

# New data and evidence on the organisation of science - implications for diversity

Julia Lane, NYU

And many colleagues



# THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking

Transparency  
Humility  
Data  
Privacy  
Rigor  
Capacity



Measuring Results  
Expanding Opportunity  
Improving Lives

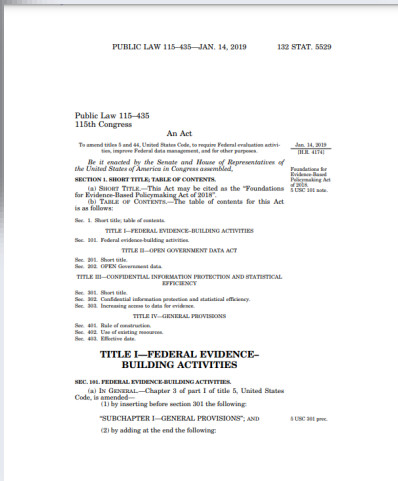
CONTRIBUTE

ABOUT MISSION APPROACH NEWS

## CHIPS+ ACT ADVANCES EVIDENCE-BASED POLICYMAKING VIA NATIONAL SECURE DATA SERVICE

By: AIF Staff

This week, President Biden signed into law **the CHIPS & Science Act of 2022**. The legislation, which was primarily focused on making the United States more competitive with China, also included an important provision establishing a National Secure Data Service (NSDS) demonstration project at the National Science Foundation.



### Advisory Committee on Data for Evidence Building: Year 2 Report

October 14, 2022



## EDITORIAL

### Wanted: Better Benchmarks

**H**ow much should a nation spend on science? What kind of science? How much from private versus public sectors? Does demand for funding by potential science performers imply a shortage of funding or a surfeit of performers? These and related science policy questions tend to be asked and answered today in a highly visible advocacy context that makes assumptions that are deserving of closer scrutiny. A new "science of science policy" is emerging, and it may offer more compelling guidance for policy decisions and for more credible advocacy.

All developed and many developing nations today have accepted the need to support technical education and research as keys to future economic strength. Studies from the 1990s show that U.S. investment in R&D development led to greater economic productivity, and that information technology, in particular, has been a major factor in sustaining U.S. productivity growth. The question is not whether R&D investments are important, but what investment strategies are most effective in the rapidly changing global environment for science. Here, ideas diverge.

Take the issue of the technical workforce. Sharply differing opinions exist regarding the production of U.S. scientists to meet possible impending shortages.<sup>6</sup> The differences turn on the interpretation of "benchmark" data regarding the numbers of degree holders produced in the United States and other countries, particularly China and India. In the latter countries, the rates of growth in the numbers of scientists are high, although actual numbers are small relative to those in the United States. Advocates for increased production of U.S. scientists point to our low graduation rates, whereas critics emphasize limited short-term job opportunities for graduates and postdocs. Resolution of this issue requires a broader understanding of socioeconomic factors in a number of nations that would allow us to attach probabilities to different future scenarios. Optimal strategies for large mature economies such as that of the United States will doubtless differ from those for smaller or developing economies. Here, as elsewhere in policy debates, the benchmarks do not speak for themselves.



sciencemag.org on October 4, 2012

# The Science of Science Policy

A HANDBOOK

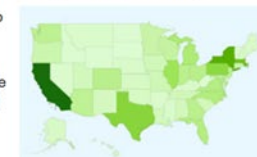
EDITED BY  
KAYE HUSBANDS FEALING,  
JULIA I. LANE,  
JOHN H. MARBURGER III,  
AND STEPHANIE S. SHIFF



## Recovery Act

### OVERVIEW OF THE AMERICAN RECOVERY AND REINVESTMENT ACT OF 2009 (RECOVERY ACT)

The American Recovery and Reinvestment Act of 2009 (Recovery Act) was signed into law by President Obama on February 17th, 2009. It is an unprecedented effort to jumpstart our economy, create or save millions of jobs, and put a down payment on addressing long-neglected challenges so our country can thrive in the 21st century. The Act is an extraordinary response to a crisis unlike any since the Great Depression, and includes measures to modernize our nation's infrastructure, enhance energy independence, expand educational opportunities, preserve and improve affordable health care, provide tax relief, and protect those in greatest need.



View Map of NSF ARRA Awards by State

### ACCELERATION OF ARRA EXPENDITURES

Context

Data

Evidence

Policy



## NSB passes resolutions to address Missing Millions & deliver research benefits across America

March 4, 2021

NSB's *Vision 2030* emphasizes the urgent need for greater participation of women and other underrepresented groups in the U.S. science and engineering enterprise and ensuring that research benefits reach all Americans. Last week, the National Science Board (NSB) passed two resolutions to advance both goals. One resolution aims to address unconscious biases and improve the preparedness of proposal reviewers. The second seeks to increase the potential of proposals' Broader Impacts (BI) to benefit society.

"The Board is committed to working with NSF to find new ways to building America's workforce and ensuring its innovation leadership," said NSB Chair Ellen Ochoa. "We trust in Director Panofsky's leadership to find a way to implement the policies we outline in the resolutions and to maximize their impact."

<https://www.ai.gov/nairrtf/#ABOUT-NAIRRTF>

HOME ABOUT STRATEGIC PILLARS DOCUMENTS RESOURCES NEWS

↑ Top

### A NATIONAL AI RESEARCH RESOURCE

The NAIRR is envisioned as a shared computing and data infrastructure that will provide AI researchers and students across scientific fields and disciplines with access to compute resources and high-quality data, along with appropriate educational tools and user support. The goal for such a national resource is to democratize access to the cyberinfrastructure that fuels AI research and development, enabling all of America's diverse AI researchers to participate in exploring innovative ideas for advancing AI, including communities, institutions, and regions that have been traditionally underserved.

