EITM and (Incentivized) Experiments

Jonathan Woon
University of Pittsburgh

EITM @ University of Houston
June 20, 2017

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)

Why model?

- Partial representations of the world
- Explicit assumptions about what we think matters
- Logically coherent and consistent implications
- Sharp predictions, clear guidance for empirical tests

Important insights

- Proposal power
  - Legislative bargaining (Romer and Rosenthal, Baron and Ferejohn)
  - Gridlock (Krehbiel)
- Information transmission
  - Theories of organization (Gilligan and Krehbiel)
  - Conflict (Fearon)
- Commitment problems
  - Domestic transitions (Acemoglu and Robinson)
  - Interstate rivalries (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."

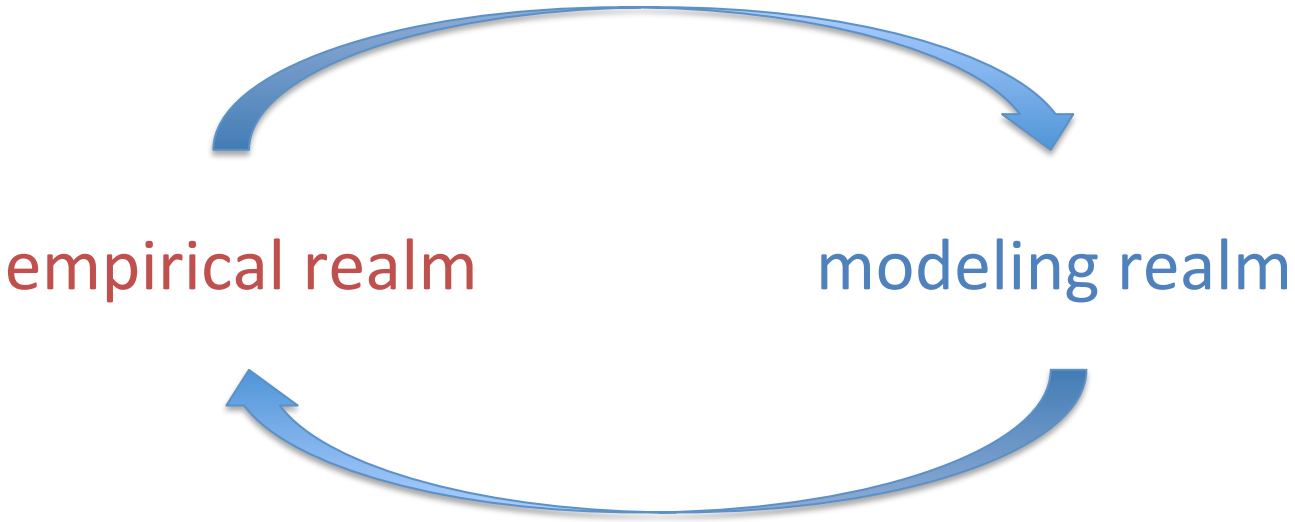
Roger Myerson. 1992. "On the Value of Game Theory in Social Science"

Modeling dialogue

- "Simplifying assumptions must be tested and challenged."

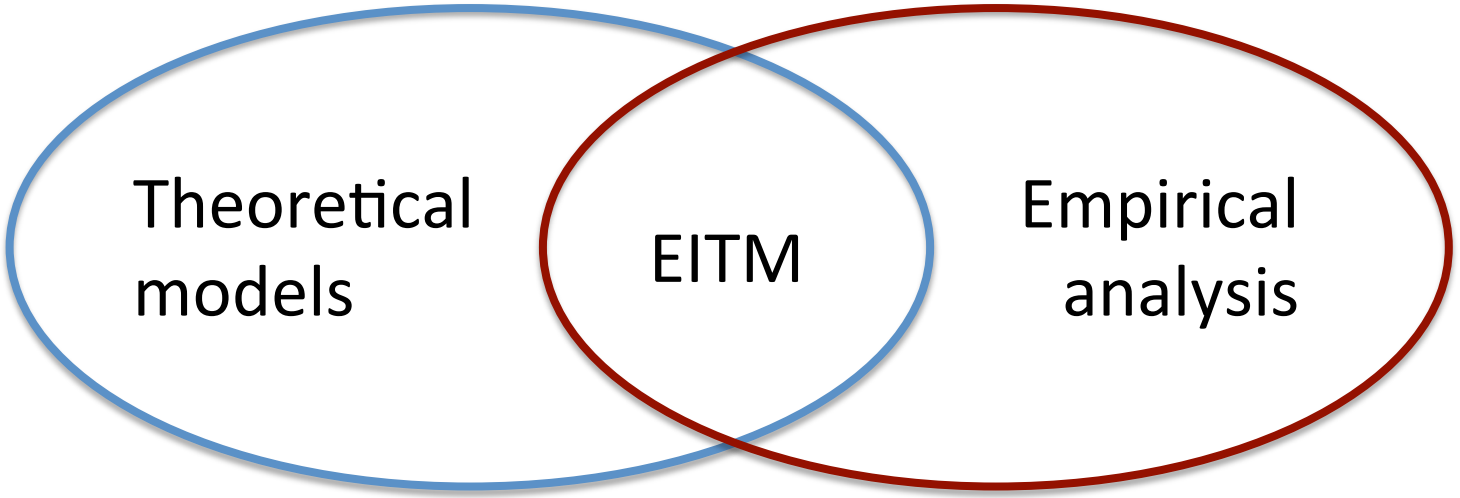
- "We must constantly compare the predictions of our simple models with what we know about the real world and **ask whether the appropriate simplifications have been made**."

Modeling dialogue



THIS IS EITM!

Methodological intersection



THIS IS EITM, TOO!

EITM

EITM is a misnomer – we really want **linkages** between theoretical models and empirical analysis

- Indirect: Modeling dialogue (to determine appropriate simplifications)

- Direct: Empirical methods that directly incorporate elements of theoretical models (to estimate theoretically interesting quantities)

How do experiments fit into this framework?

Let's play!

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

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

- Session name: **woon4**

Let's play!

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

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

- Session name: **woon5**

Let's play!

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

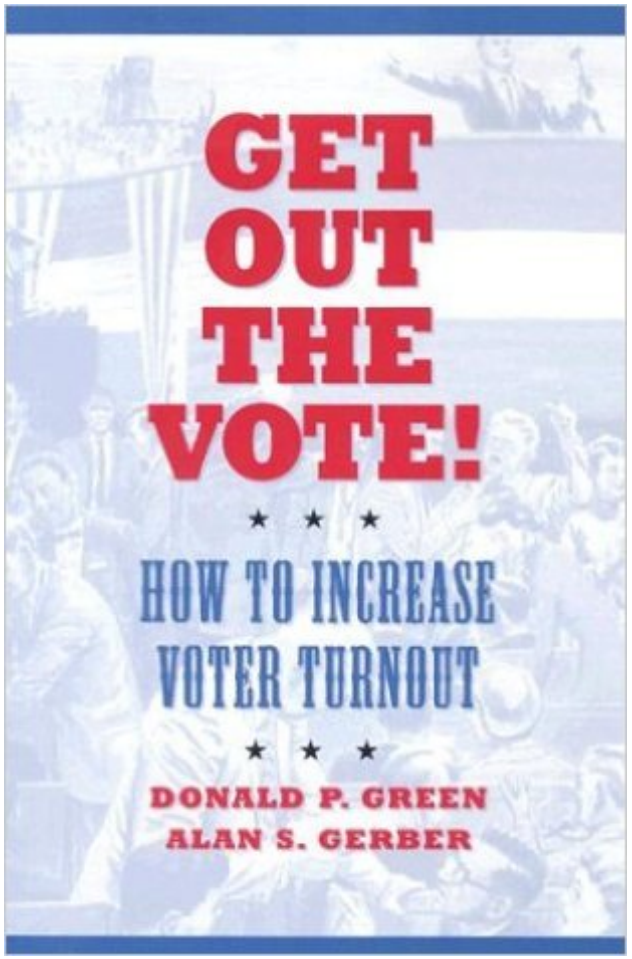
- http://veconlab.econ.virginia.edu/login1.php

