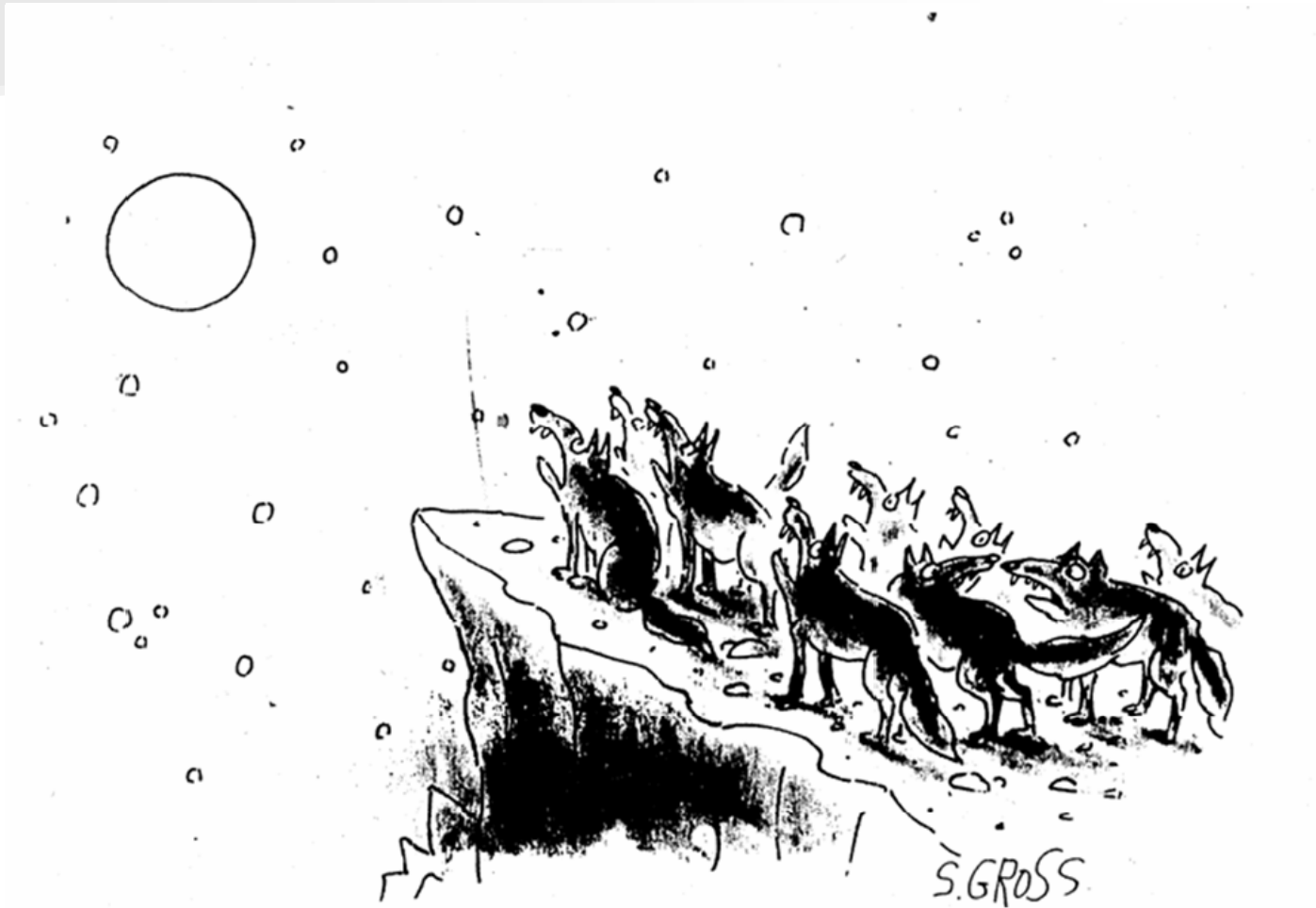


# The Case for Impact Evaluation: Working with a Private Bank in India

Innovations in Evaluations for Development Assistance  
IFC, New Delhi  
December 11th, 2007

