Methodological Advances in Measuring the Effectiveness of Behavioral Nudges on Participation in Agri-Environmental Programs

Paul J. Ferraro

Johns Hopkins University

CBEAR

USDA, Washington DC

4 April 2018

[also broadcast and recorded via Zoom]

Informed by Behavioral Science · CBEAR · Tested with Experimental Design

The Center for Behavioral and Experimental Agri-Environmental Research

Evidence for Action · epic · Action for Evidence

Environmental Program Innovations Collaborative

Johns Hopkins Bloomberg School of Public Health

Johns Hopkins Carey Business School

Johns Hopkins Whiting School of Engineering

Informed by Behavioral Science
CBEAR
Tested with Experimental Design
The Center for Behavioral and Experimental Agri-Environmental Research

USDA United States Department of Agriculture
Economic Research Service

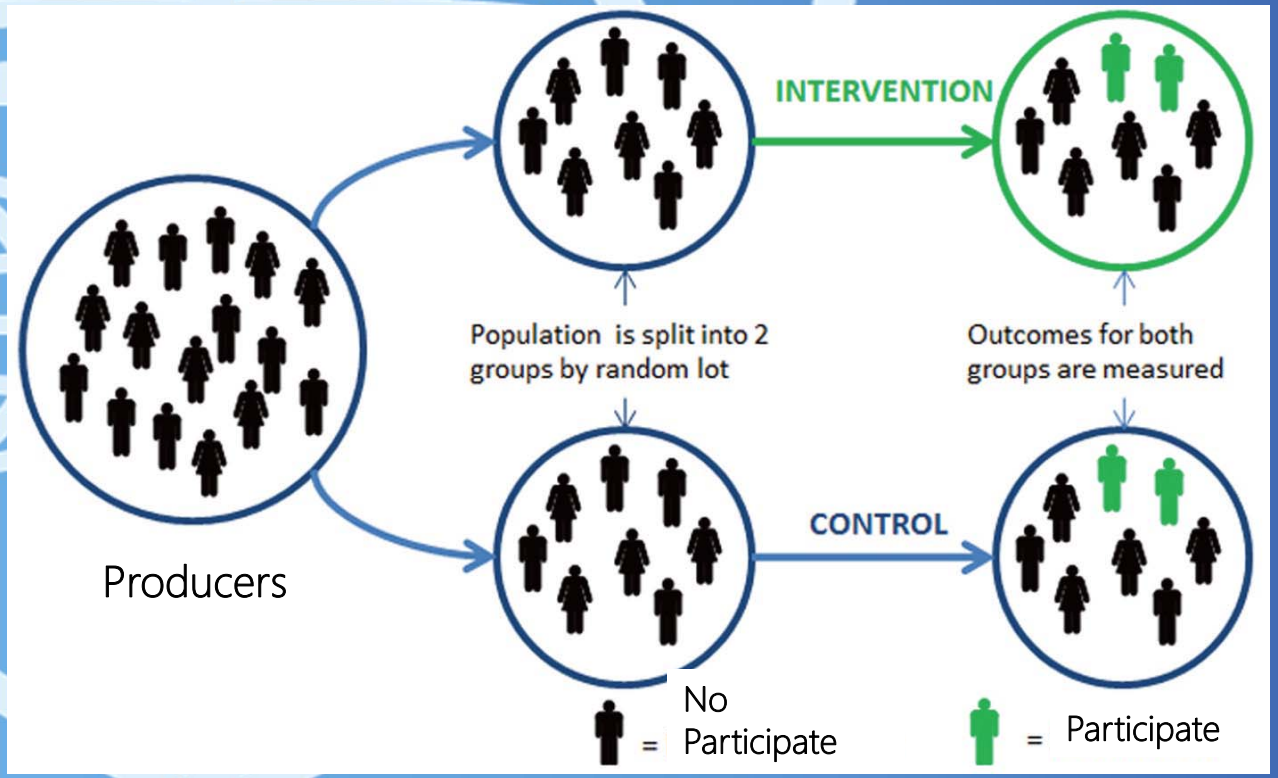
USDA United States Department of Agriculture
National Institute of Food and Agriculture

WILLIAM PENN FOUNDATION

Bring insights from the behavioral sciences to agri-environmental programs

Create a culture of experimentation in agri-environmental programs

USDA runs 1000s of uncontrolled experiments every year.



Experimental Designs Make Learning Easier

Non-operator Landowners and Soil Health

Non-operator Landowner

+

Operator

=

Soil Health

Counties with highest rented land and nitrogen pollution

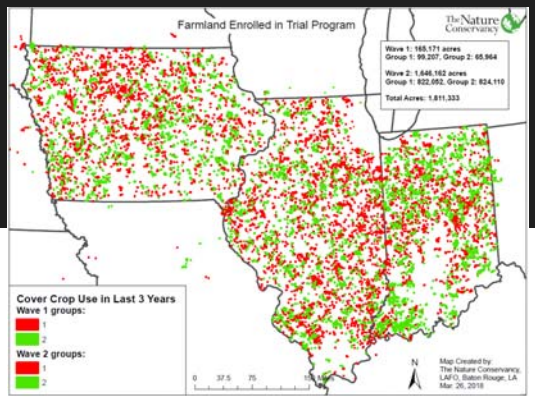


Photo credits: wfan.org, nrcs.usda.org, farm3.staticflicker.com

The Nature Conservancy
Protecting nature. Preserving life.®

PURDUE UNIVERSITY

Informed by Behavioral Science
CBEAR
Tested with Experimental Design
The Center for Behavioral and Experimental Agri-Environmental Research

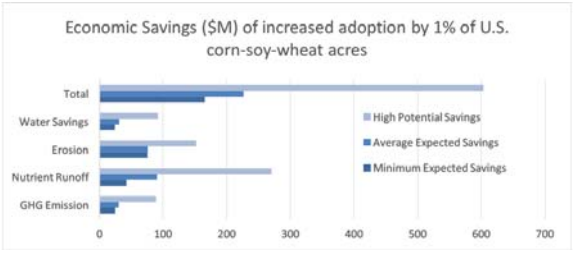
Implement trial program

Randomized controlled trial of incentives and nudges targeting barriers.

Testing ways to overcome barriers to soil health and cover crops on rented lands by providing:

- **A: Example lease language** requiring cover crops and specifying how they will be paid for (e.g., cost-share reduced rental rate)
- **B: Financial incentive** to motivate and enable landowner to require or support cover crops by providing cost-sharing or a reduced rental rate
- **A and B combined**

Control: Info, discussion guide, testimonial



A: Add Lease Insertion Language (Nudge)

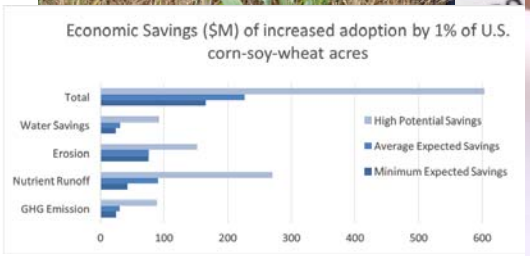
B: Add Financial Incentive to nudge

CBEAR-NACD-USDA Collaboration?

Information, Technical assistance, and Financial Incentives

What about one-on-one technical assistance?