- Session name: **woon6**

Today's objectives

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

Political science experiments



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

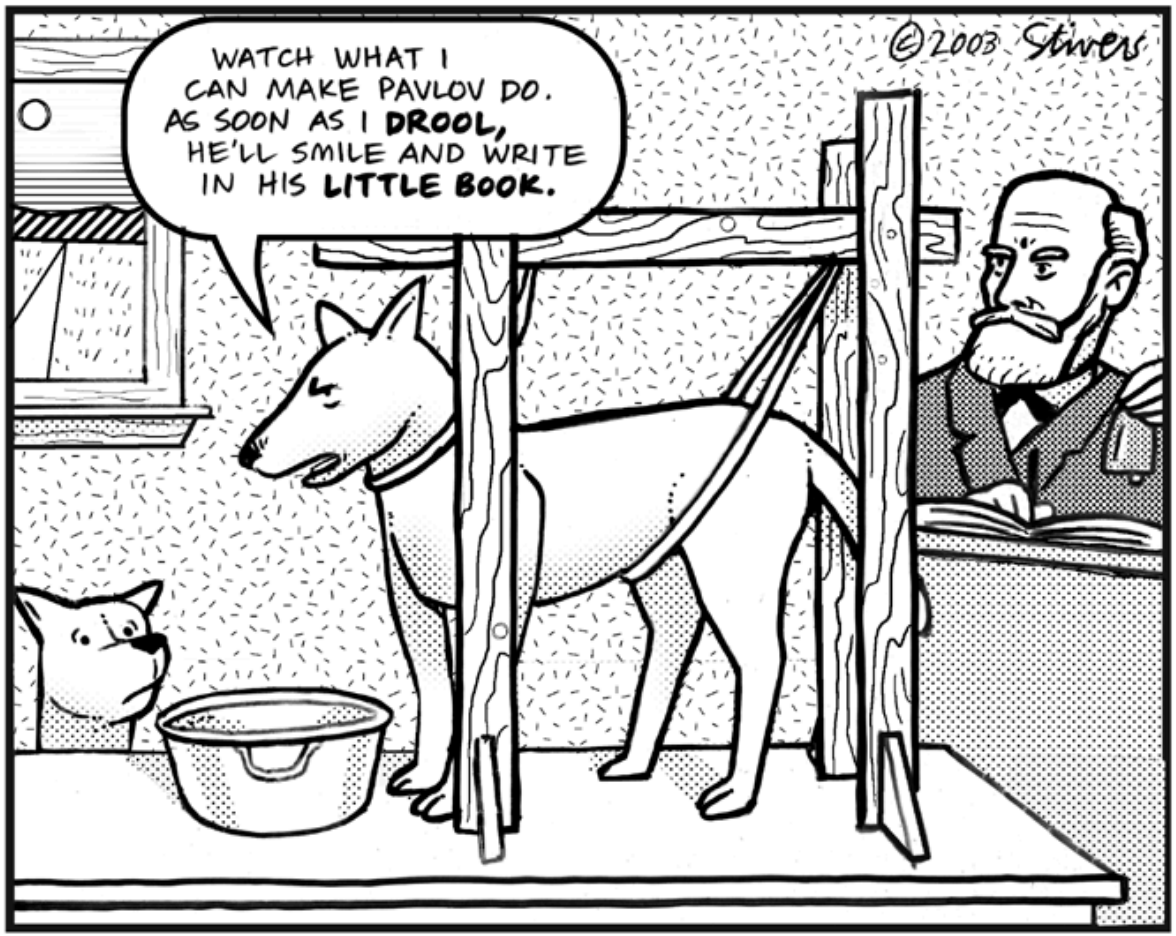
Would you want to participate in a RCT to test the effect of parachutes on mortality?

Galileo’s inclined plane

Newton’s prism

Mendel's peas

Pavlov's dogs



Experiments are defined by control

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

(Morton and Williams 2010)

Incentivized laboratory experiments

- Implement games and decisions that closely match theoretical models
- Pay subjects monetary rewards based on choices and outcomes of the game – actions have real consequences

Control of decision context, game form

- Alternatives, sequence of actions, information as described in theoretical model - **NO DECEPTION**!
- Try to make rules and payoffs "common knowledge," so need to be explained clearly
- Quizzes and examples help to ensure comprehension (but tradeoffs)
- Describe the game, but don't tell anyone what to do!

Control of preferences

"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)

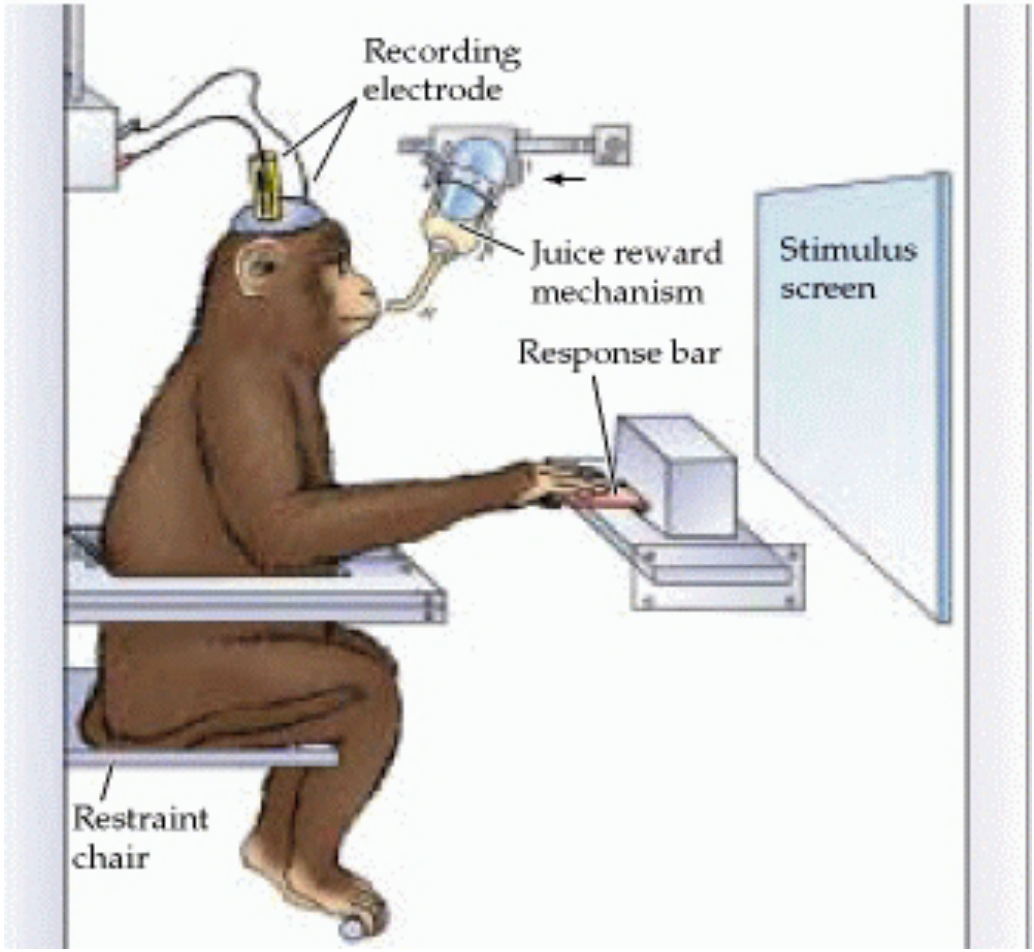
Left	Right
$0	$10

Left	Right
$10	$0

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

Implementation

- Computerized interfaces
  - Helps maintain anonymity, automate computations, facilitate randomization
  - z-tree software is widely used, but any general programming language will do
- Pencil and paper – can be easier to implement for simple games and decisions

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

Roth's typology of experiments

Speaking to theorists

Searching for facts

Whispering in the ears of princes

Roth (1995)

Uses of experiments

- Theory testing, discriminating between competing theories
- Establishing empirical regularities
- Measuring otherwise unobservable quantities
- Wind tunnels – creating situations or counterfactuals that don't occur naturally

Experiments and theory testing

- If we use experiments to test theory, what are we testing?

