EITM and (Incentivized) Experiments

Morning Session

Jonathan Woon
University of Pittsburgh

EITM @ University of Houston
June 24, 2016

Outline

1. Some thoughts on EITM

2. Let’s play!

3. Intro to $$$ experiments

4. Examples

What is science?

Does science need models?

Why “test” theoretical models?

What do experiments offer?

Feynman on science

https://youtu.be/OL6-x0modwY

Feynman on science

“If it disagrees with experiment, it’s wrong. In that simple statement is the key to science.”

What is a model?

"Models are a **constrained, best effort** to capture what the modeler believes to be the **essence** of a complex empirical phenomenon or at least **an important aspect of it**."

(Powell 1999)

What is a model?

- Inspired by real-world, empirical phenomena
- Partial representations – abstract, simplified versions of reality
- Identifies key features – explicit assumptions about what we think matters
- Need not be formal, mathematical – but needs to be clearly defined, coherent, logically consistent
- Formal models generate sharp predictions, testable empirical implications

Theoretical models in political science

- Predominantly rational choice, game theory

- Important insights in all subfields
  - Proposal and veto power (Romer and Rosenthal, Baron and Ferejohn)
  - Information and communication (Gilligan and Krehbiel, Fearon)
  - Commitment problems (Acemoglu and Robinson, Powell)

Should we “test” models?

- Clarke and Primo say NO: models are like “maps” and their “usefulness” depends on their purpose

- How do we know if a model is “useful”?

Is this map useful?

Should we "test" models?

- Clarke and Primo say NO: models are like "maps" and their "usefulness" depends on their purpose

- How do we know if a model is "useful"?

- But Clarke and Primo are right that H-D is a narrow way of thinking about science

Modeling dialogue

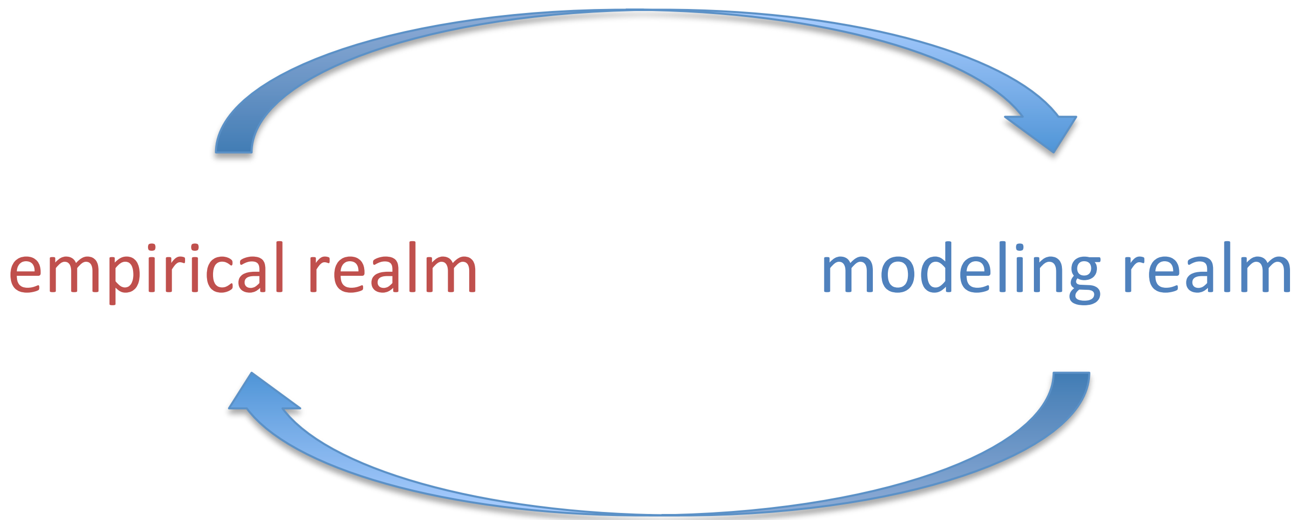
"Process in which **theorists and empiricists work together interactively** on the difficult task of finding tractable models that capture and clarify important aspects of real situations."

Roger Myerson. 1992. "On the Value of Game Theory in Social Science"

Modeling dialogue

- “Theoretical output…guide[s]…search for patterns in empirical data.”
- “Empiricists must help theorists refine and extend their models.”
- “Simplifying assumptions must be tested and challenged.”
- “We must constantly compare the predictions of our simple models with what we know about the real world and **ask whether the appropriate simplifications have been made**.”

Modeling dialogue



THIS IS EITM!

Today's objectives

- Basic principles of experiments: control and incentives
- Examples of dialogue between theory and experiments, design considerations
- Whet your appetite for behavioral experiments

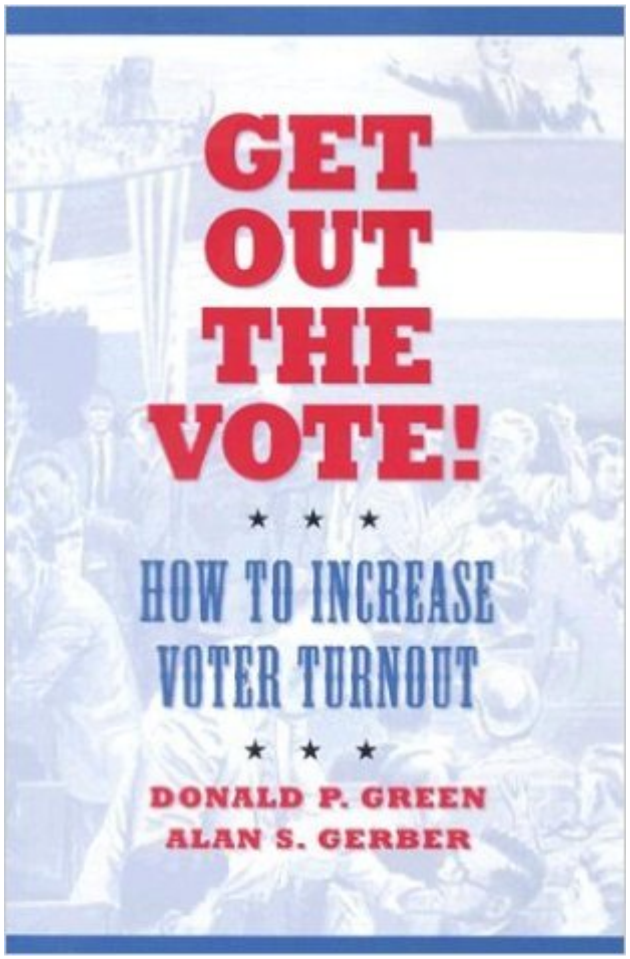
Let’s play!

- Before we talk about experiments, let’s participate in some experiments

- http://veconlab.econ.virginia.edu/login1.php

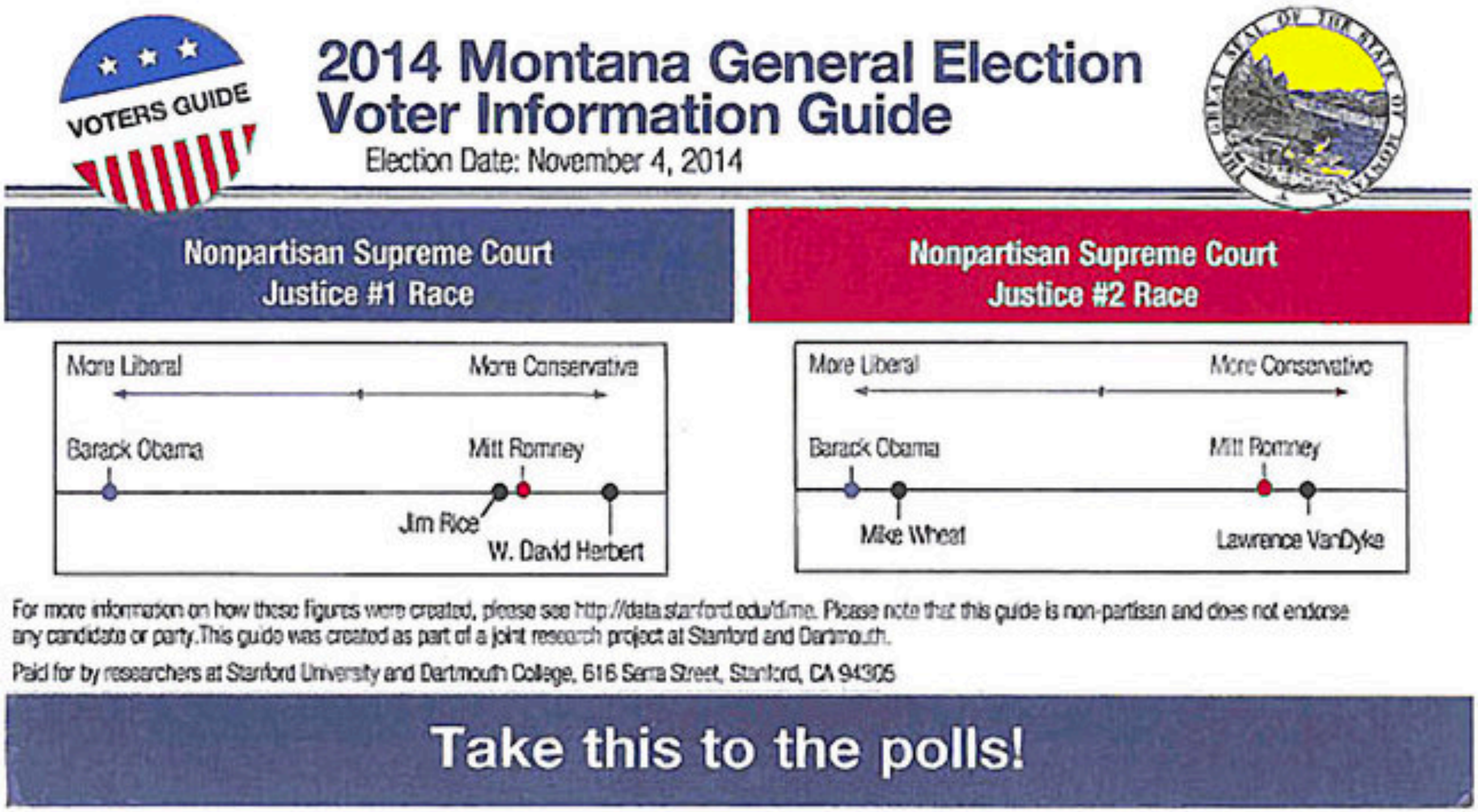
- Session name: woon1

Political science experiments



$$E[\delta] = E[Y_1 - Y_0]$$

Political science experiments



Common objections to lab experiments

Artificial and unrealistic!

