Class 3: Experiments in social choice

Outline

- Testing a simple hypothesis
  - Controlled Randomized Tests
- Measuring the effect of a variable
- Testing a more complex hypothesis
  - Group vs. individual behavior
- Testing and comparing behavioral models
- Field experiments
- Natural experiments
- Story time

Putting a theory to the test

"Testing a theory means **checking some predictions of the theory** and that is what we do in this article."

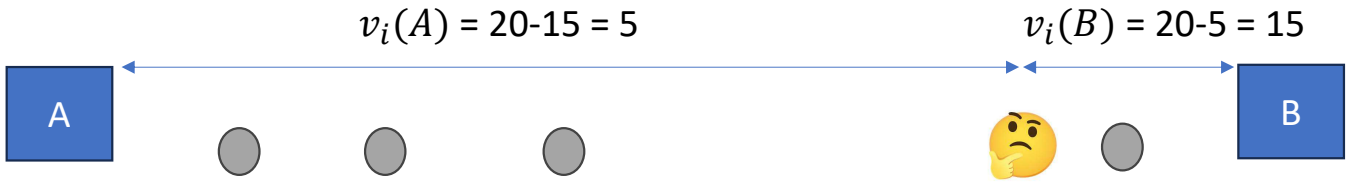
Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Putting a theory to the test

Theory: People vote rationally (maximize expected utility)

(simple) setting: two candidates, Majority voting.

Each voter can vote (at a cost of 1) or abstain. Gets $v_i(winner) - cost$.



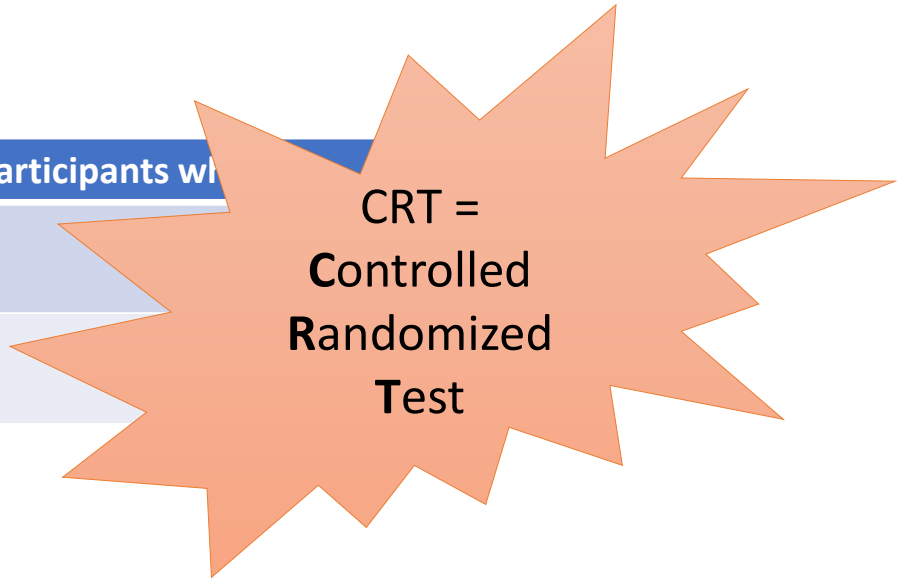
$$\underbrace{u_i(\text{vote}) - u_i(\text{abstain})}_{\text{“DiffVote”}} = [\![i \text{ is pivotal}]\!]\ (v_i(B) - v_i(A))\ - 1$$

Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Putting a theory to the test

- A voter should vote if and only if DiffVote is positive
- Do they?

	% of participants wh
DiffVote positive (17% of votes)	
DiffVote negative (82% of votes)	



Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Putting a theory to the test

- A voter should vote if and only if DiffVote is positive
- Do they?

	% of participants who vote
DiffVote positive (17% of votes)	**82%**
DiffVote negative (82% of votes)	

Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Putting a theory to the test

- A voter should vote if and only if DiffVote is positive
- Do they?

	% of participants who vote
DiffVote positive (17% of votes)	**82%**
DiffVote negative (82% of votes)	**71%**

"The poor performance … may be due to the fact that subjects are not very good at predicting other voters' behavior, but **their decision may be consistent with their perceptions**."

Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Putting a theory to the test

How can we know this?

- A **consistent** voter should vote if and only if her **subjective DiffVote** is positive
- Do they?

(subjective perception)	% of participants who vote
DiffVote positive (31% of votes)	**72%**
DiffVote negative (69% of votes)	**76%**

Can you think of an alternative explanation?

Blais, André, et al. "To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections." Electoral studies 36 (2014): 39-50.

Measuring the effect of a variable

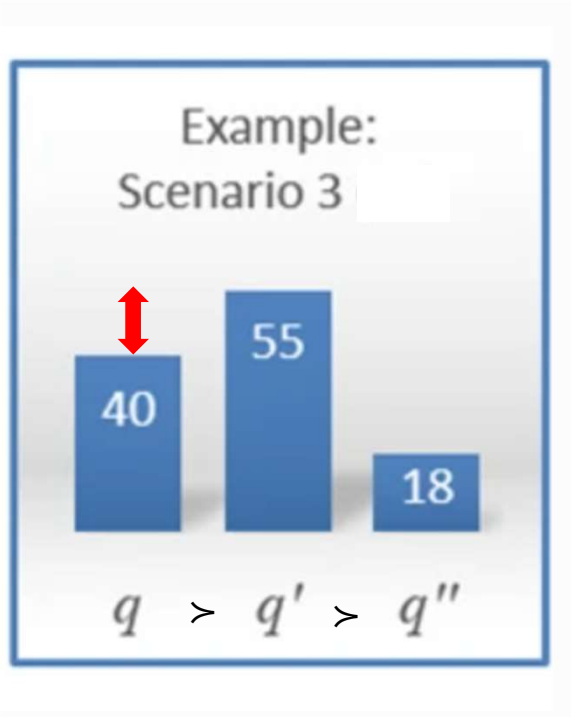
Theory: People vote rationally

Derived hypothesis I:

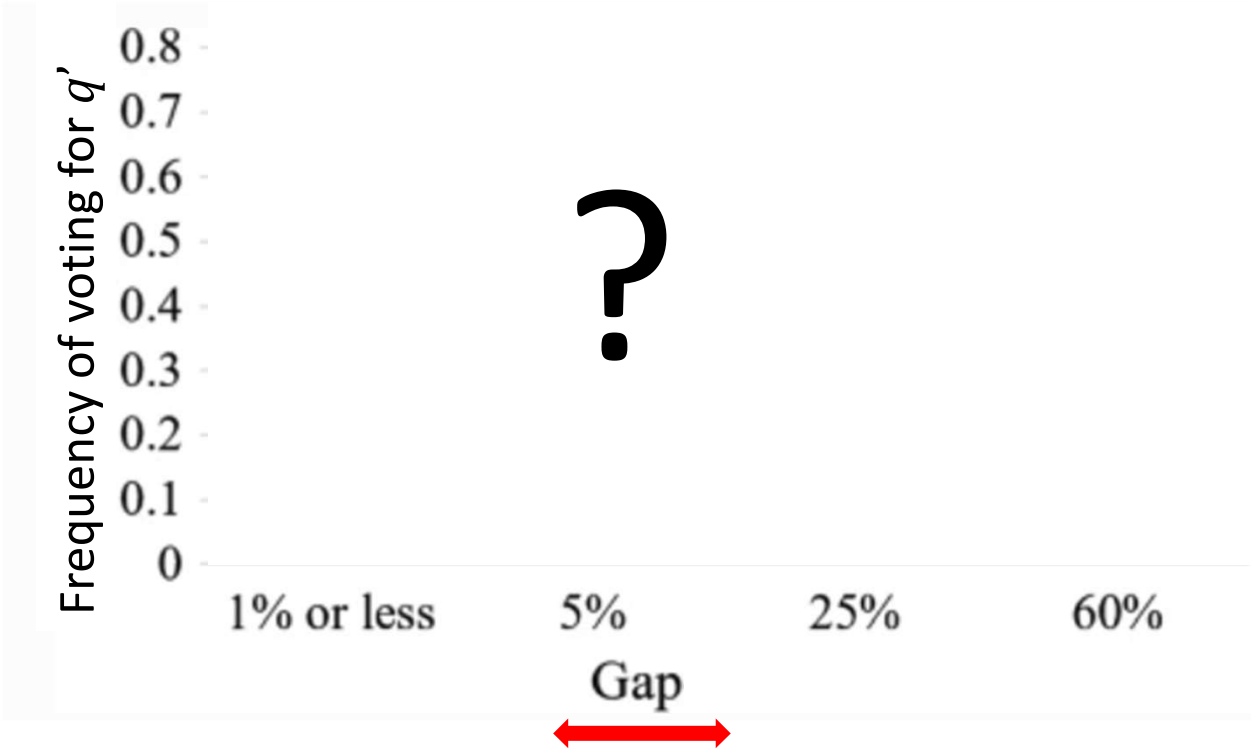
"always vote for q in Scenario 3"

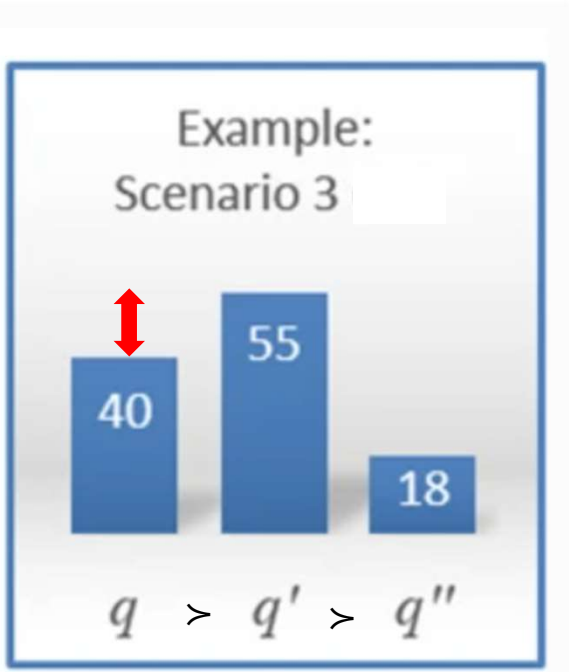
Derived hypothesis II:

"score gap should not matter"



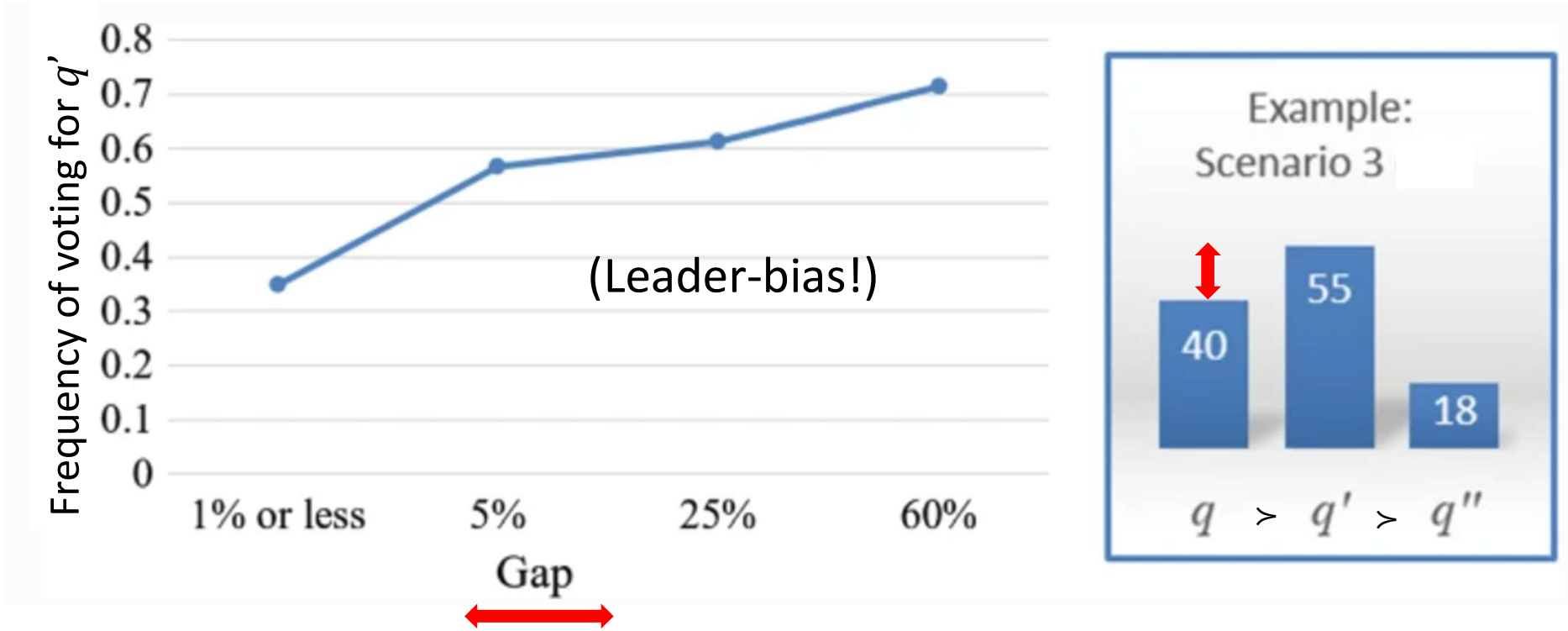
Measuring the effect of a variable





Strategic voting in the lab: compromise and leader bias behavior, R Meir, K Gal, M Tal , JAAMAS 2020.

Measuring the effect of a variable



Another example

Hypothesis: Candidates converge to the median voter's position

With full information: Downs-Hotelling model*

More surprising: also true (theoretically!) with very limited information**

Does it hold in practice?

*Anthony Downs, "An Economic Theory of Democracy" (1957)

**McKelvey, Richard D., and Peter C. Ordeshook. "Elections with limited information: A fulfilled expectations model using contemporaneous poll and endorsement data as information sources." *Journal of Economic Theory* 36.1 (1985a): 55-85.

Voting-on-a-line experiment

- Several “voter subjects” with single-peak preferences are placed on a line. Their positions and preferences are **private**.

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

Voting-on-a-line experiment

- Several “voter subjects” with single-peak preferences are placed on a line. Their positions and preferences are **private**.
- Two “candidate subjects” A and B select positions (also privately)
  - It is only announced which candidate is Left and which is Right

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

Voting-on-a-line experiment

- Several “voter subjects” with single-peak preferences are placed on a line. Their positions and preferences are **private**.
- Two “candidate subjects” A and B select positions (also privately)
  - It is only announced which candidate is Left and which is Right
- All voters vote for A or B

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

Voting-on-a-line experiment

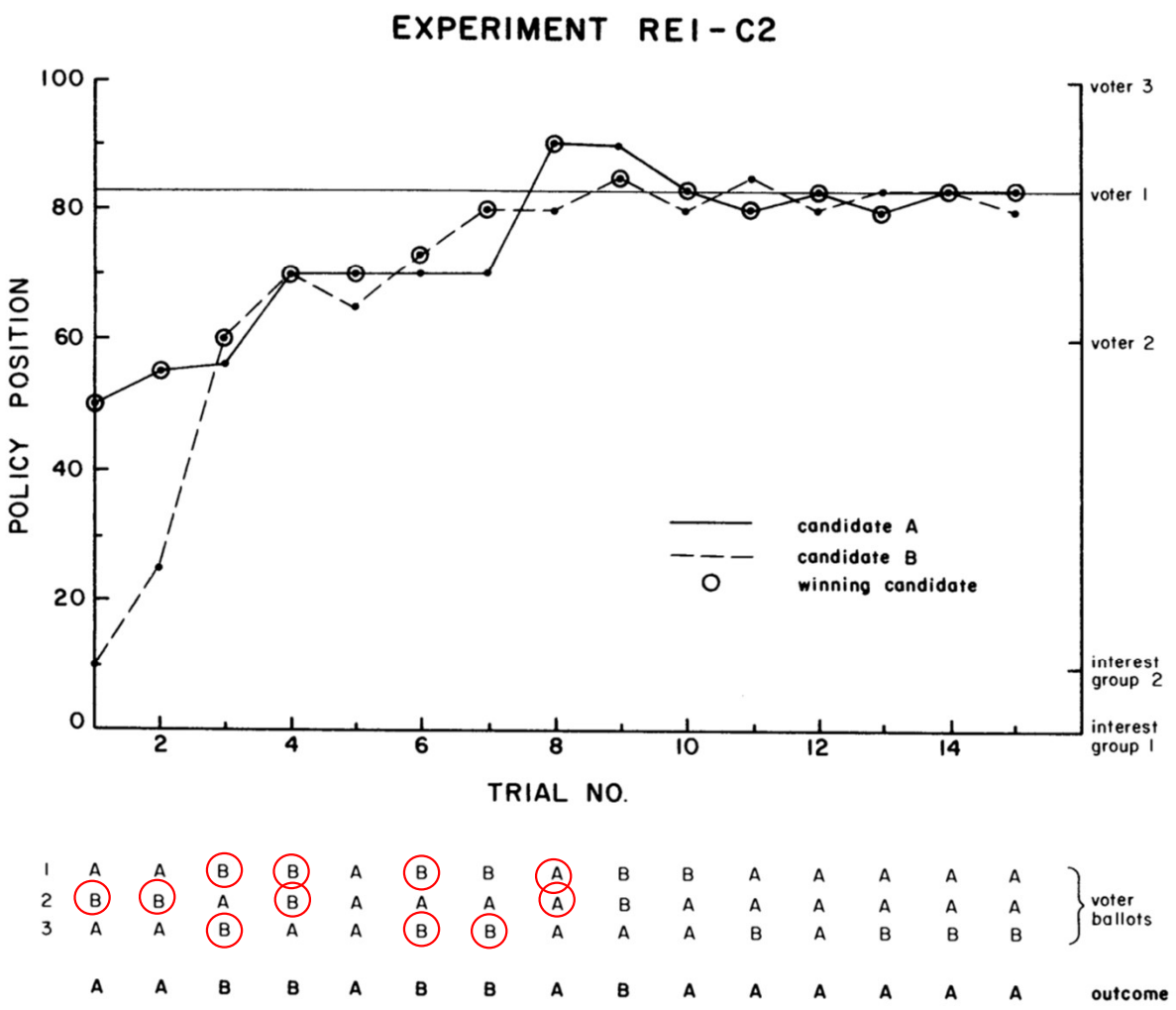
- Several “voter subjects” with single-peak preferences are placed on a line. Their positions and preferences are **private**.
- Two “candidate subjects” A and B select positions (also privately)
  - It is only announced which candidate is Left and which is Right
- All voters vote for A or B
- Winner’s position and margin are announced
- Repeat for up to one hour

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

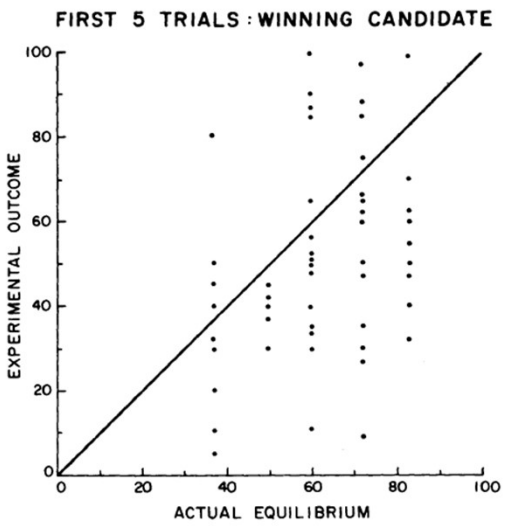
Voting-on-a-line experiment

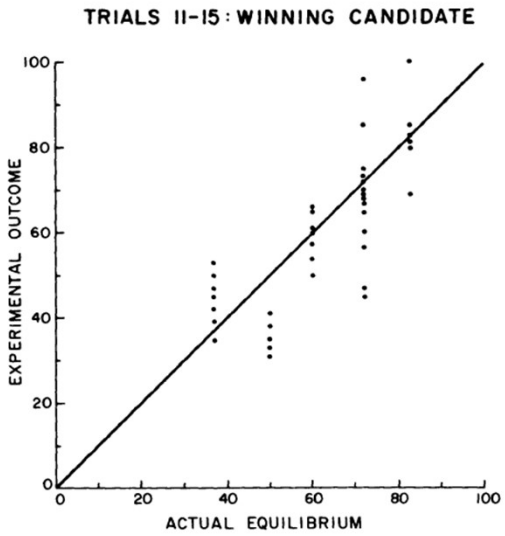
- Prediction: The two candidates will converge to the median voter position

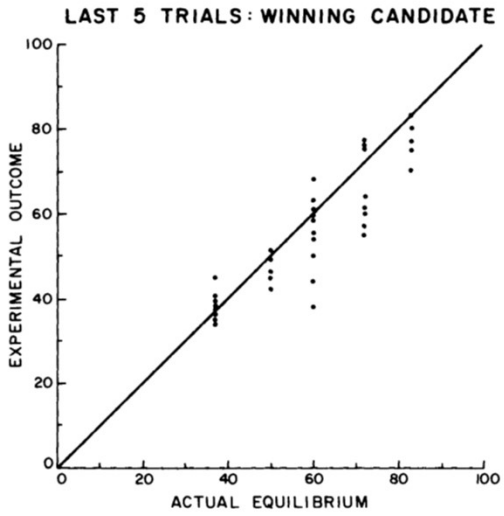
McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.