↑ Top



JANUARY 20, 2021

## Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government

BRIEFING ROOM PRESIDENTIAL ACTIONS

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered:

**Section 1. Policy.** Equal opportunity is the bedrock of American democracy, and our diversity is one of our country's greatest strengths. But for too many, the American Dream remains out of reach. Entrenched disparities in our laws and public policies, and in our public and private institutions, have often

Context

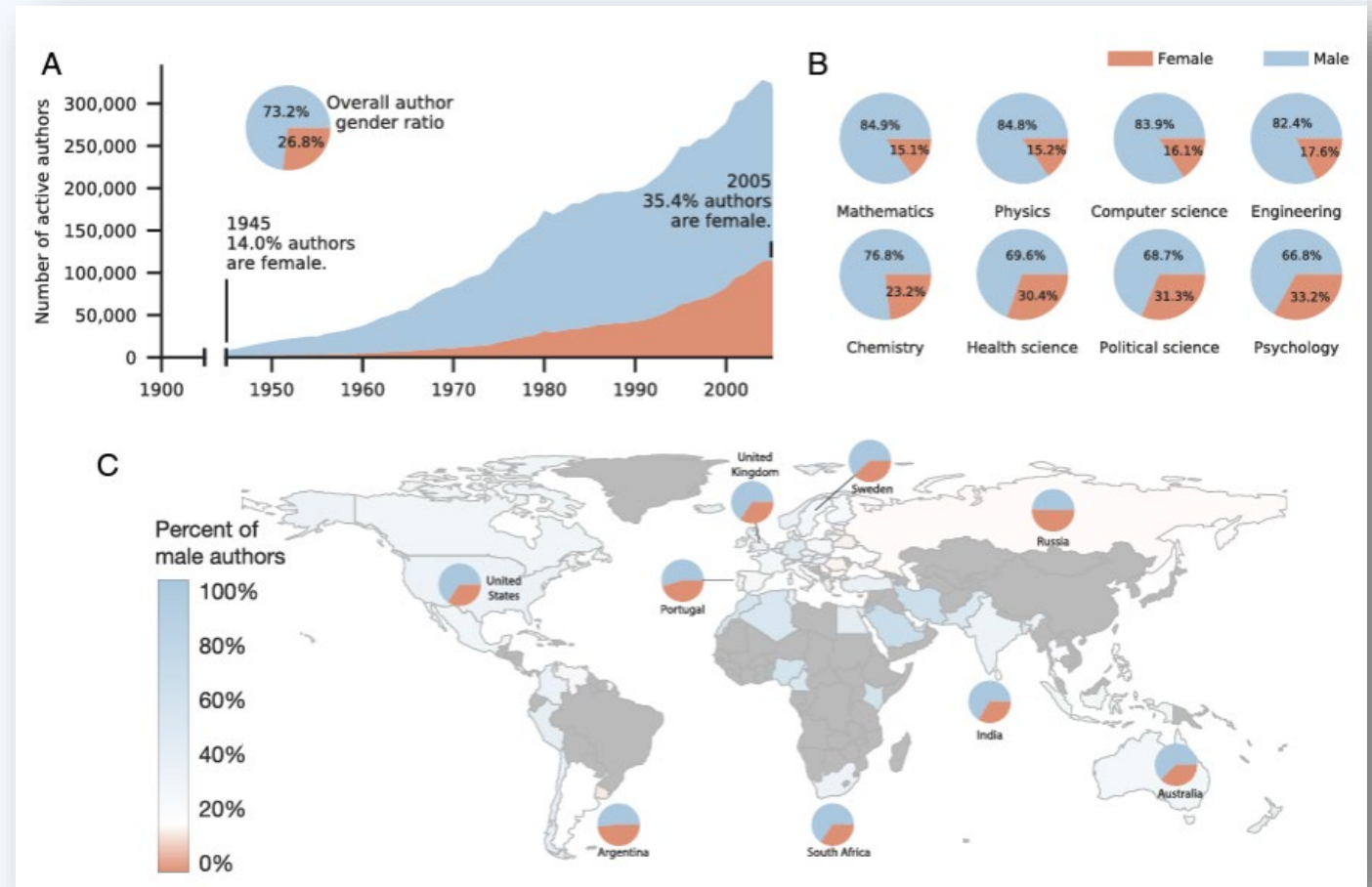
Data

Evidence

Policy

# Women publish and patent less than men

- Women make up only 12% of patent inventors, even though they make up 28% of scientific and technical workers (Toole et al., 2016; Ding et al., 2006; Thebaud and Charles 2018)
- Relative to men, women also tend to publish less, with estimates using data from NSF's Survey of Doctorate Recipients suggesting 5-year productivity differences of roughly 20 percent (Ceci et al., 2014)



Historical comparison of gender inequality in scientific careers across countries and disciplines. Huang et al., 2020 PNAS <https://doi.org/10.1073/pnas.1914221117>

# The state of science and technology policy (ARRA)



unavailable. The ITG found that:

- There is a well developed body of social science knowledge that could be readily applied to the study of science and innovation.
- Although many Federal agencies have their own communities of practice, the collection and analysis of data about the science and scientific communities they support is heterogeneous and unsystematic.
- Agencies are using very different models, data and tools to understand their investments in science and technology.
- The data infrastructure is inadequate for decision-making.

**F**

All dev  
as keys to  
greater eco  
productivit  
most effect

Take th  
to meet po  
numbers o  
China and  
are high, a  
Advocates  
rates, wher  
ates and po  
socioecon  
probabilitie  
economies suc...

for smaller or developing economies.  
the benchmarks do not speak for them

## THE SCIENCE OF SCIENCE POLICY: A FEDERAL RESEARCH ROADMAP



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

## ERC FRONTIER RESEARCH LEAVES 73% BREAKTHROUGHS OR MAJOR

11-09-2017

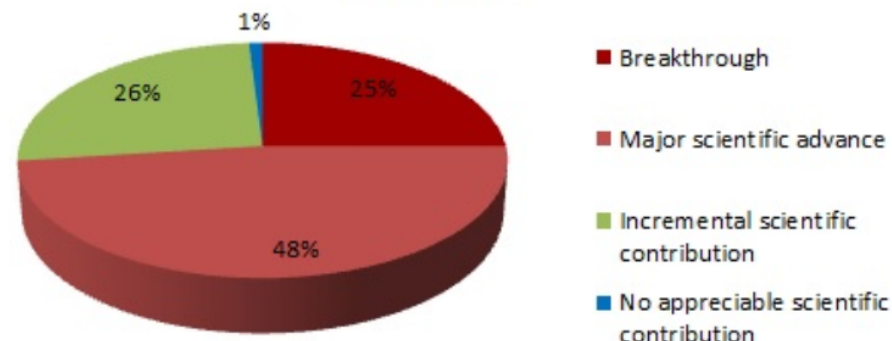


The European Research Council's blue sky research continues to pay off, triggering real breakthroughs. The study shows that the European Union's investment, through the ERC, is clearly leaving its mark on science and is already making a difference in countless ways, as this impact study strongly suggests. The report found that the ERC is achieving its goals of funding high-risk and interdisciplinary frontier research.