One-on-one consultation on lease, business and conservation plan



A: Enhanced Information Only

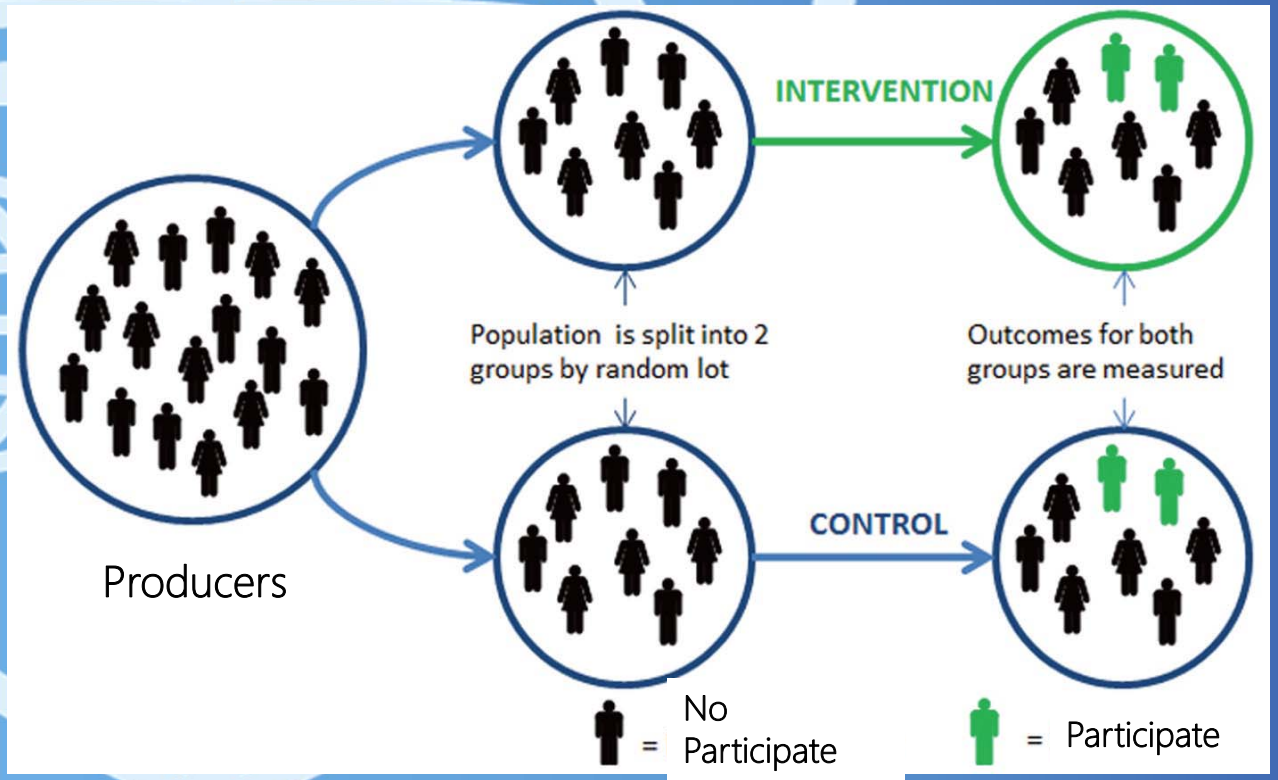
B: Information + Technical Assistance

C: Information + Financial Incentive

D: A+B+C

Status Quo (Control)

We propose a collaboration to contrast the cost-effectiveness of popular approaches to owner and operator engagement in the soil health context



Experimental Designs Make Doing Credible Science Easier

The Center for Behavioral and Experimental Agri-Environmental Research

Conference on Behavioral & Experimental Agri-Environmental Research: Methodological Advancements & Applications to Policy (CBEAR-MAAP)

October 14-15, 2017

Shepardstown, WV

Common Issues

1. Low power designs (and no power analyses)

2. Multiple comparisons
   i) multiple treatments; (ii) multiple outcome variables; and (iii) tests of heterogeneous treatment effects (subgroup effects). Richer ≠ Better

3. Lack of clarity about which estimands are identified by randomization and which are not

4. Lack of clarity about the difference between causal inference questions (Does X cause Y and by how much?) and predictive inference questions (For which subgroups does X cause Y and by how much?), and the implications for methods

5. Lack of clarity about identification issues and statistical inference issues (leading to lower precision)

ERS and NIFA need to push higher standards for all research.

Incentives: how they are presented matters

Can perform up to 50 action units (e.g., acres placed in riparian buffers).

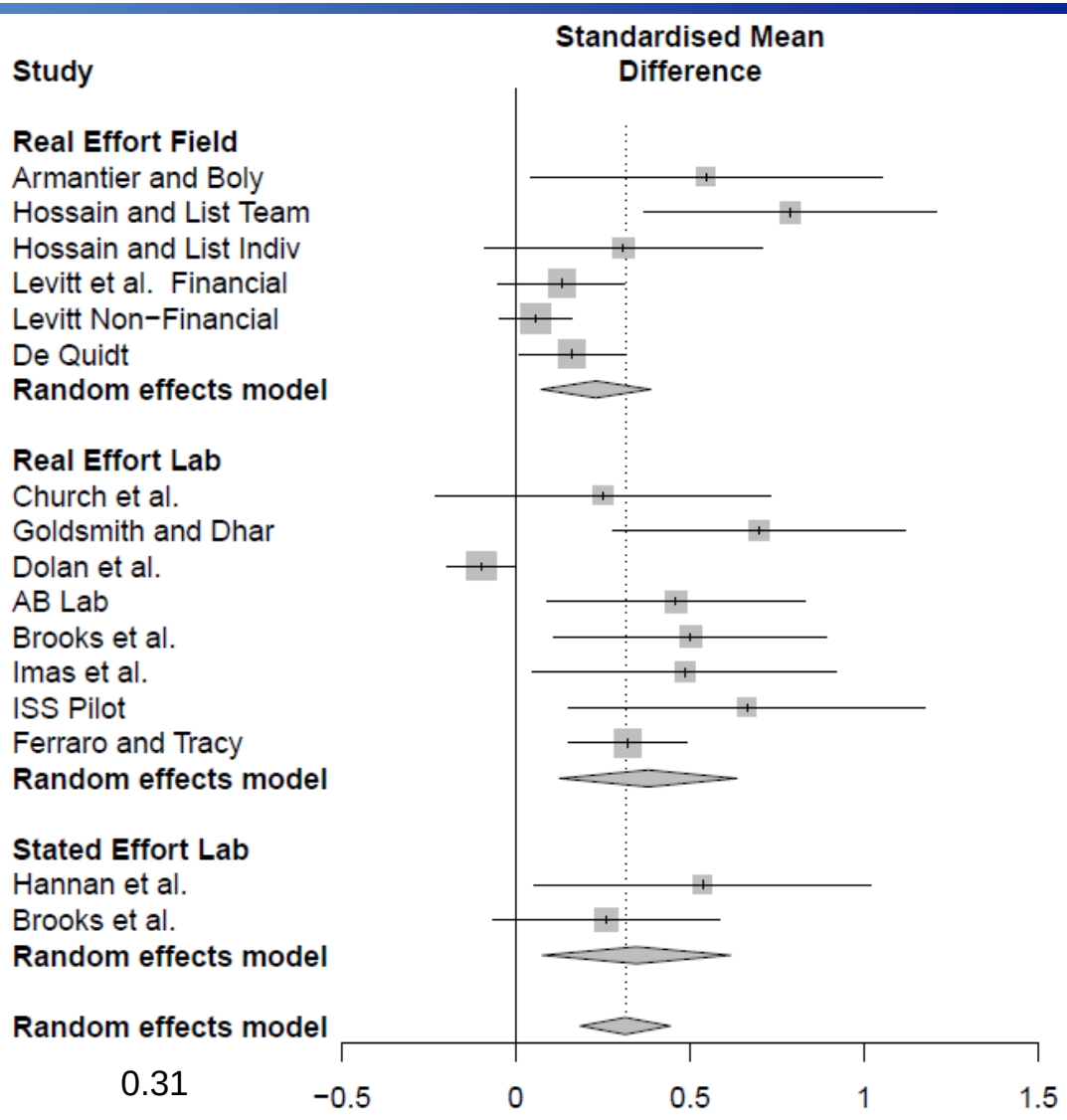
- Gain-Frame Contract: Start with $0. “For every action you perform, you receive $100, up to $5000.”
- Loss-Frame Contract: Start with $5000. “For every action you do not perform, you lose $100.”

If losses are weighed more heavily than equivalent gains by many people (est. 1.5-2X), then Loss-Frame Contract could induce greater total effort.