EXPERIMENT RE1-C2
POLICY POSITION
100
80
60
40
20
0
2
4
6
8
10
12
14
TRIAL NO.
voter 3
voter 1
voter 2
interest group 2
interest group 1
candidate A
candidate B
winning candidate
1 A A B B A B B A B B A A A A A
2 B B A B A A A A B A A A A A A
3 A A B A A B B A A A B A B B B
A A B B A B B A B A A A A A A
voter ballots
outcome

Results







- The prediction is corroborated!
  - But…

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

Group-level vs. Individual Behavior

- Many different dynamics may lead to the same outcome
- Outcome alone (e.g. convergence/equilibrium) does not tell us what voters **did**
- We can form and test explicit hypotheses
- **H1:** voters vote as if they have full information
  - Consistent with ~82% of votes
  - Less on first trials
  - But not possible!
- **H2:** voters form beliefs on candidates' positions using regression on recent rounds
  - Consistent with ~85% of votes

What else can voters do?

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential elections with limited information." *American Journal of Political Science* (1985b): 480-512.

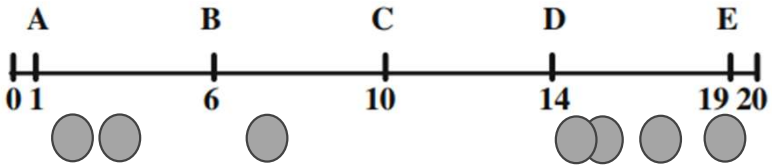
Back to Individual Behavior

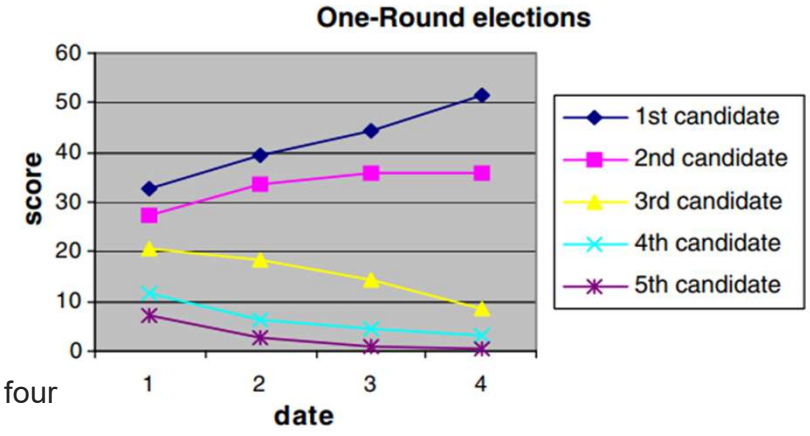
- We saw in second class several “theories” of strategic voting
  1. Truthful (i.e. non-strategic)
  2. Heuristic (say, 2-pragmatist, Laslier’s Leader Rule)
  3. Rational (say, Calculus of Voting)

- Which best describes voters’ behavior?

Comparing theories

- Setting: five candidates on a line
- Voters are placed at random **known** positions
- Vote simultaneously four times on consecutive days
- Voting rules: 1R Plurality, 2R Plurality, STV, Approval





Van der Straeten, Karine, et al. "Strategic, sincere, and heuristic voting under four election rules: an experimental study." *SCW* 35 (2010): 435-472.

Results

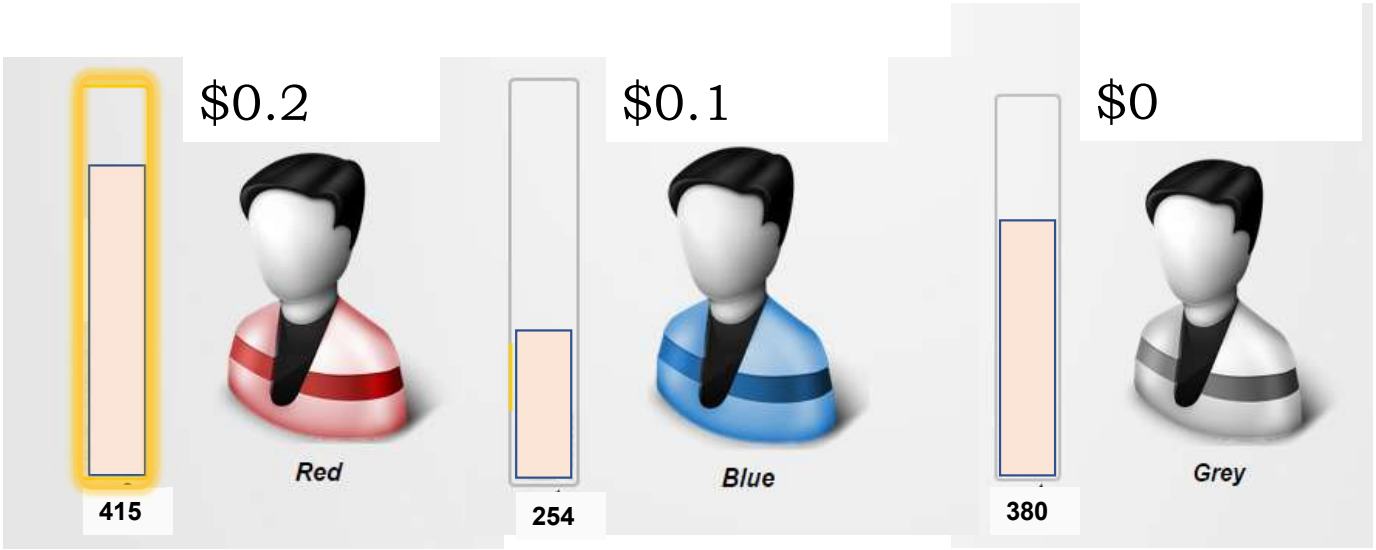
- STV: >90% votes consistent with truthful voting
- Approval: ~87% of votes consistent with the Leader Rule
- Impressive!
- No competing theories
- What about Plurality?

1R: correct predictions	Sincere	Strategic	Top-Two	Top-Three
$t = 1$ (%)	68.7	53.8	49.7	67.5
$t = 2$ (%)	54.8	64.2	60.7	71.2
$t = 3$ (%)	48.7	74.6	75.3	69.5
$t = 4$ (%)	44.7	86.7	80.1	66.8
All dates	54.2	66.7	66.5	68.5
(Testable, all dates)	2647	1968	2775	2667

- No theory is consistent with the votes
- Why?

Voting experiments (one shot)

Vote only once.
Winner determined based on simulated votes.