ERC President Jean-Pierre Bourguignon commented: "I am pleased to see that the ERC delivers even slightly better results than the already very uplit. It shows that the European Union's investment, through the ERC, is clearly leaving its mark on science and is already making a difference in countless ways, as this impact study strongly suggests. The report found that the ERC is achieving its goals of funding high-risk and interdisciplinary frontier research."

The pilot study, commissioned by the ERC Scientific Council, is seen as a novelty as few funding agencies invest in assessing the impact of the research they support. Subsequently, the Scientific Council will commission an independent qualitative evaluation.

Impact study of ERC projects  
New results



High-level peer reviewers evaluated a random sample of 155 concluded Starting and Advanced Grant projects. They found that a large majority generated very high scientific value: 73% of the projects have already made **breakthroughs or major scientific advances**. About 27% of the projects were incremental or, in a very few cases, did not make an appreciable scientific contribution. These findings by and large confirm the results of the pilot exercise, yet with a slight improvement. The study concludes that the ERC indeed funds **high-risk/high-gain** projects, in accordance with its mission, and that such projects are more likely to lead to breakthroughs. It also highlights the interdisciplinarity of many ERC projects.

# What not to do: A

www.shanghairar



Home

Home>

Ranking

- Sele
- Ran
- Defi
- Data

■ Sele

ARWU  
papers  
Citation  
univers  
and 80

■ Ran

Univers  
Nobel p  
major c  
scoring  
The dis  
technic

## NEWS



### Big is beautiful for merging universities

By Nic Mitchell

25 November 2015 Business



Sorbonne lecture: Paris could have a mega-university, if those broken up after 1968 are re-united

Universities across Europe are talking about merging or forming alliances like never before.

Almost 100 mergers have taken place since the beginning of the century. The European University Association (EUA), representing universities in 47 countries, is mapping this changing landscape with an [interactive merger map](#).

And the pace is accelerating, with eight super-universities or clusters identified in 2012; 12 in 2013 and 14 in 2014.

So what's driving the merger mania?

Is it a way of climbing world university rankings by concentrating the best brains and resources to attract more students and bigger research grants?

Or is it a way of responding to funding cuts?

Thomas Estermann, director for governance, funding and public policy development at the EUA, says bigger numbers of staff and students give these

#### Global education

Birmingham to open campus in the Gulf

Apple's Italian job for finding top talent

'Meresearch' - when study really is all about me

How a university became a battle for Europe's identity

five Conference S

d Researchers, or  
ers indexed by Science  
n total, more than 1300  
ranked between 501

alumni and staff winning  
ence, papers indexed in  
licator, the highest  
stage of the top score.  
dard statistical  
d as shown below to

Context

Evidence

Policy

# What not to do: Bur

## Overview of the assessment

### 2011-2012 Preparation

The UK funding bodies appointed the REF expert panels, based on nominations from academic and other organisations. The funding bodies and REF panels consulted with the sector and published detailed criteria and guidelines for the exercise.

### 2012-2013 Submissions

Each institution decided which UOAs to submit in, and prepared their submissions. Submissions were made by the deadline of 29 November 2013.

## REF 2014 cost almost £250 million

Accountability review finds cost of assessment equates to 2.4 per cent of funding bodies' expected spend over next six years

July 13, 2015



By Holly Else  
Twitter: @HollyElse

### All quality profile and average missions



	4*	3*	2*	1*	U
	30	46	20	3	1
	22.4	49.5	23.9	3.6	0.6
	44.0	39.9	13.0	2.4	0.7
	44.6	39.9	13.2	2.2	0.1

Universities spent about £4,000 for each researcher they submitted to the research excellence framework, a report has revealed.

The estimate of institutions' own total spend on the REF exceeds £230 million, of which £55 million went on preparing impact statements and £19 million for panellists' time.



REF 2014 impact case studies: government policy cited most

The cost to the four UK funding councils is estimated at £14 million, according to the report by policy advisory group Technopolis, putting the total cost of REF 2014 at an estimated £246 million. This is almost four times that of an



