louis Leon (L.L.) Thurstone

- U. Chicago psychologist
- Pioneered the idea that measurements could be obtained from pairwise comparisons
 - Example: quality of handwriting samples
 - Simplified by rank-orderings, under transitivity

Application to economics:

“The indifference function” (1931) *J. Social Psych*

Let's Evaluate This Paper

1. What are the treatments (if any)?
2. What does this design *really* test?
 - Does that match the stated hypotheses?
3. Proper incentives?
4. Subject pool?
5. Proper sample size? Statistical power?
6. What's the main result?
7. Are the statistical tests appropriate?
8. Are the results conclusive?
9. Robustness?
10. Extensions?



Louis Leon (L.L.) Thurstone

“The indifference function” (1931) *J. Social Psych*

- Suggested by economist Henry Schultz
- Measures “satisfaction” of a good (= utility) via pairwise choice
- Assumes $u(x_i) = k_i \log(x_i) + c$ for each good i
 - Fechner’s law! Perception is logarithmic
- $k_1 \log(x_1) + k_2 \log(x_2) = \log(\bar{u})$ at utility level \bar{u}
- $x_1^{k_1} \cdot x_2^{k_2} = \bar{u}$
- Teaches reader about indifference curves

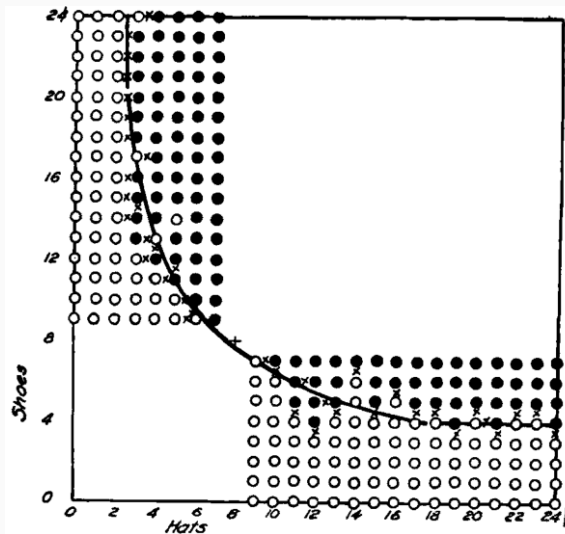
Thurstone (1931)

- Choices: Hat-shoe combinations
- Multiple Price List (MPL) against a fixed bundle!

8 hats and 8 shoes
6 hats and 9 shoes
8 hats and 8 shoes
4 hats and 15 shoes
8 hats and 8 shoes
9 hats and 3 shoes

- Single subject
- Hypothetical choices
- MANY choices

Thurstone (1931)



x's: "median" switch point.

Thurstone (1931)

Hats (x_1), shoes (x_2), & overcoats (x_3)

$$k_1 \log(x_1) + k_2 \log(x_2) + k_3 \log(x_3) = \log(\bar{u})$$

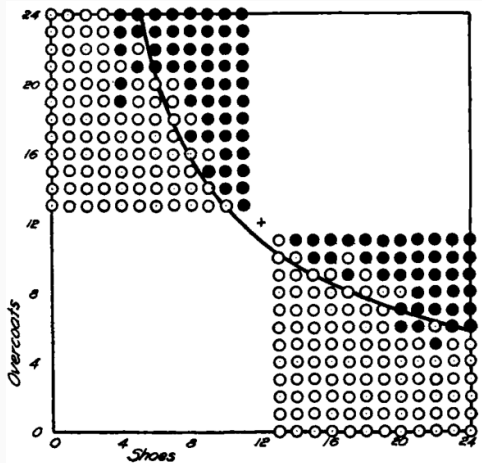
- Fix $k_1 = 1$
- Hats vs. shoes: $k_2^* = 1.26$
- Hats vs. overcoats: $k_3^* = 1.32$
- Now predict shoes vs overcoats

Thurstone (1931)