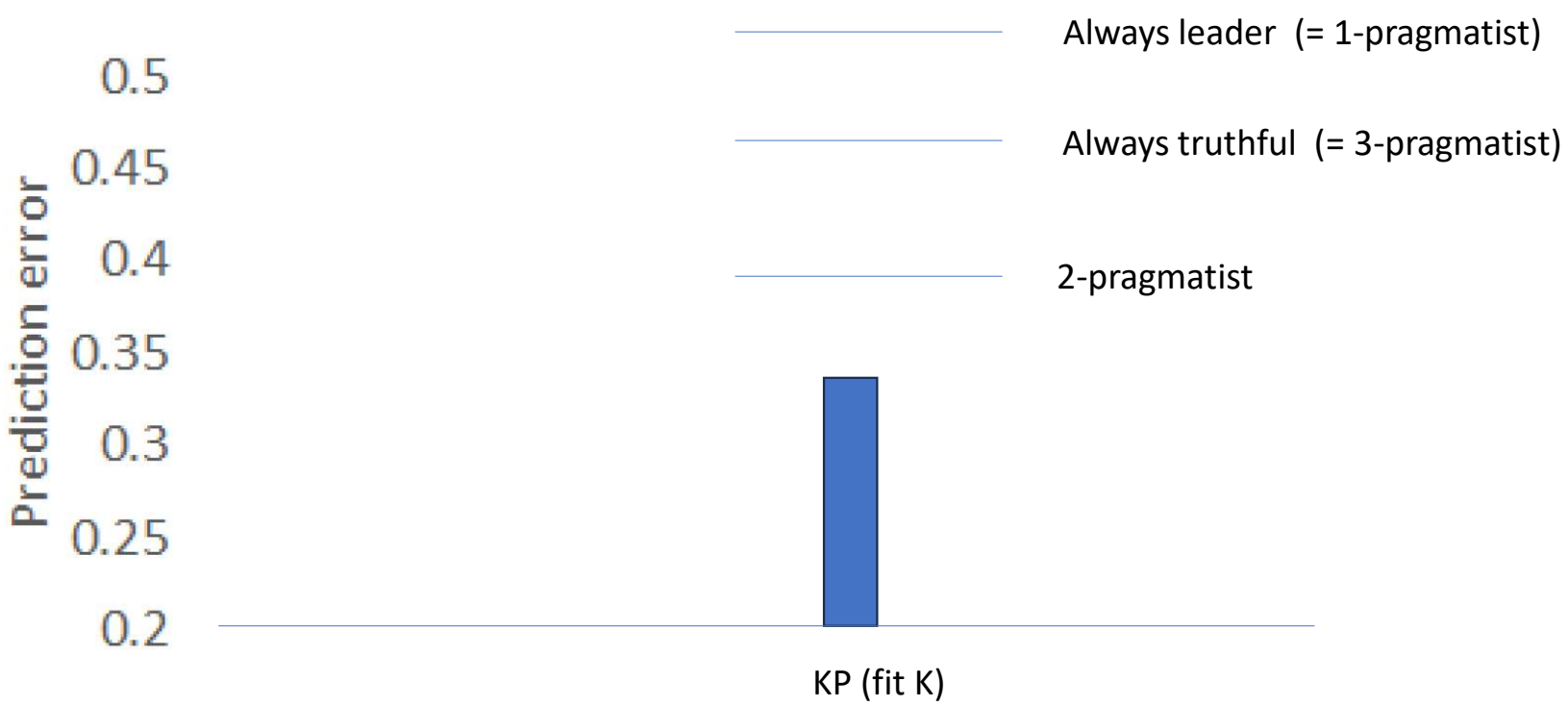


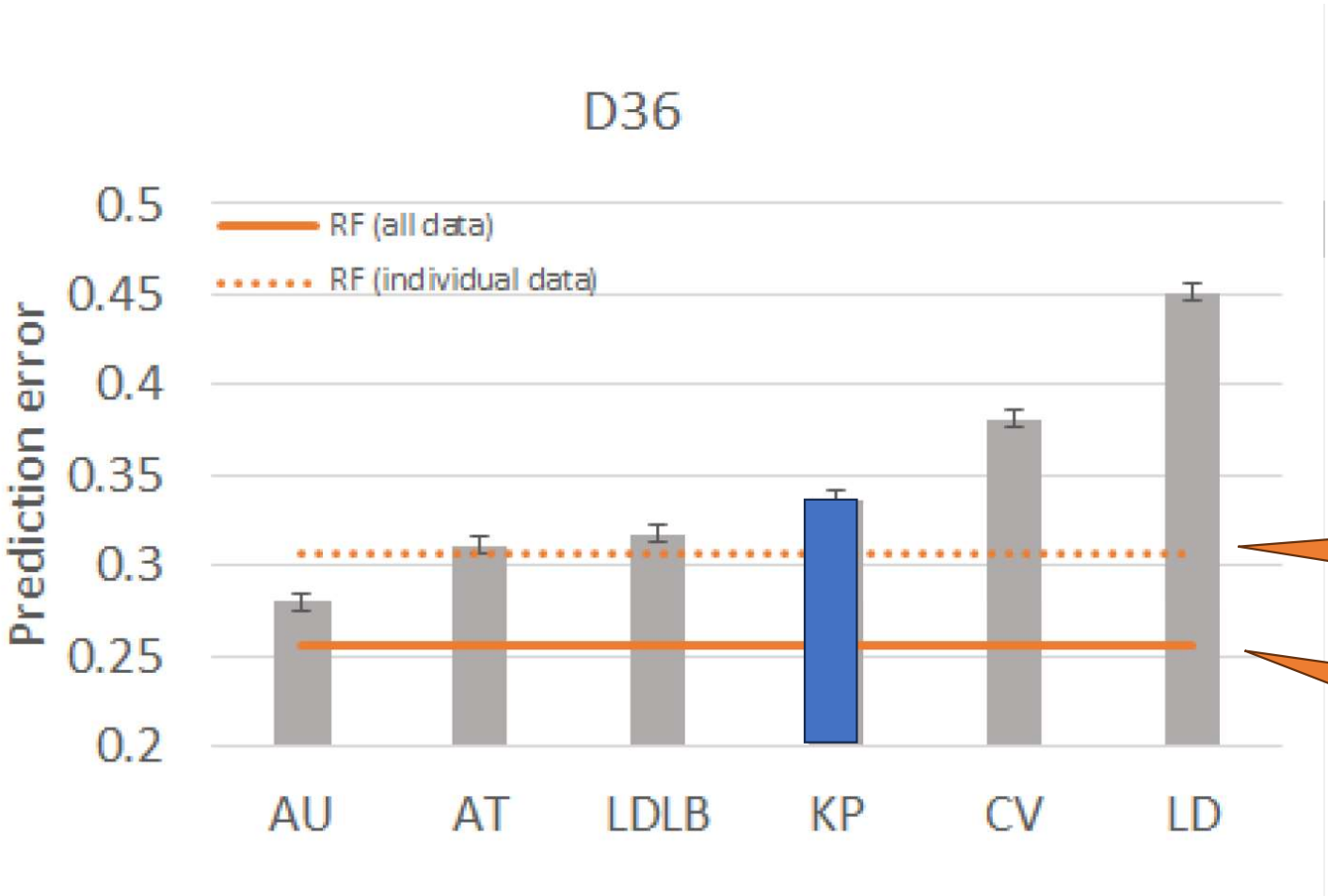
Strategic voting in the lab: compromise and leader bias behavior, R Meir, K Gal, M Tal , JAAMAS 2020.
Fairstein, Roy, et al. "Modeling people's voting behavior with poll information." *, AAMAS 2019.*

Testing individual behavior

- Each participant voted in 20-40 different polls
- Experimental design: a single-player game
- For every heuristic/model:
  - Learn voter's parameter(s) from samples
  - **Predict** remaining samples
  - Measure accuracy with 10-fold cross validation

- Which models are most predictive?


Always leader (= 1-pragmatist)
Always truthful (= 3-pragmatist)
2-pragmatist
Prediction error
0.5
0.45
0.4
0.35
0.3
0.25
0.2
KP (fit K)

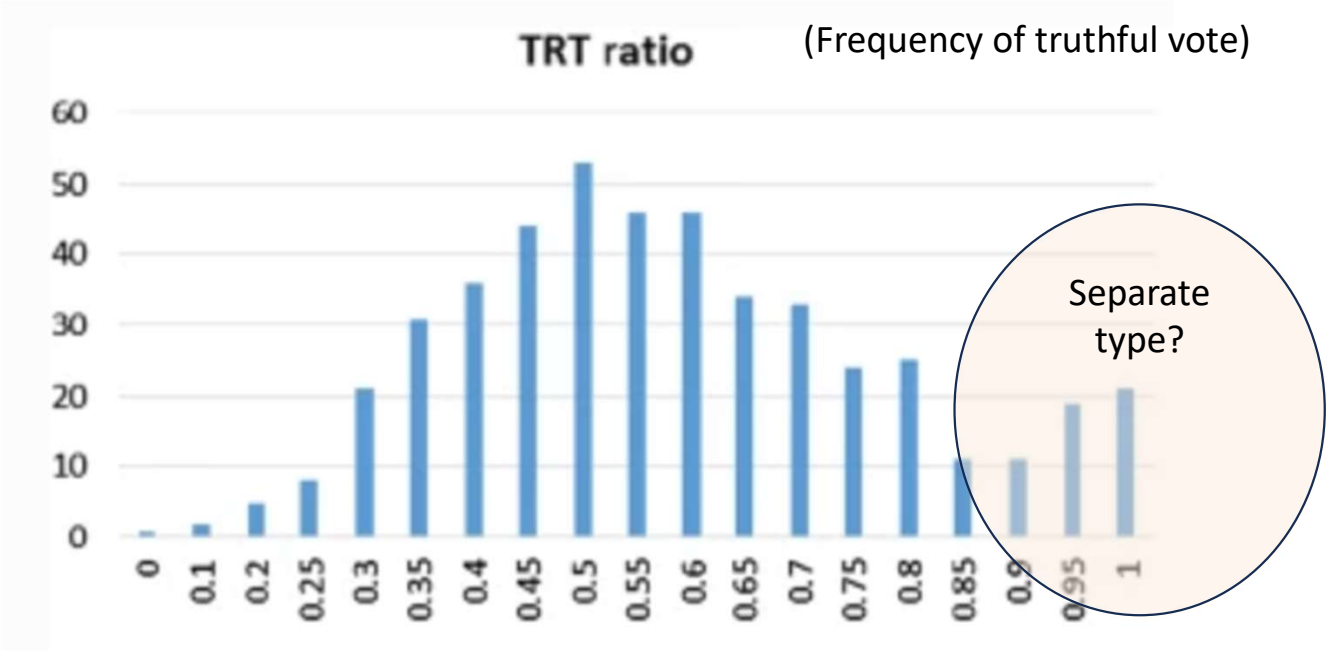


Benchmark 2: machine learning with access to this subject's data

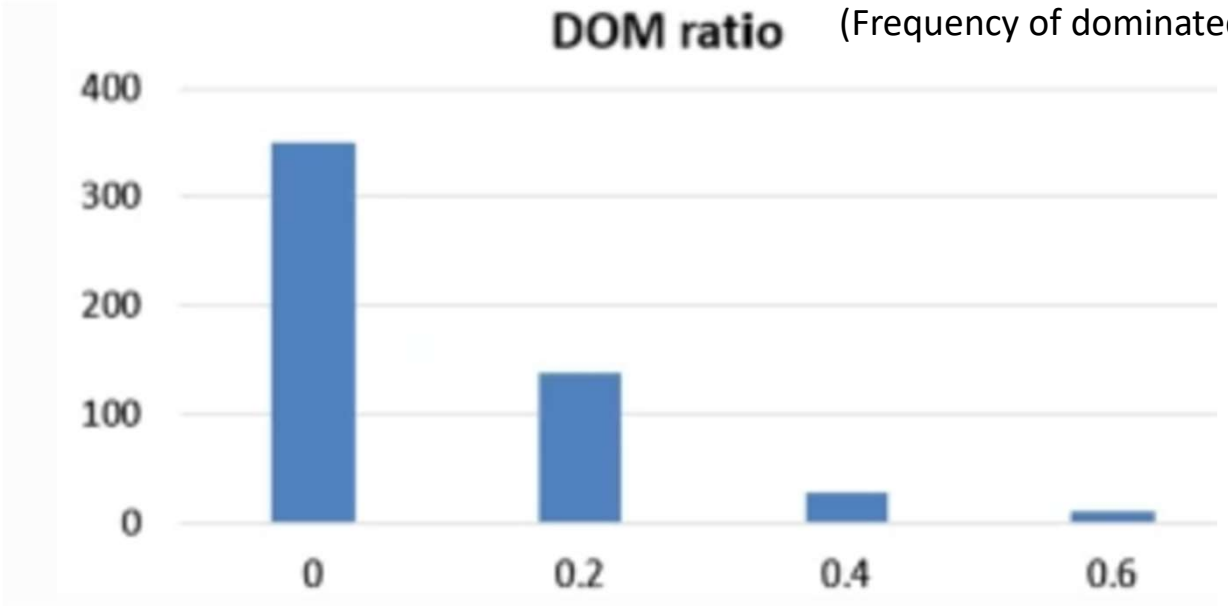
Benchmark 1: machine learning with access to all data

(all these models discussed on Thursday!)