Small stakes!

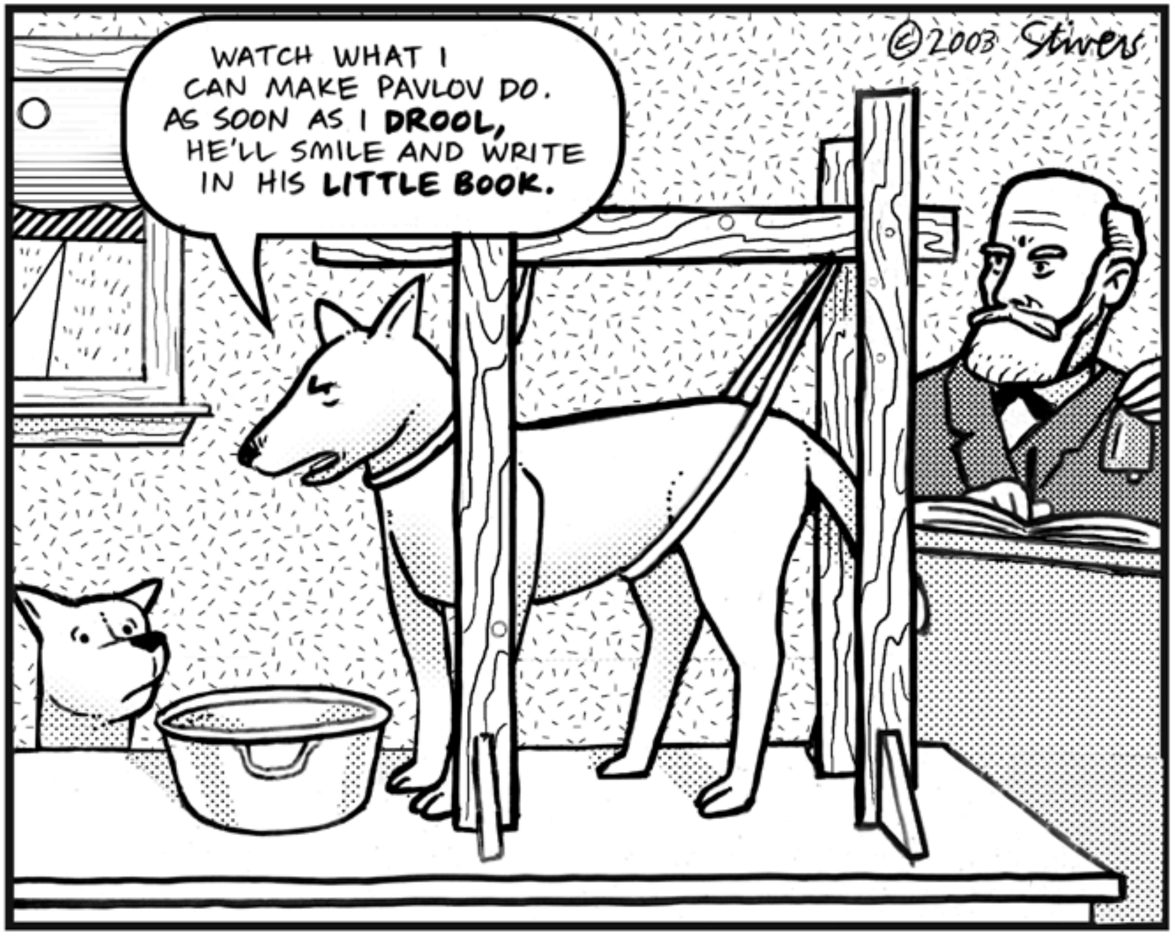
Student samples!

Galileo’s inclined plane

Newton’s prism

Mendel’s peas

Pavlov's dogs



Control


YOU ARE IN CONTROL!

Broad definition of experiment

"In an experiment, the researcher intervenes in the data generating process by purposely manipulating elements of the environment."

(Morton and Williams 2010)

Roth’s typology of experiments

Speaking to theorists

Searching for facts

Whispering in the ears of princes

Roth (1995)

Experiments for speaking to theorists

- Directly test game-theoretic predictions under conditions closely resembling the theoretical model
- Artificiality and control are "features" not "bugs"

Game theory's behavioral assumptions

- Rational choice: choices consistent with preferences (e.g., dominance)
- Expected utility: vN-M continuity and independence
- Best responses: rational in strategic settings given beliefs
- Nash equilibrium: mutually consistent beliefs and actions
- Forward-looking, sequentially rational
- Bayesian: rational beliefs and learning

Left	Right
$0	$10

Induced value theory

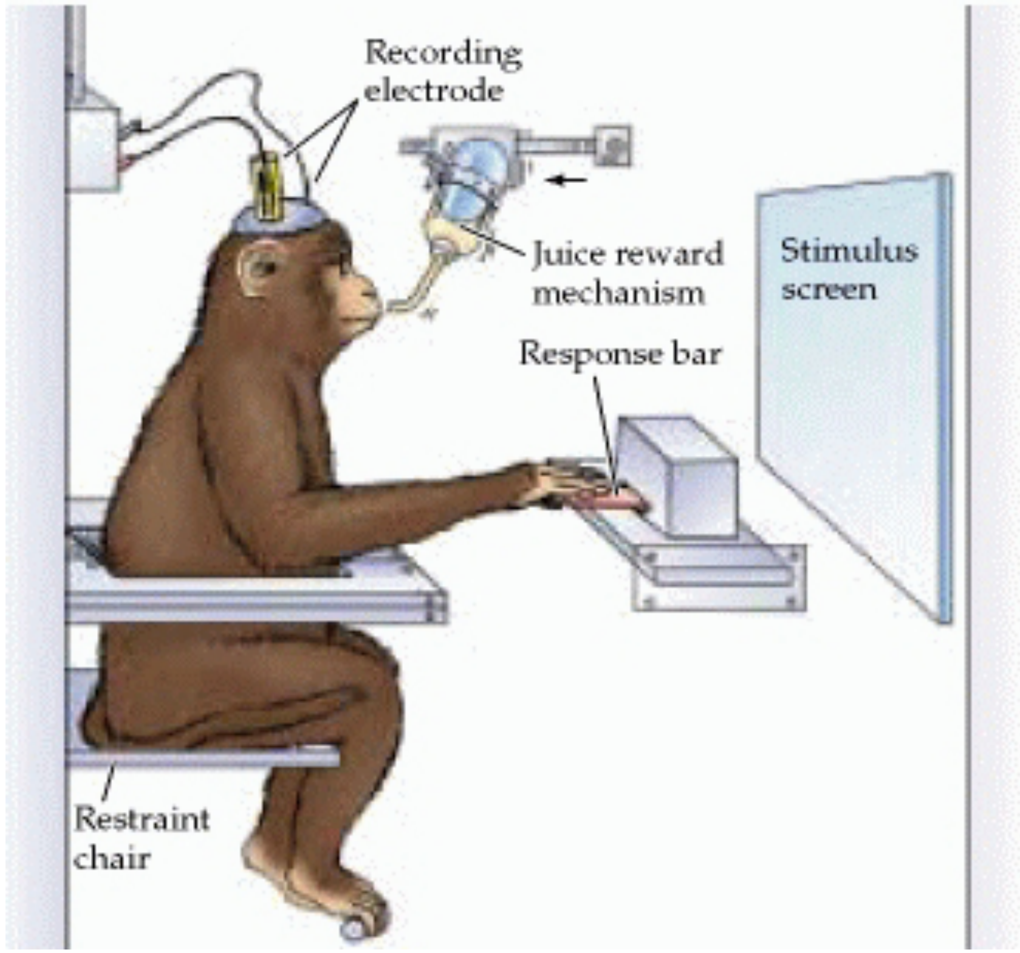
“Control can be exercised by using a reward structure… to induce prescribed monetary value on (abstract) outcomes.” (Vernon Smith 1982)

“Proper use of a reward medium allows an experimenter to induce pre-specified characteristics in experimental subjects.” (Friedman and Sunder 1994)

Induced value theory

Sufficient conditions for control of preferences

1. Monotonicity	More reward preferred to less (non-satiation)
2. Salience	Rewards depend on subjects' actions
3. Dominance	Utility from reward crowds out other, subjective motivations


Recording
electrode
Juice reward
mechanism
Stimulus
screen
Response bar
Restraint
chair


F
Y

But we are not monkeys…or are we?

Implementation

- Computerized interfaces – automate computation of payoffs, maintain anonymity, facilitate randomization, group assignment and matching

- Pencil and paper – easier to implement for simple games and decisions

Control of game form

- Experiment should implement the theoretical game or decision context as closely as possible
- Rules and payoffs need to be clearly explained – common knowledge – and **NO DECEPTION**!
- Use quizzes and examples to ensure comprehension (but tradeoffs)
- Describe the game, but don't tell anyone what to do!

Sample instructions

You will be paid in cash for your participation, and the exact amount you receive will be determined during the experiment and will depend partly on your decisions, partly on the decisions of others, and partly on chance. You will be paid your earnings privately, meaning that no other participant will find out how much you earn. These earnings will be paid to you at the end of the experiment along with the $7 participation payment.

Sample instructions

This experiment consists of several parts. Each part consists of a series of elections, and we will explain the instructions for each part before beginning that part. There are a total of 50 elections divided into two parts.

We will **randomly select one election to count** for payment from the entire session. Each election is equally likely to be selected. The points you receive from that election will be used to calculate your payment for the experiment, and points will be converted to cash at the rate of $1 for every 10 points. More specifically, we will take the total number of points you earned in the election that counts, divide by 10, and then round this amount to the nearest quarter. You should think of each election as a separate decision task.

Sample instructions

Matching

For each election, you will be randomly matched against one other candidate. This matching is independent across elections.

Campaign Stage

In the campaign stage, you choose a whole number from 1 to 200. This number is a "campaign promise" and you can think of it as a position or stance on a particular policy issue that both voters and candidates care about. All candidates choose their campaign promises at the same time. If a candidate wins the election, then the winning candidate's campaign promise will determine the payoffs for both participants involved in that election.

Sample instructions

<u>Voting Stage</u>

In the voting stage, a “computer voter” chooses the winner of the election. The computer voter is like a robot programmed to always vote for the candidate whose campaign promise gives it the higher payoff value. As described next, this is the promise closest to the computer voter’s favorite position. If both candidates offer the computer voter the same payoff, then the computer voter will cast its vote randomly between the two candidates, with votes for each candidate equally likely.

Sample instructions

<u>Payoffs</u>

In each round, you will be assigned a “favorite position” and you will earn points based on how close the winning candidate’s campaign promise is to your favorite position.

