Persuasion Bias in Science: An Experiment

Arianna Degan (University of Quebec at Montréal, CIRPÉE)
Ming Li (Concordia University, CIREQ)
Huan Xie (Concordia University, CIREQ, CIRANO)

Workshop on the economics of strategic communication and persuasion: Application to ethics and incentives in scientific research
October 28, 2017

Questions and Motivations

- Investigate the impact of conflicts of interests between researchers and evaluators
- Asymmetric information between researchers and evaluators
- Game theoretical model not replying on reputation or social preference
    - Do researchers have incentives to cheat?
    - Can evaluators predict the bias and correct their evaluation accordingly?
    - What are the impacts on welfare?

Literature

- Blume, Lai and Lim (2017): Survey of experiments and theoretical foundations on strategic information transmission
- Experimental studies on persuasion
- Our experiment is based on a simplified model of Selective Sampling in Di Tillio, Ottaviani and Sørensen (2017a)

Model: Di Tillio, Ottaviani and Sørensen (2017)

- Use a game-theoretical framework to model randomized controlled trial (RCT)
- Three cases of possible manipulation by researchers
  - Selective sampling: non-randomly select sample ⇒ undermine the external validity
  - Selective assignment: non-randomly assign subjects into treatment ⇒ undermine the internal validity
  - Selective reporting ⇒ challenge both internal and external validity

Model: Basic Elements

- Two risk-neutral players: Researcher and Evaluator
- Researcher sets up an experiment.
- Evaluator observes the experiment outcome and decides whether to grant Researcher a desired acceptance (e.g., a funding award or a journal publication).
- The aim of the experiment is to estimate the effect of a treatment (e.g., by a new drug or a new policy).
- Evaluator only grants acceptance if the average treatment effect is strong enough.
- Researcher always benefits from acceptance.

Model: Treatment Effects

- The experiment can be conducted in one of two locations: Left or Right.
- Population is equally divided between the two locations.
- For simplicity, assume all individuals in one location have the same treatment effect: $\beta_L, \beta_R \in \{0, 1\}$
- β_L, β_R are i.i.d. across locations:
  $\Pr(\beta_L = 1) = \Pr(\beta_R = 1) = q$
  $\Pr(\beta_L = 0) = \Pr(\beta_R = 0) = 1 - q$
- Average Treatment Effect for the entire population:
  $\beta_{ATE} = (\beta_L + \beta_R)/2$

Model: Experiment Outcome/Evidence

- Location where the experiment is conducted: $t = L, R$
- Baseline experiment outcome: 0
- Experiment outcome under treatment conducted at location t: $v = \beta_t$
- From previous assumption β_L, β_R are i.i.d.
  - $\Pr(v = 1) = q$
  - $\Pr(v = 0) = 1 - q$
- Evaluator only observes the experiment outcome under treatment v, but not the location t where the experiment is conducted.

Model: Timing of the Game

- Non-manipulation
  - Both players observe the Evaluator's cost of acceptance k.
  - Researcher selects one location $t \in \{L, R\}$ to conduct the experiment.
  - Evaluator chooses to accept or reject after observing the experiment outcome v.

Model: Timing of the Game

- Non-manipulation
  - Both players observe the Evaluator's cost of acceptance k.
  - Researcher selects one location $t \in \{L, R\}$ to conduct the experiment.
  - Evaluator chooses to accept or reject after observing the experiment outcome v.
- Manipulation
  - Both players observe the Evaluator's cost of acceptance k.
  - Researcher observes the treatment effect in one location, β_A, $A \in \{L, R\}$.
  - Researcher selects one location $t \in \{L, R\}$ to conduct the experiment.
  - Evaluator chooses to accept or reject after observing the experiment evidence v.

Model: Payoffs

- Researcher's payoff:
    - 1 if acceptance is granted
    - 0 otherwise

Model: Payoffs

- Researcher's payoff:
  - 1 if acceptance is granted
  - 0 otherwise
- Evaluator's expected payoff:
  - $E(\beta_{ATE}|v) - k$ if acceptance granted
    - $E(\beta_{ATE}|v)$: posterior expectation of the average treatment effect after observing experiment outcome v
  - 0 otherwise

Model: Payoffs

- Researcher's payoff:
  - 1 if acceptance is granted
  - 0 otherwise
- Evaluator's expected payoff:
  - $E(\beta_{ATE}|v) - k$ if acceptance granted
    - $E(\beta_{ATE}|v)$: posterior expectation of the average treatment effect after observing experiment outcome v
  - 0 otherwise
- Evaluator's best response:
  - accept if $E(\beta_{ATE}|v) \geq k$
  - reject otherwise