- **What are the "behavioral" assumptions of rational choice and game theory?**

Rational choice

- Set of alternatives

$$C = \{\text{Clinton, Trump, Stein}\}$$

- Preferences over alternatives

$$U(\text{Stein}) > U(\text{Clinton}) > U(\text{Trump})$$

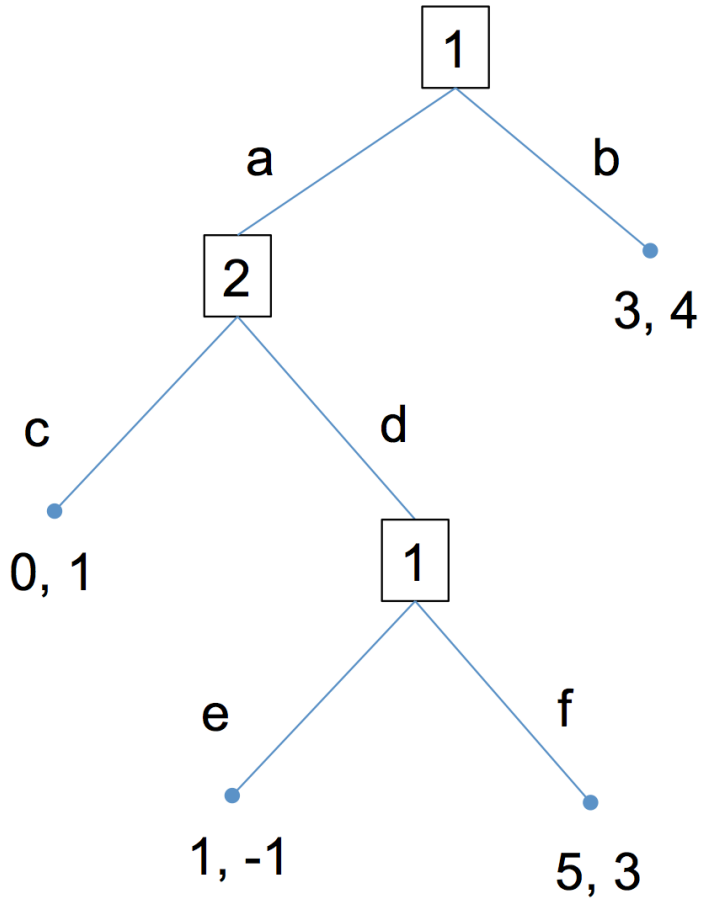
- **What do we have to assume to predict behavior?**

Simultaneous move games

		Column		
		L	**C**	**R**
Row	**U**	-1, 2	0, 1	2, 3
	M	0, 0	-1, -1	3, -1
	D	1, 2	0, 2	4, 0

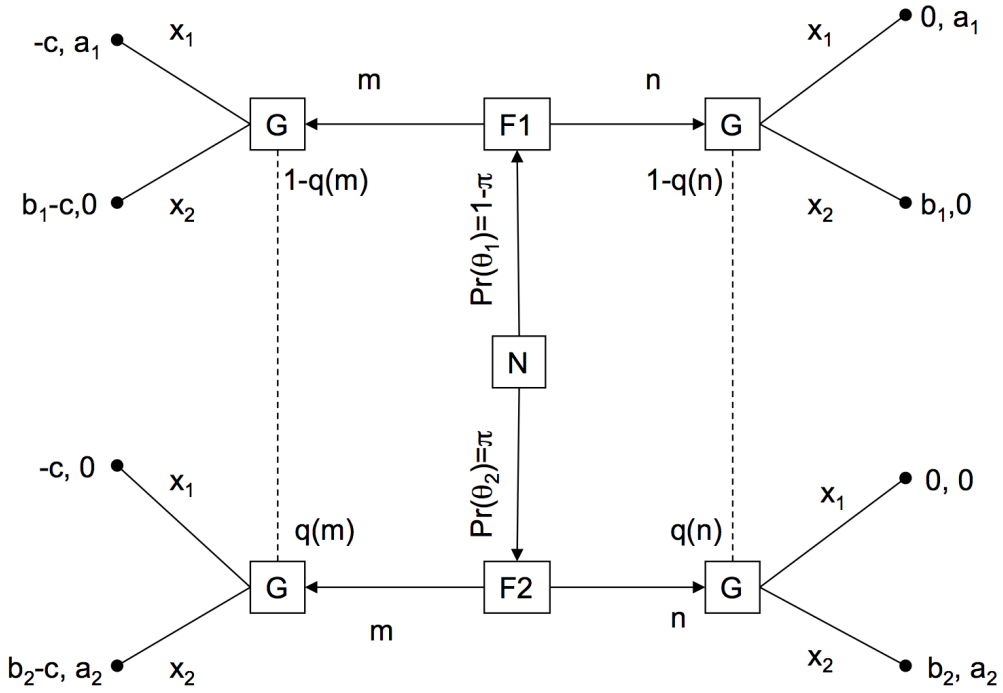
- How do we solve these games?
- **How should we interpret Nash equilibrium in terms of behavior?**
- **What must be true of people to think they will choose Nash equilibrium actions?**

Sequential games



What additional behavioral assumptions do we use to solve extensive form games?

Games of imperfect information



- What's the solution concept for a signaling game?
- **What additional behavioral assumptions are needed?**

Standard RCGT behavioral assumptions

- Complete and transitive preferences
- Choices consistent with preferences
- Expected utility: When uncertain, preferences over gambles satisfy continuity and independence
- In dynamic settings, forward looking

Standard RCGT behavioral assumptions

- In strategic settings, choose best responses
- Nash equilibrium: Best responses mutually consistent; correct expectations about others' beliefs and behavior
- Sequential games: Rational anticipation, credible threats
- Bayesian: Draw correct inferences from limited information
- Common knowledge of the game, beliefs, rationality

Why are incentivized experiments good for testing theories of behavior?

Model of behavioral inference

$$\underbrace{G \wedge P \wedge B}_{\text{Theoretical model}} \Rightarrow \underbrace{H}_{\text{Empirical implication}}$$

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

P = Preferences (utility function)

B = Behavior

H = Hypothesis

See Woon (2012)

Model of behavioral inference

¬ H

What if the hypothesis is falsified?

Model of behavioral inference

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

We infer only that at least one assumption must be false – there is an **identification problem**

Model of behavioral inference

Experimental control of G and P increases confidence in the inferences we can make about B

Model of behavioral inference

Experimental control of G and P increases confidence in the inferences we can make about B

…but more difficult if uncertain about control of P

Model of behavioral inference

If we maintain behavioral assumptions, can use experimental methods to measure preferences, beliefs

Towards behavioral models of politics

- Experiments can help to refine – not discard – theoretical models so that "appropriate simplifications" be made
- Identify conditions when standard RCGT assumptions work
- Develop new models that more accurately reflect the range of human behavior in the political domain
  - Insights from psychology, cognitive science, behavioral game theory
  - Bounded rationality, beliefs, learning, adaptation, limited reasoning

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

Common objections

Artificial and unrealistic!

Small stakes!

Undergraduates!

No method is perfect

- Not a substitute for observational data
- Results from specific populations or contexts may not generalize – but this is true for observational research, too
- Often stylized or context-free; relevant real-world factors sometimes difficult or impossible to reproduce in the lab
- Control is never 100% complete

Experiments and EITM

- Experiments are models, too – purposeful simplifications, but behavior is observed rather than assumed
- Good for behavioral inference – artificiality and control are "features" not "bugs"
- Establish direct and indirect linkages between theoretical models and empirical research
- Speeds up modeling dialogue – help generate new models of behavior before testing in the field
- Experiments complement analysis of observational data

Outline of examples

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

The matrix game

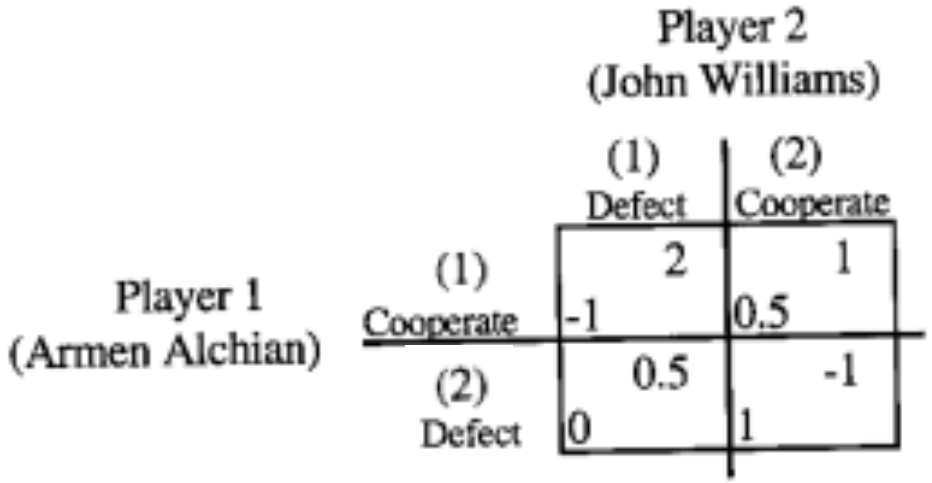
		Column: Left	Column: Right
Row	**Up**	2, 2	0, 4
	Down	4, 0	1, 1

What game is this?