Predicted equations for the indifference curves (shoes-overcoats) with four standard combinations	Standard Shoes	combination Overcoats
$1.26 \log x_2 + 1.32 \log x_3 = 1.94$	6	6
$1.26 \log x_2 + 1.32 \log x_3 = 2.26$	8	8
$1.26 \log x_2 + 1.32 \log x_3 = 2.51$	10	10
$1.26 \log x_2 + 1.32 \log x_3 = 2.72$	12	12

Four different reference bundles

Thurstone (1931)



“agreement is... quite satisfactory”

Let's Evaluate This Paper

1. What are the treatments (if any)?
2. What does this design *really* test?
 - Does that match the stated hypotheses?
3. Proper incentives?
4. Subject pool?
5. Proper sample size? Statistical power?
6. What's the main result?
7. Are the statistical tests appropriate?
8. Are the results conclusive?
9. Robustness?
10. Extensions?

Wallis-Friedman Critique #1: Incentives



W. Allen Wallis & Milton Friedman (1942)

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

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

Wallis-Friedman Critique #2: Identifiability

They argue indifference curves can't be identified!

- Consumer choice has 3 components
 - Physical quantities of goods
 - “Taste factors” shape preferences.
 - “Opportunity factors” shape what's available. Budgets.
- e.g., prices may affect tastes *and* opportunities
- Need independent variation in one factor
- They argue that's impossible!

Why not study choices without opportunity factors?? Like Thurstone?

- The resulting data would be a “hodgepodge of psychology and economics”
- Meaning: inferred utility may not generalize to constrained choice settings

Rousseas & Hart (1951)

Stephen Rousseas (PhD student) & Albert Hart (Prof, Columbia U.)

- Wanted real incentives... plates of eggs & bacon!!

9a

NAME JONES, JOHN J.
Last First Middle
(Please Print)

Choice	Bacon	Eggs	Order of Pref.
A	1 $\frac{3}{4}$	1 $\frac{1}{4}$	
B	2 $\frac{3}{4}$	1 $\frac{3}{4}$	
C	2	2	

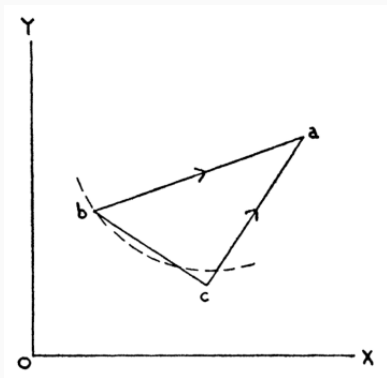
Instructions:

1. List your order of preference using the numbers 1, 2 and 3.
2. If you are indifferent to two or all three of the choices offered write "N" in the appropriate two or all three boxes.

Rousseas & Hart (1951)

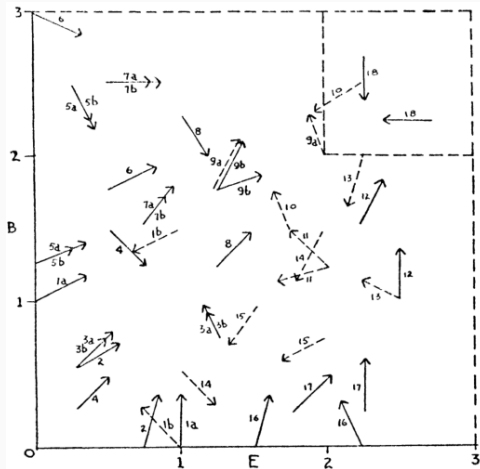
- Rank 3 plates
- Subjects ate their most-preferred plate
 - Eggs were scrambled
- Intransitivity not allowed
- Subjects repeat the task 1 month later
- 67 subjects
 - Class of graduate sociology students
 - Recruitment issues! They wanted 72

Rousseas & Hart (1951)