Loss-framed Incentive Contracts

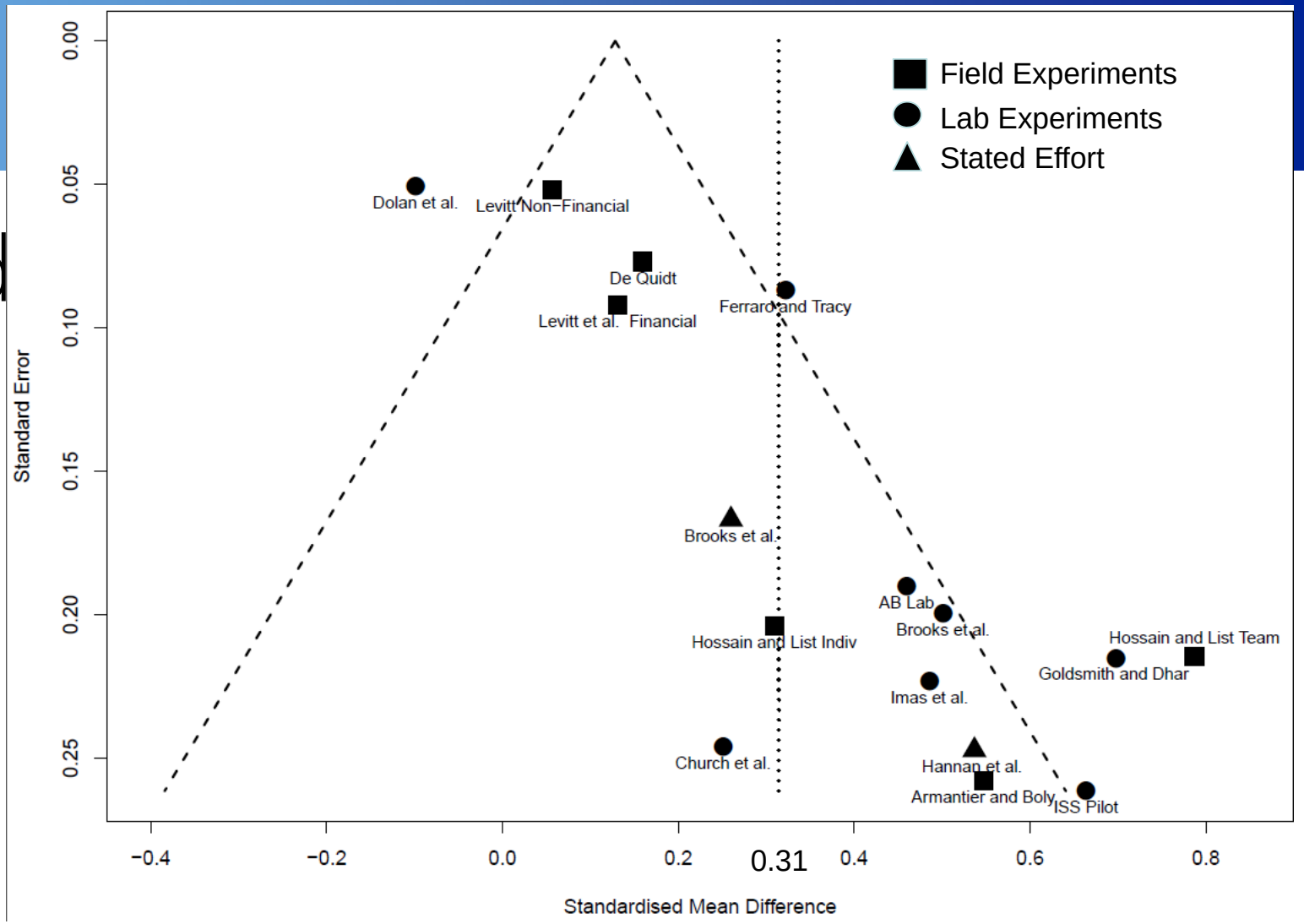
16 experiments imply that loss-framed contracts, on average, increase effort (success) at the incentivized task

Meta-analysis yields an overall weighted average effect of 0.31 SD [95%CI 0.18, 0.44]