How did you play?

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	2	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 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)

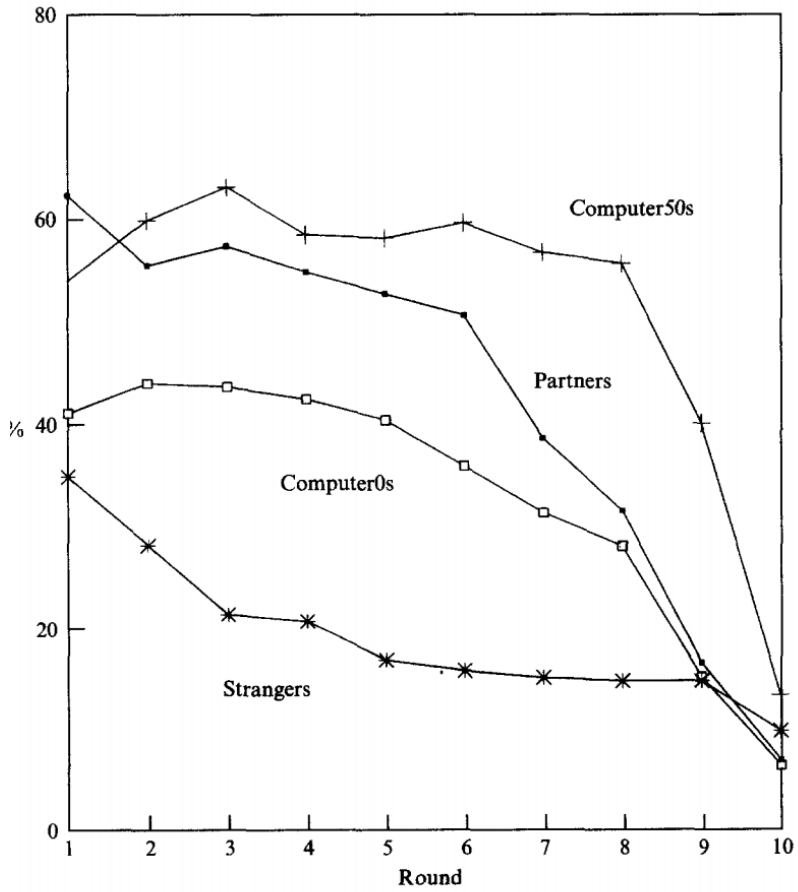
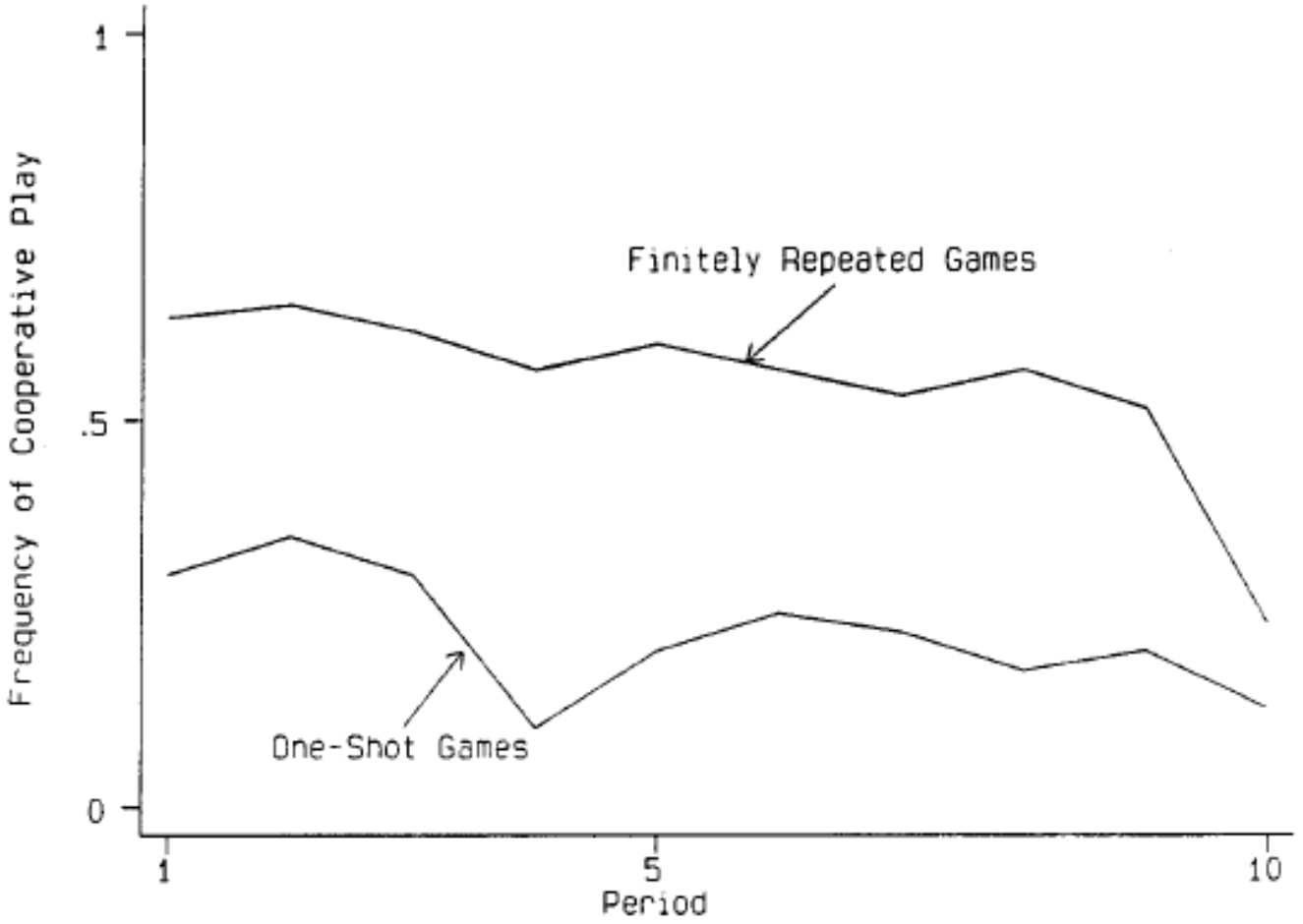


Fig. 2. Percent cooperation by round. Averaged over all 20 10-period games.

Cooper et al (1996)



Cooper et al (1996)