# What not to do: manual reporting

www.recovery.gov/Pages/home.aspx



This site can't be reached

**www.recovery.gov**'s server D.N.S. address could not be found.

DNS\_PROBE\_FINISHED\_NXDOMAIN

Reload

the end of each fiscal quarter starting on October 10<sup>th</sup>

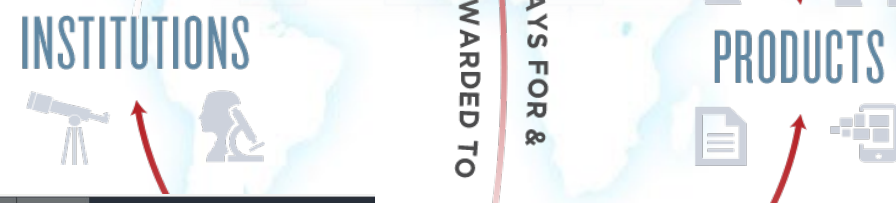
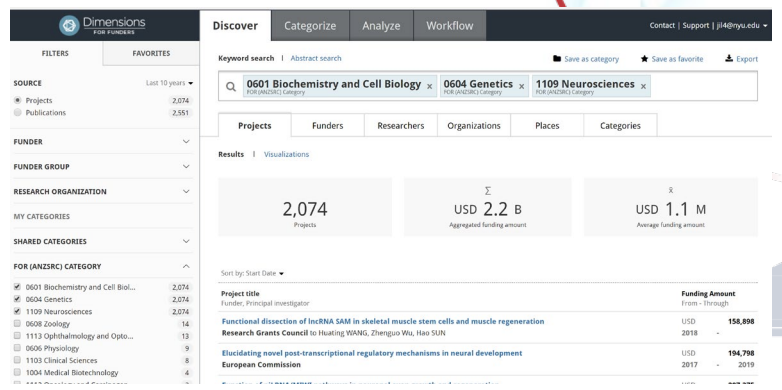
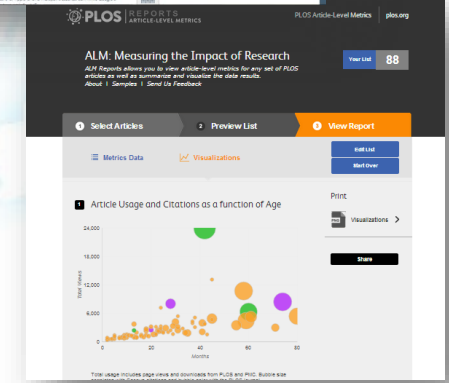
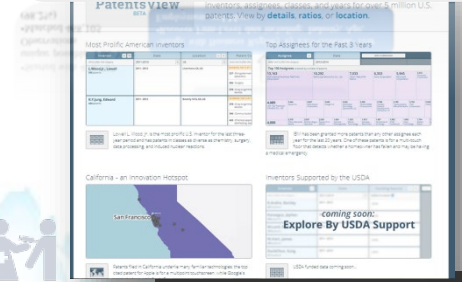
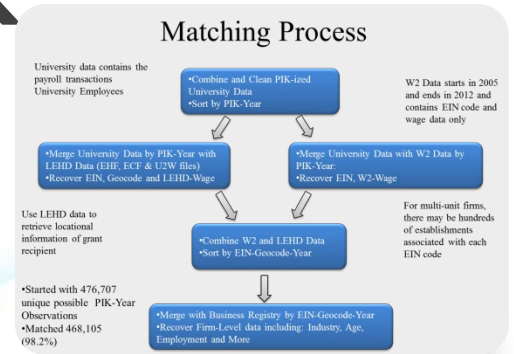
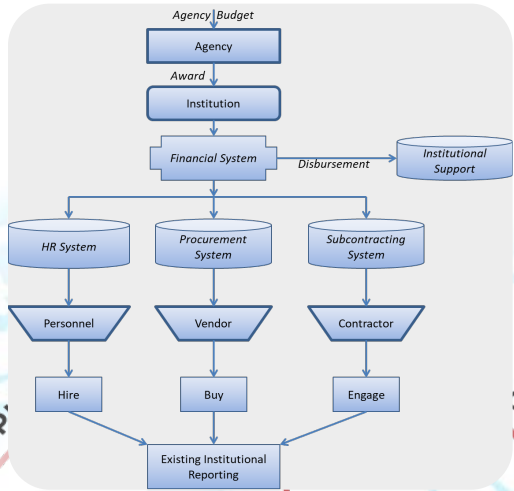
Context

Data

Evidence

Policy

# The state of data in 2022: Build a sensible framework





**IRIS:** IRIS links member data with other datasets, protects the security of the data, and produces and documents de-identified research data.



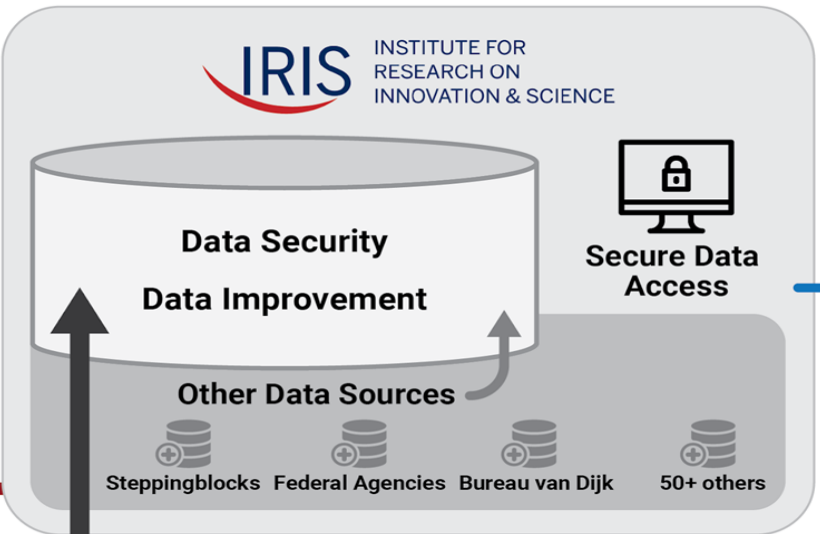
### MEMBER

**MEMBERS:** Universities contribute data, support IRIS infrastructure, and contribute to IRIS governance



### DATA REPORTS & TOOLS

**DATA REPORTS & TOOLS:** Members receive campus-specific reports and tools showing the internal and external impact of their research enterprise



**USERS:** Approved users securely access de-identified datasets



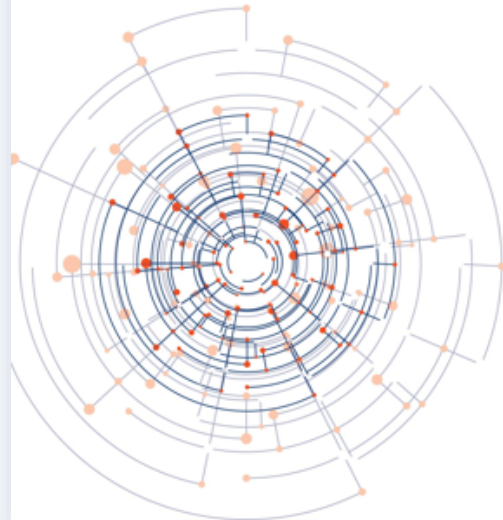
### USER



**PARTNERS:** Approved partners such as the U.S. Census Bureau and NCSSES receive data from IRIS which they improve and make accessible through their own secure systems

# IRIS/UMETRICS Dataset

- Researchers from around the world have used IRIS-UMETRICS data in nearly 100 published papers
- Fields: economics, higher education, science of science policy, operations engineering, and more...