The closer the winning campaign promise is to your favorite position, the more points you will earn. Specifically, we will compute the absolute difference between the winning campaign promise and your favorite position and then subtract this amount from 200. This is described by the following formula:

Points = 200 - |Winning campaign promise – Your favorite position|

Comprehension check

Before we begin the experiment we would like you to answer a few questions to make sure you understand how the election experiment works. You will answer these questions on your computers and will receive immediate feedback once you answer all of the questions. We will then begin the experiment when everyone has answered these questions.

1. If your favorite position is 100 and the winning candidate's campaign promise is 140, how many points would you earn?

2. If your favorite position is 20 and the winning candidate's campaign promise is 90, how many points would you earn?

3. If your favorite position is 165, your campaign promise is 150, and you win the election, how many points would you earn?

Why not deceive?

- Seems convenient and psychologists do it, but experimental economists have very strong norms against it

- **Deception amounts to a loss of control**
  - If subjects don't believe they are playing the game you describe, they might form their own ideas about what the game is about
  - This creates a mismatch between their actions and rewards, hence incentives lose their salience

Model of theory testing

$$\underbrace{G \wedge P \wedge B}_{\text{Theory}} \Rightarrow H$$

G = Game form (actions, histories, information sets)

P = Preferences (utility function)

B = Behavior (Nash equilibrium)

H = Hypothesis

Woon (2012)

Model of theory testing

$$\neg H \Rightarrow \underbrace{\neg G \vee \neg P \vee \neg B}_{\text{Theoretical implications}}$$

Experimental control of G and P (to the extent possible) increases confidence in the inferences we can make about B

Model of theory testing

$$\neg H \Rightarrow \neg G \vee \neg P \vee \neg B$$

Theoretical implications

Experimental control of G and P (to the extent possible) increases confidence in the inferences we can make about B

Advantages of experiments

- Experiments especially well-suited for making controlled comparisons and for studying decision-making and behavior

- Testing and developing theory – experiments as models – control over key features of the environment (game form, information, payoffs)

- Design new treatments to isolate and tease out causes of a theory's failure

Advantages of experiments

- Investigate behavior in the context of institutions that don't exist in the real world
- Elicit and measure normally unobserved concepts (e.g., risk preferences, beliefs)
- Well-designed experiments can reduce reliance on complicated econometric modeling and assumptions

But no method is perfect

- Can't substitute for your own thinking
- Can't theorize or generate hypotheses
- Not a substitute for observational data
- Control is never 100% complete
- Experiments are best thought of as complements to analysis of observational data

Outline of examples

- Cooperation in social dilemmas
- Strategic sophistication
- Risk preferences
- Electoral accountability
- Gender and candidate emergence

The matrix game

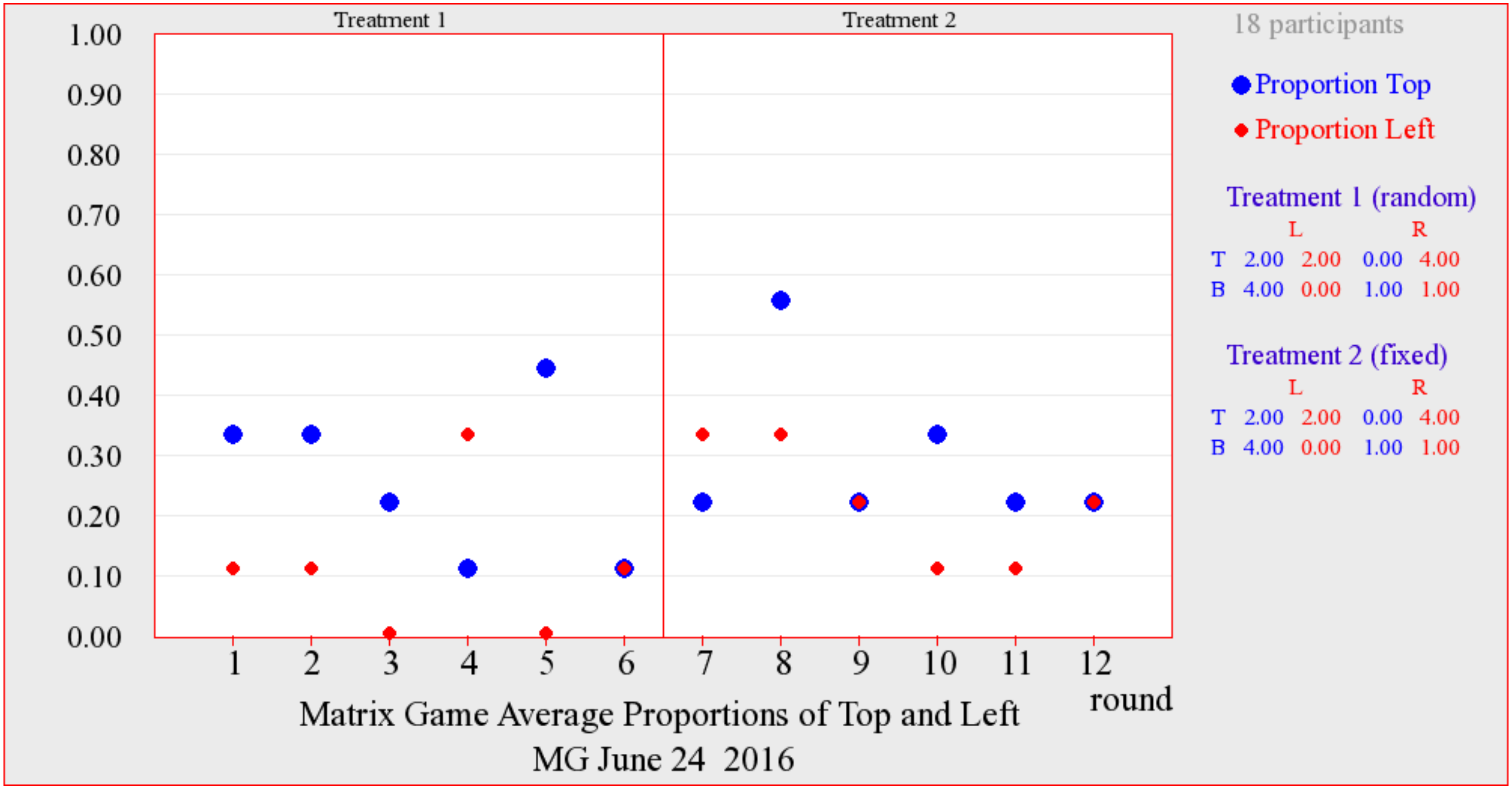
Column

Row

	Left	Right
Up	2, 2	0, 4
Down	4, 0	1, 1

What game is this?

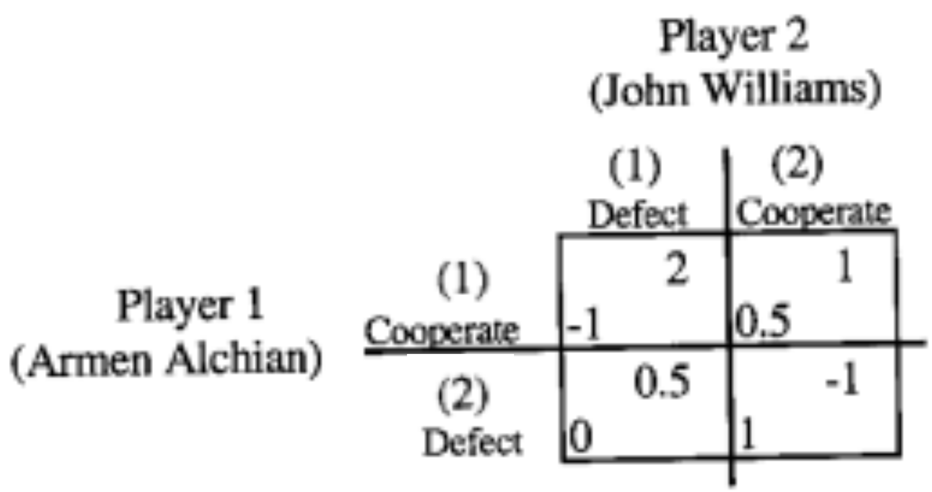
How did you play?



Treatment 1
Treatment 2
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6
7
8
9
10
11
12
round
18 participants
Proportion Top
Proportion Left
Treatment 1 (random)
L R
T 2.00 2.00 0.00 4.00
B 4.00 0.00 1.00 1.00
Treatment 2 (fixed)
L R
T 2.00 2.00 0.00 4.00
B 4.00 0.00 1.00 1.00
Matrix Game Average Proportions of Top and Left
MG June 24 2016

Early PD experiment

RAND mathematicians, Merrill Flood and Melvin Dresher, sought to test Nash's equilibrium concept using a non-zero sum game played 100 times by two of their acquaintances



Flood (1958)

Early PD experiment

(C, C)

TABLE 3

Strategy Frequencies

AA	JW 1	JW 2	JW Total
1	8	60	68
2	14	18	32
Total	22	78	100

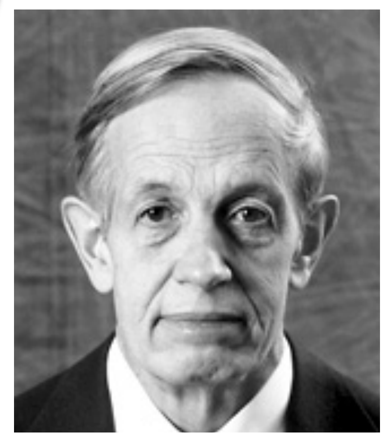
(D, D)

Nash's response

"The flaw in this experiment as a test of equilibrium point theory is that the experiment really amounts to having the players play one large multimove game. One cannot…think of the thing as a sequence of independent games…there is too much interaction."

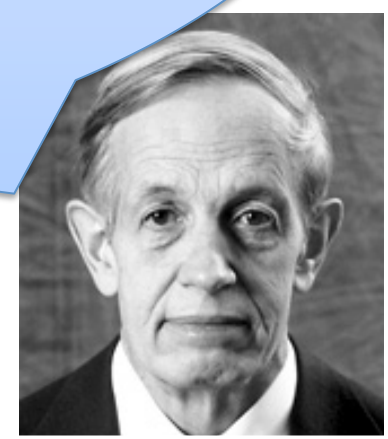
Nash's response

"It is really striking, however, how inefficient AA and JW were in obtaining the rewards. One would have thought them more rational."



Nash's response

"If this experiment were conducted with various different players rotating the competition and with no information given to a player of what choices the others have been making until the end of all trials, then the experimental results would have been quite different, for this modification of procedure would remove the interaction between the trials."