Voter types


TRT ratio
(Frequency of truthful vote)
60
50
40
30
20
10
0
0
0.1
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
Separate type?

Voter types



(Frequency of dominated vote)

- Laslier, Myerson – repeated voting games
- Field experiments?
- Tal, Meir, Gal – lab experiments

Field experiments

- Theory: more information leads to more efficient outcomes

Submissions | Reviews | Status | Paper Bidding | Events | Support

AAMAS 2024 Paper Bidding

Explanation for Choices

Choice	Explanation	Papers
yes	I want to review this paper	11
maybe	I can review it	4
no	I prefer not to review it	2
conflict	I have a conflict of interest	0

Submissions

~1000 submissions

Choice	#	Submission
yes maybe no conflict	4	(anonymous). **Designing an Adaptive Learning Module to Teach Software**
yes maybe no conflict	5	(anonymous). **This is a cool title** (details)
yes maybe no conflict	6	(anonymous). **Something on game theory** (details)
yes maybe no conflict	7	(anonymous). **AI for dummies** (details)

“Welcome to the AAMAS 2024 program committee… **Please bid positively on at least 40 papers…”**

Field experiments

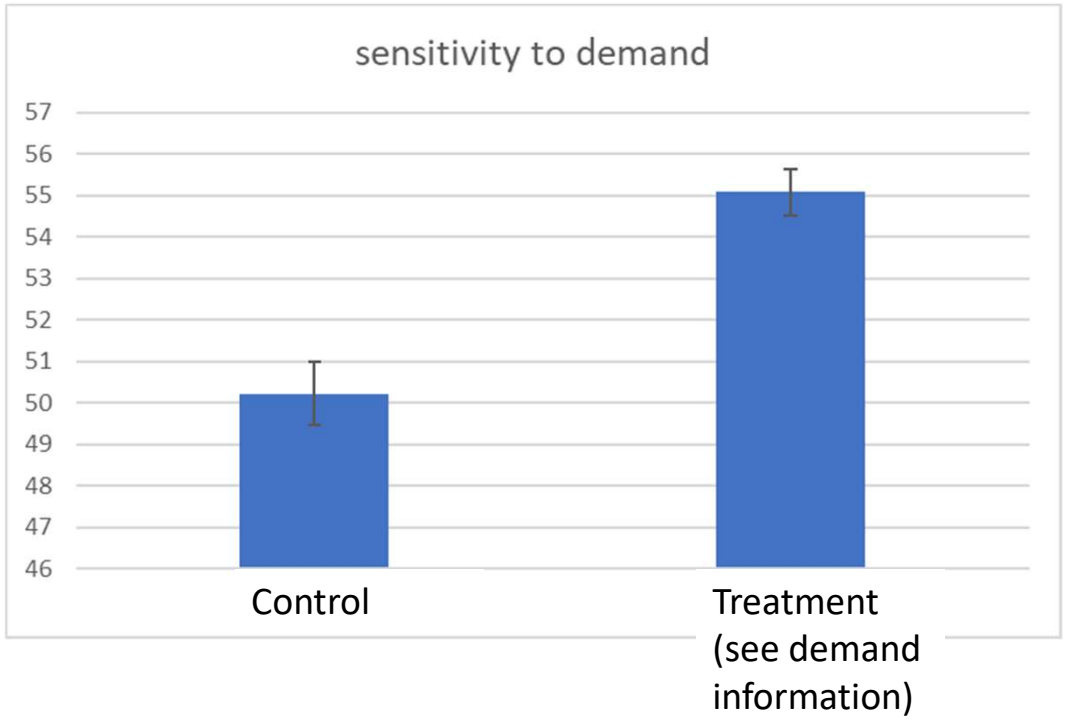
- Theory: more information leads to more efficient outcomes
- Derived hypothesis:
  - Revealing dynamic information on demand will incentivize bidders to pick low-demand papers*
- Partition bidders in a conference into two groups:
  - Control group bids as usual
  - Treatment group see additional information on demand
- Preliminary results in lab experiments and a small workshop**
- A Large experiment at ECAI-2023

*Meir, Reshef, et al. "A market-inspired bidding scheme for peer review paper assignment." *AAAI* 2021.

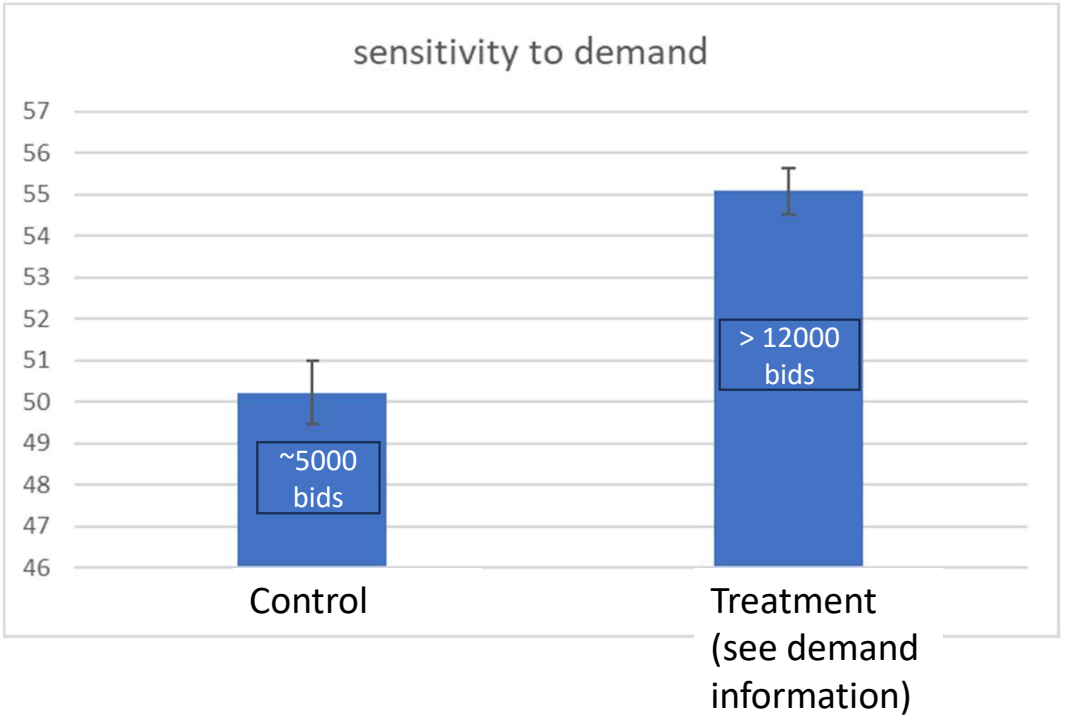
**Rozenzweig, Inbal, et al. "Mitigating Skewed Bidding for Conference Paper Assignment." *AAMAS* 2023.

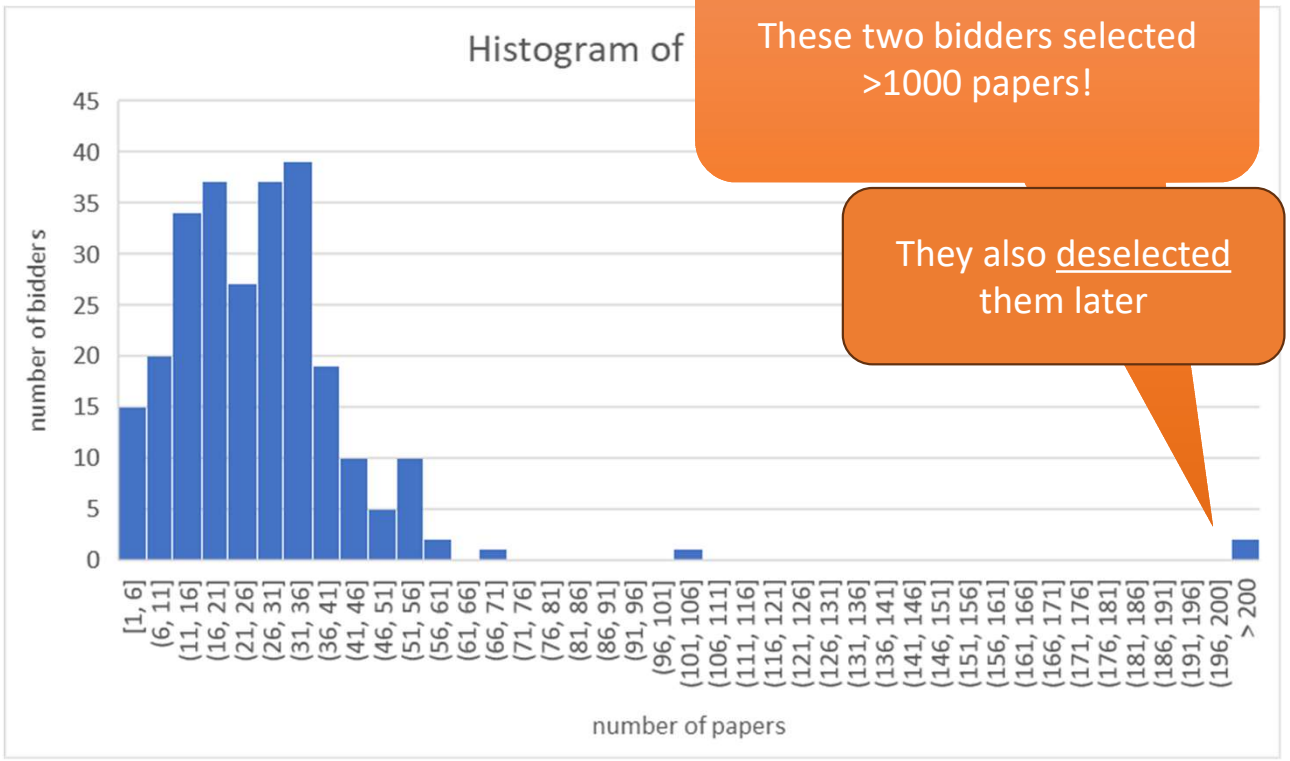
Results

- We average over all bids (papers selected by user) in each group



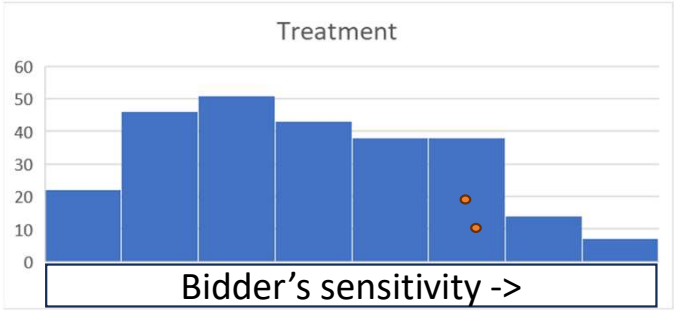
- Looks like hypothesis is true!
- But wait
  - Treatment group was only slightly larger
  - Number of bids is **more than twice**
  - Something is suspicious