## 2022 Data Release INSTITUTE FOR RESEARCH ON INNOVATION & SCIENCE

Nearly **480,000** funded awards at **80+** member campuses



**\$107 billion** in sponsored project spending



Payments to more than **880,000** vendors



Data collected from IRIS member universities from FY 2001-21



Wages to about **790,000** employees

# Using data to understand diversity Why do women publish/patent less?

## They really are less productive

Family responsibilities (e.g. Fox 2005; Hunter 2006)

Career paths and promotion differences

Less welcoming work environments

- Different positions in laboratory
- Different supervision

Sorting into research areas, interests, and specialization (e.g. Leahey 2006)

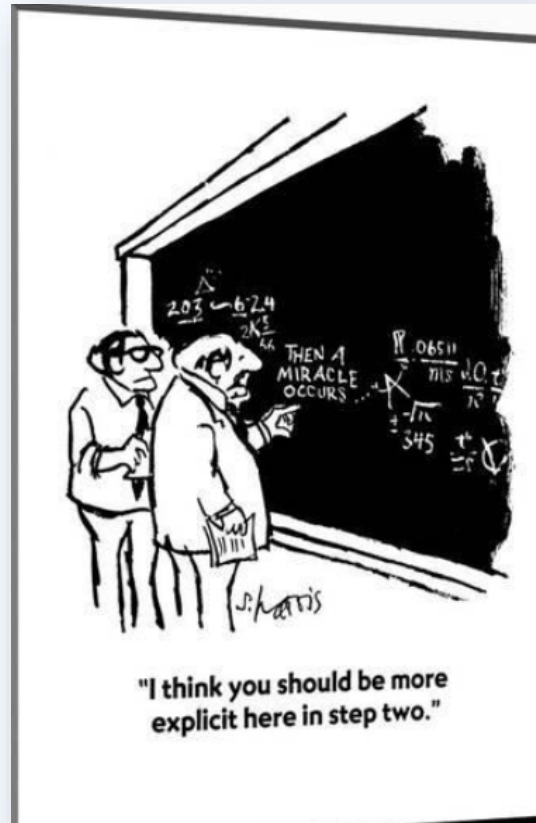
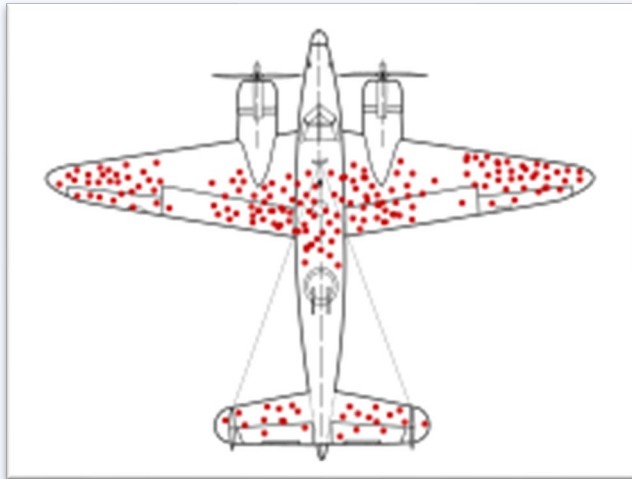
Response to failure (Subramani, Aneja, and Reshef 2021)

## They may be undervalued

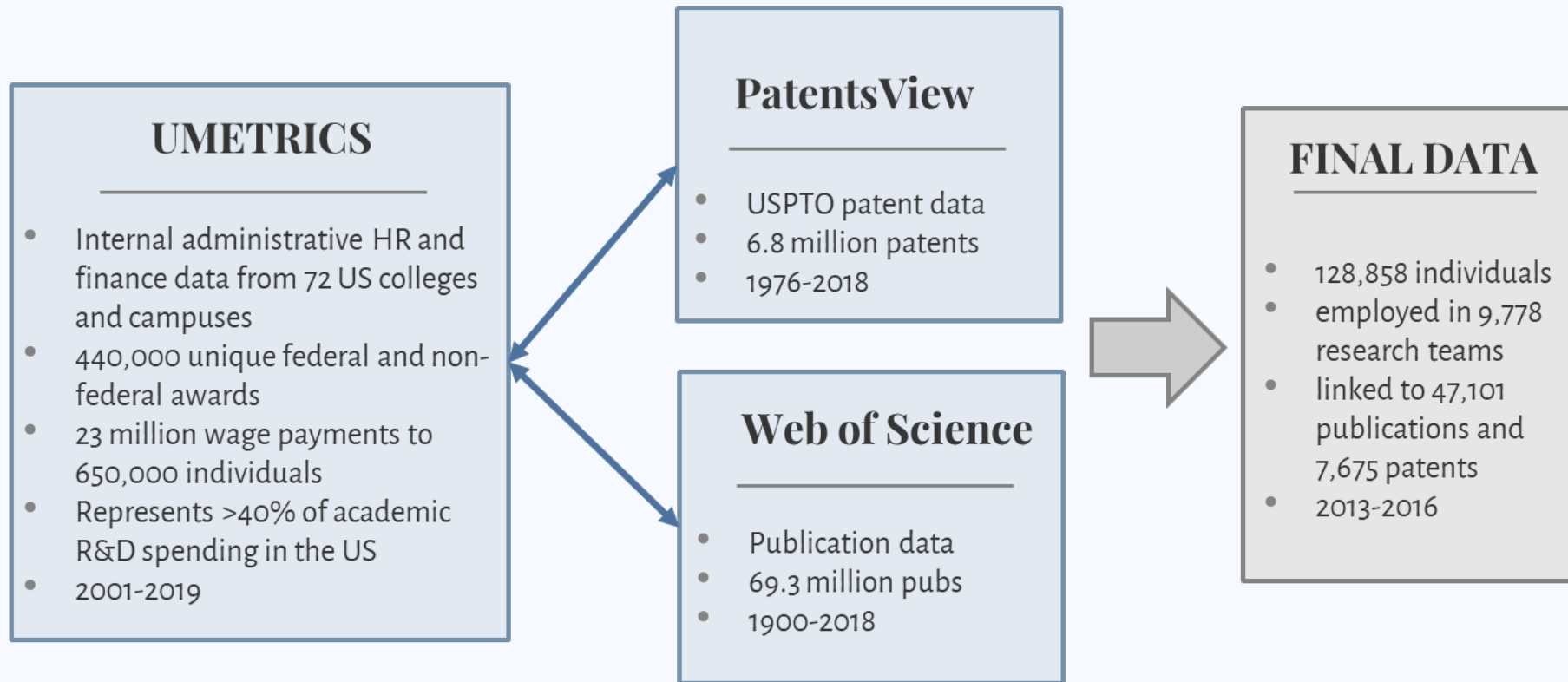
The screenshot shows the Nature journal website interface. At the top, the 'nature' logo is on the left, and navigation links for 'View all journals', 'Search', and 'Login' are on the right. Below the logo, there are links for 'Explore content', 'About the journal', and 'Publish with us'. On the right side, there are links for 'Sign up for alerts' and 'RSS feed'. The main content area displays the article title 'Women are credited less in science than men' with authors 'Matthew B. Ross, Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg & Julia L. Lane'. It includes publication details: 'Nature 608, 135–145 (2022)' and 'Published: 22 June 2022'. Metrics are shown as '65k Accesses | 15 Citations | 2414 Altmetric | Metrics'. A 'Download PDF' button is visible. On the right sidebar, there is an 'Associated Content' section titled 'The science of inequality' with sub-sections for 'Abstract', 'Main', 'Attribution and administrative data', 'Attribution and survey data', 'Attribution and qualitative data', 'Discussion', and 'Ethical approval'.