Remarks

- How close is the connection between theory and experiment? Blame the theory? Blame the experiment?
- Generates new theory: Distinction between one-shot and repeated games
- Advances in experimental methodology: Designs with repetition that reduce feedback and interdependence

Design trade-offs

- Experimentalists spend a lot of time thinking about design choices (auxiliary conditions) that can (and often do) affect the results

- One-shot or repetition?

  Allow subjects to learn and gain experience, but possible repeated game effects

- How much feedback?

  No feedback about others would make it difficult to determine best response in games that are not dominance-solvable

- Pay for all rounds or one?

  Paying all rounds might introduce wealth effects, increasing dependence between trials

Alternative explanations

- Altruism or “warm glow”: Players receive non-monetary utility from choosing to cooperate
- Reputations (Kreps et al 1982): Incomplete information about others’ altruism gives rational players incentives to imitate altruistic players early, but defect in later rounds
- How can these explanations be tested? How can we discriminate between competing theories?

Compare matching protocols

- All subjects anonymous (e.g., identified by ID numbers)
- Number of games N known in advance by all subjects
- Perfect strangers ("turnpike") matching: Play game N times, exactly once against each opponent
- Partners (fixed) matching: Play all N games against the same opponent

Andreoni and Miller (1993)

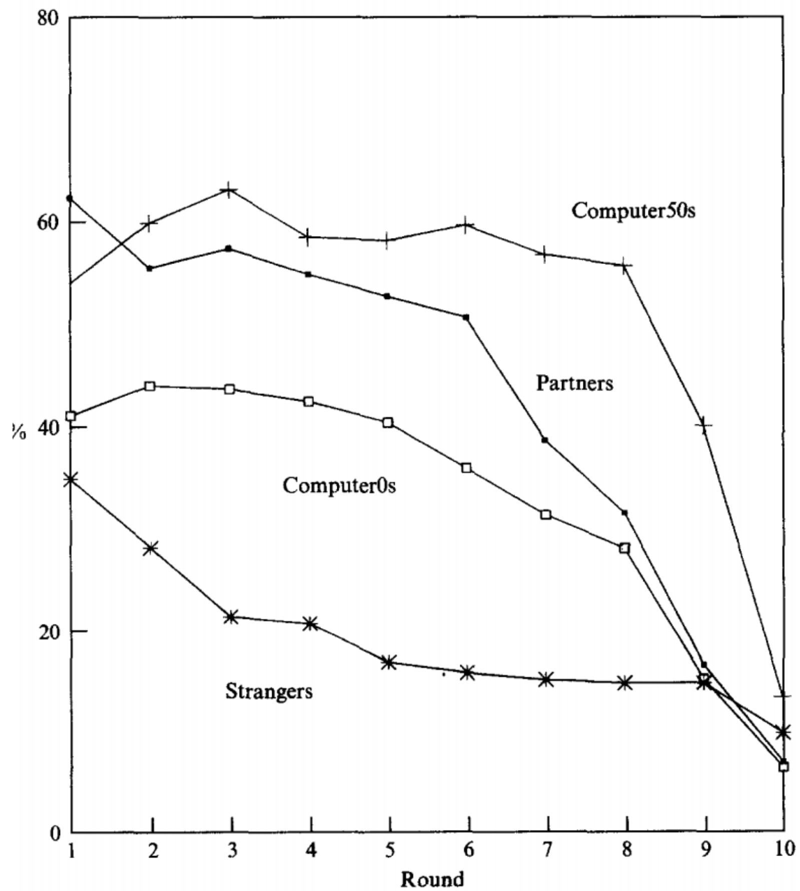
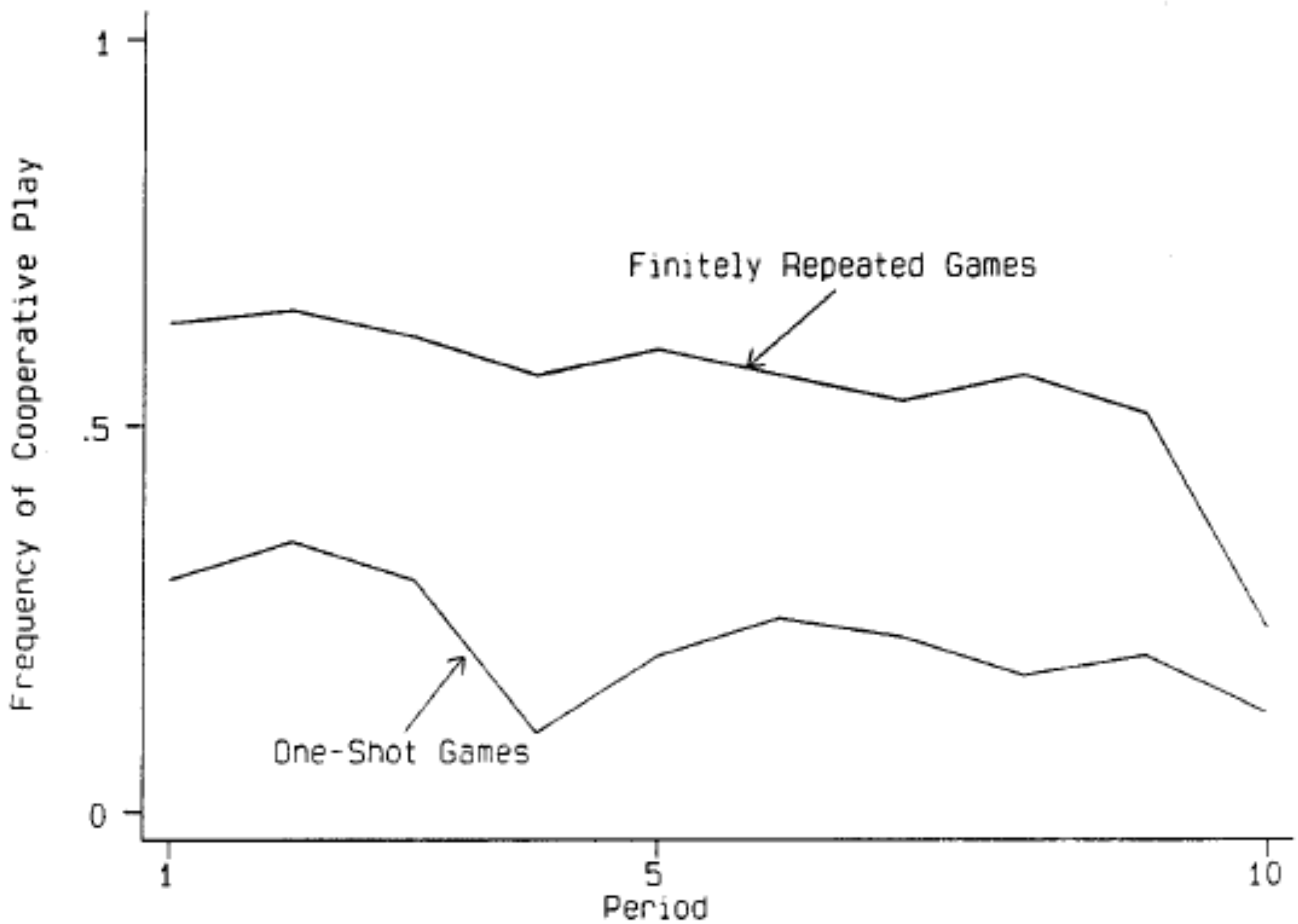


Fig. 2. Percent cooperation by round. Averaged over all 20 10-period games.

Cooper et al (1996)



Cooper et al (1996)

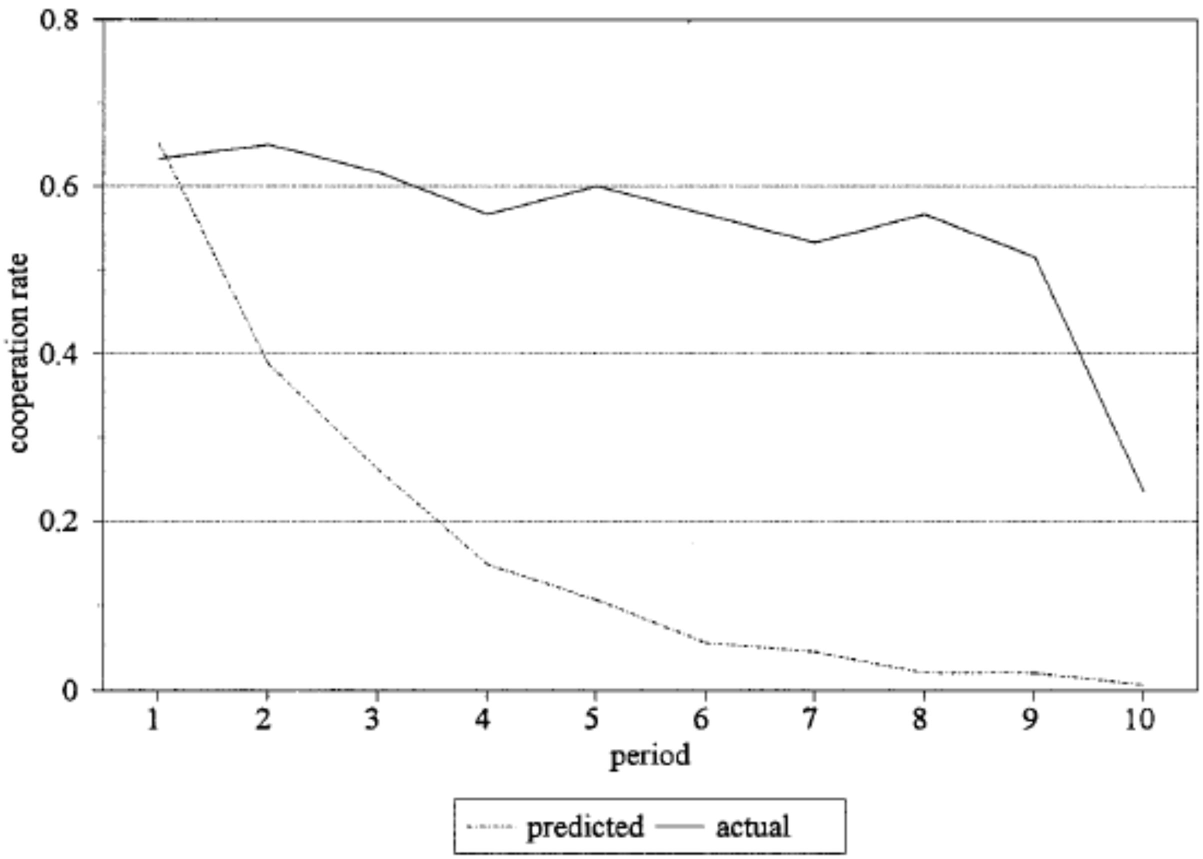


FIG. 3. Predicted and actual cooperation rates for PD-FR.