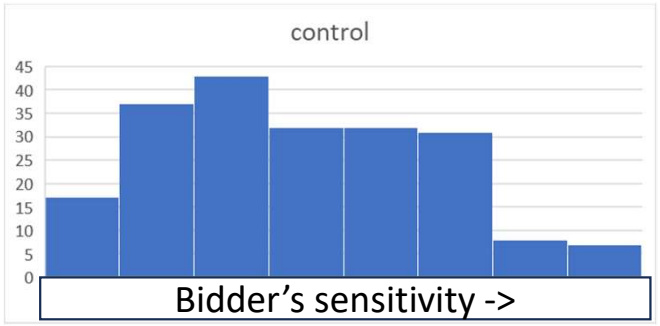




These two bidders selected >1000 papers!

They also deselected them later

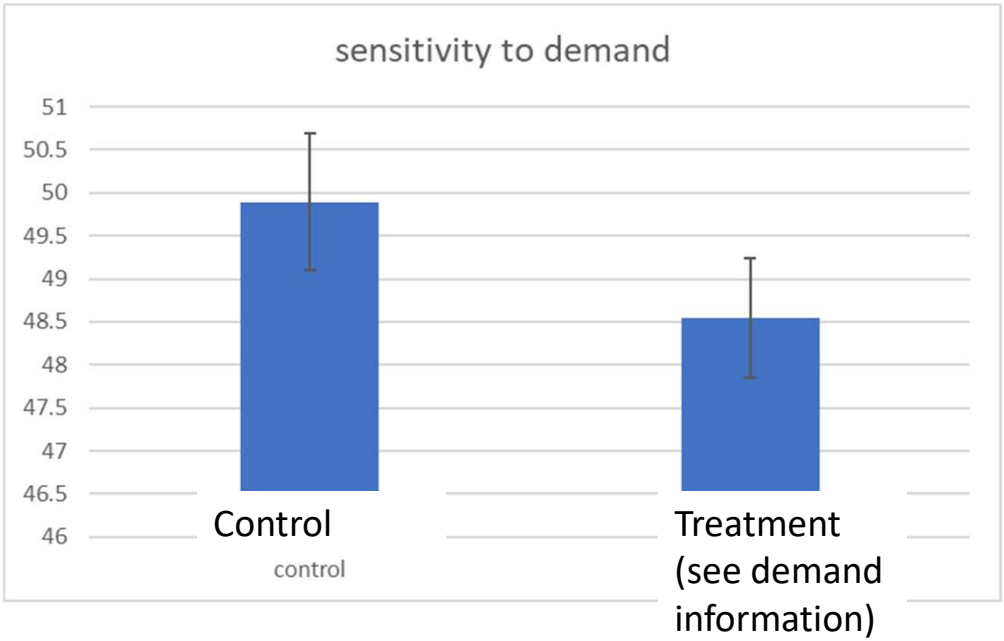




Let’s re-do the analysis

- We now only average over all bids **that were not later removed** in each group

- Effect is gone ☹




I WANT TO
BELIEVE

Natural experiments

- Theory: geographical distance from the polling station affects turnout
- Hypothesis: moving the station farther will reduce turnout

- Option 1: analyze correlation between distance and turnout
  - Correlation ≠ causation!!!
- Option 2: Run a controlled experiment where we move some **random** polling stations and measure the effect
  - Good luck with that!

A *Natural experiment* takes advantage of some difference between populations that are expected to be otherwise similar

Natural experiment (cont.)

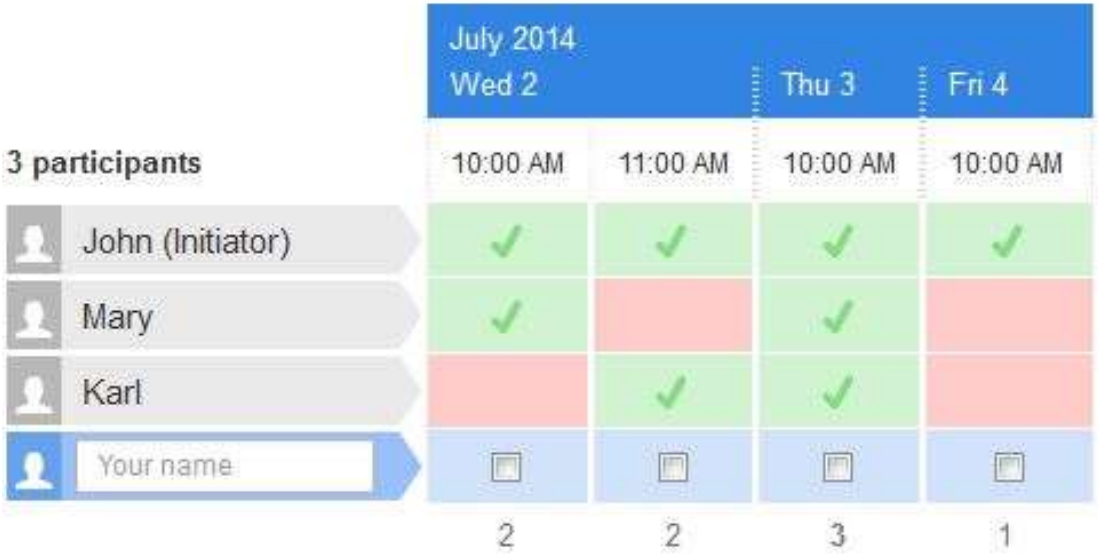
- Theory: geographical distance from the polling station affects turnout
- Hypothesis: moving the station farther will reduce turnout

- In October 2003, LA county consolidated **some** polling stations
  - These stations became farther from some of their voters
- First need to show that affected populations are similar to those unaffected
  - (mostly) uncorrelated with other factors that predict turnout
- Then measure difference in turnout between groups

What if there is correlation?

Brady, Henry E., and John E. McNulty. "The costs of voting: Evidence from a natural experiment." *Annual Meeting of the Society for Political Methodology, Palo Alto, CA*. 2004.

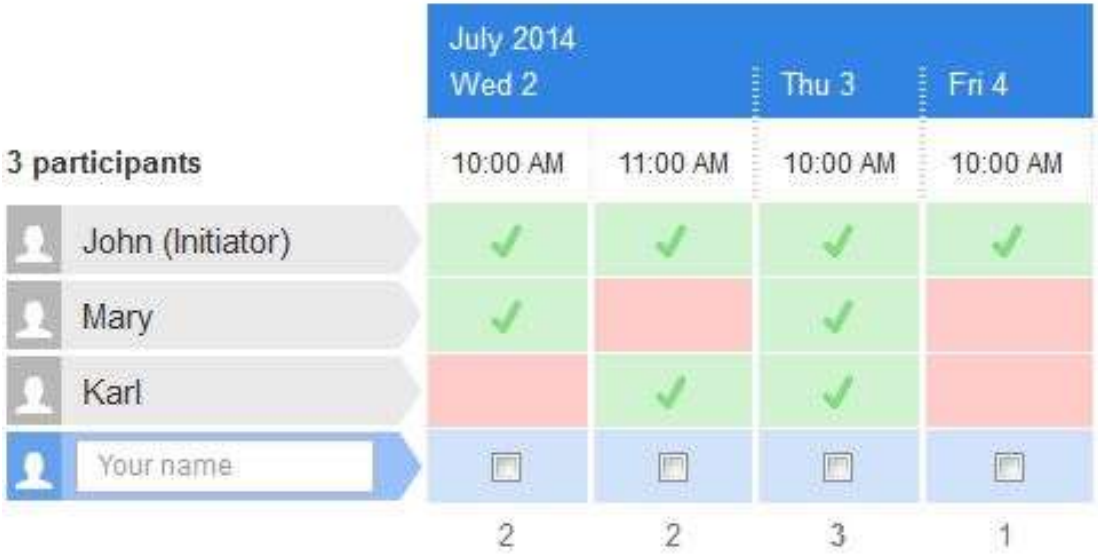
Doodle example



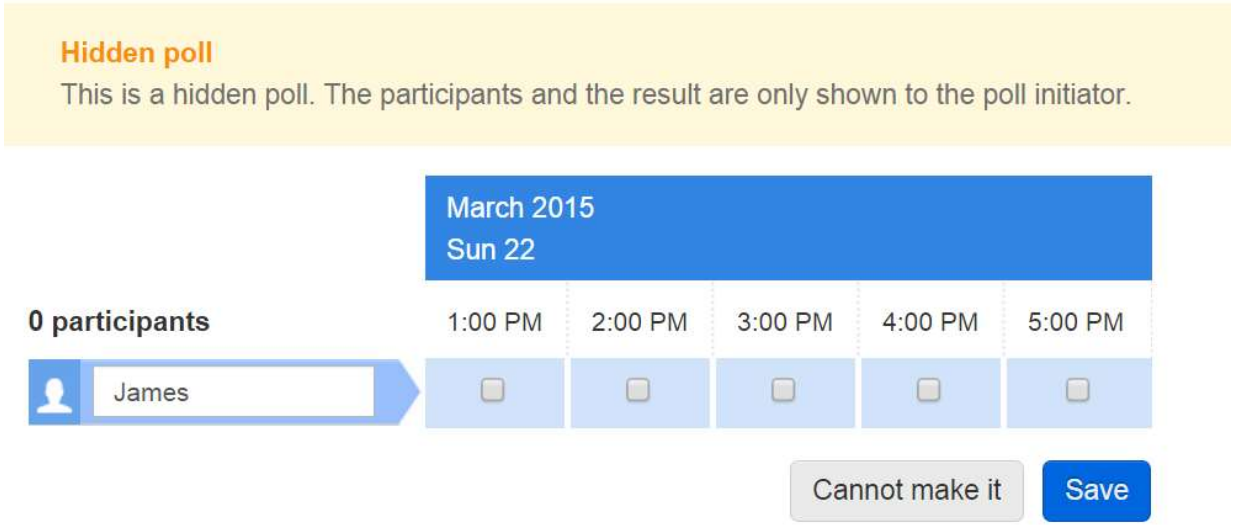
How do people coordinate?