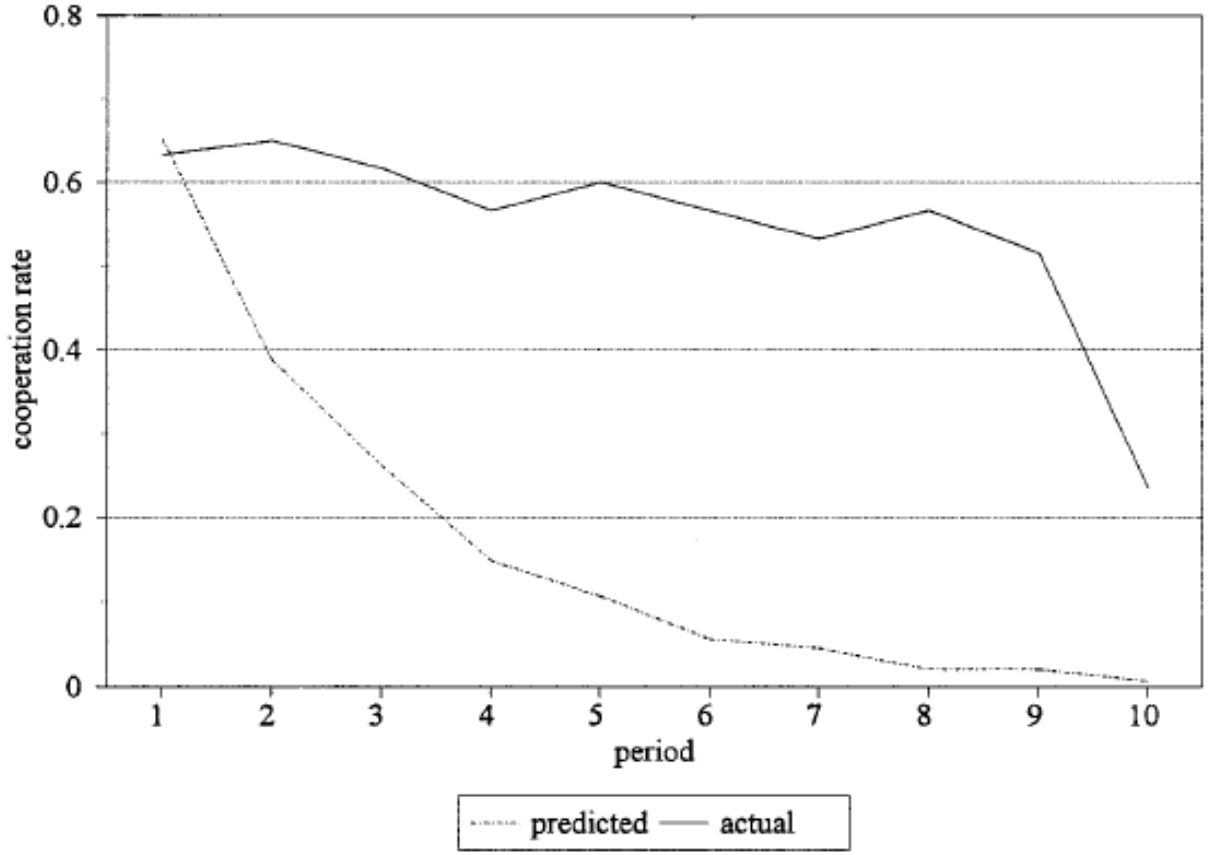


FIG. 3. Predicted and actual cooperation rates for PD-FR.

Collective action and public goods

- What kinds of institutions solve collective action problems?

- Experimenters can push and pull a variety of institutional levers

Communication

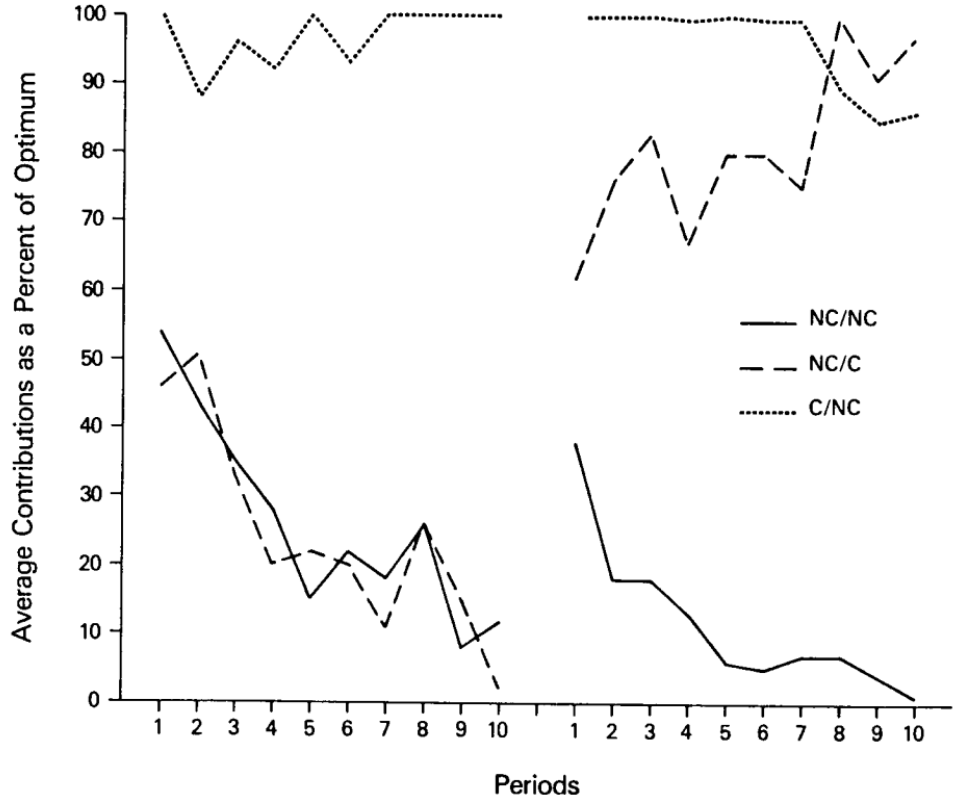


FIGURE 1
Design I Experiments

Isaac and Walker (1988)

Punishment

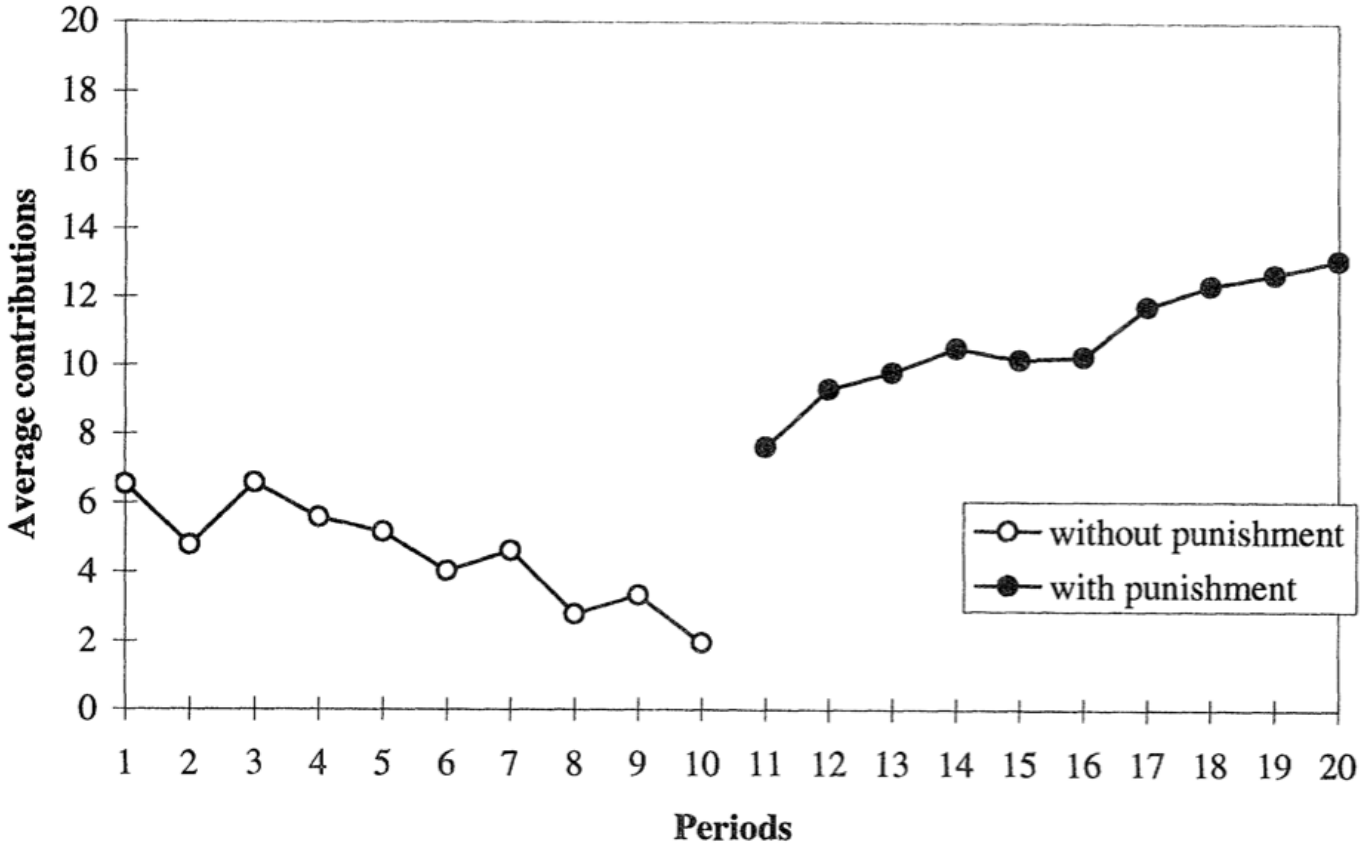
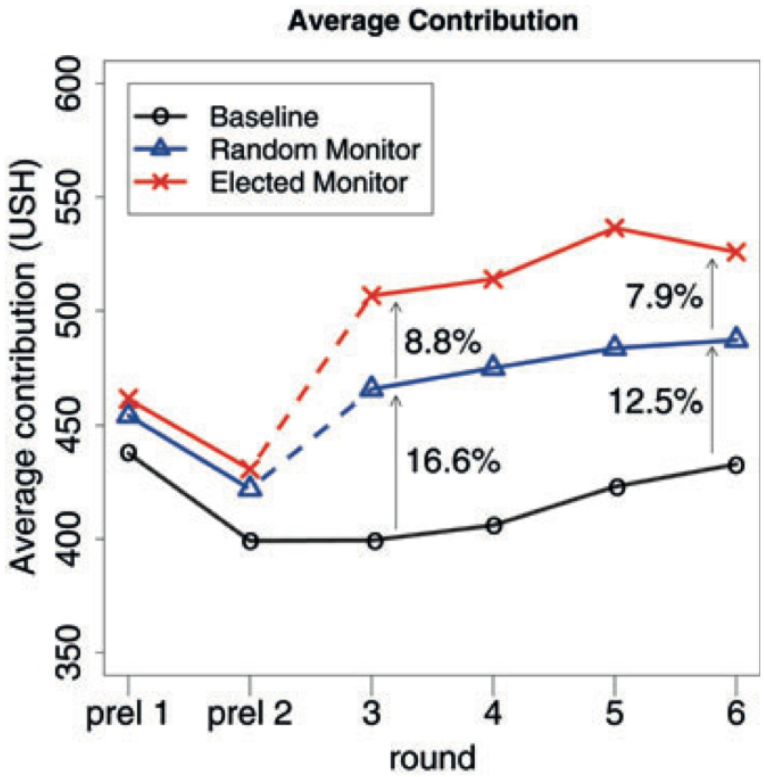