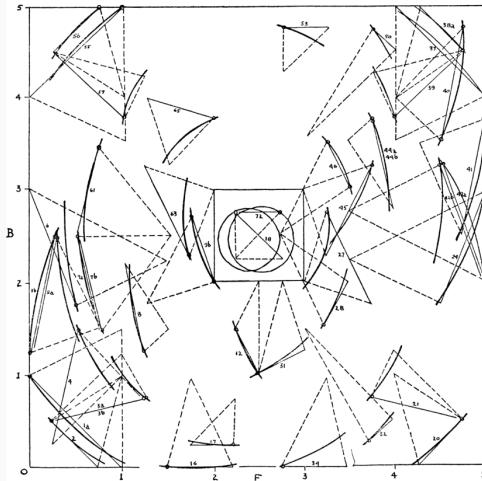
Construct "saturation" vector (gradient of u)

Rousseas & Hart (1951)



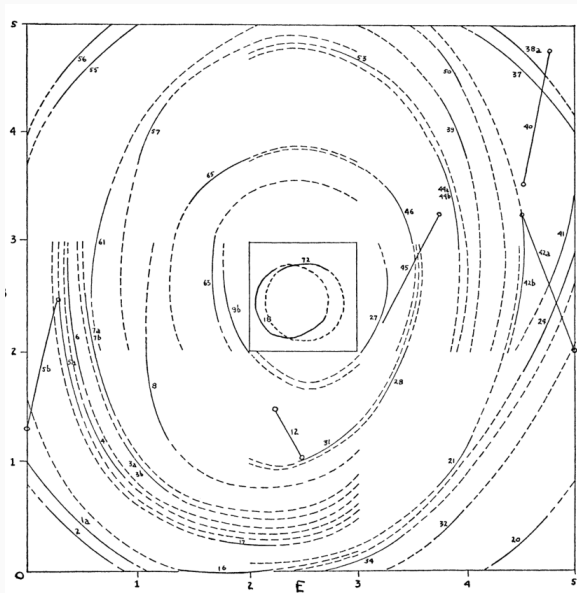
Heterogeneity

Rousseas & Hart (1951)



Combined subjects with choices consistent with a certain saturation point

Rousseas & Hart (1951)



Let's Evaluate This Paper

1. What are the treatments (if any)?
2. What does this design *really* test?
 - Does that match the stated hypotheses?
3. Proper incentives?
4. Subject pool?
5. Proper sample size? Statistical power?
6. What's the main result?
7. Are the statistical tests appropriate?
8. Are the results conclusive?
9. Robustness?
10. Extensions?

Expected Utility



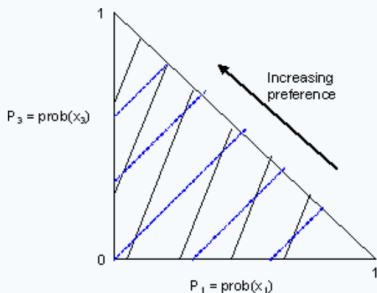
John von Neumann & Oskar Morgenstern

- 1920s: U. Vienna: von Mises advised Hayek (1921) and Oskar Morgenstern (1925)
- 1935: Morgenstern wrote “Perfect Foresight and Economic Equilibrium”, criticizing price-taking theories
- Colleague pointed him to von Neumann’s 1928 “Zur Theorie der Gesellschaftsspiele”
- The two meet at Princeton
- 1944: they publish *The Theory of Games & Economic Behavior*
- Opening chapters: Expected Utility Theory

Expected Utility

- Choice objects: objective lotteries \mathcal{L}
- \succeq over \mathcal{L}
 1. Complete & transitive
 2. Continuous
 3. Linear & parallel indifference curves (“independence”)

Machina-Marshak triangle:



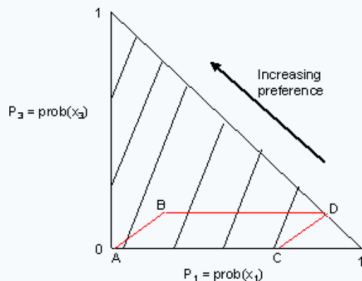
Allais Paradox

1953: Maurice Allais survey of colleagues & friends

- **Option A:** 100% chance of \$1M
- Option B: 10% \$5M, 89% \$1, 1% \$0

and

- Option C: 11% \$1M, 89% \$0
- **Option D:** 10% \$5M, 90% \$0



Mosteller & Noguee (1951)

