Incentive compatible experiment design

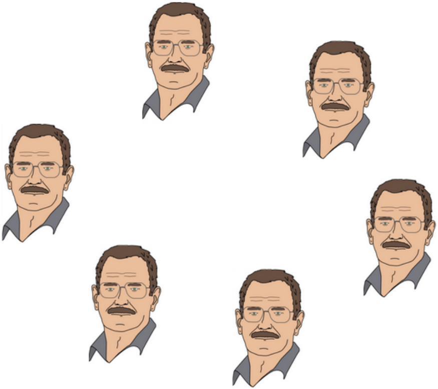
presented by
Panos Toulis ptoulis@fas.harvard.edu

Joint work with DC Parkes (Harvard), E Pfeffer (Pursway),
J Zou (MSR), G Gildor (Pursway)

October 10, 2014

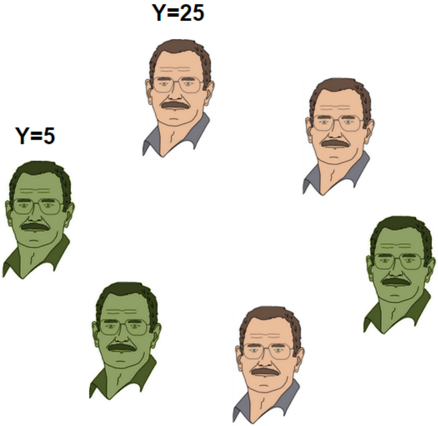
Experiment design

Agents in the experiment



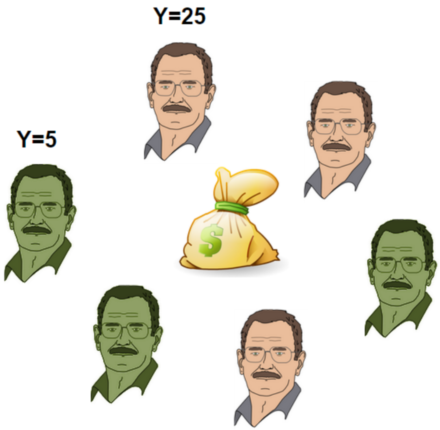
Experiment design

Assignment of treatment to avoid bias and minimize random errors



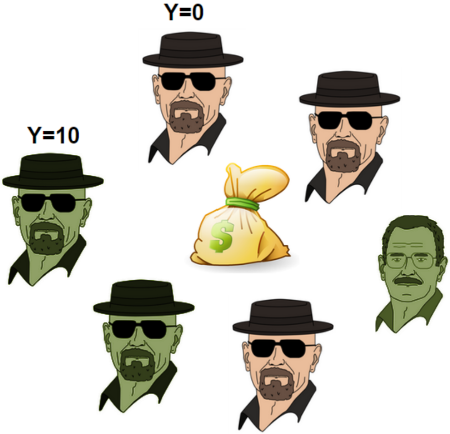
Experiment design

...introducing incentives



Experiment design

...introducing incentives may alter the experiment outcomes



Research agenda

Incentive compatible experiment design

- Problem: Strategic behavior by agents interferes with the experiment.
- Typical experiment design wishes to avoid systematic biases and the impact of random errors.
- The goal of **incentive compatible experiment design** is

  to design an experiment in a way to elicit "natural actions" from the agents i.e., the actions that agents would take in the absence of competition.

Research agenda

Application in viral marketing

- A company (*experimenter*) wishes to target a population with a new product through a marketing campaign. The initial target set is called the *seed set*.
- Two viral marketing companies (*agents*) claim knowledge of a hidden social network so as to better choose the seed set and maximize product adoption.
- Our research question is the following: "How to design an experiment that will decide which agent is better able to pick a seed set so as to maximize product adoption?"

The ideal experiment

- Ideally, the experimenter wants to test the two agents on two disjoint but otherwise *identical* populations.
- This is hard to achieve because
  (a) may be impossible to split the target population into two halves having no **interactions** between them,
  (b) the experimenter is unaware of the factors that are important to pick two identical (or even balanced) populations,
  (c) geography-specific, or other events (e.g. sports) that could affect the outcome are hard to control for.