Collective action and public goods

- What kinds of institutions solve collective action problems?
- Experimenters can push and pull a variety of institutional levers

Communication

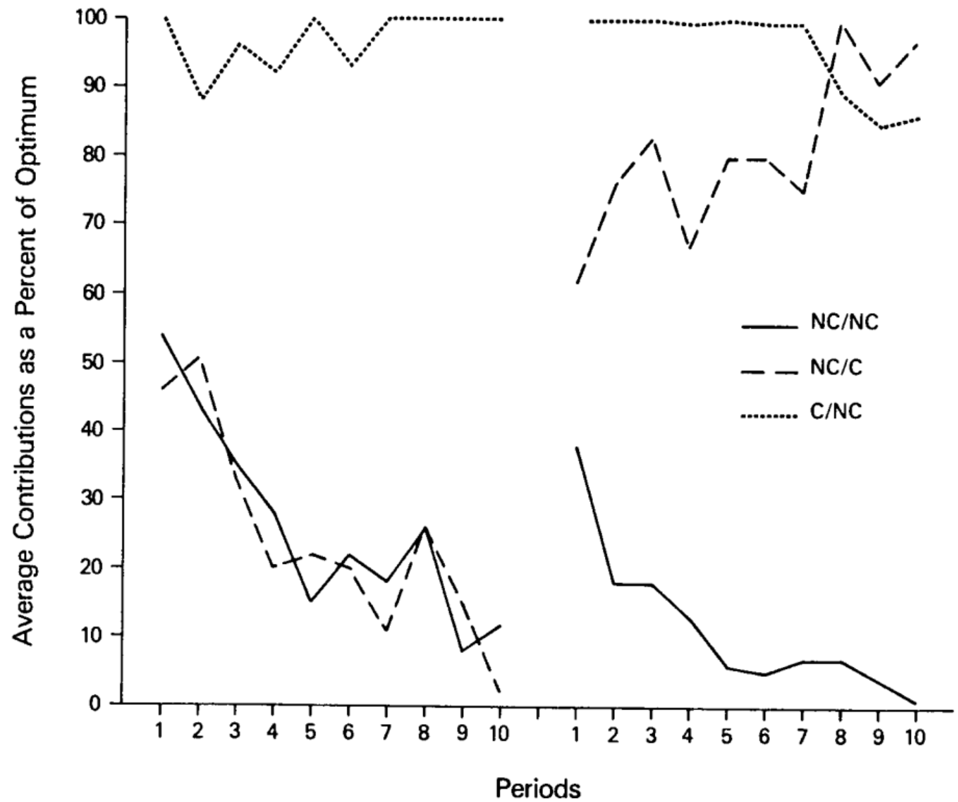


FIGURE 1
Design I Experiments

Isaac and Walker (1988)

Punishment

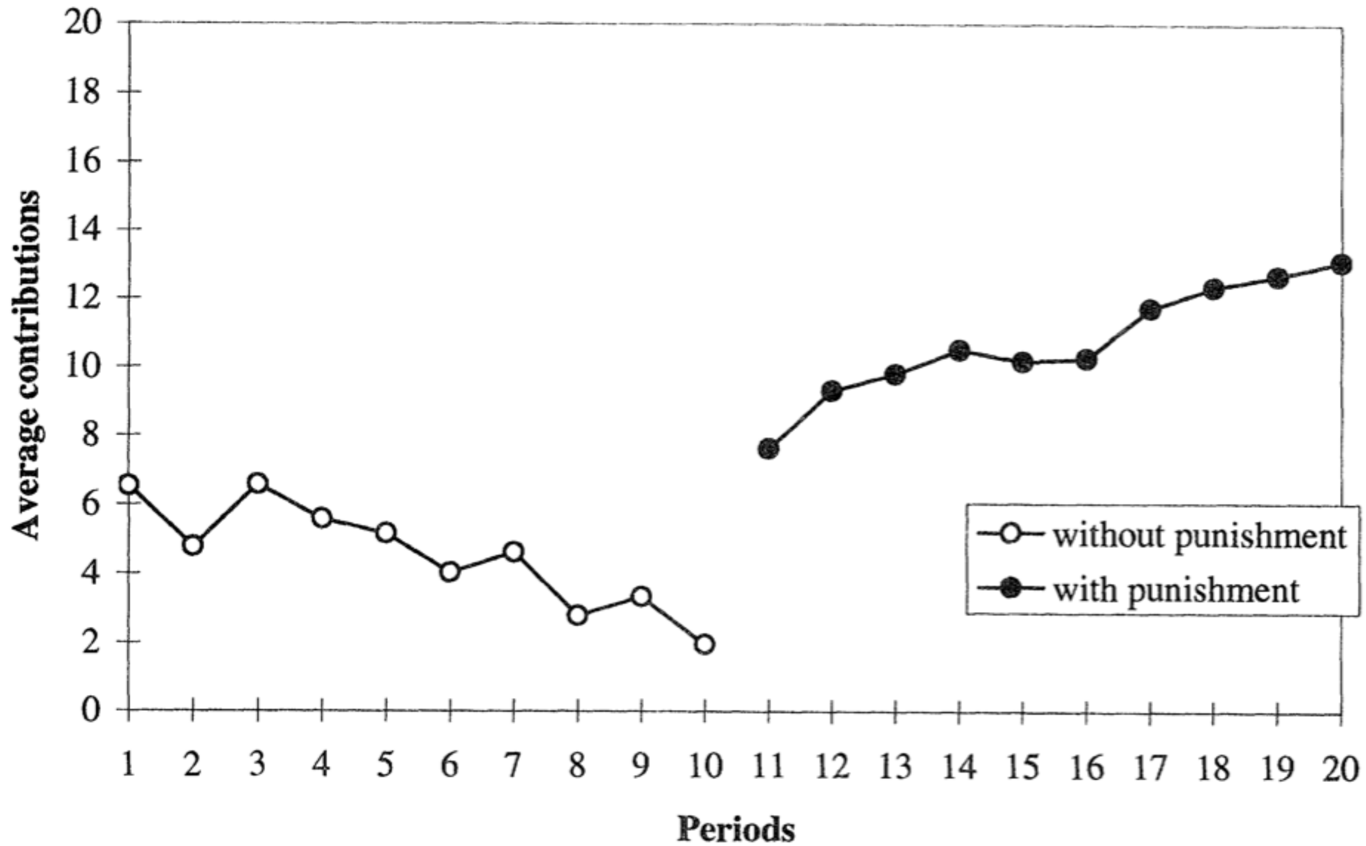
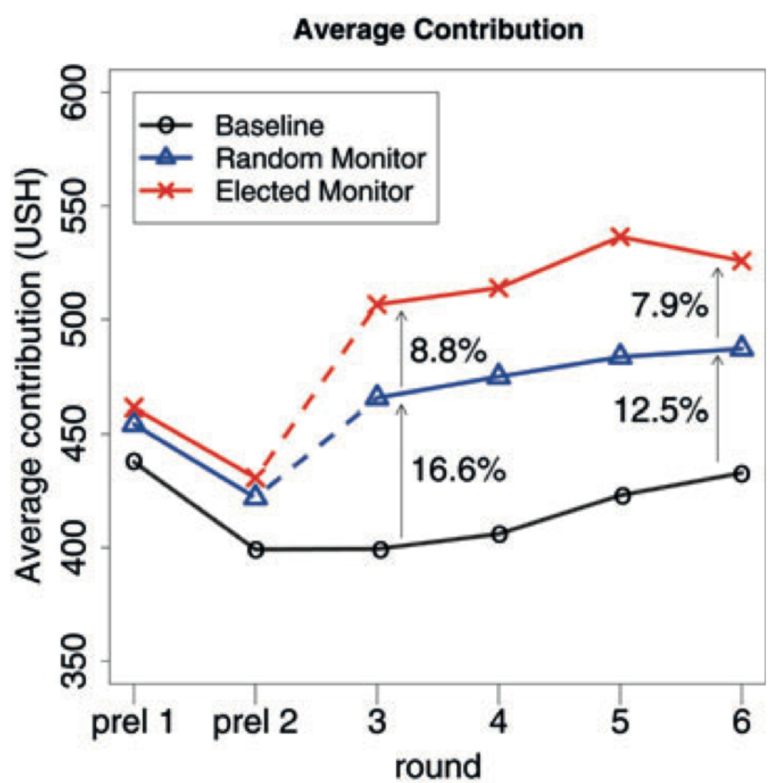


FIGURE 1B. AVERAGE CONTRIBUTIONS OVER TIME IN THE STRANGER-TREATMENT (SESSION 3)

Fehr and Gachter (2000)

Elected leaders

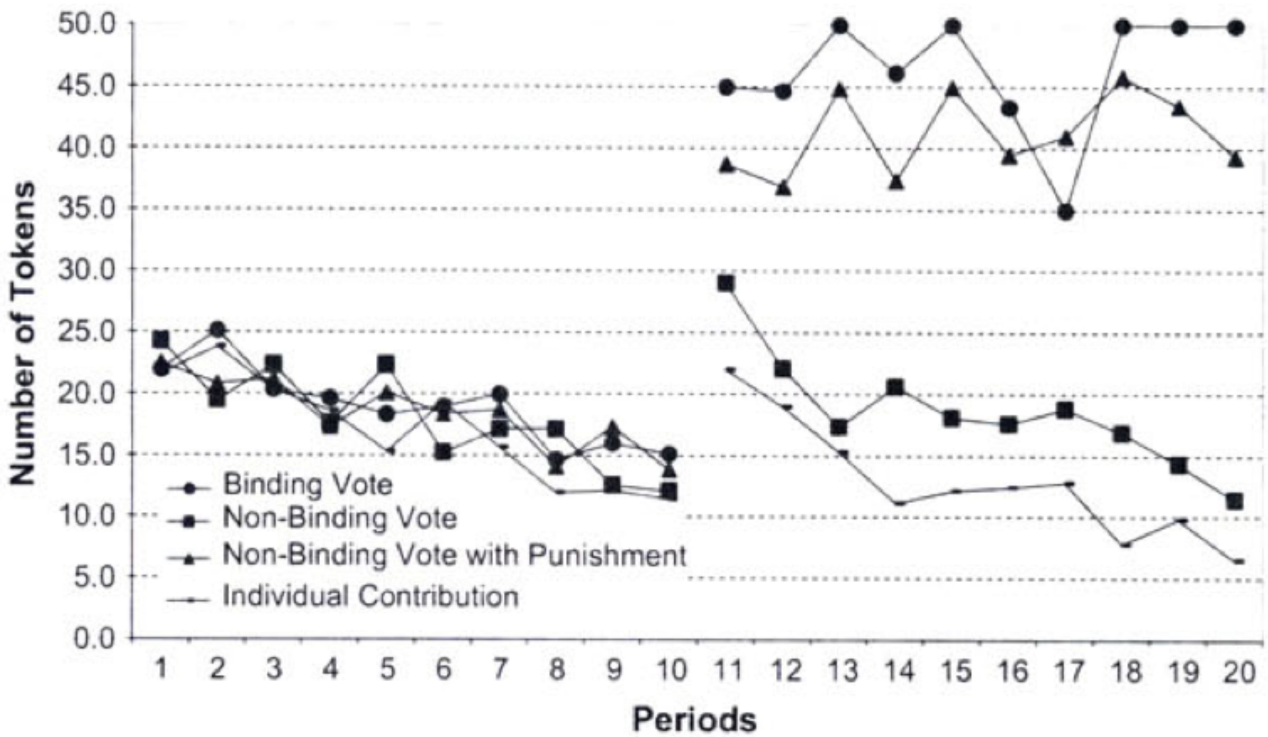
FIGURE 1 Average Contribution to the Public Good by Treatment