FIGURE 1B. AVERAGE CONTRIBUTIONS OVER TIME IN THE STRANGER-TREATMENT (SESSION 3)

Fehr and Gachter (2000)

Elected leaders

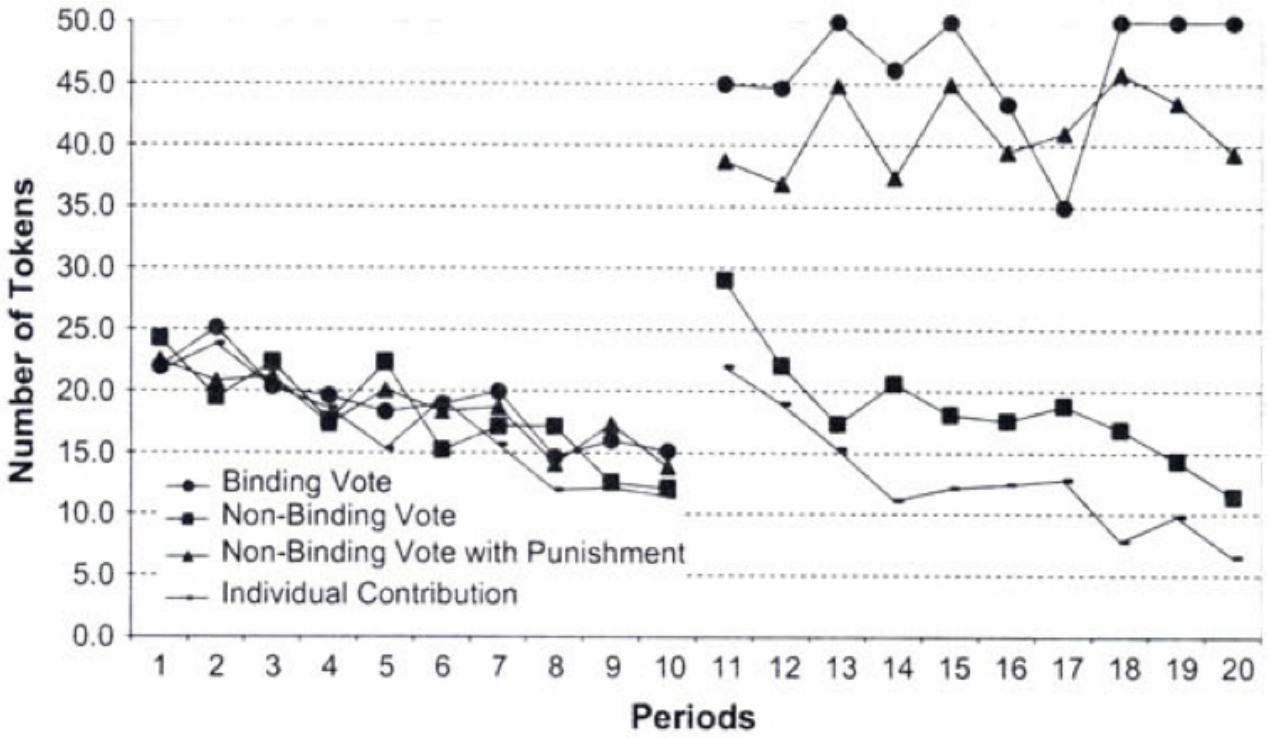
FIGURE 1 Average Contribution to the Public Good by Treatment



Grossman and Baldassarri (2012)