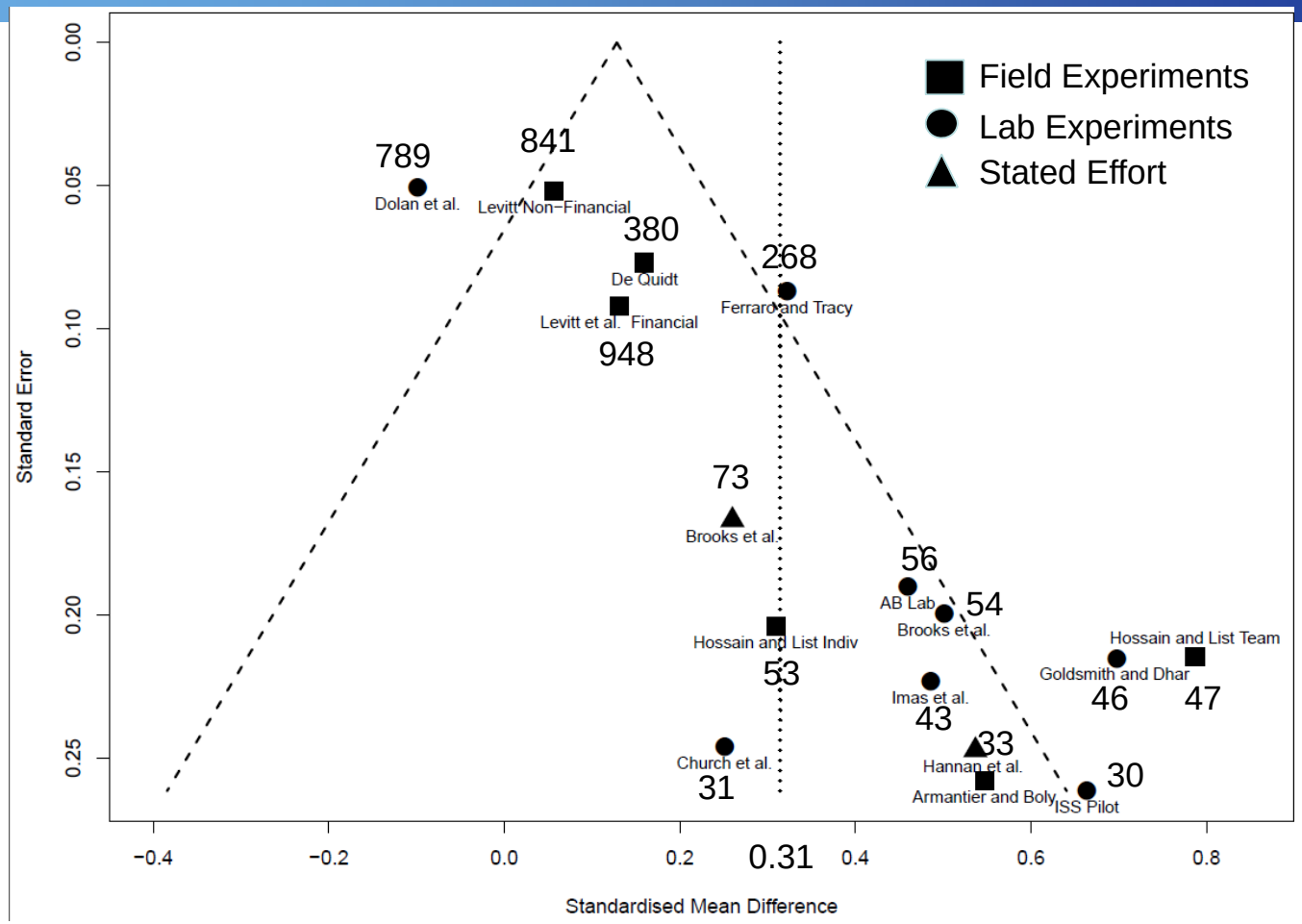


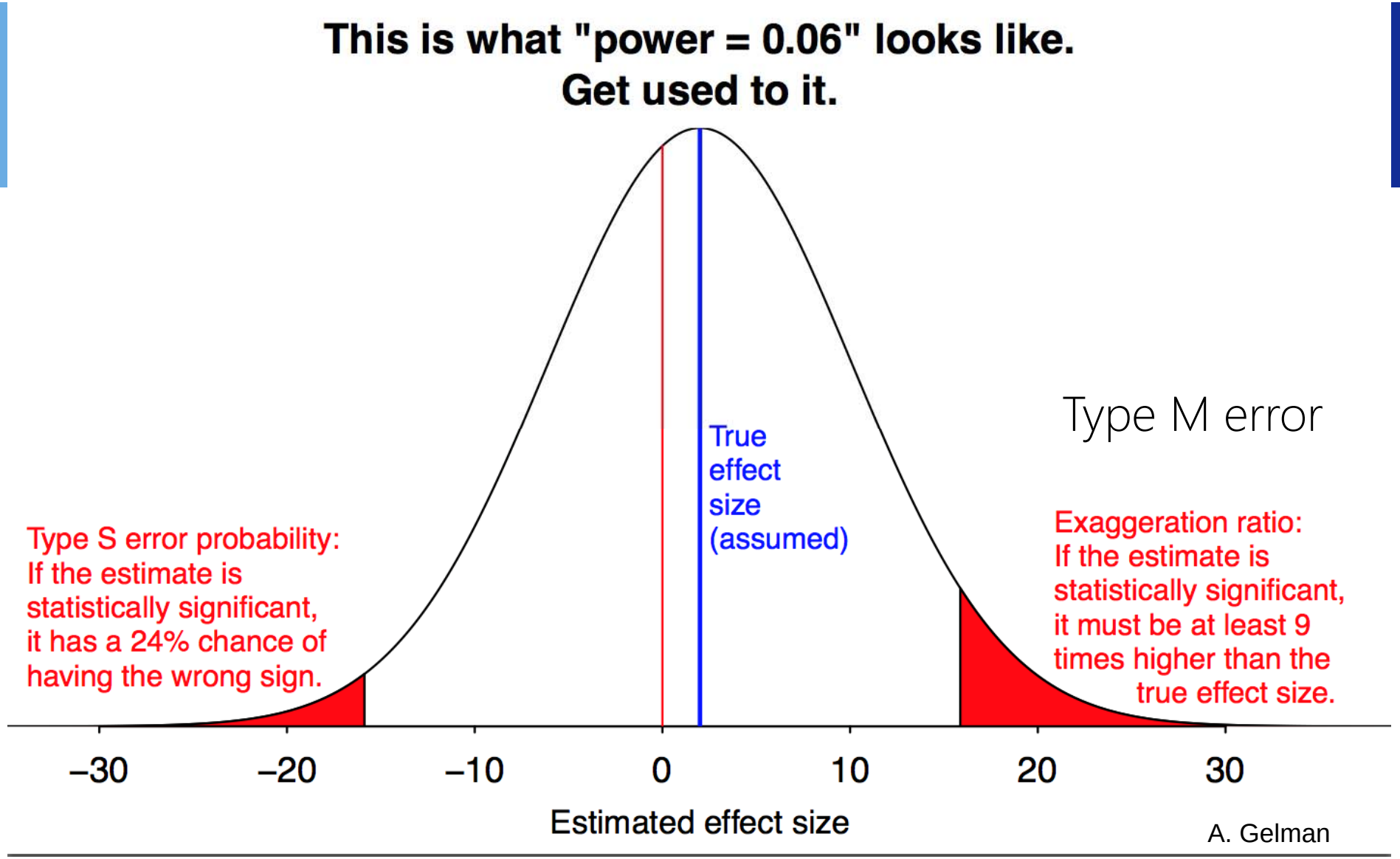
Ferraro and Tracy, unpublished

Loss-framed Incentive Contracts



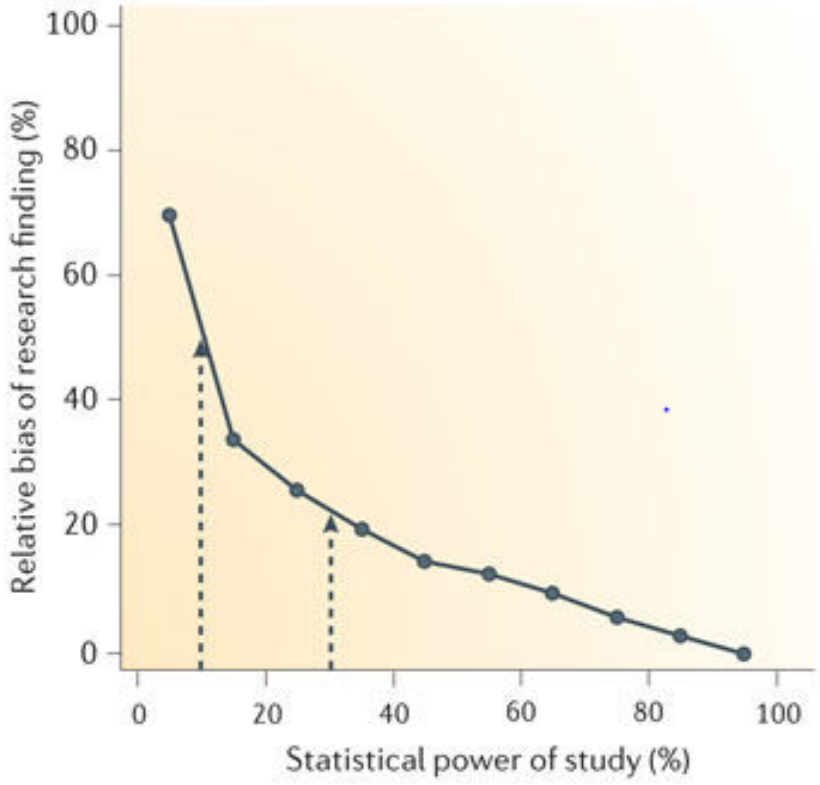
Sample Sizes




This is what "power = 0.06" looks like.
Get used to it.
Type S error probability:
If the estimate is
statistically significant,
it has a 24% chance of
having the wrong sign.
True
effect
size
(assumed)
Exaggeration ratio:
If the estimate is
statistically significant,
it must be at least 9
times higher than the
true effect size.
–30
–20
–10
0
10
20
30
Estimated effect size
A. Gelman

Type M error

Figure 5: The winner's curse: effect size inflation as a function of statistical power.



Nature Reviews | Neuroscience

Govinddelhi

THE ECONOMIC JOURNAL

The Economic Journal, **127** (*October*), F236–F265. Doi: 10.1111/ecoj.12461 © 2017 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

THE POWER OF BIAS IN ECONOMICS RESEARCH*

John P. A. Ioannidis, T. D. Stanley and Hristos Doucouliagos

We investigate two critical dimensions of the credibility of empirical economics research: statistical power and bias. We survey 159 empirical economics literatures that draw upon 64,076 estimates of economic parameters reported in more than 6,700 empirical studies. Half of the research areas have nearly 90% of their results under-powered. The median statistical power is 18%, or less. A simple weighted average of those reported results that are adequately powered (power ≥ 80%) reveals that nearly 80% of the reported effects in these empirical economics literatures are exaggerated; typically, by a factor of two and with one-third inflated by a factor of four or more.

Addressing Participant Inattention in Federal Programs: A Field Experiment with the Conservation Reserve Program

Steven Wallander, P

N= 46,823 (producers with expiring CRP contracts)