Influence model

- Each seed set is associated with a parameter λ which is the rate of edges originating from that set to the rest of the network.
- Each customer i(=node in test set) has $n_i = \#$ incoming edges.
- The outcome $Y_i \in \{0, 1\}$ depends on n_i i.e., the actual #incoming edges to node i.
- The quality p_j of the agent j is a parameter that identifies those edges with noise (e.g., $N_{ij} \sim Binomial(n_i, p_j)$).
- Implications:
  - **(a)** Better agents can pick seed sets with higher intensities λ.
  - **(b)** Higher rates λ mean higher product adoption Y_i.
  - **(c)** Better agent quality p_j means better prediction of outcomes.
  - **(d)** Action space when agent j picks a seed is $(\lambda_{j1}, \lambda_{j2}) \in \mathbb{R}_+^2$.

Experiment designs

We consider the following operational constraints for all designs.

1. **Pick seed set.** Either the experimenter or the agents pick the seed set(s).
2. **Pick test set.** Agents pick at least one test set. The test set is used to measure performance of the agent.
3. **Outcomes.** Experimenter targets individuals in the seed set(s). Outcomes $Y_i \in \{0, 1\}$ are realized (product adoption from customers) in the test sets.
4. **Evaluation.** Total adoption from customers (in the test set) is the agent's score.

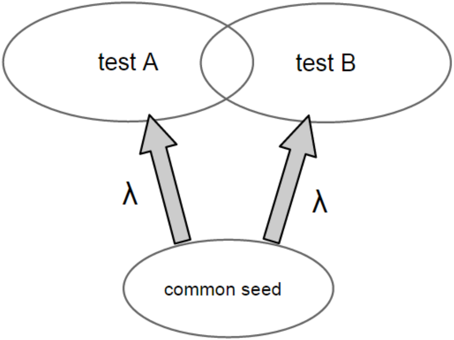
Experiment designs

We consider the following operational constraints for all designs.

1. **Pick seed set.** Either the experimenter or the agents pick the seed set(s).
2. **Pick test set.** Agents pick at least one test set. The test set is used to measure performance of the agent.
3. **Outcomes.** Experimenter targets individuals in the seed set(s). Outcomes $Y_i \in \{0, 1\}$ are realized (product adoption from customers) in the test sets.
4. **Evaluation.** Total adoption from customers (in the test set) is the agent's score.

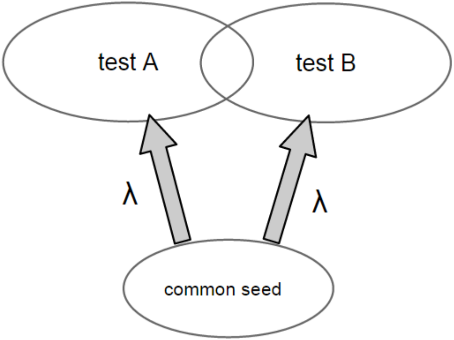
Note: We will focus on two agents, namely A and B. We will assume $p_A > p_B$ i.e., A is a better-quality agent.

Fixed-seed, One-test: M_0



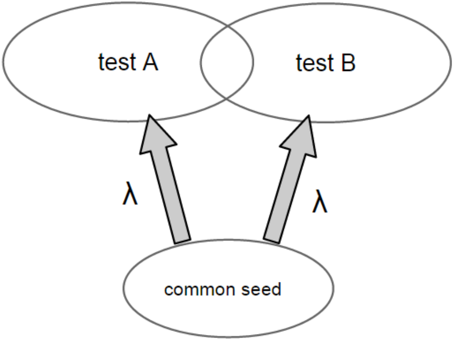
- The experimenter picks a common seed set.
- Each agent picks one test from the rest of the population (possibly overlapping with each other).
- Agent scores are calculated by measuring production adoption in their respective test sets.

Fixed-seed, One-test: M_0



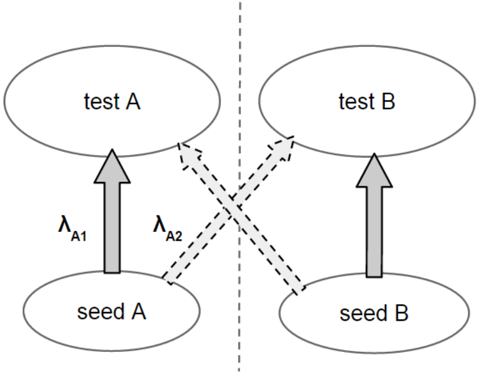
- **Result.** *It is a dominant strategy for an agent to pick the optimal test set in a straightforward way.*
- Straightforward play = maximize expected product adoption.
- Agents could opt for a more risky strategy but straightforward play is stochastically dominant.