Non-manipulation Benchmark

- Researcher: choose one location randomly
- Evaluator's inferences:
  - $v = 0$
    $\Rightarrow \beta_t = 0$ and $\beta_{-t} \in \{0, 1\}$
    $\Rightarrow E(\beta_{ATE} | v = 0) = q/2$
  - $v = 1$
    $\Rightarrow \beta_t = 1$ and $\beta_{-t} \in \{0, 1\}$
    $\Rightarrow E(\beta_{ATE} | v = 1) = (1 + q)/2$

Manipulation (Selective Sampling)

- Researcher's equilibrium strategy (Intuitive Strategy):
  - If $\beta_A = 1$, choose $t = A$.
  - If $\beta_A = 0$, choose $t = -A$.

Manipulation (Selective Sampling)

- Researcher's equilibrium strategy (Intuitive Strategy):
  - If $\beta_A = 1$, choose $t = A$.
  - If $\beta_A = 0$, choose $t = -A$.
- The Evaluator's inferences given the Intuitive Strategy:
  - $v = 0$
    $\Rightarrow \beta_t = 0$, and $\beta_{-t} = \beta_A = 0$
    $\Rightarrow E(\beta_{ATE} | v = 0) = 0$

Manipulation (Selective Sampling)

- Researcher's equilibrium strategy (Intuitive Strategy):
  - If $\beta_A = 1$, choose $t = A$.
  - If $\beta_A = 0$, choose $t = -A$.
- The Evaluator's inferences given the Intuitive Strategy:
  - $v = 0$
    $\Rightarrow \beta_t = 0$, and $\beta_{-t} = \beta_A = 0$
    $\Rightarrow E(\beta_{ATE} | v = 0) = 0$
  - $v = 1$
    $\Rightarrow$ case 1: $\beta_t = \beta_A = 1$ and $\beta_{-t} \in \{0, 1\}$ (w.p. q)
    case 2: $\beta_{-t} = \beta_A = 0$ and $\beta_t = 1$ (w.p. $q(1 - q)$)
    $\Rightarrow E(\beta_{ATE} | v = 1) = 1/(2 - q)$

Effect of Manipulation

	Non-manipulation		Manipulation	
	$E(\beta_{ATE}\|\cdot)$	w. p.	$E(\beta_{ATE}\|\cdot)$	w. p.
$v=0$	$q/2$	$1-q$	0	$(1-q)^2$
$v=1$	$(1+q)/2$	q	$1/(2-q)$	$q(2-q)$

- Increase the probability of positive experiment outcome
- Decrease conditional expectation of ATE, $E(\beta_{ATE}|\cdot)$
- Therefore, the effect of manipulation on players' welfare is NOT monotonic.
- If Evaluator is naive, $E(\beta_{ATE}|\cdot)$ under manipulation same as non-manipulation $\Rightarrow$ Researcher's welfare will improve

Equilibrium when $q = 1/2$

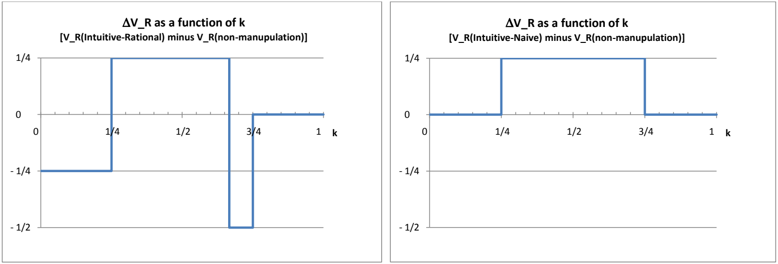
Evaluator's BR under Non-manipulation

	$k \leq 0.25$	$0.25 < k \leq 0.75$	$k > 0.75$
$v = 0$	accept	reject	reject
$v = 1$	accept	accept	reject

Evaluator's BR under Manipulation

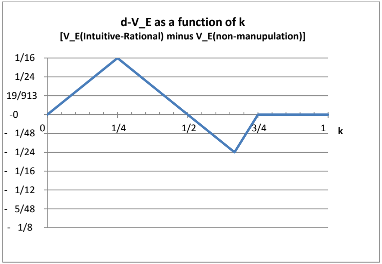
	$k \leq 0.67$	$k > 0.67$
$v = 0$	reject	reject
$v = 1$	accept	reject

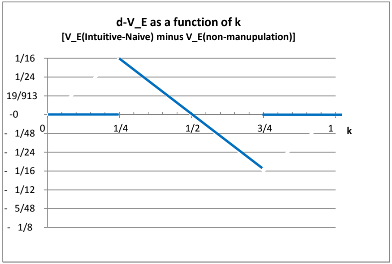
Welfare Analysis: Researcher



- Researcher's expected payoff under manipulation minus that under non-manipulation, as a function of k
- Left panel: rational Evaluator
- Right panel: naive Evaluator