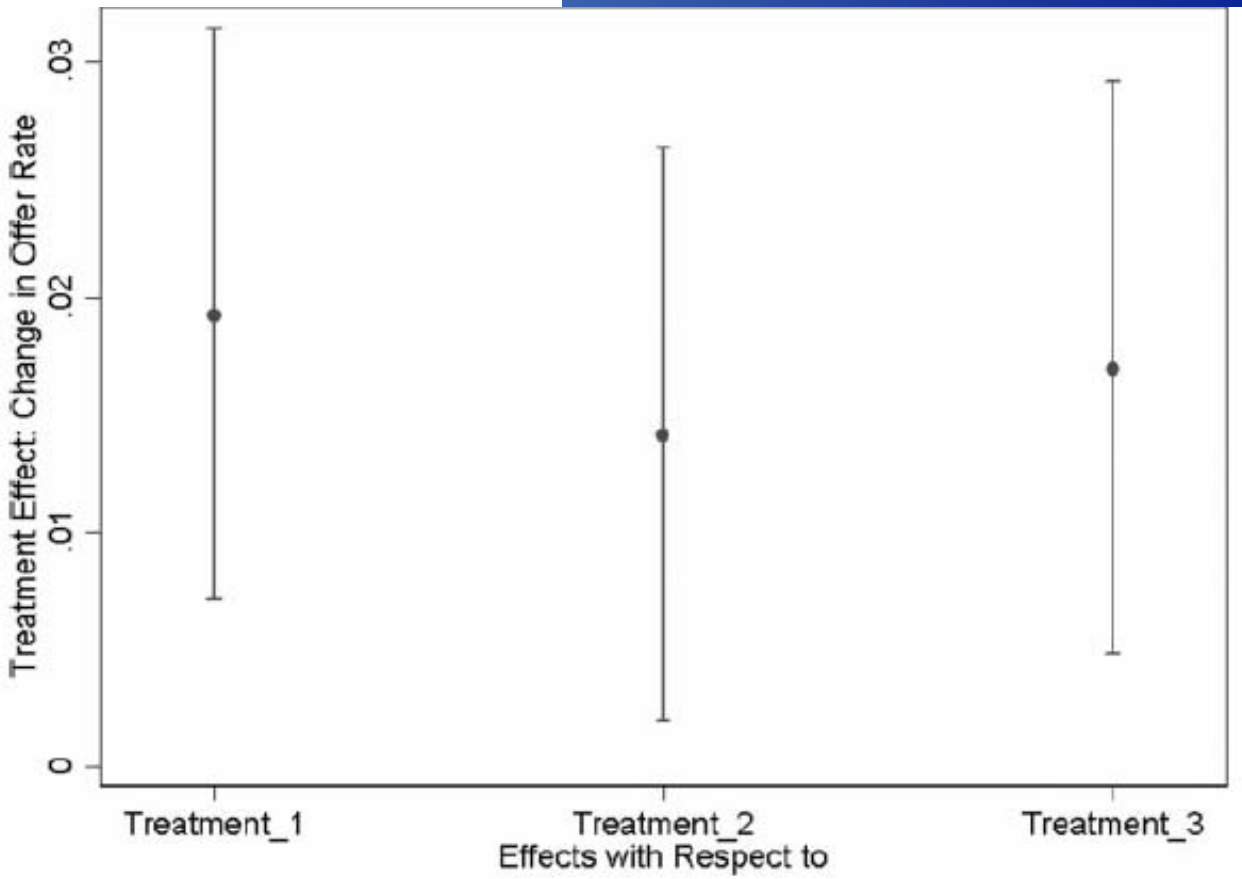


Figure 3. Treatment effect estimates for population of farms with expiring contracts

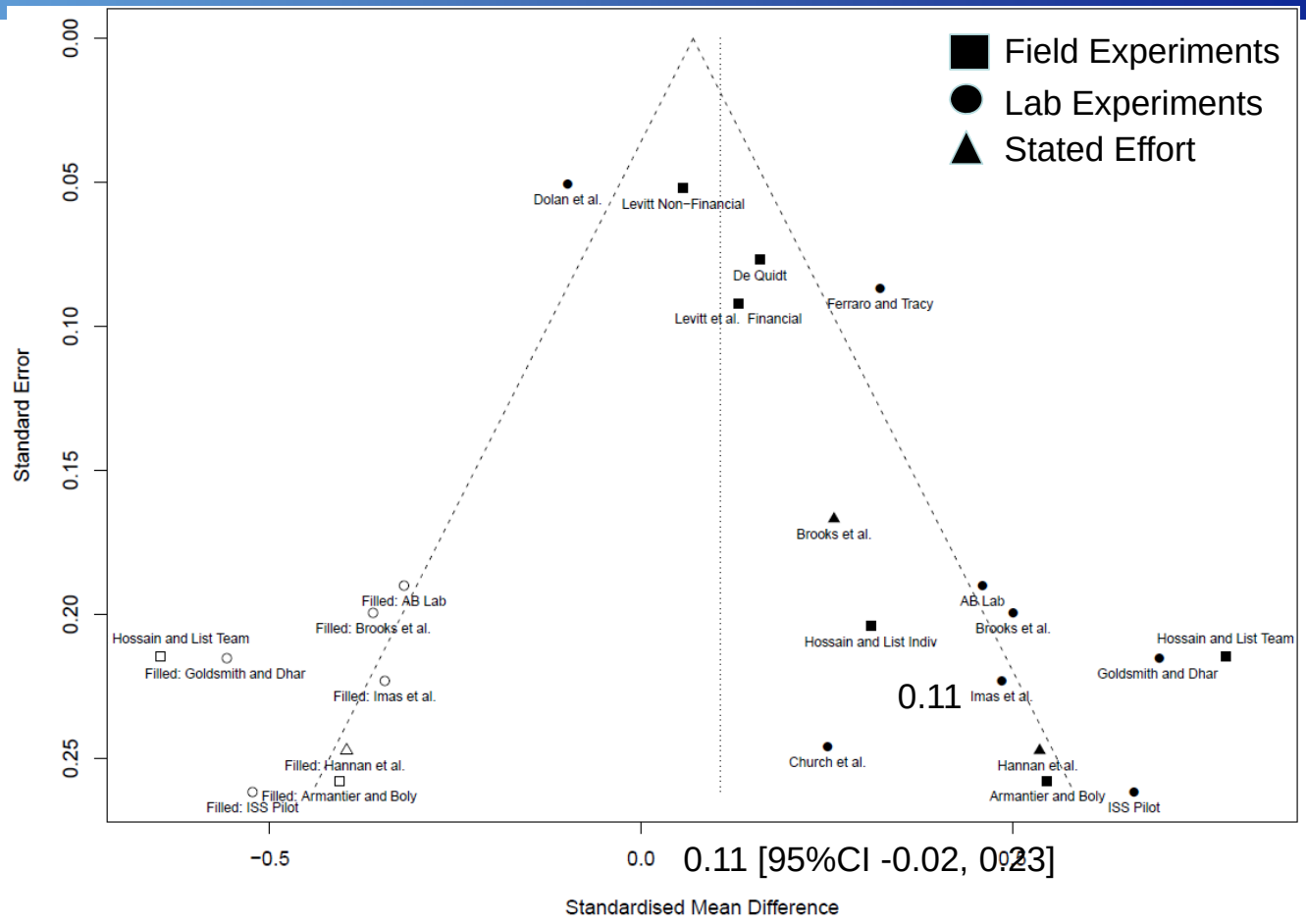
AAEA Agricultural & Applied Economics Association
American Journal of Agricultural Economics

"Reviewer 3 finds the small/no impacts of the treatment to reduce the contribution of this paper."

"Reviewer 1 and 2 would also like to see more exploration of the types of farms and regions where the treatment had a bigger impact."

Loss-framed Incentive Contracts

But there's more….endogenous sample selection, p-hacking, wishful discarding of outliers, deliberate fraud



We should not expect large treatment effects

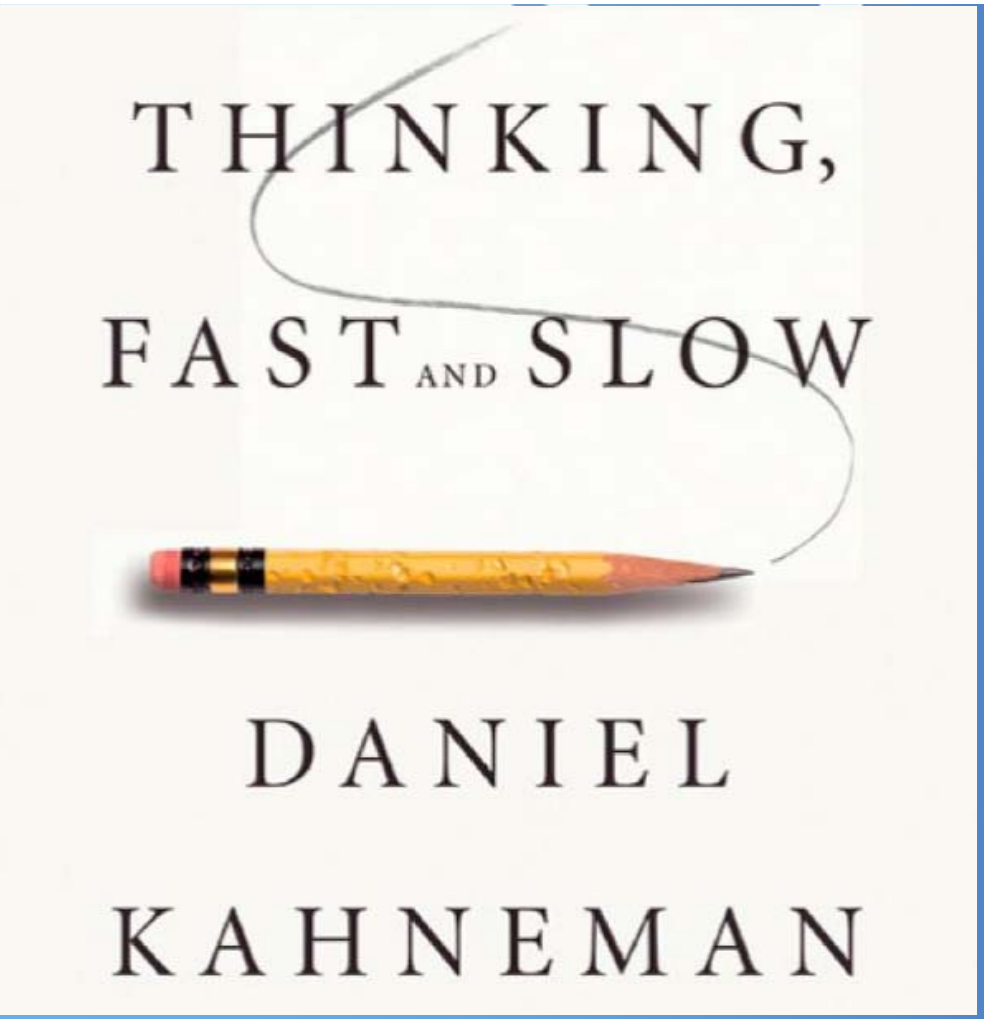
P. Rossi. *The Iron Law of Evaluation and Other Metallic Rules* (1987)

The Iron Law of Evaluation: The expected value of any net impact assessment of any large scale social program is zero.

