Persuasion Bias in Science: An Experiment

Paper by Arianna Degan, Ming Li and Huan Xie

Discussion by Christopher Cotton (Queen's University)

Overview of paper

- Simple game of strategic sample selection
- Researcher draws one piece of evidence and then decides whether to report that evidence or to draw a second piece of evidence to report
- Evaluator sees reports, updates beliefs and decides whether to support the research / fund the project

- Test of the Theory using an Experiment

Overview of paper

Theory

- Researcher should report initially good results, and keep searching when initial results are bad
- Evaluator should account for this when choosing whether to accept the research
- Ability to manipulate can improve payoffs for sophisticated evaluator

Experiment

- Researcher often but not always plays the intuitive strategy
- Evaluator doesn't always behave in a sophisticated way
- Expected payoffs don’t appear to be improved by the presence of manipulation

Possible stories

Not a story in which the researcher and evaluator both value learning the truth

Researcher cares about funding, not doing quality research

- Information collection is a pilot study. But, researcher for some reason doesn't care about whether his research is worthwhile, so long as he gets funding

Evaluator values only certain results, not learning

- Journal only interested in publishing surprising results. Organization only interested in funding research that supports their agenda.

Strategic sample selection

- In which university's lab should we run our experiment?
- What sort of recruitment procedure should we follow?
- Survey locations. Time of day to collect data.
- Health history of those to include in the sample for a health study.

Does this happen? Are researchers transparent about it?

Must not be able to sample from both locations

If producing evidence from both locations is sufficiently inexpensive, then there will be an equilibrium in which failure to produce evidence from both locations is seen as evidence that bad evidence was found.

Journal submission, robustness checks, and requests to run additional treatments.

Coarse evidence production technology

Researcher has only limited control over evidence search

If the researcher had more precise control over evidence production, there may not be cases where the manipulation of evidence leads to the evaluator being better off.

How sure are we about the p-values?

Some questions about the p-values

Sample size of 18 and often getting very precise p-values of 0.001, 0.002, 0.003...

How do you calculate a p-value for the following difference:

- Theoretical prediction of frequency: 100% of the time.
- Observed frequency: 90.5% of the time.
- p-value for difference of 0.046

Non-manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.905	1	0.893	1	0.537	1
$v = 0$	0.612	1	0.302	0	0.071	0

Manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.921	1	0.896	1	0.443	0
$v = 0$	0.415	0	0.091	0	0.086	0

Non-manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	**0.905**	1	**0.893**	1	0.537	1
$v = 0$	0.612	1	0.302	0	0.071	0

Manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	**0.921**	1	**0.896**	1	0.443	0
$v = 0$	0.415	0	0.091	0	0.086	0

Observation 1: About 10% of people are really confused. They reject even when they've seen evidence that guarantees it is a good decision.

Non-manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.905	1	0.893	1	0.537	1
$v = 0$	0.612	1	0.302	0	0.071	0

Manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.921	1	0.896	1	**0.443**	0
$v = 0$	**0.415**	0	0.091	0	0.086	0

Observation 2: Some evidence suggests that 40% of people are naive. They accept where a sophisticated person should reject.

Non-manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.905	1	0.893	1	0.537	1
$v = 0$	0.612	1	0.302	0	0.071	0

Manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.921	1	0.896	1	**0.443**	0
$v = 0$	**0.415**	0	0.091	0	0.086	0

But, are they naïve, confused or uncertain?

Non-manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.905	1	0.893	1	**0.537**	1
$v = 0$	**0.612**	1	**0.302**	0	0.071	0

Manipulation

	K = 10		K = 40		K = 70	
	Data	Theory	Data	Theory	Data	Theory
$v = 1$	0.921	1	0.896	1	**0.443**	0
$v = 0$	**0.415**	0	0.091	0	0.086	0

Similar percentages of people are screwing up in other places where the computation isn't straightforward

Alternatives to Naïve Evaluators

1. Perhaps people just randomize when they are unsure what to do, confused, or uncertain.

   Play seemingly consistent with 20% of people flipping a coin when correct strategy is straightforward, and 80% of people flipping a coin when correct strategy is difficult to calculate

   Does the data show any evidence that the players are randomizing, or are they consistent in their choices across rounds?

Alternatives to Naïve Evaluators

2. Heterogeneity in risk preferences and other-regarding preferences

   In some cases, the evaluator must choose whether to give up 10 in expected payoffs to increase other player's payoff by 100.

   There are big social welfare gains from unselfish behavior. What do we know about people's willingness to make this tradeoff in similar environments?

Alternatives to Naïve Evaluators

3. Differences in endowment effects and loss aversion

   Evaluators were endowed with an amount, that they then had to give up to implement the project. This framing could have a significant impact on behavior.

   The loss of 25 to take a risk on a potential gain of 25… Some people might see that loss of 25 as much more costly than the similar gain of 25.

Alternatives to Naïve Evaluators

4. Players may know they should update beliefs but may make mistakes in systematic ways

   Putting too much weight on prior… overweighting new evidence … not understanding small probabilities … thinking that a red ball in one draw makes it more likely that a different color ball is drawn next… and so on …