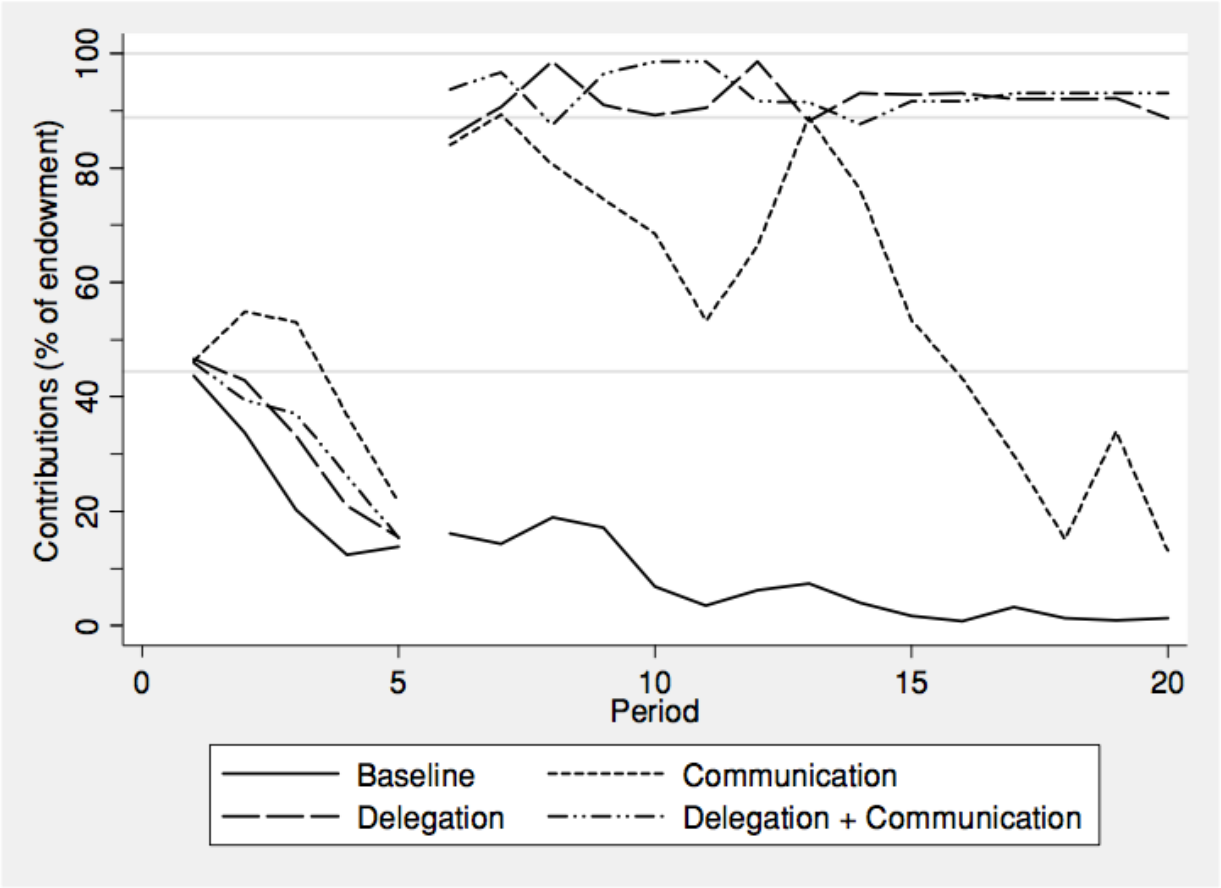
Voting

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

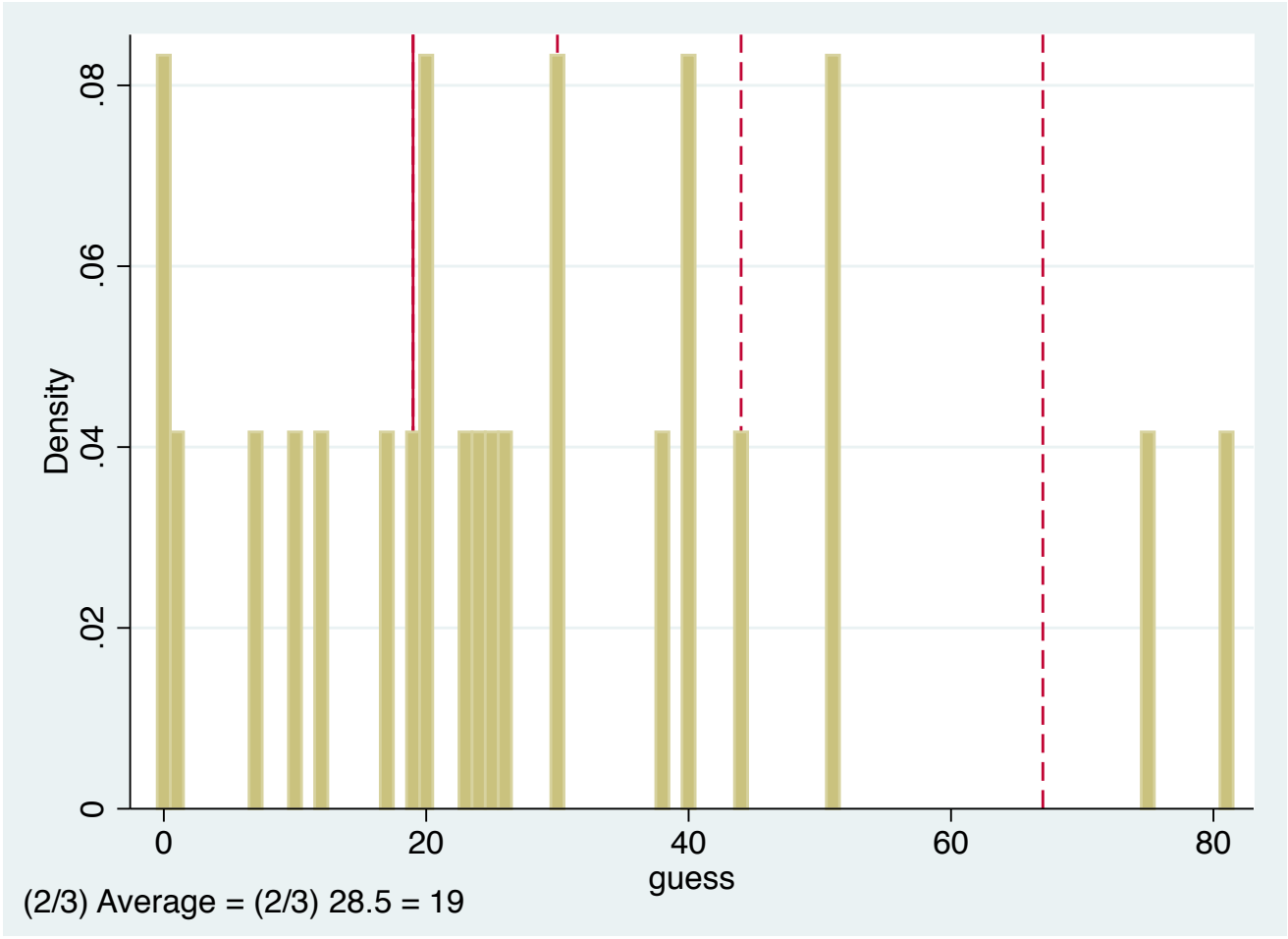
  - Men tend to trust more than women (Croson and Gneezy 2009)
  - Attractive people believed to be more trustworthy, but aren't (Eckel and Wilson 2006)
  - Darker-skinned people believed to be less trustworthy, but are more so (Eckel and Wilson 2008)

Lessons

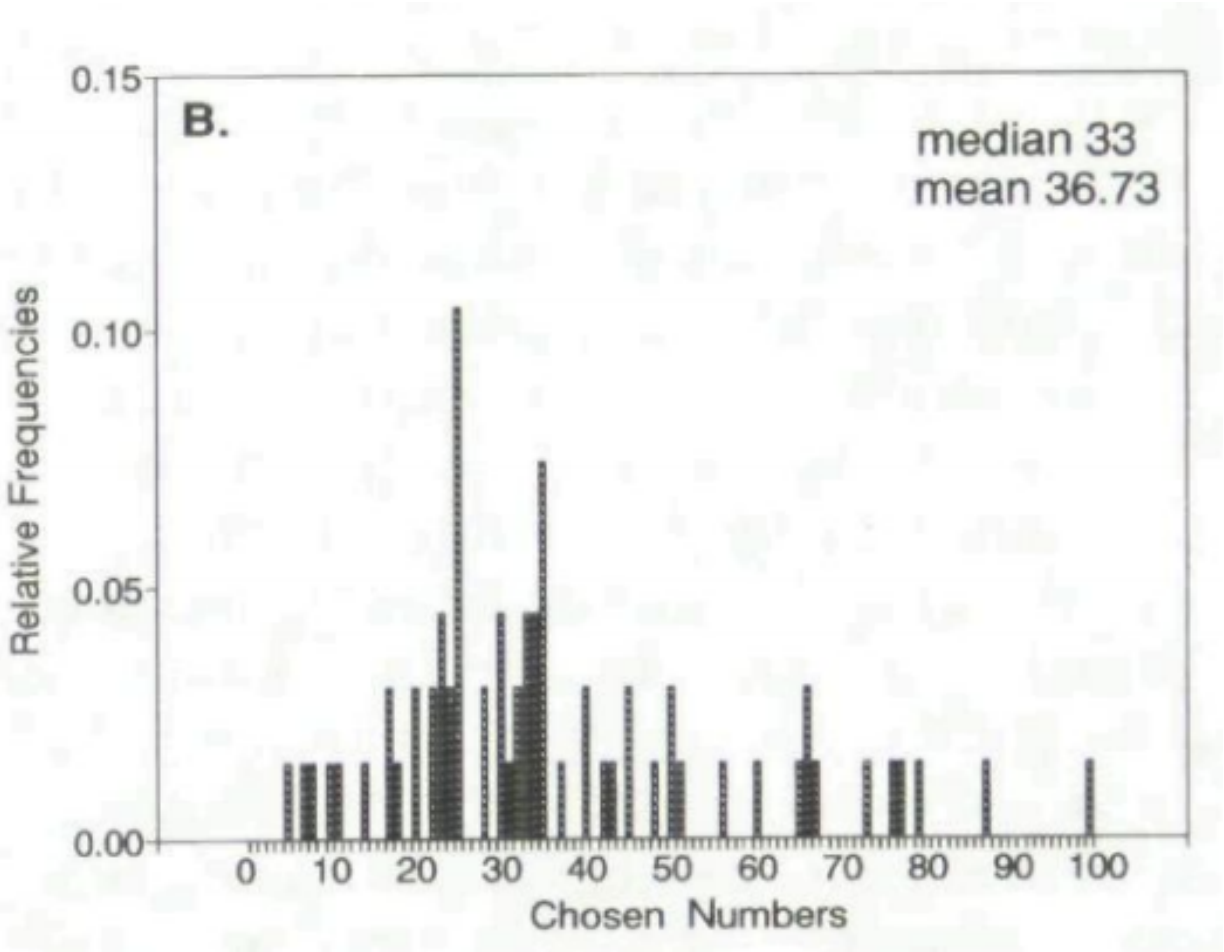
- Rejection of game-theoretic predictions in distributional games suggests violation of dominance, imperfect control of preferences
- Generates new theories of social preferences (altruism, inequality aversion, spite, kindness)
- Use observed behavior in games as measures of unobserved preferences (i.e., 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?