The Stainless Steel Law of Evaluation: The better designed the impact evaluation of a social program, the more likely is the resulting estimate of net impact to be zero.

Curb your enthusiasm

Of 13,000 RCTs conducted by Google and Microsoft to evaluate new products or strategies in recent years, 80-90 percent have reportedly found no statistically significant effects (Arnold Foundation report, 2018)

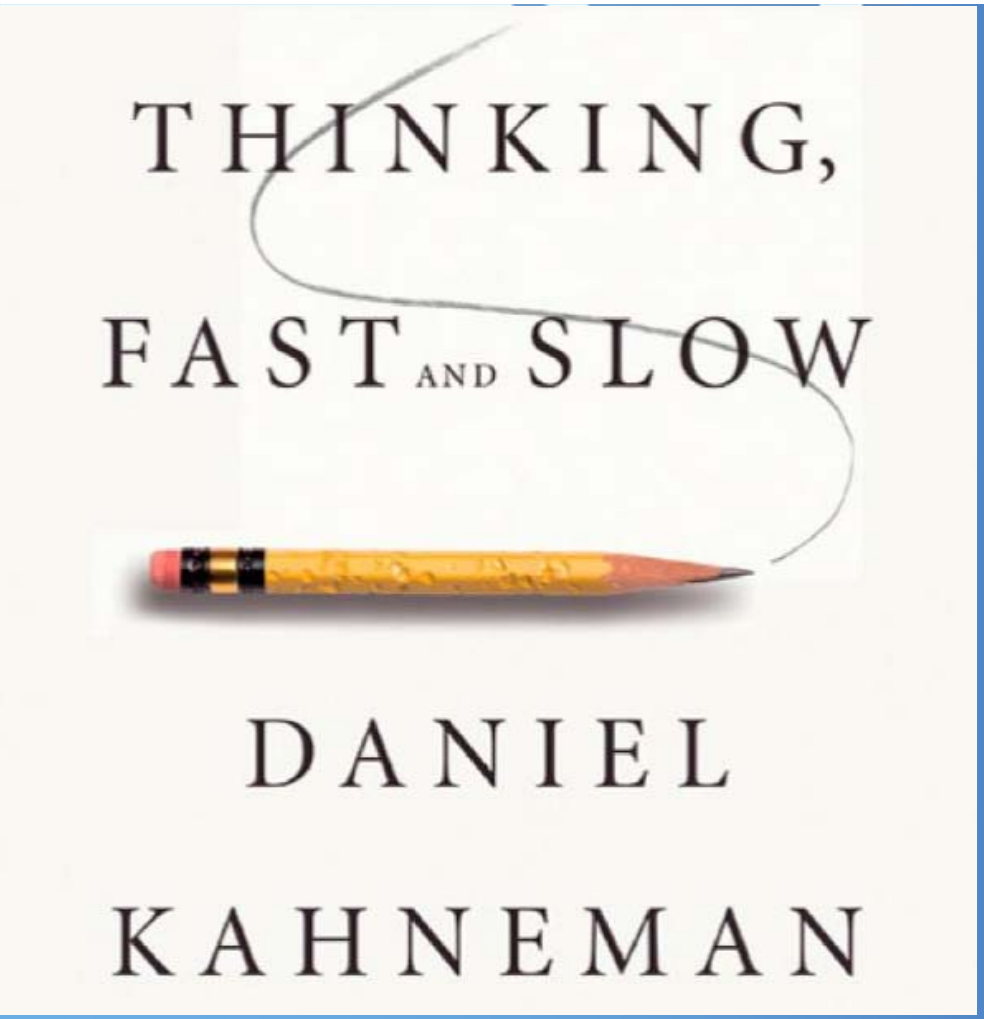


Anchoring

Tversky & Kahneman: Roulette wheel rigged to fall on either 10 or 65. TK spins, subjects write down #, and then asked

1. Is the percentage of African nations among UN members larger or smaller than this number?
2. What is your best guess of the percentage?

Subjects who received 65 anchor had average estimate almost double the estimate of subjects who receive 10



Anchoring

Anchoring "occurs when people consider a particular value for an unknown quantity before estimating that quantity. What happens is one of the most reliable and robust results of experimental psychology: the estimates stay close to the number that people considered – hence the image of an anchor."

Kahneman (2013)

“Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences

Dan Ariely, George Loewenstein, Drazen Prelec

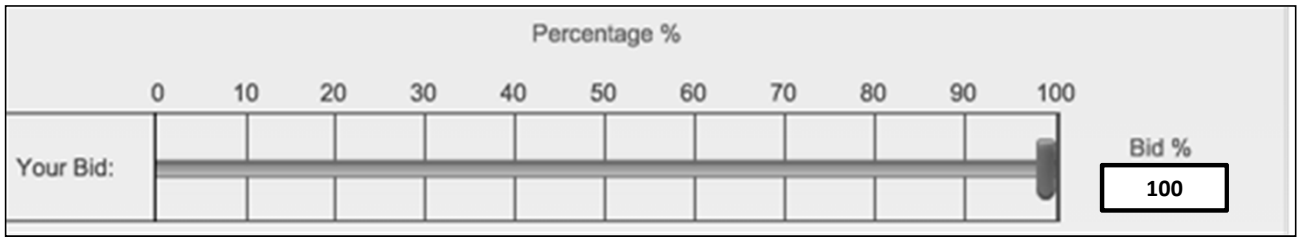
The Quarterly Journal of Economics, Volume 118, Issue 1, 1 February 2003, Pages 73–106,

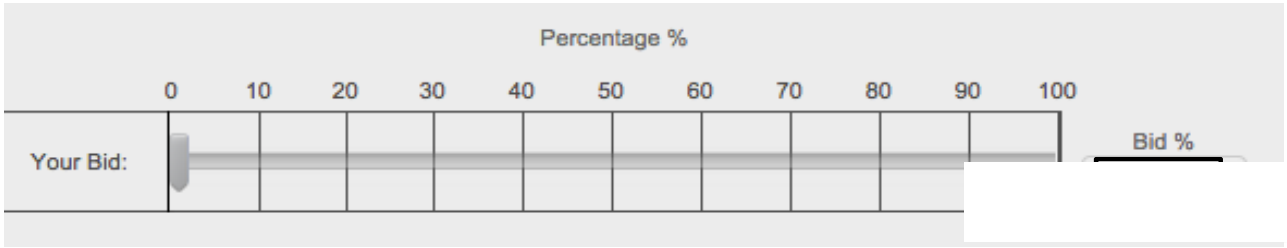
Results imply that people’s preferences are characterized by a very large degree of arbitrariness. In particular, they provide evidence that subjects’ preferences for an array of goods and hedonic experiences are strongly affected by normatively irrelevant cues, namely anchors.