## Why data important?

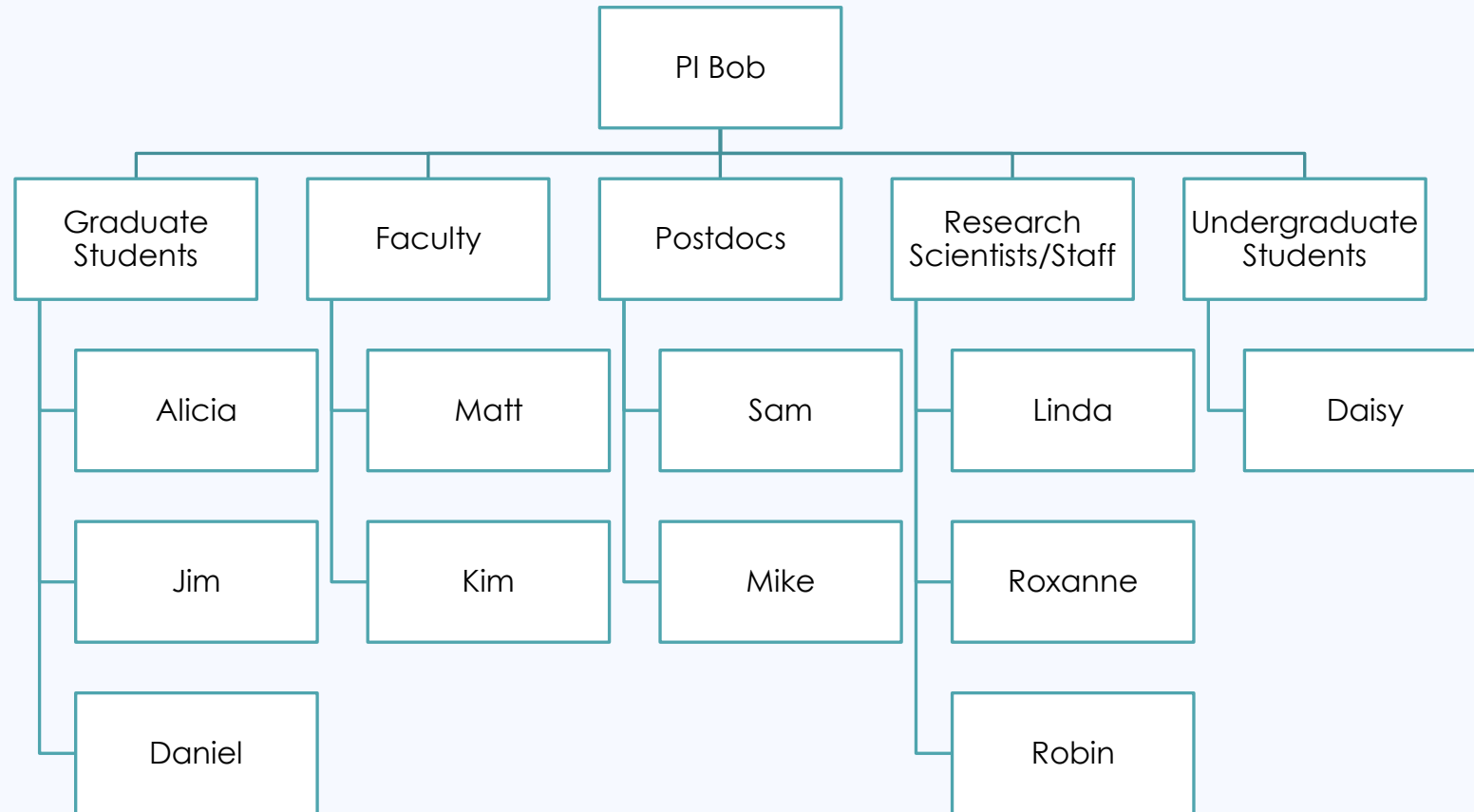
There is an inherent problem in relying on *observed* output to infer gender gaps in scientific productivity