Fixed-seed, One-test: M_0



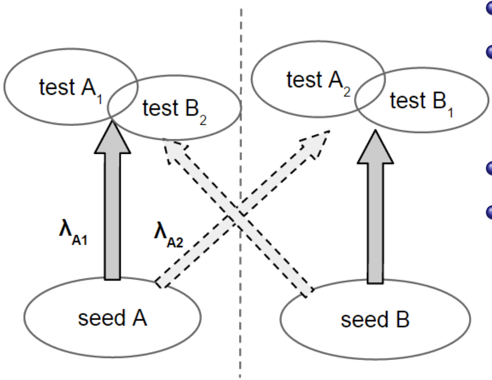
- **Statistical problem.** M_0 does not estimate the ability of the agent to use its knowledge to pick a seed set jointly with a test set.
- Experimenter has limited knowledge to pick a good seed set so the test has limited power.

Split, variable-seed, one-test: M_1



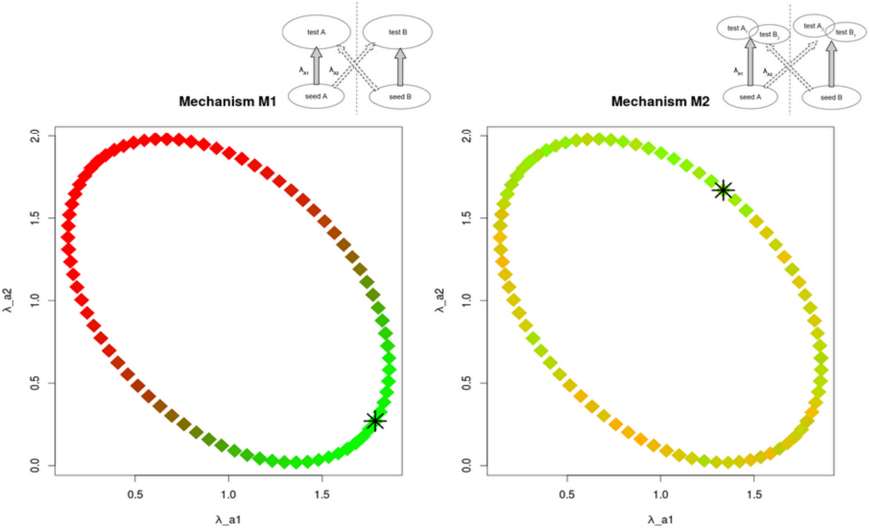
- Population split in half.
- In each half a agent picks a seed set and a test set.
- There is interference because a seed set on one side is contributing in the outcomes of the other side.
- An agent can *free-ride* on the other agent.

Split, variable-seed, two-test: M_2



- Population split in half.
- An agent picks a seed set and a test set on its half, and one test set on the other half.
- There is multiple interference.
- Mitigates incentive problems because the better agent can also "free-ride" on its own seed set.

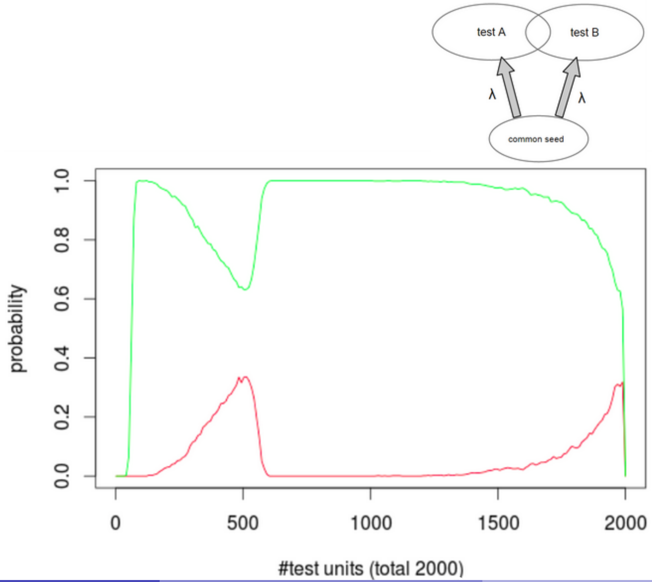
Comparison between M_1 and M_2



green $= P(\text{A wins})$, **ellipse** $=$ action space $(\lambda_{A1}, \lambda_{A2})$, ***** $=$ best response.

Goldfish paradox

Power of test does not (monotonically) increase with sample size



Summary & future work

- We introduce the problem of incentive compatible experiment design. Rich research problem: game theory for statistics, and statistics for game theory.
- Current work: develop a game-theoretic analysis for M_1 and M_2.
- Thanks to Pursway Inc., we are currently working on a dataset in viral marketing.
- Long-term: analyze classical experiment designs (blocking, factorial) in our incentive compatible design framework.