50-200% changes in WTP and WTA as anchor changes

AgVISE (Agricultural Values, Innovation, and Stewardship Enhancement)
Default Starting Bid in Auction

Farm operators bidding on cost-share conservation contracts (e.g., riparian buffers, remove abandoned poultry houses, feral hog trapping systems – i.e., impure public goods)



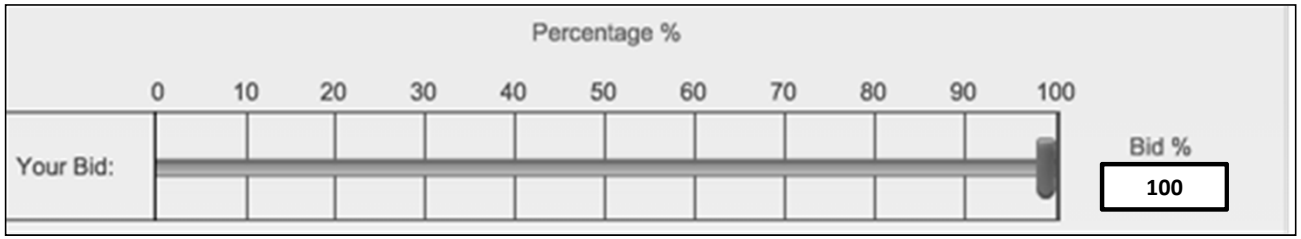


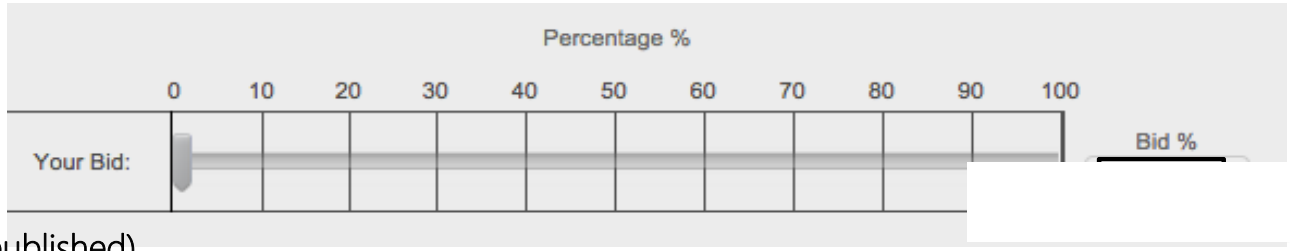
(Ferraro and Messer, unpublished)

AgVISE (Agricultural Values, Innovation, and Stewardship Enhancement)
Default Starting Bid in Auction

Bids 10 percentage points higher if assigned 100% starting bid. Equivalent to forgoing ~USD 1400

Out of 537 total participants, 178 participants placed bids.

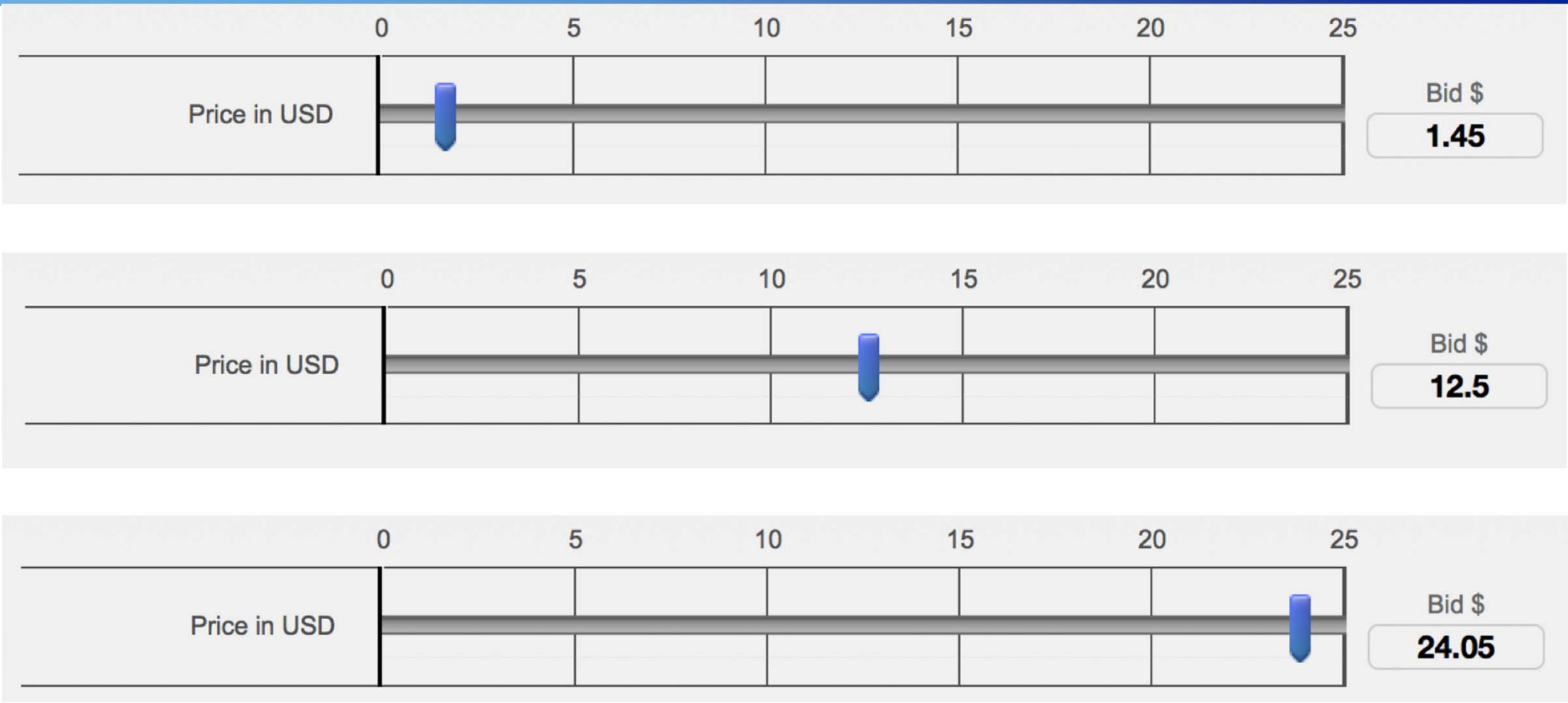




(Ferraro and Messer, unpublished)

HomeVISE: Homeowner Value, Innovation, and Stewardship Enhancement

HomeVISE: Homeowner Value, Innovation, and Stewardship Enhancement



HomeVISE: Homeowner Value, Innovation, and Stewardship Enhancement

HomeVISE 1 (2016)

Each of the 336 adult participants placed five bids (one for each item). Each was randomized to one anchor: $0 to $25 anchors

(So 26 anchors with ~13 subjects per anchor value)

When anchor goes from $0 to $15, the average bid increases by ~40%

(95% CI goes from ~5% to ~75%)

HomeVISE: Homeowner Value, Innovation, and Stewardship Enhancement

HomeVISE 2 (2017)

Each of the 1200 adult participants placed four bids (one for each item). Each subject was randomized to one of only two anchors: $0 or Full Endowment ($15). Tried to also raise salience of anchor (as a treatment).

When anchor goes from $0 to $15, the average bid increases by ~5% (95% CI goes from ~2% to ~8%). Without the salience treatment, it's ~0%

“Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences

Dan Ariely, George Loewenstein, Drazen Prelec

The Quarterly Journal of Economics, Volume 118, Issue 1, 1 February 2003, Pages 73–106,

One Swallow Doesn’t Make a Summer:

New Evidence on Anchoring Effects

By Zacharias Maniadis, Fabio Tufano and John A. List[1]