# We combine three datasets to generate a rich dataset linking research teams to their scientific output



# A hypothetical team



Context

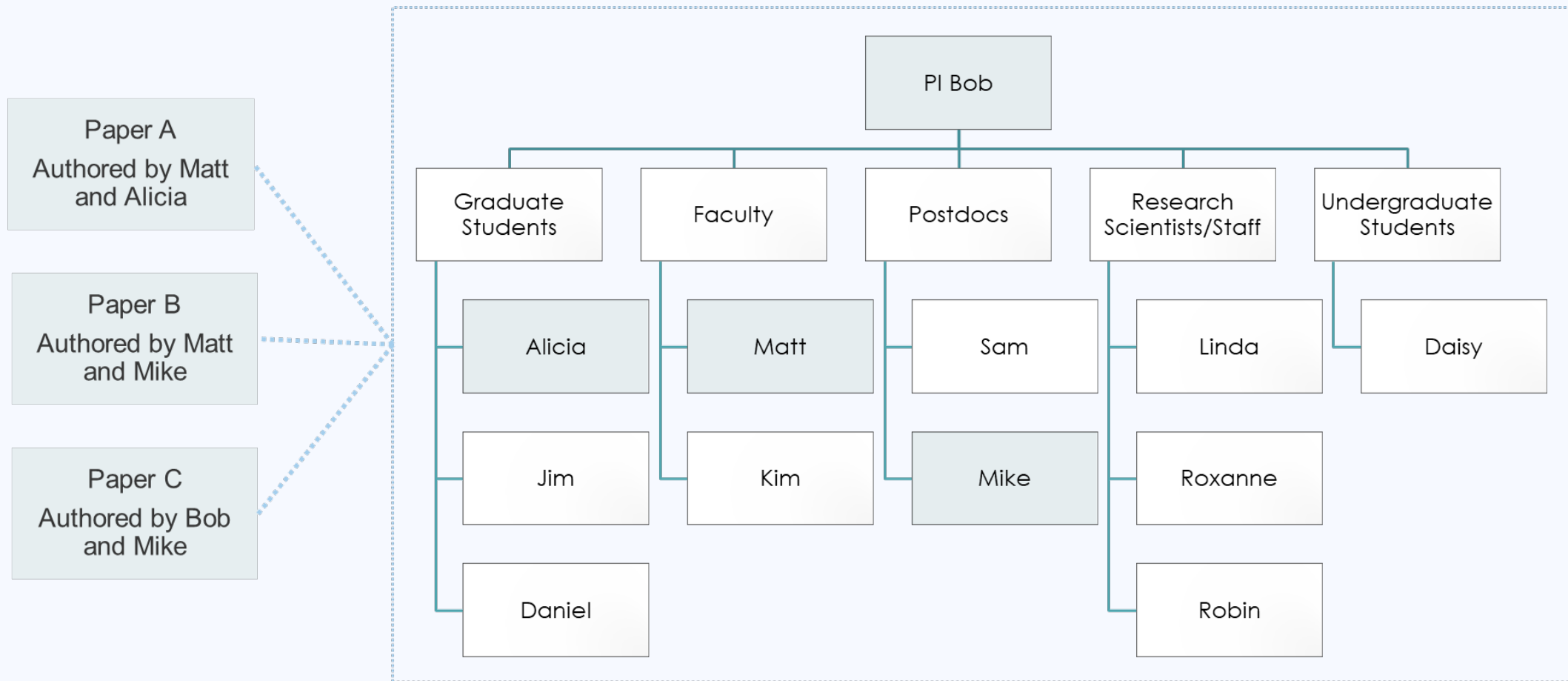
Data

Evidence

Policy



All members of this team are “potential authors”, but only some become “realized authors” on any given paper/patent



Context

Data

Evidence

Policy

## Different measures of attribution in the raw data

**“Ever Author”**: Ever cited as an author on any article or patent produced by research lab

- 17% of research team members are “ever authors”
- 21% for men; 12% for women.

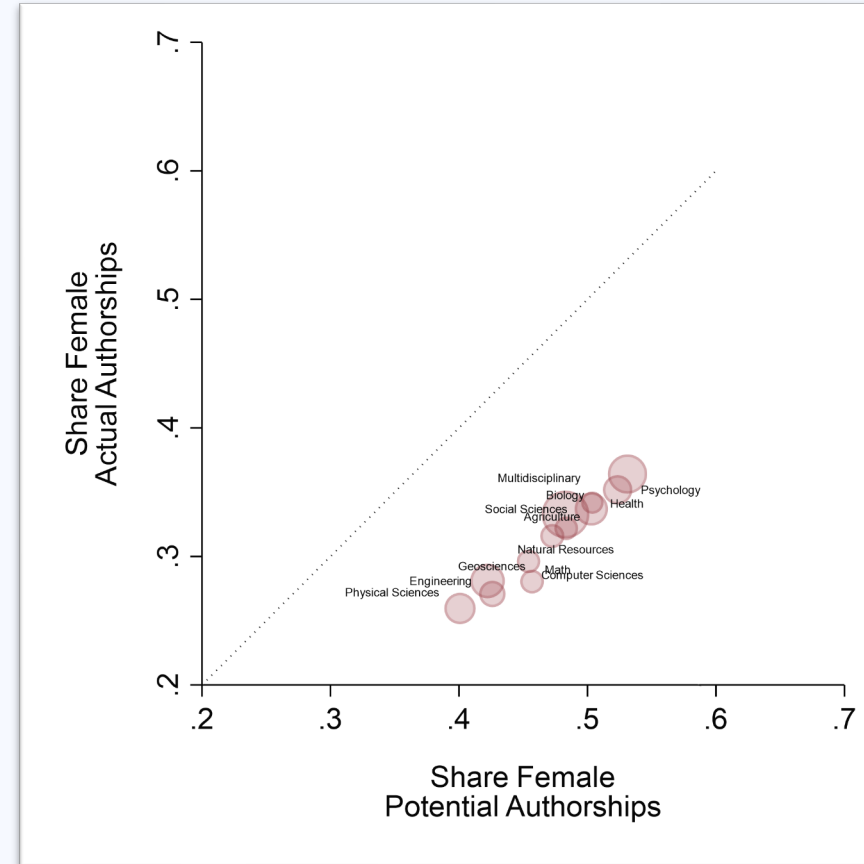
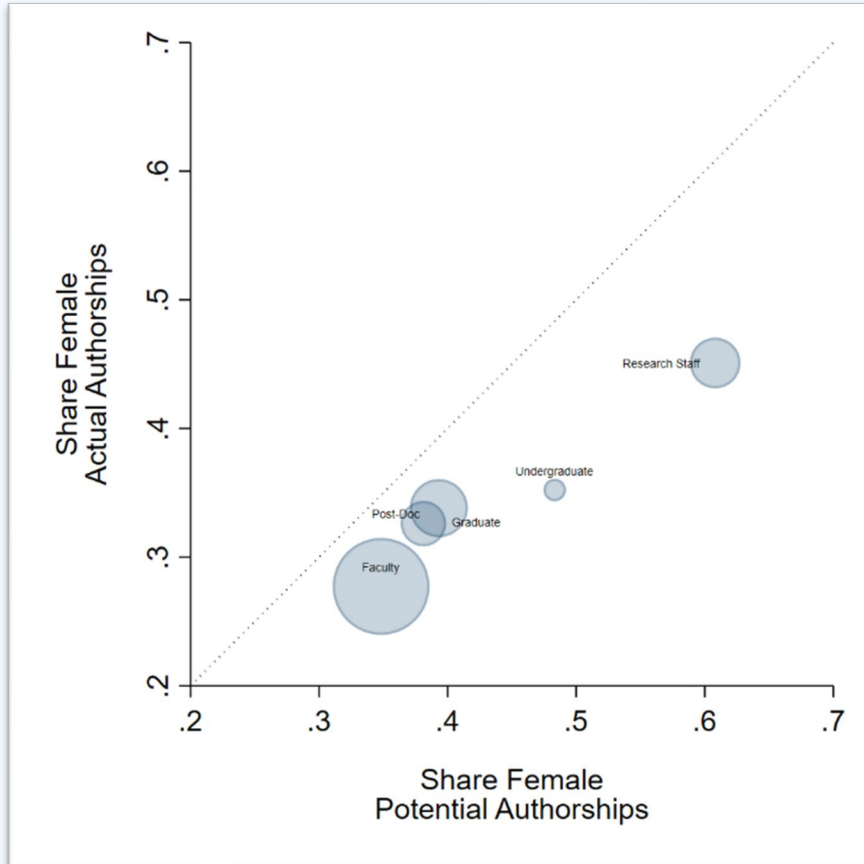
**“Ever Authorship”**: Cited on any given paper or patent