- Scheduling as a form of group coordination.
- Each participant balances her own interests with the group interest.
- Do people behave strategically?
  - Problem: We don’t know their preferences!
- Idea:
  - Compare behavior to a situation where there is no opportunity to strategize

(open) Doodle example



(hidden) Doodle example



Assumption 1: Populations on both conditions are similar
Assumption 2: People did not strategize on the "hidden" condition

We had no way of testing these assumptions ☹

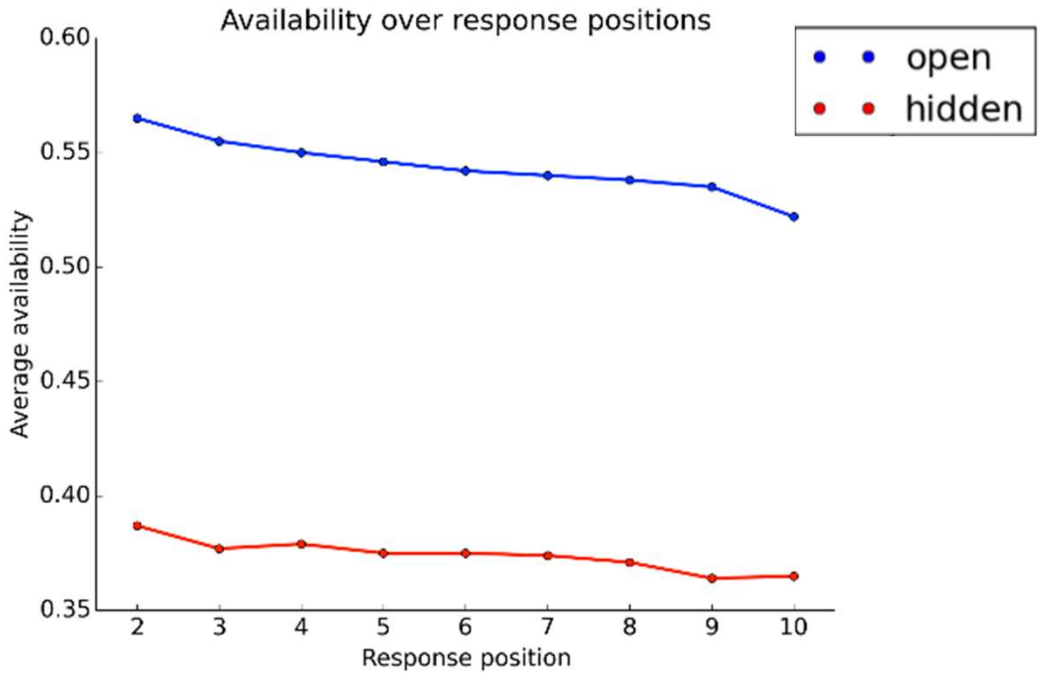
Making conjectures

- In which condition more time slots get approved?
- How does availability behave as time passes?

- Try to guess the result *before* doing the analysis!
- Then check and compare

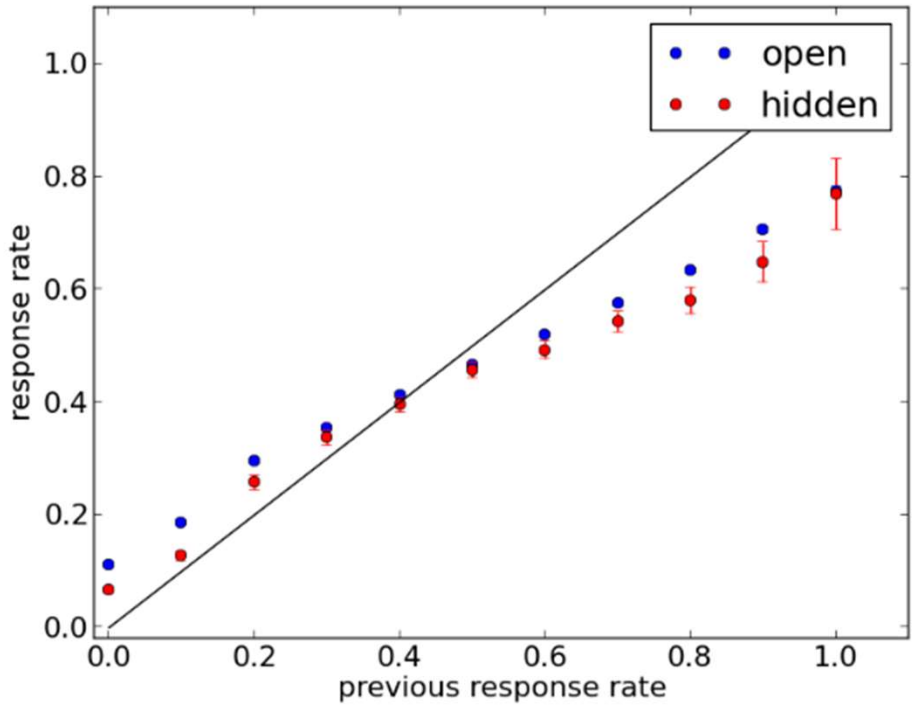
Availability over response positions

The availability of a voter is the fraction of slots that she approves.
The response position of a voter is the order that she participates in.



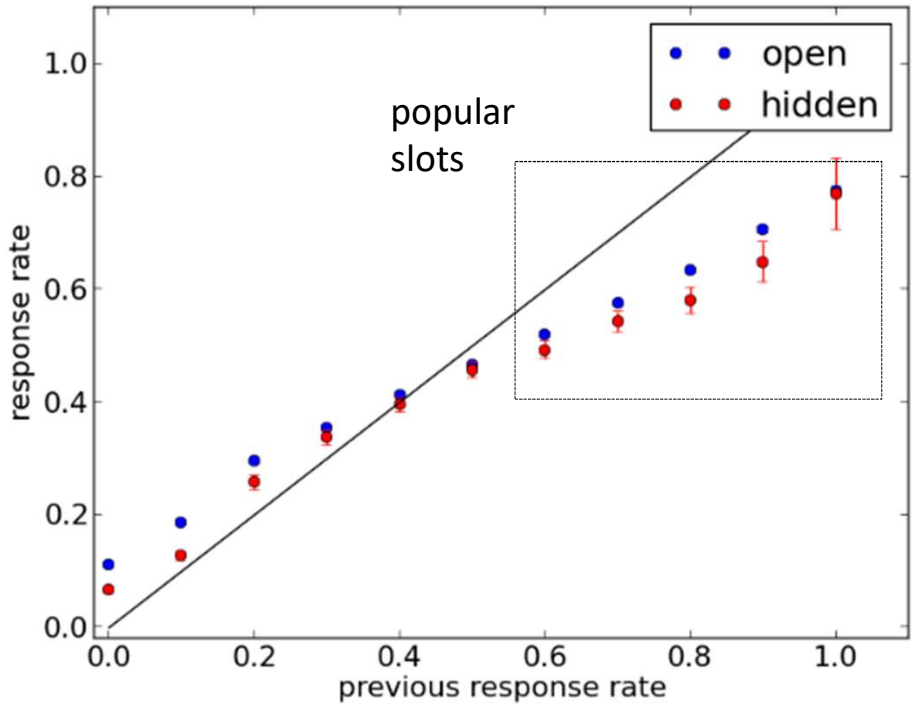
Response curve at 11

Conditioned on a slot approved by x of the first 10 voters, what is the probability that the 11th voter approves it?



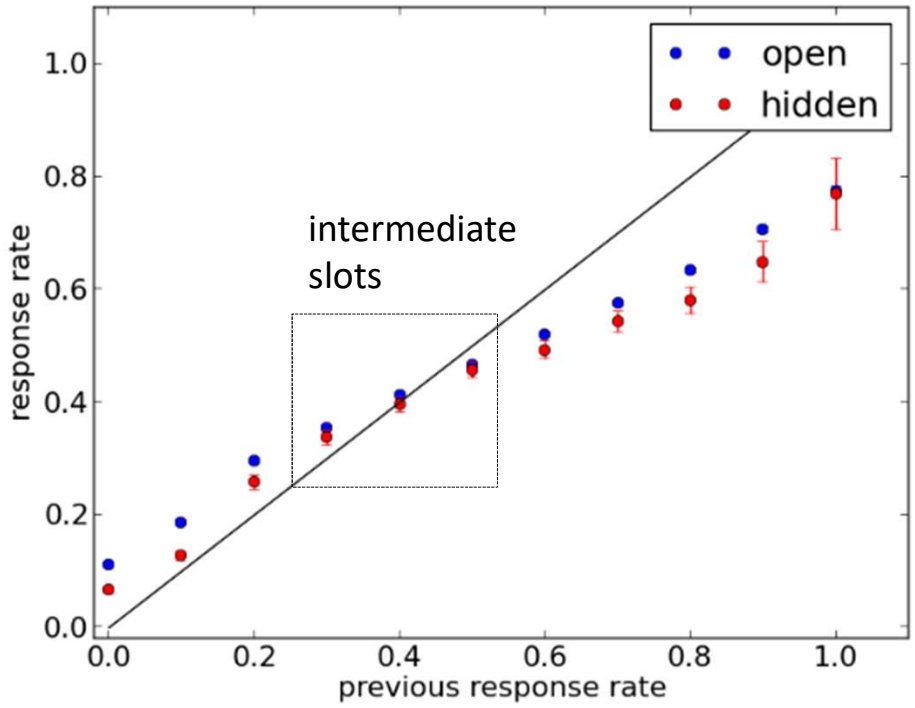
Response curve at 11

Conditioned on a slot approved by x of the first 10 voters, what is the probability that the 11th voter approves it?