Professor Antoinette Schoar, MIT and NBER

# One View of Impact Evaluation



*My question is: Are we making an impact?*

# My Take on Evaluation

- Question behind any impact evaluation:
  - Did a particular program work?
  - What was its impact?
- Benefits of evaluation:
  - The stakes are too high to screw up!
  - Given limited development resources, projects should have the highest cost-benefit ratio
  - Not only find out what works, but quantify the impact
  - Donor organizations will be more willing to provide funding if there is proven impact

# Basic Problem of Impact Evaluation

- What would the world be absent the program?
- How do we compare a world in which a program happens to a hypothetical world in which it doesn't?
  - We usually cannot create a world in which the program never happened
  - So how do we measure this “Counterfactual”?
- Every impact evaluation establishes a counterfactual
  - We can try to identify group with “similar” shocks.
  - And measure their outcomes.

# Outline of the Argument

- **Argument 1**

Randomized evaluations are most robust way to measure impact: prevents bias in selection and subjective evaluations

- **Argument 2**

Randomized evaluations are far more feasible than people think

- **Argument 3**

Many of the problems of randomized evaluations are problems of *any* impact evaluation

# Wason Selection Task

- Hypothesis: If a card has a vowel on one side, then it has an even number on the other side
  - What do you turn over?

A

D

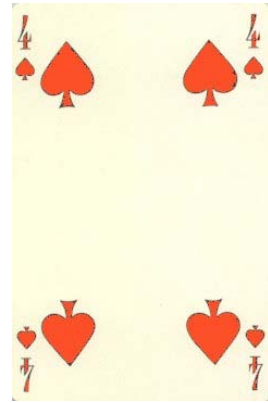
4

7

# Modal Answers

- A and 4 (approx. 50%)
- A (approx. 30%)
- Few chose A and 7

# What's going on here?



- Confirmatory bias
  - Large body of evidence
  - Well established tendency for people to seek information confirming a belief
  
- Beliefs can persevere in the face of contradictory evidence
  - Look at the card..closely.

# Stark Example

- Ancient medicines
  - Devised through experimentation and experience
  - Some cases (e.g. Chinese, Roman and Indian medicines), a written knowledge was carried forward.
- Most ideas that persisted were ineffective
  - Several were downright harmful
  - For example, “rub dirt on a wound”, “cool the heat of a fever by dousing a person in material”

# What does this have to do with evaluation?

- Basic problem of evaluation: I can “see” the program works.
  - Yet we see what we want to believe
  - Magnified by basic desire to see programs work
- In short: we are over-optimistic about which programs work
  - Contrast optimism about one's *own program* against overall pessimism about unresolved challenges
  - Note: This kind of optimism is rampant in every walk of life. It is needed to function. But it needs to be tempered when allocating decisions

# Fundamental Problem of Evaluation

- People/Firms differ on many *unobservable* characteristics
  - How good a comparison group can “matching” make?
- Selection problems
  - Programs select participants on unobservable characteristics such as motivation/predisposition
  - More motivated people self-select into program
- Confirmatory bias
  - Psychology evidence: we see what we want to believe
  - Basic desire to see programs work
  - Salience of positive outcomes

# Traditional Evaluation Methods

- Before-After Evaluation
  - Changes in outcomes of the treated group
  - Problem: no comparison group, how do we know if effect is due to program
- Matching on Observables
  - Comparison group of untreated “similar” units
  - Problem: Difficult to find perfect matching group
- Difference-in-Difference
  - Change in outcomes *relative* to comparison group
  - Problem: Treated clients might be on different trajectory

# Outline of Impact Evaluation

- Collect baseline data: Necessary in all impact evaluations
- Implement program in control and treatment
  - Key point: Resources are always limited. Allocating them randomly is not withholding resources
  - Lotteries: Considered fair when limited resources
  - Expansion strategies: Indifference between some sets of units
  - Advertise to increase take-up
- Collect follow-up data on relevant outcomes