Baldassarri and Grossman (2012)

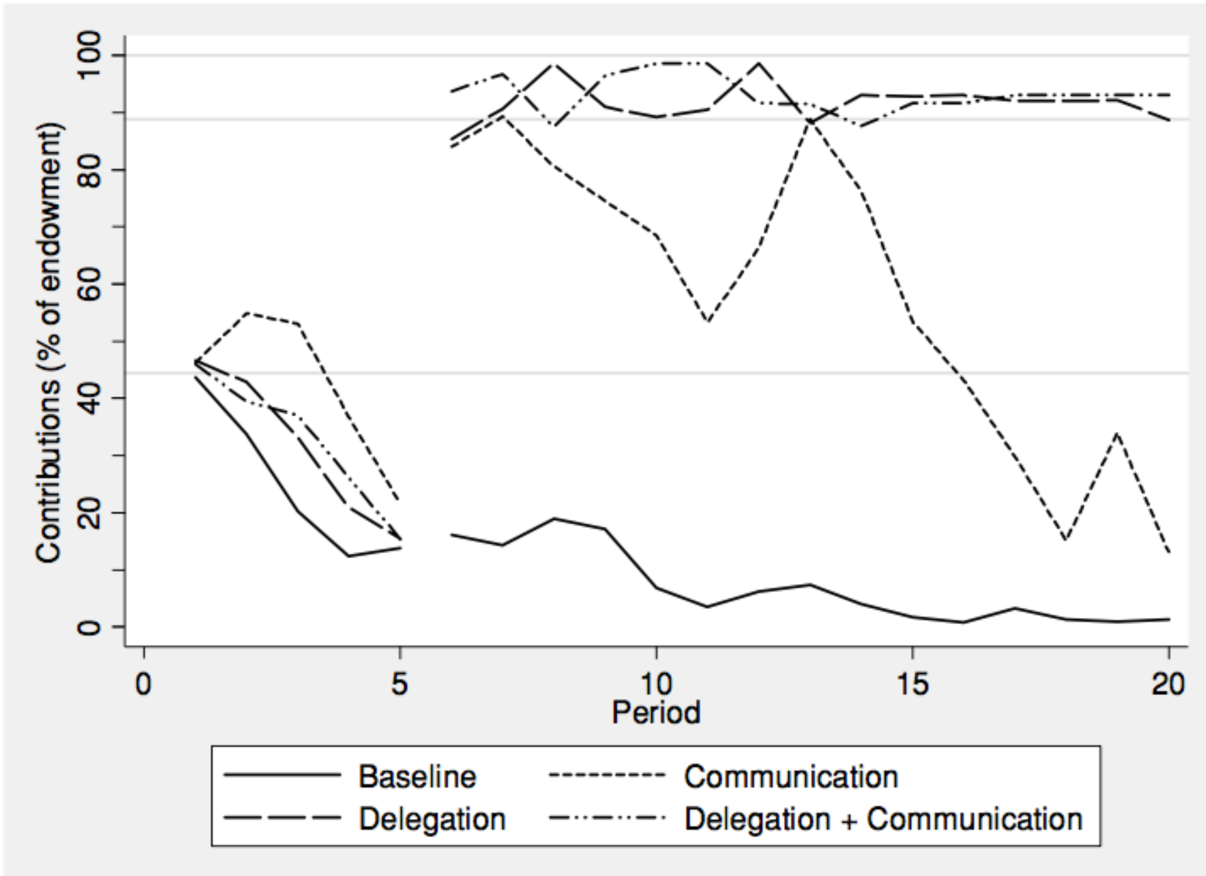
Voting

FIGURE 1
Aggregate Group Contributions



Kroll, Cherry, and Shogren (2007)

Delegation



Hamman, Weber, and Woon (2011)

Bargaining and distributional games

- Ultimatum game
  - Average offers typically 30-40%
  - Offers 40-50% rarely rejected, below 20% more often
  - Variation across cultures correlated with degree of market integration (Henrich et al 2004)

- Dictator "game"
  - Average offers around 20%, implying fairness preferences
  - Direct comparison of ultimatum and dictator game implies ultimatum proposals partly strategic, partly altruistic

See Camerer (2003) for a review.

Bargaining and distributional games

- Trust game
  - Trust: Send ~ 50% endowment
  - Trustworthiness: Return barely more than investment

  - Men tend to trust more than women (Croson and Gneezy 2009)
  - Attractive people believed to be more trustworthy, but aren't (Eckel and Wilson 2006)
  - Darker-skinned people believed to be less trustworthy, but are more so (Eckel and Wilson 2008)

Lessons

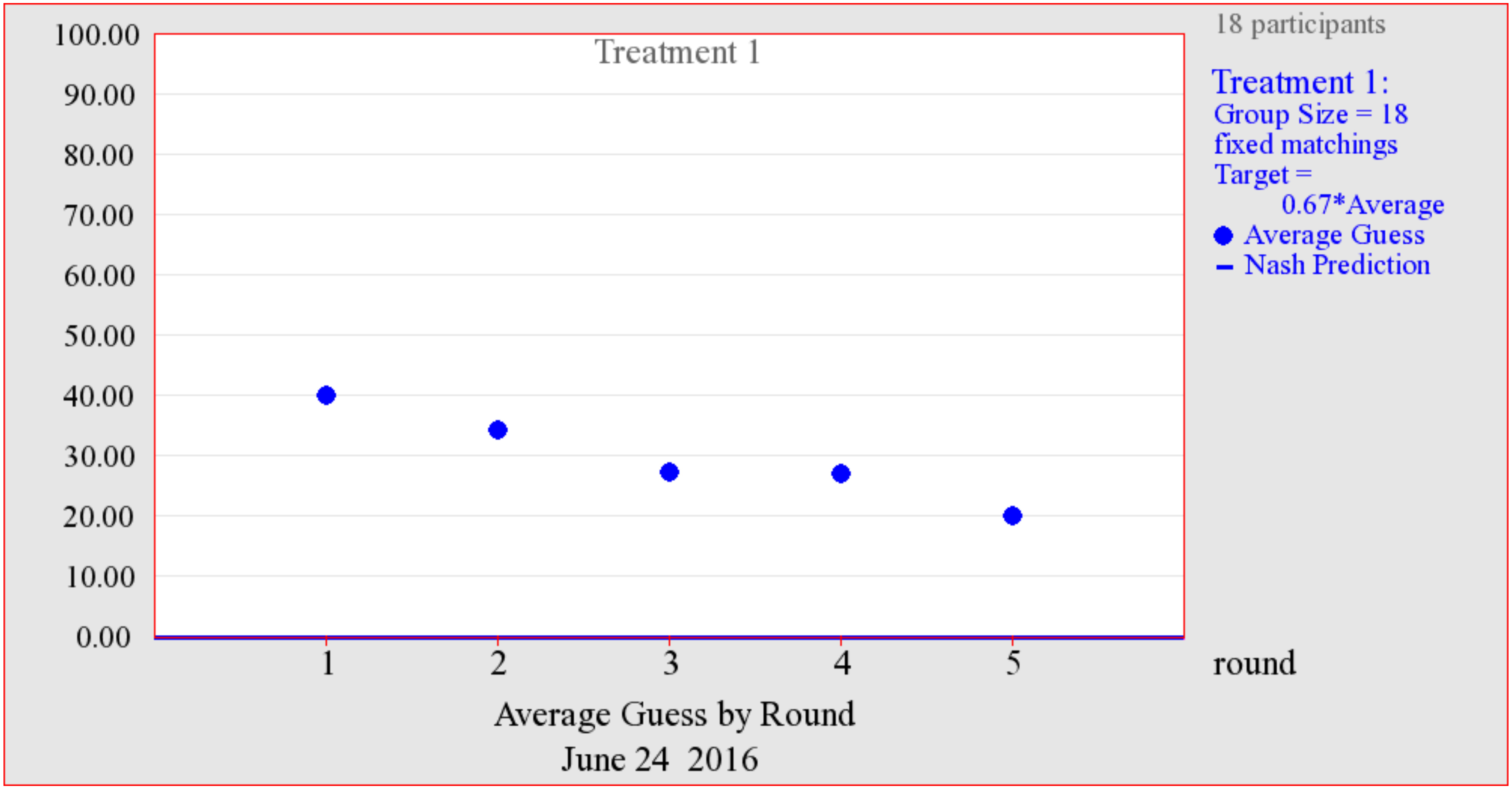
- Rejection of game-theoretic predictions in distributional games suggests violation of dominance, imperfect control of preferences

- Generates new theories of social preferences (altruism, inequality aversion, spite, kindness)