But first let's rewind 2 years

Mosteller & Noguee "An Experimental Measurement of Utility" *JPE*

- 10 Harvard undergrads + 7 Mass. National Guardsmen
- Gathered data on their financial situation & aspirations
- ≈ 3 1-hour sessions per week for 10 weeks
 - 3 drop-outs
- \$1 endowment for gambling (or psych. tests)
- The gamble: poker dice
 - 5 dice rolled to create a "hand"
 - Hands are ranked
 - A baseline hand H is shown
 - Wager: win \$ x if you roll a hand $\succ H$
 - Wager costs 5 cents
 - Choice: Bet or not bet for each wager
 - Around 2,000 observations

Mosteller & Noguee (1951)

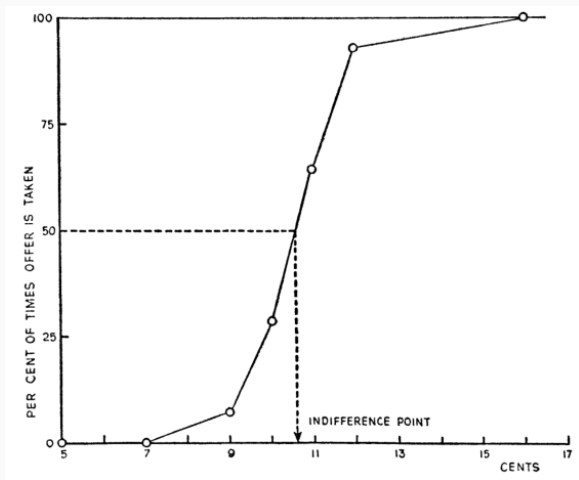
- Training sessions: Probabilities not calculated
- Known sessions: Probabilities given
- Doublet sessions: two baseline hands w/ different rewards
 - 20 cents if you beat 22263
 - 3 cents if you beat 66431 (but not 22263)
- Goal: Estimate cardinal utility from Known sessions, predict Doublet
- How to get utility estimates?
 1. Find prize that gives indifference. Call that A
 2. Under EU, calculate $u(A)$

$$pu(A) + (1 - p)u(-5) = u(0)$$

$$pu(A) + (1 - p)(-1) = (0)$$

$$u(A) = \frac{1 - p}{p}$$

Mosteller & Noguee (1951)



$$p = 0.332 \Rightarrow U(10.6) = (1 - p)/p \approx 2$$

Mosteller & Noguee (1951)

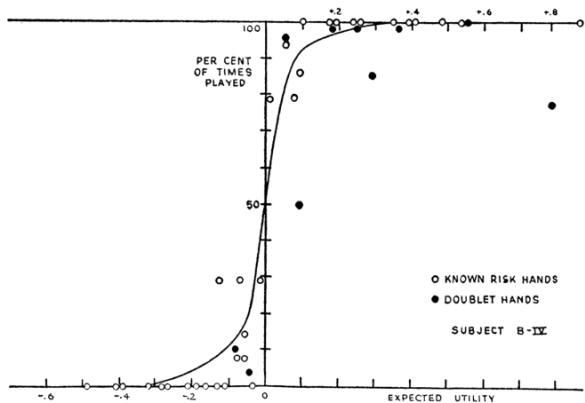


FIG. 4b.—Percentage times played plotted against expected utility for subject B-IV. Curve is freehand fit to Known-risk points. The fit of this curve to the Doublet points shows roughly how good predictions of subsequent behavior are.

The fit of EU

TABLE 14
 NUMBER AND PERCENTAGE OF CASES* CORRECTLY PREDICTED AS A FUNCTION
 OF DIFFERENCE IN EXPECTED UTILITY AND OF DIFFERENCE IN
 EXPECTED MONEY VALUE

PREDICTION BY EXPECTED UTILITY VALUE				PREDICTION BY EXPECTED MONEY VALUE			
Difference in Expectancies in Utiles of the Two Offers	All Groups Combined			Difference in Actuarial Values (in Cents) of Two Offers	All Groups Combined		
	No. Right	No. Wrong	Per Cent Right		No. Right	No. Wrong	Per Cent Right
0.01-0.15	37	32	54	\$0.02-\$0.50 . . .	34	46	42
0.16-0.30	21	7	75	0.55- 1.55 . . .	17	13	57
0.31-0.45	13	3	81	1.80- 2.20 . . .	9	6	60
0.46 and above.	14	1	93	2.50 and above	10	5	67

* A "case" represents the preference displayed by one subject for one pair of offers.