Welfare Analysis: Evaluator





- Evaluator's expected payoff under manipulation minus that under non-manipulation, as a function of k
- Left panel: rational Evaluator
- Right panel: naive Evaluator

Parameterization

- The probability of positive treatment effect in each location: $q = 0.5$
- Under manipulation, the probability that Researcher observes private information from each location: $m = 0.5$
  - Evaluator is not informed of the experiment location $\Rightarrow$ The value of m does not affect players' decision.
  - The value of m is not explicitly told to subjects.
- Payoffs and cost of acceptance multiplied by 100
- k =10, or 40, or 70
  - In theory k is revealed to both Researcher and Evaluator.
  - We choose to test the theory given several fixed k values rather than drawing k from a distribution every round.

Parameterization cont'd

- The values of k are chosen to satisfy the following predictions:

		$k_1 = 10$	$k_2 = 40$	$k_3 = 70$
$v = 0$	Manipulation	**reject**	reject	reject
	Non-Manipulation	**accept**	reject	reject
$v = 1$	Manipulation	accept	accept	**reject**
	Non-Manipulation	accept	accept	**accept**

- The predictions not only hold for risk-neutral Evaluators, but also hold for risk-aversive Evaluators who have CRRA utility function u^r with $r = 0.5$.

Experimental Design

- Treatments: Non-manipulation vs. Manipulation
- Within-subject design
- 30 rounds under Non-manipulation, followed by 30 rounds under Manipulation

  $\Rightarrow$ We choose this order for subjects to learn first in a simpler environment
- Instructions for Manipulation treatment only distributed upon the time to play
- 3 practice rounds before each treatment starts
- 12 subjects each session, 6 Researchers and 6 Evaluators, without changing player roles
- Each round Researchers and Evaluators randomly and anonymously paired with each other

Assignment of k to Evaluators

- Each Evaluator experiences all three k values.
- In order to facilitate learning, in each treatment, each Evaluator experiences the same k value for 10 consecutive rounds, called a block.
- Evaluators randomly assigned to three cohorts. In each treatment
    - Cohort 1: k_1 block, followed by k_2 block, followed by k_3 block
    - Cohort 2: k_2 block, followed by k_3 block, followed by k_1 block
    - Cohort 3: k_3 block, followed by k_1 block, followed by k_2 block
- Given random matching, in each round Researchers always face the same distribution of k.

Implementation of the Game in a Round

Game environment:

- There are 50 balls in the Left Bin and 50 balls in the Right Bin.
- All balls in the same bin are of the same color.
- In each bin, the color of the balls is Red w.p. 50% and Blue w.p. 50%.
- Red balls have a value of 1 point and Blue balls have no value.

Implementation of the Game in a Round Cont'd

Game in the round:

- Both players observe k for the round. (k is described as Player B's endowed income.)
- If in the Manipulation treatment, Player A receives a private message about the color of the balls in one bin.
- Player A chooses one bin, Left or Right.
- The color of the balls in the chosen bin is shown to both players.
- Player B chooses whether to choose Implement the project.
    - If yes, Player B receives the value of the project, which equals the total number of red balls in the two bins, but has to give up the endowed income k. Player A receives 100 points.
    - If no, Player B receives k points. Player A receives nothing.

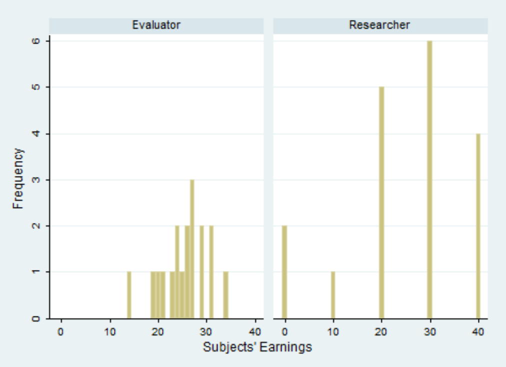
Payment

- At the end of the experiment, 2 rounds in each treatment are chosen for actual payment. In total, 4 rounds are paid.
- In every round, subjects are shown the history of play and previous payoffs from each round in that treatment.
- Points are converted to Canadian dollar at 10 points=$1.
- Show-up fee: $10
- If in the end, subjects' total earning, including show-up fee, is less than $15, then they receive $15.