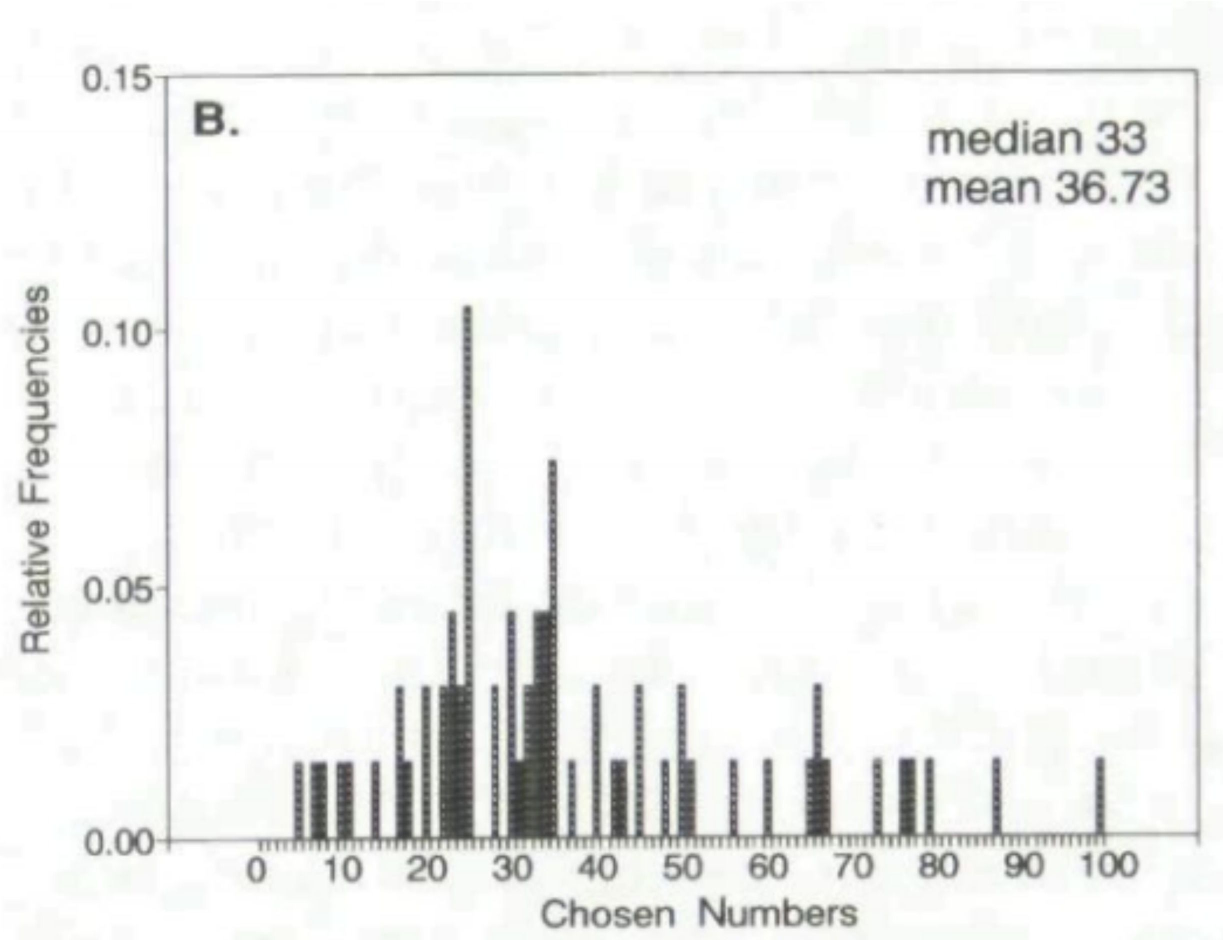
- Use observed behavior in games as measures of unobserved preferences (i.e., revealed preferences)

Guessing game

- Players choose numbers between 0, 100
- Player whose number is closest to 2/3 of the average receives a prize, others get nothing
- Game is competitive and dominance solvable…
- How did you play?


18 participants
Treatment 1
Treatment 1:
Group Size = 18
fixed matchings
Target =
0.67*Average
Average Guess
Nash Prediction
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
1
2
3
4
5
round
Average Guess by Round
June 24 2016

Guessing game



Nagel (1995)

Interpretation

- Falsify prediction of the unique, dominance solvable Nash equilibrium – why?
- If people don't, is choosing 0 "rational"?
- Winning the game depends on **beliefs** about what others will choose

Keynes' Beauty Contest

"It is not a case of choosing those [faces] that, to the best of one's judgment, are really the prettiest, nor even those that average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees." (Keynes, General Theory of Employment, Interest and Money, 1936)

Level-K model

- Nagel (1995) and others developed theory of iterated reasoning to explain the experimental data
- Beliefs reflect level or degree of strategic thinking
  - Level 0 guesses randomly, average 50
  - Level 1 chooses (2/3)50 = 33
  - Level 2 chooses (2/3)33 = 22
  - Level K best responds to one level below (K-1)
- Most players exhibit 1-2 levels of iterated reasoning (and rarely more than 3)

Risk elicitation methods

- People vary in their risk preferences, but the shape of one's utility function cannot be observed directly (if it even exists)

- How can we measure degree of risk aversion?

- Choice between gambles where risk preferences imply differing patterns of behavior

Lottery choice task

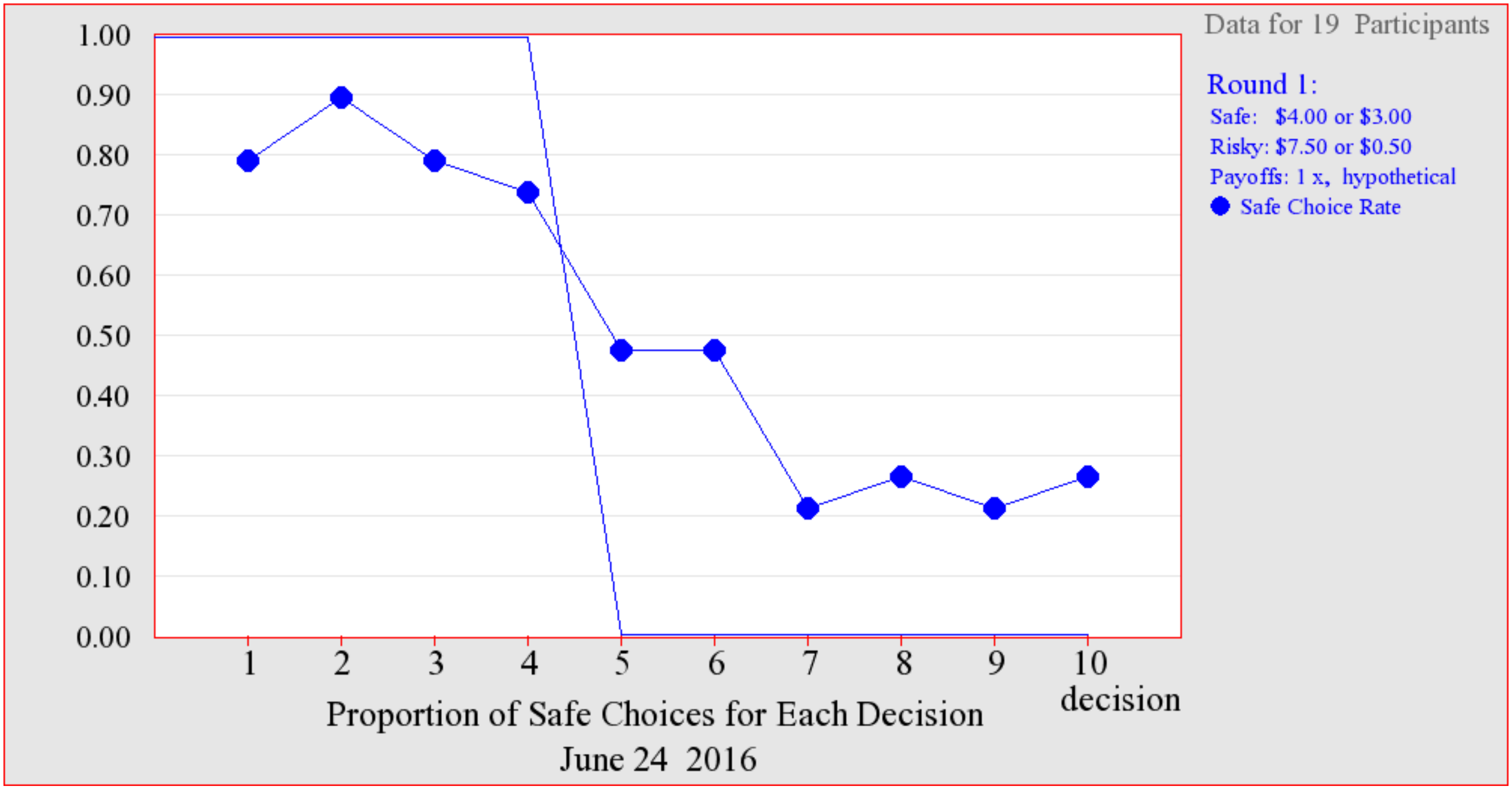
Decision #	Option A	Option B
1	1/10 $4.00, 9/10 $3.00	1/10 $7.50, 9/10 $0.50
2	2/10 $4.00, 8/10 $3.00	2/10 $7.50, 8/10 $0.50
3	3/10 $4.00, 7/10 $3.00	3/10 $7.50, 7/10 $0.50
4	4/10 $4.00, 6/10 $3.00	4/10 $7.50, 6/10 $0.50
5	5/10 $4.00, 5/10 $3.00	5/10 $7.50, 5/10 $0.50
6	6/10 $4.00, 4/10 $3.00	6/10 $7.50, 4/10 $0.50
7	7/10 $4.00, 3/10 $3.00	7/10 $7.50, 3/10 $0.50
8	8/10 $4.00, 2/10 $3.00	8/10 $7.50, 2/10 $0.50
9	9/10 $4.00, 1/10 $3.00	9/10 $7.50, 1/10 $0.50
10	10/10 $4.00, 0/10 $3.00	10/10 $7.50, 0/10 $0.50

Lottery choice task

Decision #	Safe Choice	Risky Choice
1	1/10 $4.00, 9/10 $3.00	1/10 $7.50, 9/10 $0.50
2	2/10 $4.00, 8/10 $3.00	2/10 $7.50, 8/10 $0.50
3	3/10 $4.00, 7/10 $3.00	3/10 $7.50, 7/10 $0.50
4	4/10 $4.00, 6/10 $3.00	4/10 $7.50, 6/10 $0.50
5	5/10 $4.00, 5/10 $3.00	5/10 $7.50, 5/10 $0.50
6	6/10 $4.00, 4/10 $3.00	6/10 $7.50, 4/10 $0.50
7	7/10 $4.00, 3/10 $3.00	7/10 $7.50, 3/10 $0.50
8	8/10 $4.00, 2/10 $3.00	8/10 $7.50, 2/10 $0.50
9	9/10 $4.00, 1/10 $3.00	9/10 $7.50, 1/10 $0.50
10	10/10 $4.00, 0/10 $3.00	10/10 $7.50, 0/10 $0.50