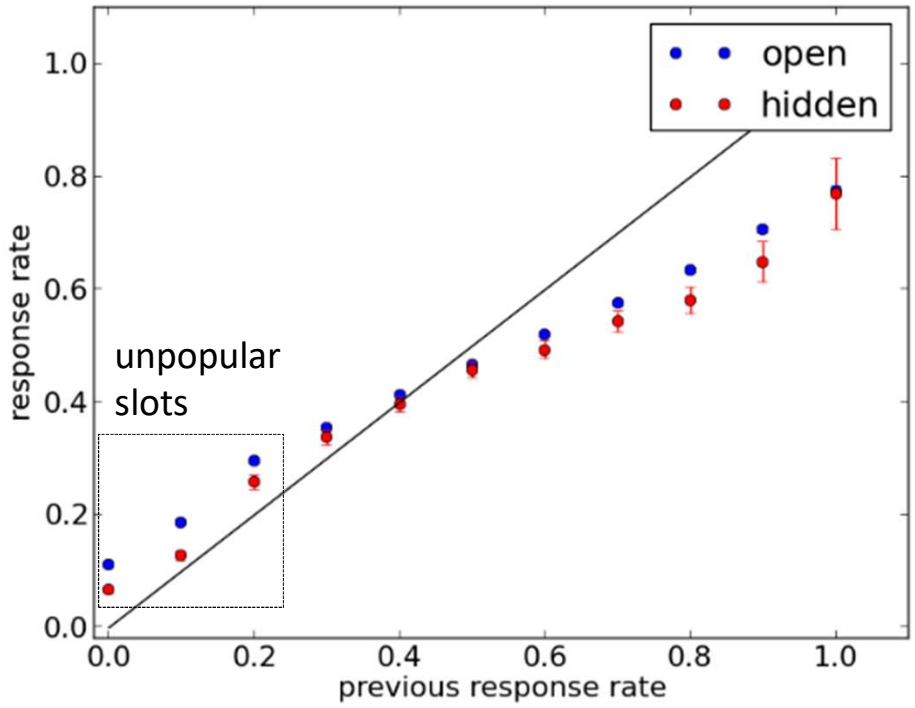
Response curve at 11

Conditioned on a slot approved by x of the first 10 voters, what is the probability that the 11th voter approves it?



Response curve at 11

Conditioned on a slot approved by x of the first 10 voters, what is the probability that the 11th voter approves it?



Stepping up: testing varyin

Finally, there is the method of extrapolating from real outcomes obtained under one voting procedure the likely outcomes, *ceteris paribus*, that would be obtained under other voting procedures. Two difficulties are associated with this method. Firstly, a prerequisite for conducting extrapolations from an observed procedure to other procedures is that the voters' preference orderings among the candidates under the observed procedure are known. Since most

Felsenthal, Dan S., Zeev Maoz, and Amnon Rapoport. "An empirical evaluation of six voting procedures: do they really make any difference?." British Journal of Political Science 23.1 (1993): 1-27.

The most common of these methods is to carry out computer simulations.[3] The main drawback of this method is that all possible preference orderings that voters may have among the competing candidates are considered to be both complete and equally likely, or if there are too many of them, the ones under investigation are assumed to constitute a random sample from a well-defined population. Clearly, these simplifying assumptions do not necessarily reflect reality.

A second method is to conduct controlled laboratory experiments where voters' preference orderings are held constant and their behaviour under various voting procedures is observed.[4] The main problem with this research method is that it must be limited to small (and usually unrepresentative) samples, where the voters' preference orderings are induced artificially.

A third method is to conduct survey research in which a representative sample of voters are asked how they would vote under various procedures for a given

They collect data from various non-Plurality voting instances, estimate preferences under various assumptions, and run alternative voting rules.

In view of the deficiencies of these four methods, we must conclude that in the absence of mathematical proofs all four methods are complementary and should be employed – under various simplifying assumptions – in order to assess the degree of robustness of the results. If the same conclusion holds under various sets of assumptions and different research methods, confidence in the conclusion increases. Thus the present study begins by investigating

Simulations:

Felsenthal, Dan S., Zeev Maoz, and Amnon Rapoport. "The Condorcet-efficiency of sophisticated procedures." *Behavioral Science* 35.1 (1990): 24-33.

Forsythe, Robert, et al. "An experimental study of voting rules and polls in three-candidate elections." International Journal of Game Theory 25 (1996): 355-383.

Recap: No silver bullets

Hypothesis: “Borda leads to better outcomes than 2-Approval and STV”

Method
pros
cons

Felsenthal, Dan S., Zeev Maoz, and Amnon Rapoport. "An empirical evaluation of six voting procedures: do they really make any difference?." British Journal of Political Science 23.1 (1993): 1-27.

Recap: No silver bullets

Hypothesis: "Borda leads to better outcomes than 2-Approval and STV"

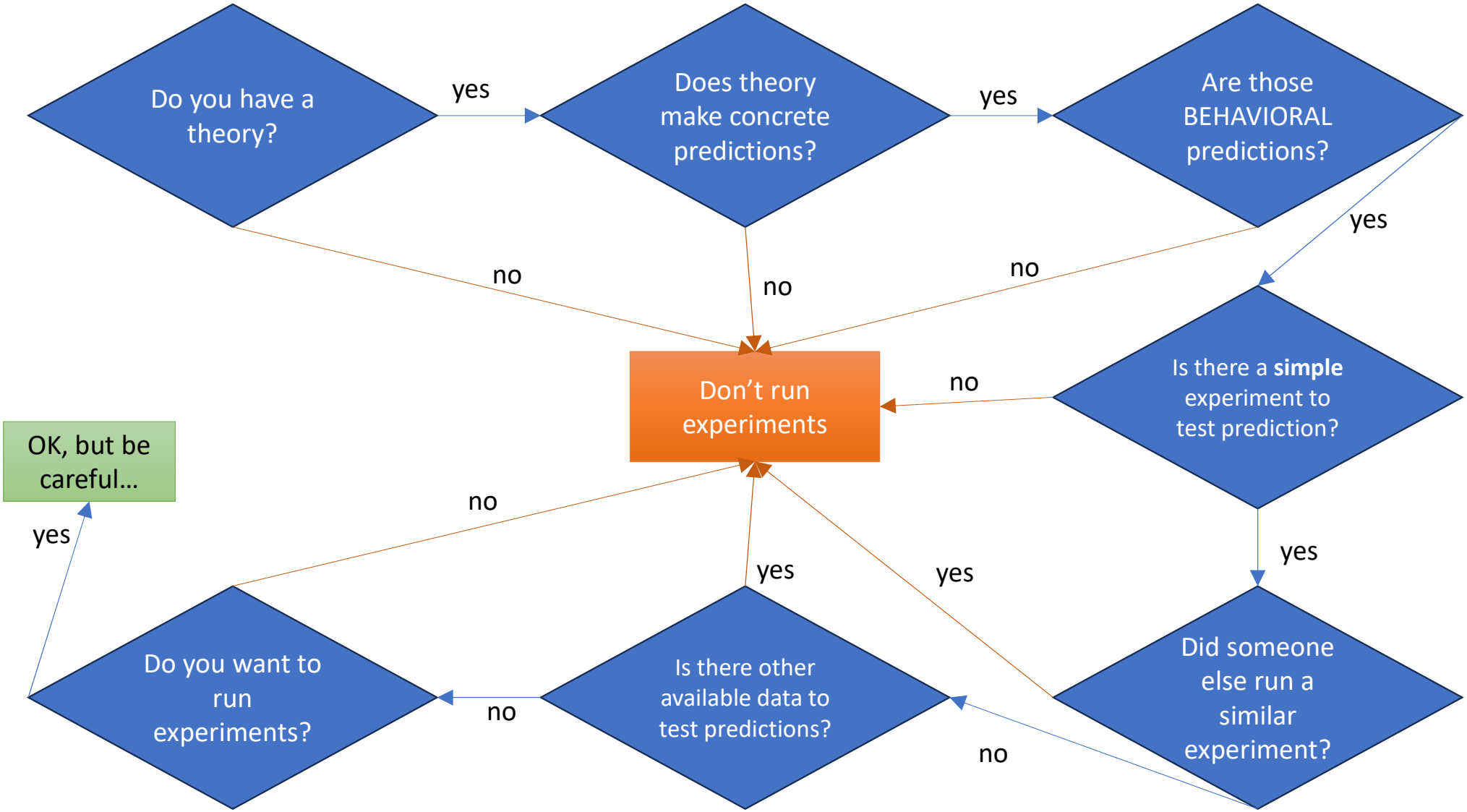
Method	Simulations	Lab experiment	Survey	Extrapolation
pros	Cheap, thorough	Control preferences	Real preferences	Real preferences
cons	Arbitrary assumptions on preferences and strategic behavior	Small, expensive, preferences arbitrary	Cannot trust hypothetical answers, cooperation	Lack of ranked data, Assume truthful vote

"… we must conclude that … all four methods are complementary and should be employed"

Felsenthal, Dan S., Zeev Maoz, and Amnon Rapoport. "An empirical evaluation of six voting procedures: do they really make any difference?." British Journal of Political Science 23.1 (1993): 1-27.

(some) Considerations in experiment design

- What are the treatments?
  - Between / within subjects
- Which conditions to control?
  - Between / within subjects
- Who are the subjects?
- What is the interface?
- What information subject get?
  - Truthful / deceitful
- The order of conditions
- How to set incentives?
- How to explain and/or test understanding?
- IRB
- …


Do you have a theory?
yes
Does theory make concrete predictions?
yes
Are those BEHAVIORAL predictions?
yes
no
no
no
Don't run experiments
no
Is there a **simple** experiment to test prediction?
yes
OK, but be careful…
yes
no
yes
yes
Do you want to run experiments?
no
Is there other available data to test predictions?
no
Did someone else run a similar experiment?