Density
0
.02
.04
.06
.08
0
20
40
60
80
guess
(2/3) Average = (2/3) 28.5 = 19

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

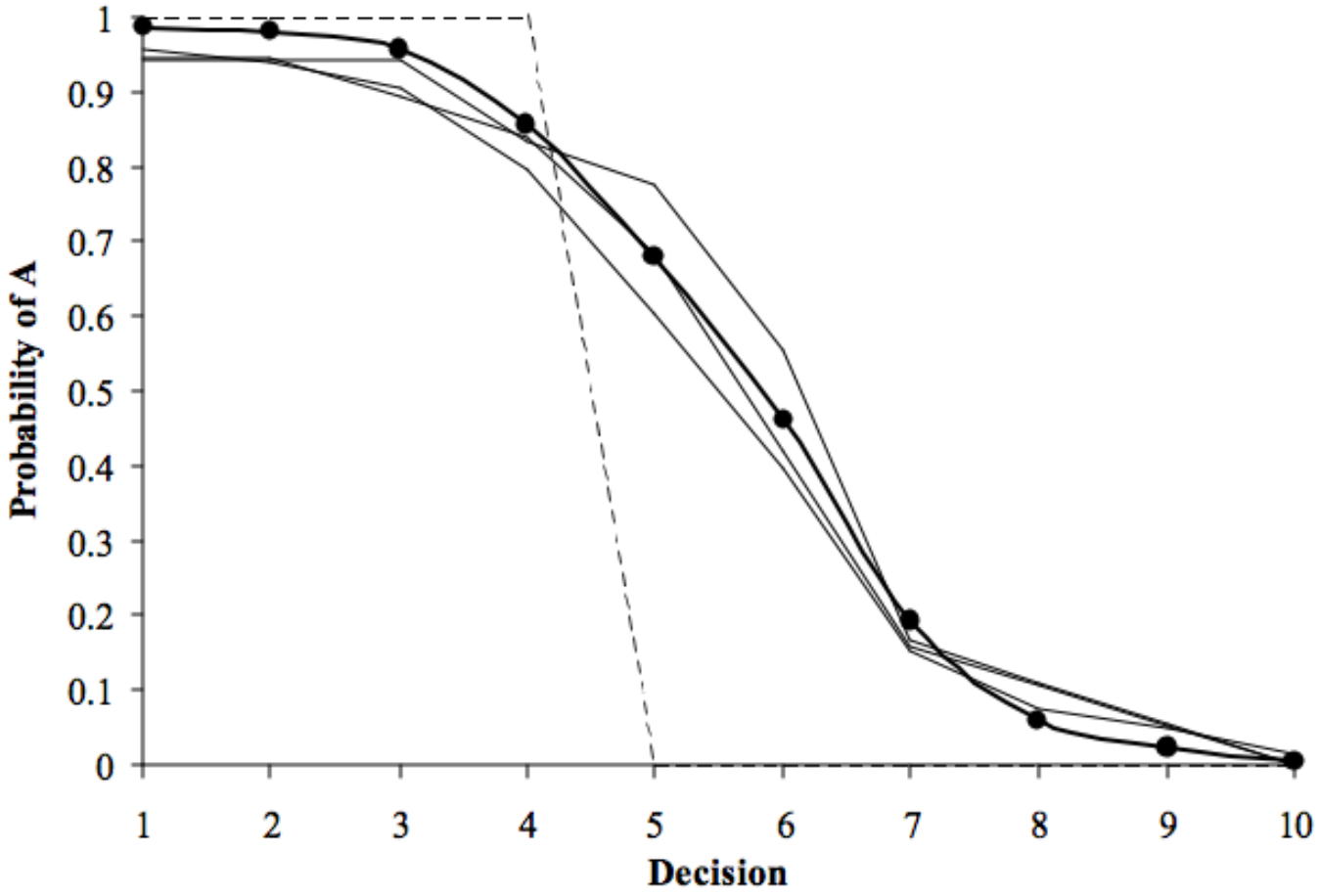
Lottery choice task

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

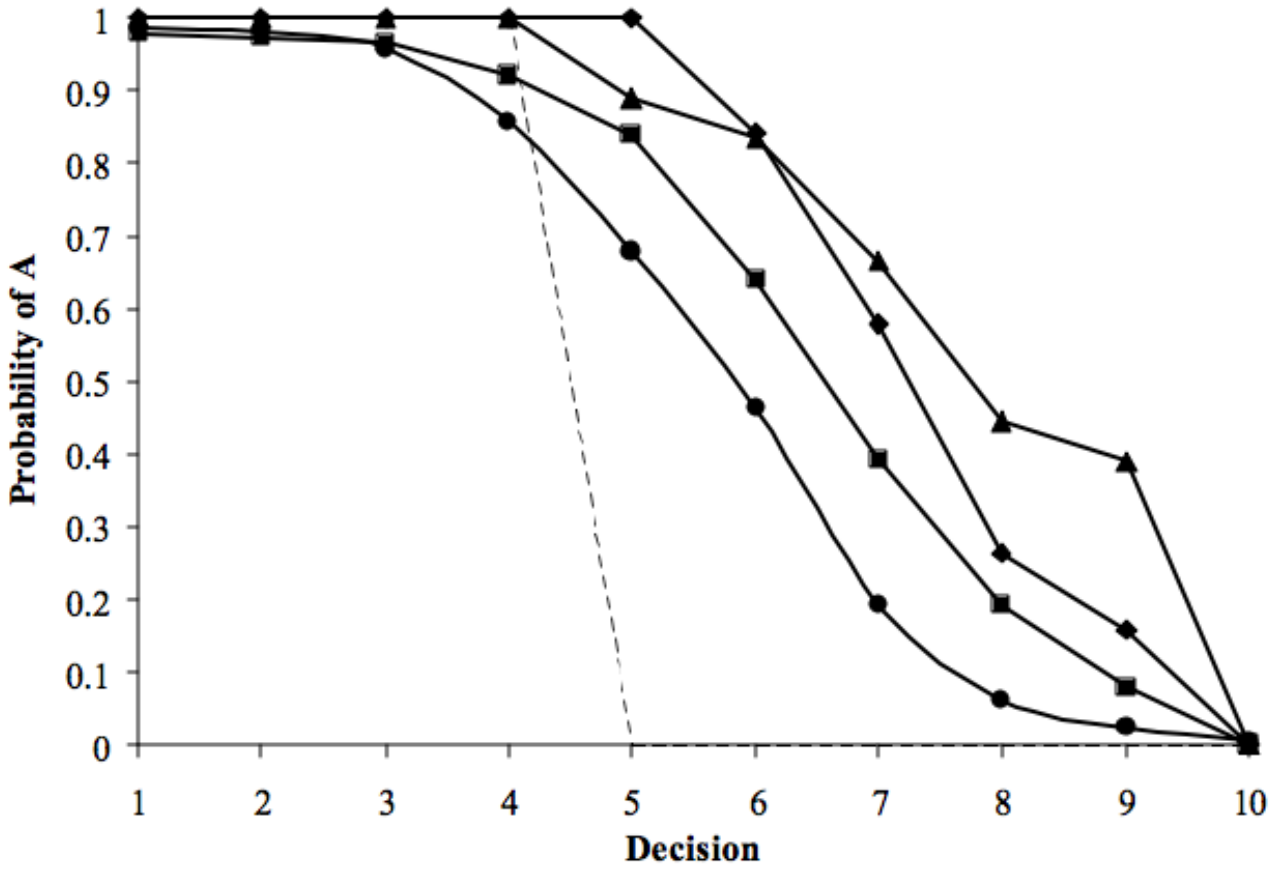
Lottery choice task

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

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