Overall prediction accuracy

Mosteller & Nogee (1951)

Other notes

- Also include paired-choice tasks
- Even look at response times! But not shown

Let's Evaluate This Paper

1. What are the treatments (if any)?
2. What does this design *really* test?
 - Does that match the stated hypotheses?
3. Proper incentives?
4. Subject pool?
5. Proper sample size? Statistical power?
6. What's the main result?
7. Are the statistical tests appropriate?
8. Are the results conclusive?
9. Robustness?
10. Extensions?

Game Theory

The Flood-Dresher Experiment

- Merrill Flood & Melvin Dresher, RAND Corporation, 1950

		Player 2 (John Williams)	
		(1) Defect	(2) Cooperate
Player 1 (Armen Alchian)	(2) Cooperate	2 -1	1 0.5
	(1) Defect	0 0.5	1 -1

Albert Tucker: 1950 "Prisoners' Dilemma"

- Play 100 times
- Sum of payoffs
- No communication, but log your thoughts

Flood-Dresher

<i>Game</i>	<i>Moves</i>		<i>Armen Alchian's comments</i>	<i>John Williams's comments</i>
	<i>AA</i>	<i>JW</i>		
1	D	C	JW will play D—sure win. Hence if I play C—I lose.	Hope he's bright.
2	D	C	What is he doing ???	He isn't but maybe he'll wise up
3	D	D	Trying mixed ?	Okay, dope.
4	D	D	Has he settled on D ?	Okay, dope.
5	C	D	Perverse!	It isn't the best of all possible worlds.
6	D	C	I'm sticking to D since he will mix for at least 4 more times.	Oh ho! Guess I'll have to give him another chance.
7	D	C		Cagey, ain't he ? Well...
8	D	D		In time he could learn, but not in ten moves so:

I can guarantee myself a gain of 5, and guarantee that Player AA breaks even (at best). On the other hand, with nominal assistance from AA, I can transfer the guarantee of 5 to player AA and make 10 for myself, too. This means I have control of the game to a large extent, so player AA had better appreciate this and get on the bandwagon. With small amounts of money at stake, I would (as above) try (by using C) to coax AA into mutually profitable actions. With large amounts at stake I would play D until AA displayed some initiative

Flood-Dresher

17	C	D		The stinker.
18	C	D		He's crazy. I'll teach him the hard way.
19	D	D	I'm completely confused. Is he trying to convey information to me ?	Let him suffer.
20	C	D		
21	C	C		Maybe he'll be a good boy now.
22	D	C		Always takes time to learn.

Flood-Dresher

90 C C

91 C C

When will he switch as a last minute grab of D ? Can I beat him to it as late as possible ?

92 C C

Good.

93 C C

94 C C

95 C C

96 C C

97 C C

98 C C

99 D C

100 D D

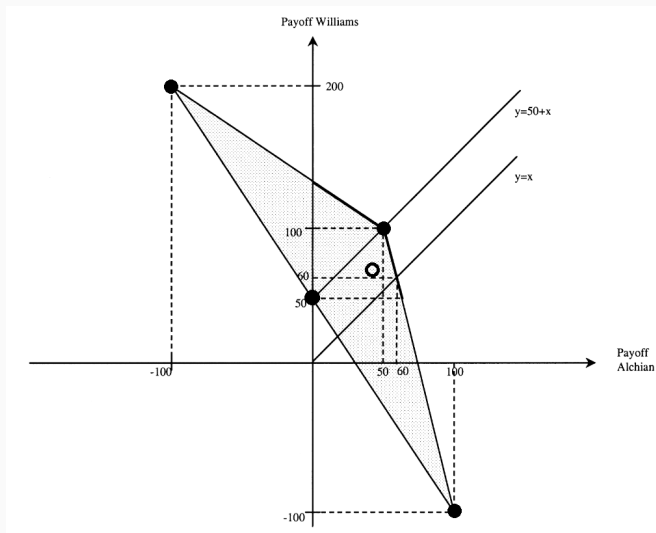
Cycles:

1. Stuck in DD
2. Williams breaks by playing C
3. Alchian responds with C
4. Alchian eventually switches back to D
5. Punishment phase
6. Williams breaks by playing C

Williams's goal: "to coax Alchian into mutually profitable actions"

Cycles did get longer each time

Flood-Dresher: De Hert (2003) analysis



Reject Nash in favor of Split-the-Difference

Nash's response (also at RAND):

- “The flaw in the experiment as a test of equilibrium point theory is that the experiment really amounts to having the players play one large multi-move game. One cannot just as well think of the thing as a sequence of independent games as one can in zero-sum cases.”
- Agrees that DD is only true equilibrium
- But anti-reciprocal strategy is “very near equilibrium” is near equilibrium of finite game, and is an equilibrium of an indefinite game
- They were irrational for not playing CC more often!
- Random rematching would remove the interaction

Further RAND Experiments

Kalisch, Milnor, Nash, and Nering (1952) advice:

1. Keep communication minimal and structured
2. Random rematching
3. Use asymmetric payoffs to avoid an “obviously fair” split



Thomas Schelling (1957)

Game 1:

- 2 players, each picks $s_i \in [0, 100]$
- If $s_1 + s_2 \leq 100$ then keep (s_1, s_2)
- If $s_1 + s_2 > 100$ then keep $(0, 0)$
- What's the modal choice?

Game 2:

- 3 players, labeled A, B, and C
- $s_i \in \{ABC, ACB, BAC, BCA, CAB, CBA\}$
- If $s_1 = s_2 = s_3 = s^*$ then:
 - Player $s^*(1)$ gets \$3
 - Player $s^*(2)$ gets \$2
 - Player $s^*(3)$ gets \$1
- Otherwise all get \$0
- What's the modal choice?

Thomas Schelling (1957)

Game 1:

- 2 players, each picks $s_i \in [0, 100]$
- If $s_1 + s_2 \leq 100$ then keep (s_1, s_2)
- If $s_1 + s_2 > 100$ then keep $(0, 0)$
- What's the modal choice? **$s_i = 50$ (90%)**

Game 2:

- 3 players, labeled A, B, and C
- $s_i \in \{ABC, ACB, BAC, BCA, CAB, CBA\}$
- If $s_1 = s_2 = s_3 = s^*$ then:
 - Player $s^*(1)$ gets \$3
 - Player $s^*(2)$ gets \$2
 - Player $s^*(3)$ gets \$1
- Otherwise all get \$0
- What's the modal choice?

Thomas Schelling (1957)

Game 1:

- 2 players, each picks $s_i \in [0, 100]$
- If $s_1 + s_2 \leq 100$ then keep (s_1, s_2)
- If $s_1 + s_2 > 100$ then keep $(0, 0)$
- What's the modal choice? $s_i = 50$ (90%)

Game 2:

- 3 players, labeled A, B, and C
- $s_i \in \{ABC, ACB, BAC, BCA, CAB, CBA\}$
- If $s_1 = s_2 = s_3 = s^*$ then:
 - Player $s^*(1)$ gets \$3
 - Player $s^*(2)$ gets \$2
 - Player $s^*(3)$ gets \$1
- Otherwise all get \$0
- What's the modal choice? $s_i = ABC$ (A:75%,B:83%,C:88%)

Early Learning Theory Experiment

Suppes & Atkinson (1960)

- How do players adapt over time?
- Most treatments: hypothetical payoffs
- Usually little knowledge of payoffs, opponent
- Lots of confounds (different games per treatment, etc)
- But... incentives did change behavior
- Precursor to 1990s–2000s boom in learning theories