Sessions

- We conducted 1 pilot and 3 sessions so far
- Total 14+36=50 subjects
- Results reported here use data from the 3 sessions
- Treat each individual as an independent observation
- Experiment conducted at CIRANO in Montreal, Canada

Earning Distributions by Type



- Average earnings excluding show-up fee: \$25.19
- Researchers: Avg. \$25, Min \$0, Max \$40
- Evaluators: Avg. \$25.39, Min \$14, Max \$34
- Wilcoxon Mann-Whitney test: $p = 0.69$, 36 obs.

Researchers' Behavior

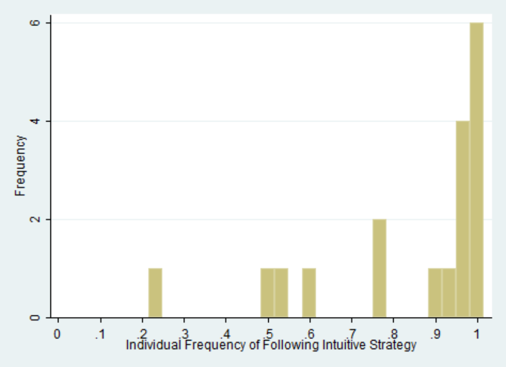
Frequency of choosing the Left Bin:

- Non-manipulation: 47.6%; Manipulation: 52.6%
- Matched-pair signed-rank test: $p = 0.5$, 18 obs.

Researchers' Behavior Cont'd

Researchers' frequency of following the Intuitive Strategy in the Manipulation treatment

- Avg. frequency 83.9%
- Distribution of the frequency of individuals



Evaluators' Behavior: Freq. of Implement

Non-manipulation (Part One)									
	$k = 10$			$k = 40$			$k = 70$		
v	Data		p	Data		p	Data		p
Red	0.905	1	0.046	0.893	1	0.046	0.537	1	*0.001*
Blue	0.612	1	*0.001*	0.302	0	*0.003*	0.071	0	0.026
Avg.	0.767			0.578			0.317		
Manipulation (Part Two)									
	$k = 10$			$k = 40$			$k = 70$		
v	Data		p	Data		p	Data		p
Red	0.921	1	0.084	0.896	1	0.084	0.443	0	*0.000*
Blue	0.415	0	*0.002*	0.091	0	0.084	0.086	0	0.084
Avg.	0.772			0.650			0.328		

Tests on Freq. of Implement

Model Prediction

v		$k_1 = 10$	$k_2 = 40$	$k_3 = 70$
Blue	Manipulation	**reject**	reject	reject
	Non-Manipulation	**accept**	reject	reject
Red	Manipulation	accept	accept	**reject**
	Non-Manipulation	accept	accept	**accept**

p-value for two-tailed matched-pair Signed Rank Tests (18 obs.)

	$k = 10$	$k = 40$	$k = 70$
Red vs. Blue (non-manipulation)	*0.003*	0.000	0.002
Red vs. Blue (Manipulation)	0.002	0.000	*0.002*
Non-manipulation vs. Manipulation (Red)	0.979	0.968	*0.184*
Non-manipulation vs. Manipulation (Blue)	*0.274*	*0.036*	0.547

Welfare Analysis for Evaluators

Non-manipulation (Part One)									
	$k = 10$			$k = 40$			$k = 70$		
v	U	U_2	p	U	U_2	p	U	U_2	p
Red	67.8	72.6	0.046	71.5	74.4	0.046	72.9	75.8	0.091
Blue	21.5	24.1	0.093	34.7	40	0.025	67.4	70	0.026
Avg.	45.9	49.7	0.017	51.9	56.1	0.005	70.3	73.1	0.004
Manipulation (Part Two)									
	$k = 10$			$k = 40$			$k = 70$		
v	U	U_2	p	U	U_2	p	U	U_2	p
Red	65.7	70.5	0.084	65.0	67.2	0.084	70.2	70	**0.930**
Blue	8.7	10	0.083	38.2	40	**0.541**	64.0	70	0.084
Avg.	48.9	52.7	0.019	56.8	58.9	0.079	68.2	70	0.510

Discussion

- Welfare analysis:
  - Comparison between non-manipulation and manipulation
  - Welfare for Researchers

Discussion

- Welfare analysis:
  - Comparison between non-manipulation and manipulation
  - Welfare for Researchers
- Evaluators' behavior:
  - Risk aversion alone cannot explain all the deviations from predictions
  - Maybe related to subjects' ability of Bayesian updating

Discussion

- Welfare analysis:
  - Comparison between non-manipulation and manipulation
  - Welfare for Researchers
- Evaluators' behavior:
  - Risk aversion alone cannot explain all the deviations from predictions
  - Maybe related to subjects' ability of Bayesian updating
- Other treatments:
  - Add a pre-stage where Researchers can choose whether to conduct the experiment: no welfare improvement for Evaluator under manipulation