# Challenges of Working with Private sector

- Private sector firms need to care about the bottom line, even when introducing products to the poor
- Research can bring resources to questions that would otherwise not be answered
- Useful creative tension

# Advantages and Disadvantages

## ■ Bottom Up

- Maximally use local knowledge
- Allows flowering of creativity
- ICICI in excellent position to do this

## ■ Top Down

- Maximally use generalized knowledge
  - Center is in better position to see latest ideas, examples of successful programs elsewhere, etc.
- IFC in excellent position to do this

# Example: Work with ICICI

- Work with field staff and do field work to identify core problems
- Bring research insights to bear
- Dialogue with field staff
  - We bring research insights
  - They bring on the ground knowledge
  - Collaborative synthesis results in novel designs
- Test these designs against each other

# Product Market Impact of Financial Constraints

- How do financial constraints affect product market behavior of firms?
  - Do constraint firms bid less aggressively in the supplier markets?
  - Offer higher/lower prices to downstream customers?
  - Volume effects?
  - Interaction of constraint and unconstraint firms
- **Challenge:** Want to test impact on individual firm behavior *and* market level effects

# Market for Used Assets in India

- Ebay India auctions off used and repossessed vehicles on a designated auction site several times a month
  - Registered car dealers can log on and participate in the auctions
  - Dealers are strictly segregated by cities
- Current market pricing
  - Used cars that were acquired from auction sell on average at 30%-40% mark up
  - Convergence of sales on auction site is on average only 70% of vehicles in the first auction

# Product Market Experiment

- Bank will provide OD facility to a random subset of dealers in each city
  - Bank will preselect the pool of dealers that can qualify for OD
  - OD will be tied to specific vehicles: Compare bidding behavior within dealers when the credit constraint is changed
- Measure the impact on individual dealers' bidding behavior and overall market pricing in each auction

# Suggestion

- Create an innovation infrastructure
  - Evaluation component already discussed
  - Idea generation component key
  
- Innovation infrastructure
  - Encourage bottom-up design through contests
  - *Most importantly* devote central resources to a few careful partnered innovation

# Key Constraints to Randomization

- Operational: Should not interfere significantly with program operations
  - Randomization is a decision that must be made before-hand
    - Incredibly valuable discipline
  - Outcomes must be specified before-hand
    - Incredibly valuable discipline
    - Prevents hindsight bias
    - Not intended to inhibit learning or unexpected outcomes. But forces us to recognize what was unexpected and what was expected
- Moral:
  - Should be perceived as unfair
  - Must be politically feasible
  - Must be ethical: no-one should be made worse off ex ante

# When Is It Not Feasible?

- Large treated units
  - An entire country changes policy
  - Macro economic policy changes
- Few treated units
  - Very few treated units (e.g. a philanthropist gives 3 big donations to 3 charities)
  - Very large individual units: costs from randomization is too big to justify, e.g. corporate loans
- Treated units are very particular
  - Sometimes the treated units are precisely chosen for their unobservable characteristics.
  - Very difficult to find measurable outcomes, e.g. a VC investment

# Problems of Impact Evaluation

- External Validity
  - Always a problem in any evaluation. It is not more of a problem here.
- Generalizability/Scalability
  - How does having a worse control group solve this problem?
- Models needed to evaluate/parse data
  - True. Models will only perform better the more accurate are the inputs.

# Conclusion

- We should always start by asking: “Why are we *not* doing a randomized evaluation here?”
  - Should be the default
- Randomized evaluation gives the most quantitative and robust results
  - Provides good discipline for evaluation
  - Is flexible enough to incorporate new knowledge and insights
- There are some good arguments why in some instances randomized evaluation are not possible, but we should carefully evaluate these

# Take Away

- If properly designed and conducted, experiments provide the **most credible** assessment of the impact of a program
- Results from experiments are easy to understand and much less subject to methodological objections
- Credibility + Ease => More likely to convince policymakers and funders of effectiveness of a program