Markets & Industrial Organization



Edward H. Chamberlain

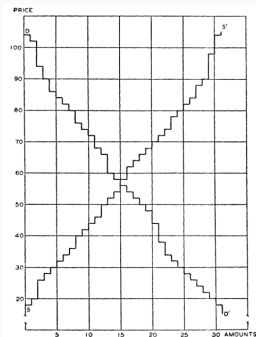
- “Father of IO”
 - Monopolistic competition
 - Product differentiation
 - Patents
- “An Experimental Imperfect Market” *JPE* (1948)
 - Classroom market experiments

“...economics is limited by the fact that resort cannot be had to the laboratory techniques of the natural sciences. On the one hand, the data of real life are necessarily the product of many influences other than those which it is desired to isolate... On the other hand, the unwanted variables cannot be held constant or eliminated in an economic ‘laboratory’ because the real world of human beings, firms, markets, and governments cannot be reproduced artificially and controlled. The social scientist who would like to study in isolation and under known conditions the effects of particular forces is, for the most part, obliged to conduct his ‘experiment’ by the application of general reasoning to abstract ‘models.’

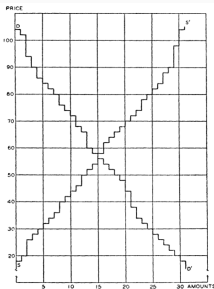
The purpose of this article is to make a very tiny breach in this position: to describe an actual experiment with a ‘market’ under laboratory conditions and to set forth some of the conclusions indicated by it.”

Edward Chamberlin

- Hypothesis: Markets will **not** equilibrate because deals can't be recontracted
- Buyers & sellers
- Single-unit supply & demand w/ induced values
- No centralized order book; move about the room
- Hypothetical payoffs



Edward Chamberlin



Main result:

- Quantity too large
 - Makes sense: inefficient matchings
- Price too low
 - Mystery! Multiple conjectures
 1. Students are used to being buyers, not sellers
 2. Buyers have money, giving outside options. Sellers don't



Vernon Smith & Charlie Plott

- Smith: first experiment 1955 at Purdue
- Inspired by Chamberlin
- Plott joined Purdue 1965, Vernon left 1967, Charlie 1970
- Charlie encouraged Vernon to formalize the theory
- Extensive exploration of market equilibration
 - Slopes of supply & demand
 - Affects path of convergence
 - Price floors & ceilings
 - Multiple simultaneous markets
 - Demand, supply, policy shocks
 - Rate of re-equilibration
- ⋮



Sidney Siegel & Lawrence Fouraker

- Siegel: Non-parametric stats, lots of econ-related experiments
 - Siegel-Tukey test: Wilcoxon but ranks=extremeness

Siegel & Fouraker (1960) *Bargaining and group decision making: Experiments in bilateral monopoly* (book)

- Bilateral bargaining (1 seller, 1 buyer) on price & quantity
- Payoff table for *own* profits at each (p, q)
- Payoff information treatments:
 - (1) Private, (2) one public, (3) both public
 - \uparrow information $\Rightarrow \uparrow$ PO & equal split
- Anonymous interactions
- Varied the size of surplus away from the PO allocation
 - \downarrow surplus away from PO $\Rightarrow \uparrow$ PO contracts
- Theory: Level of aspiration, affected by info.

Public Goods (Ledyard 1995)

Ledyard's Survey



John Ledyard

- 1967 Purdue PhD under Reiter, Plott, Smith. Overlap w/ Kagel
- Mechanism design theorist: Groves-Ledyard mechanism
- Largely focused on PG problems

The Survey: Ledyard's Summary

1. One-shot: about halfway between PO & self-interest
2. *Declines over time*
3. Face-to-face communication increases cooperation

Early Public Goods Experiments

- Robyn Dawes & John Orbell (social psychology)
 - “Social Dilemma” games
 - $x_i \in \{X, 0\}$. X pays more, but hurts everyone
 - 31% 0 w/out communication, 72% with
- Gerald Marwell & Ruth Ames (sociology)
 - Classic PG setup: tokens in private vs. public account
 - Private account pays more, public account benefits everyone
 - Result: 51–57% of tokens put in public account
 - Problem 1: discontinuous payoff function \Rightarrow multiple NE
 - Problem 2: deception about group size (80 vs 4)
 - “Economists Free Ride, Does Anybody Else?” *J.Pub.Econ* (1981)
 - 12 experiments
 - Baselines: \approx 50% contribution
 - High stakes: \approx 30%
 - Econ PhD students: 20%