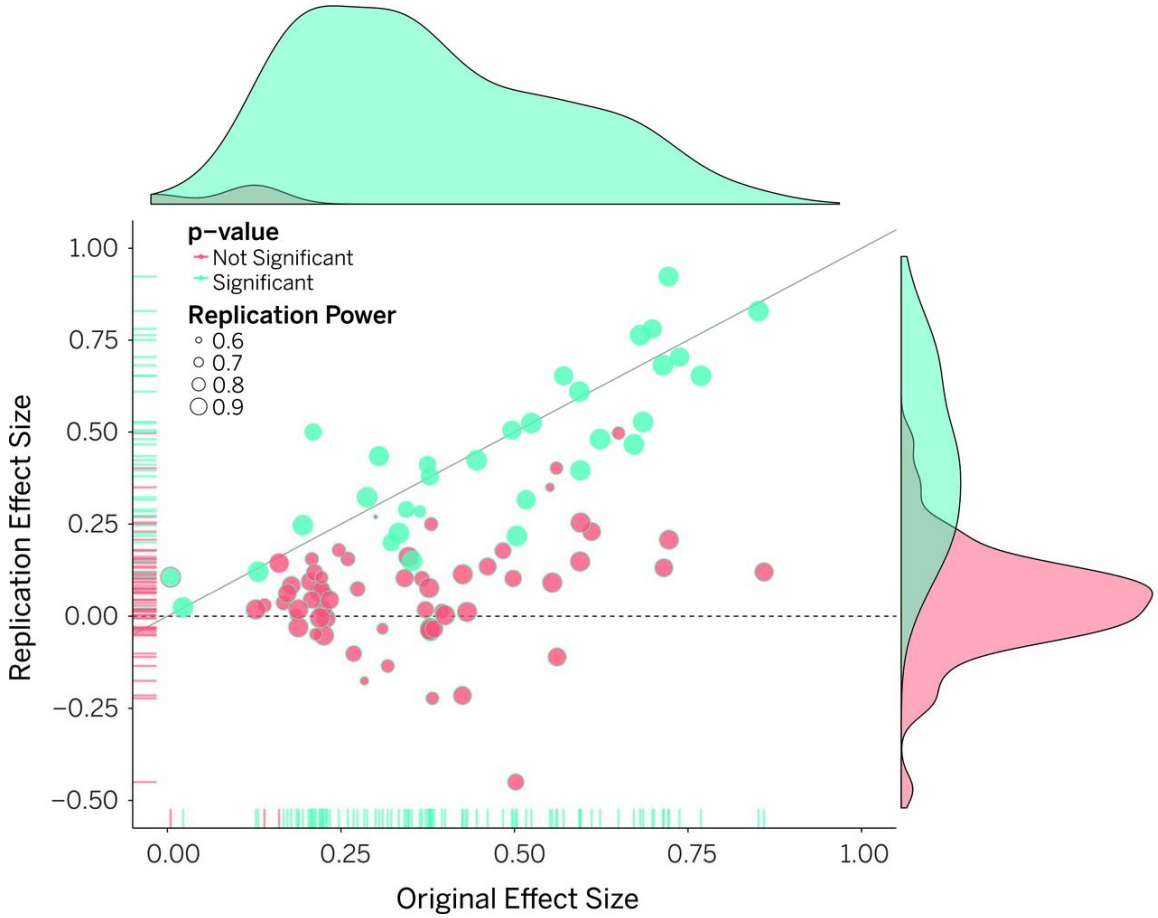
Replication of Ariely et al. found much smaller treatment effect that had debatable economic implications

Estimating the reproducibility of psychological science

Open Science Collaboration*,†
+ See all authors and affiliations

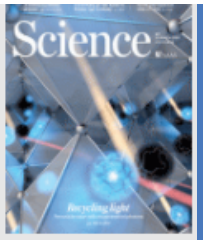
Replicated 100 studies published in 3 top journals in psychology.

Nearly all (97) of original studies reported "positive findings," but in replications, authors only found a significant effect in the same direction for 36% of these studies



Evaluating replicability of laboratory experiments in economics

Colin F. Camerer[1,*,†], Anna Dreber[2,†], Eskil Forsell[2,†], Teck-Hua Ho[3,4,†], Jürgen Huber[5,†], Ma…
+ See all authors and affiliations



18 Replications from AER and QJE (2011-2014

Found a significant effect in the same direction as in the original study for 11 replications (61%)

On average, the replicated effect size is 66% of the original.

We Need Power Analyses → Need Larger Samples and Fewer Research Questions

Recruiting farmers is expensive.

CBEAR offering $75 for a half hour survey plus an opportunity for one in ten to earn up to $3000, and we are seeing a 6% response rate (need to invite 10,000 producers)

Better to work within USDA programs that are already recruiting hundreds or thousands of participants

Pre-registration of Studies (Pre-analysis Plans)

Goal: Write the entire paper in advance, leaving out results and conclusions

Describe design in advance, including identification strategy (e.g., how you will do randomization), mode of statistical inference, sample size (including Power Analysis!), sample exclusions, outcome measures, covariates for precision, and subgroup definitions BEFORE you see outcome data.

Deviations from the plan are not prohibited, but when such deviations arise they should be highlighted and the effects on results reported.

Pre-registration of Studies (Pre-analysis Plans)

Goal: Write the entire paper in advance, leaving out results and conclusions

As new information arises or prior assumptions turn out to be incorrect, you can update plan (and clearly document as an update)

Deviations from the plan are not prohibited, but when such deviations arise they should be highlighted and the effects on results reported.

"Unexpectedly, we also found that…"
"In addition to the analyses we pre-registered we also ran…"
"We encountered an unexpected situation, and followed our Standard Operating Procedure…"

Pre-registration of Studies (Pre-analysis Plans)

1. Limits extent to which researchers can make decisions that consciously or unconsciously tilt a study toward a desired result.
2. The validity of frequentist statistical inference (SEs, CIs, p-values, significance tests) hinges on assumption that analysis follows a pre-specified strategy
3. Publicly-archived plans enable readers to see which analyses were pre-specified and to take that into account when assessing the credibility of results

Costly: You do all this work and the experiment is a "failure"! Well, that's the point of pre-registration – failures are good for science. Ideally, journal editors would accept papers before the results are seen.

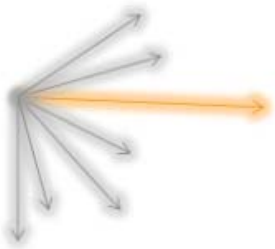
Pre-registration of Studies

AEA RCT Registry
The American Economic Association's registry for randomized controlled trials

https://www.socialscienceregistry.org

NIH U.S. National Library of Medicine
ClinicalTrials.gov

Open Science Framework
A scholarly commons to connect the entire research cycle



Your own website with credible time stamp

The Center for Behavioral and Experimental Agri-Environmental Research

Conference on Behavioral & Experimental Agri-Environmental Research: Methodological Advancements & Applications to Policy (CBEAR-MAAP)

October 14-15, 2017

Shepardstown, WV

Common Issues

1. Low power designs

2. Multiple comparisons
   i) multiple treatments; (ii) multiple outcome variables; and (iii) tests of heterogeneous treatment effects (subgroup effects). Richer ≠ Better

3. Lack of clarity about which estimands are identified by randomization and which are not

4. Lack of clarity about the difference between causal inference questions (Does X cause Y and by how much?) and predictive inference questions (For which subgroups does X cause Y and by how much?), and the implications for methods

5. Lack of clarity about identification issues and statistical inference issues (leading to lower precision)

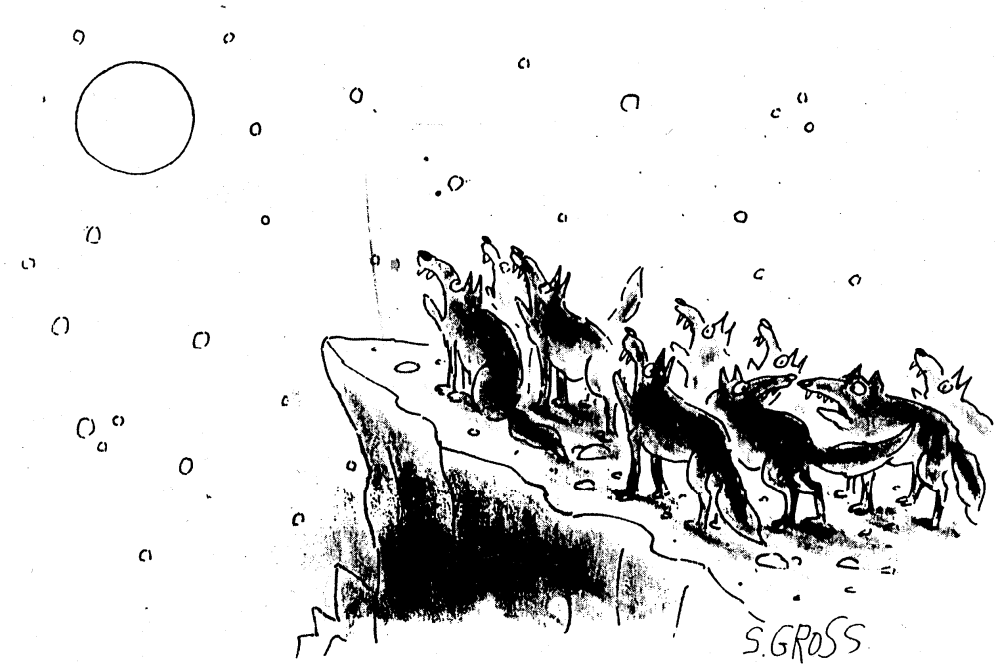
ERS, NIFA and others need to push higher standards for all research!

USDA
United States Department of Agriculture
Economic Research Service

Questions?

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

WILLIAM PENN FOUNDATION



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

CBEAR The Center for Behavioral and Experimental Agri-Environmental Research
epic ENVIRONMENTAL PROGRAM INNOVATIONS COLLABORATIVE
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH
JOHNS HOPKINS CAREY BUSINESS SCHOOL
JOHNS HOPKINS WHITING SCHOOL of ENGINEERING