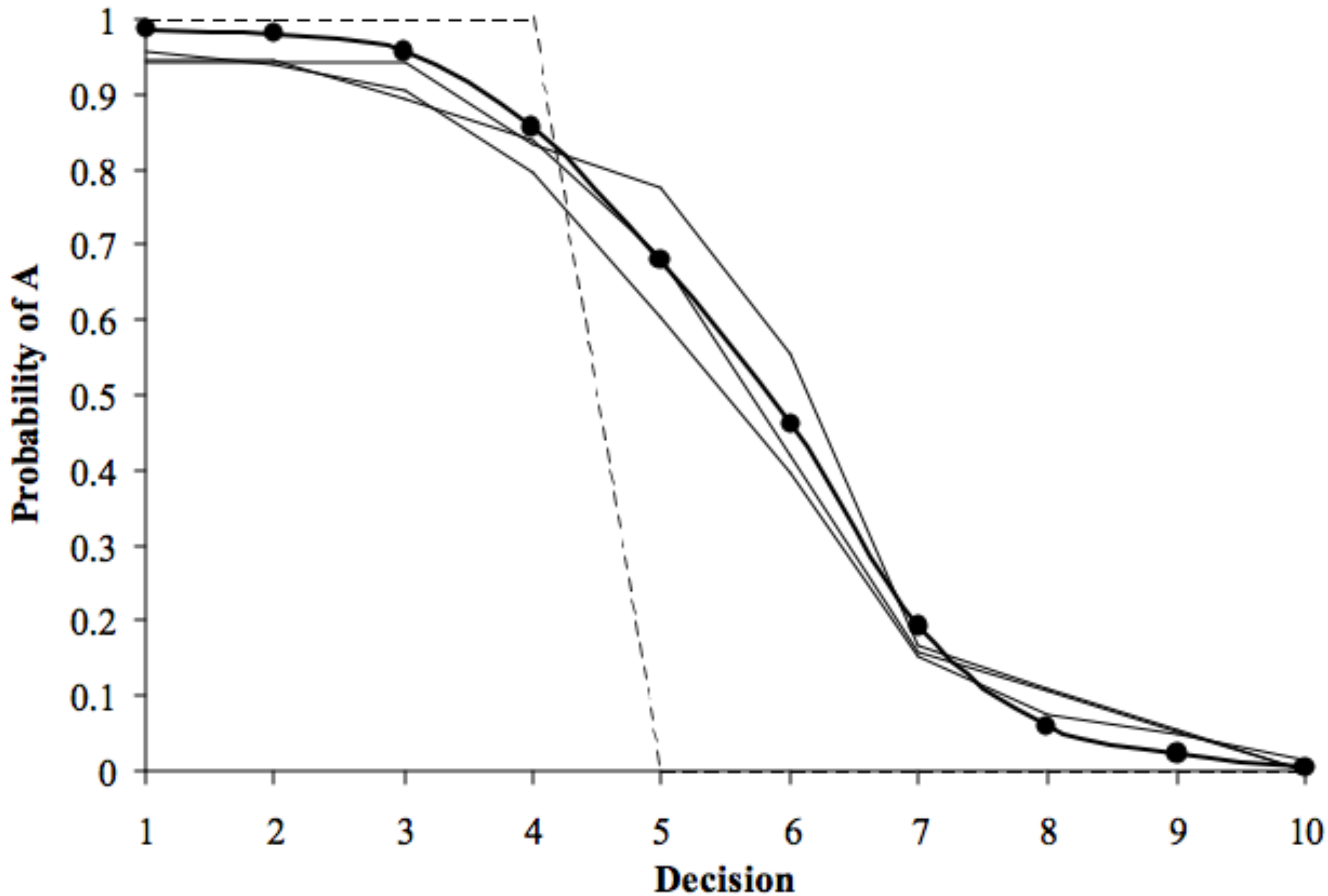
Lottery choice task

Decision #	E[A]	E[B]	E[A] – E[B]
1	$3.10	$1.20	$1.90
2	$3.20	$1.90	$1.30
3	$3.30	$2.60	$0.70
4	$3.40	$3.30	$0.10
5	$3.50	$4.00	- $0.50
6	$3.60	$4.70	- $1.10
7	$3.70	$5.40	- $1.70
8	$3.80	$6.10	- $2.30
9	$3.90	$6.80	- $2.90
10	$4.00	$7.50	- $3.50

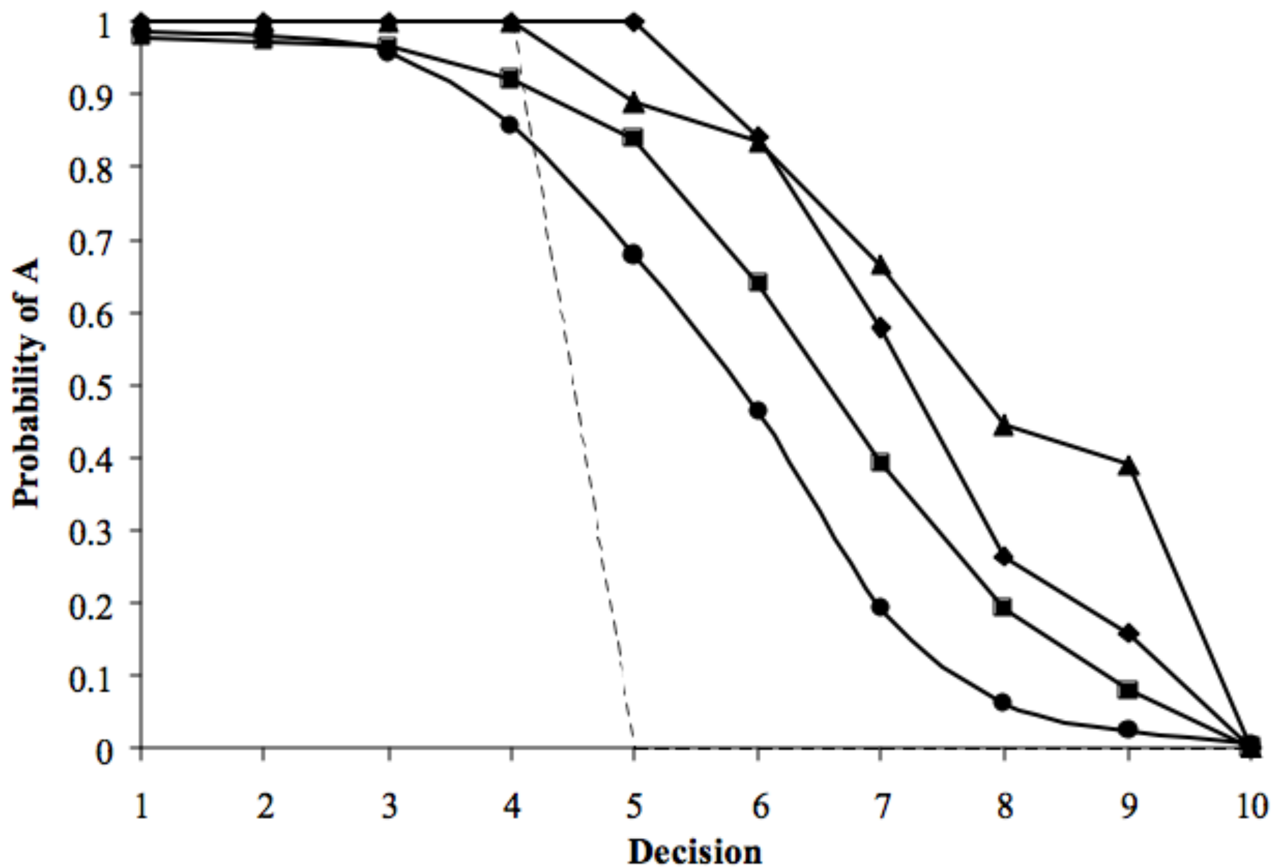


Data for 19 Participants
Round 1:
Safe: $4.00 or $3.00
Risky: $7.50 or $0.50
Payoffs: 1 x, hypothetical
Safe Choice Rate
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1 2 3 4 5 6 7 8 9 10
decision
Proportion of Safe Choices for Each Decision
June 24 2016

Hypothetical payoffs (Holt and Laury 2002)



Real payoffs (Holt and Laury 2002)



Measurement tasks

- Structure decisions so that choices will vary in known, predictable ways according to an underlying theoretical model
- Choices between gambles reveal risk preferences
- Dictator game reveals degree of altruism
- Can use lottery tasks to measure probability beliefs

Lunch!

References

Grossman, G., & Baldassarri, D. (2012). The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda.*American Journal of Political Science*, *56*(4), 964-985.

Bianco, W. T., Lynch, M. S., Miller, G. J., & Sened, I. (2006). "A Theory Waiting to Be Discovered and Used": A Reanalysis of Canonical Experiments on Majority-Rule Decision Making. *Journal of Politics*, *68*(4), 838-851.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cooper, R., DeJong, D.V., Forsythe, R. and Ross, T.W. (1996). "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games" *Games and Economic Behavior* 12: 187-218.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 448-474.

References

Eckel, C. C., & Wilson, R. K. (2009). Initiating trust: the conditional effects of skin color among strangers. Unpublished manuscript, Rice University.

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, *90*(4), 980-994.

Fiorina, M. P., & Plott, C. R. (1978). Committee decisions under majority rule: An experimental study. *American Political Science Review*, *72*(02), 575-598.

Flood, M. M. (1958). Some experimental games. *Management Science*, *5*(1), 5-26.

Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge university press.

References

Hamman, J. R., Weber, R. A., & Woon, J. (2011). An experimental investigation of electoral delegation and the provision of public goods.*American Journal of Political Science*, *55*(4), 738-752.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects.*American economic review*, *92*(5), 1644-1655.

Isaac, R. M., & Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic inquiry*, *26*(4), 585-608.

Kanthak, K., & Woon, J. (2015). Women don't run? Election aversion and candidate entry. *American Journal o*

Kroll, S., Cherry, T. L., & Shogren, J. F. (2007). Voting, punishment, and public goods. *Economic Inquiry*, *45*(3), 557-570.

References

McKelvey, R. D., & Ordeshook, P. C. (1985). Sequential elections with limited information. *American Journal of Political Science*, 480-512.

Morton, R. B. (1993). Incomplete Information and Ideological Explanations of Platform Divergence. *American Political Science Review*, *87*(02), 382-392.

Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge University Press.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313-1326.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, 1067-1101.

References

Roth, A. (1995) "Introduction to Experimental Economics" in Roth and Kagel, eds., *The Handbook of Experimental Economics.* Princeton University Press.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923-955.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*(2), 189-202.

Woon, J. (2012a). Laboratory tests of formal theory and behavioral inference. In *Experimental Political Science* (pp. 54-71). Palgrave Macmillan UK.

Woon, J. (2012b). Democratic Accountability and Retrospective Voting: A Laboratory Experiment. *American Journal of Political Science* 56:913-930.