Economists React

Mark Isaac (Caltech PhD, Arizona, FSU)

Jimmy Walker (Texas A&M PhD, Arizona, Indiana)

- Isaac, McCue & Plott (1985), Kim & Walker (1984)
 - Indefinitely repeated
 - No communication
 - Linear public good (single-period dom strat)
 - Result:
 - First period: 50% of max payoffs
 - By 5th period: 9% of max payoffs
- Isaac, Walker & Thomas (1984) & the MPCR
 - Computerized (“PLATO”) for true anonymity, control
 - Discovered the MPCR confound:

The MPCR

Marginal Per-Capita Return (MPCR) and Group Size:

- Endowment ω_i , Contribution c_i , PG $\sum_j c_j$ (shared equally)
- Price p , Benefit a , with $a > p > a/n$

Individual incentive: $\pi_i = p \cdot (\omega_i - c_i) + a \cdot (\sum_j c_j)/n$

$$\frac{\partial \pi_i}{\partial c_i} = -p + a/n < 0$$

$$\frac{a/n}{p} < 1$$

$$\text{MPCR} < 1 \Rightarrow c_i^* = 0$$

Social benefit (altruism/PO): $\sum_j \pi_j p \cdot \sum_j (\omega_j - c_j) + a \cdot n(\sum_j c_j)/n$

$$\frac{\partial \sum_j \pi_j}{\partial c_i} = -p + a > 0$$

$$a/p > 1 \Rightarrow c_i^0 = \omega_i$$

The MPCR and Group Size

What's the effect of increasing n ?

1. Change in incentives:

- Individual incentive: $MPCR = (a/n)/p$: $\uparrow n \Rightarrow \downarrow MPCR \Rightarrow \downarrow c_i$
- Social benefit (altruism): $a/p \Rightarrow$ no change

2. What if we increase a along with n ?

- Individual incentive: $MPCR = (a/n)/p$: \Rightarrow no change
- Social benefit (altruism): $a/p \Rightarrow \uparrow c_i$

Solution: ???

The MPCR and Group Size

What's the effect of increasing n ?

1. Change in incentives:

- Individual incentive: $MPCR = (a/n)/p$: $\uparrow n \Rightarrow \downarrow MPCR \Rightarrow \downarrow c_i$
- Social benefit (altruism): $a/p \Rightarrow$ no change

2. What if we increase a along with n ?

- Individual incentive: $MPCR = (a/n)/p \Rightarrow$ no change
- Social benefit (altruism): $a/p \Rightarrow \uparrow c_i$

Solution: Vary n and $MPCR$ independently. 2×2 design.

General Patterns of Results

1. Initial round: 51%, clear decline
2. Session-level variance
3. Higher *MPCR* (0.3 to 0.75) \Rightarrow much higher c_i at either n
4. Experience reduces contributions
5. At low $M = 0.3$:
 - Repetition reduces contributions
 - Larger group size increases contributions
6. At high $M = 0.75$:
 - Repetition and group size have little effect

Later work:

- Thresholds, provision points, other mechanisms
- Deeper explorations of repetition, group size, communication...
- HUGE literature



Elinor “Lin” Ostrom

- Beverly Hill High School, but poor. Mother discouraged college.
- Rejected at UCLA Econ PhD for lack of math. Got into PoliSci
- PhD work on water basins in SoCal: Tragedy of the Commons
- Community found ways to self-enforce responsible management
- Career at Indiana U. Lots of field work
- Irrigation systems in Spain, Nepal
- 8 design principles for effective self-governance
 - Defining group boundaries, inclusive decision-making, effective monitoring, scaled sanctions, dispute resolution protocols, respect from authority
- “Ostrom’s Law:” A resource arrangement that works in practice can work in theory
- 2009 Nobel prize. Passed away 2012