- 13.24% gap between men and women on article
- 58.40% gap between men and women on patent

**“High-impact Authorship”**: Cited on any given high impact paper or patent

- The higher the impact of the paper, the less likely they are to be named.
- E.g. an article with 25 citations is 20% less likely to be named

# Women are less likely to be named at all stages (L) and in all fields (R)



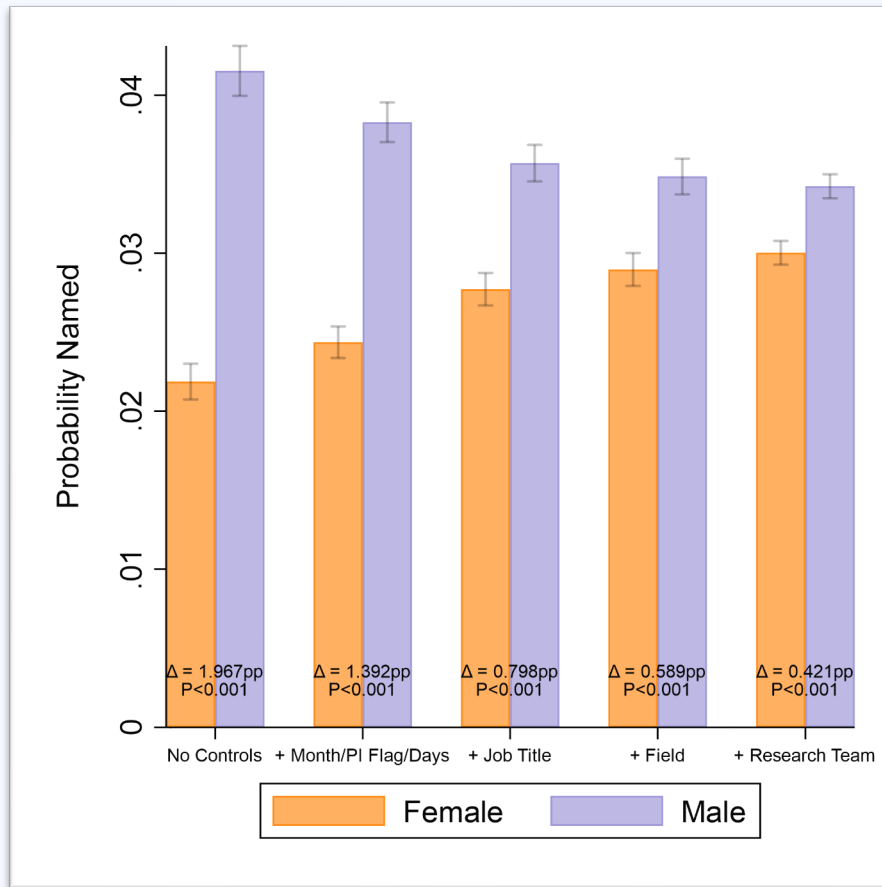
Context

Data

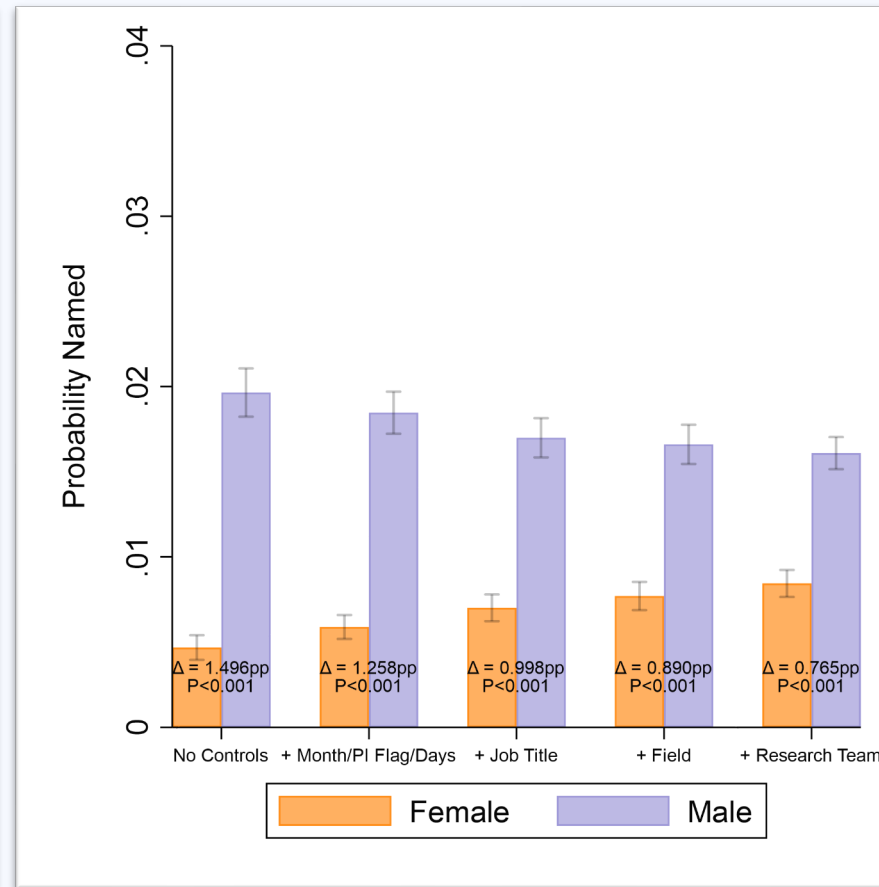
Evidence

Policy

# The gender gap in attribution persists even with controls

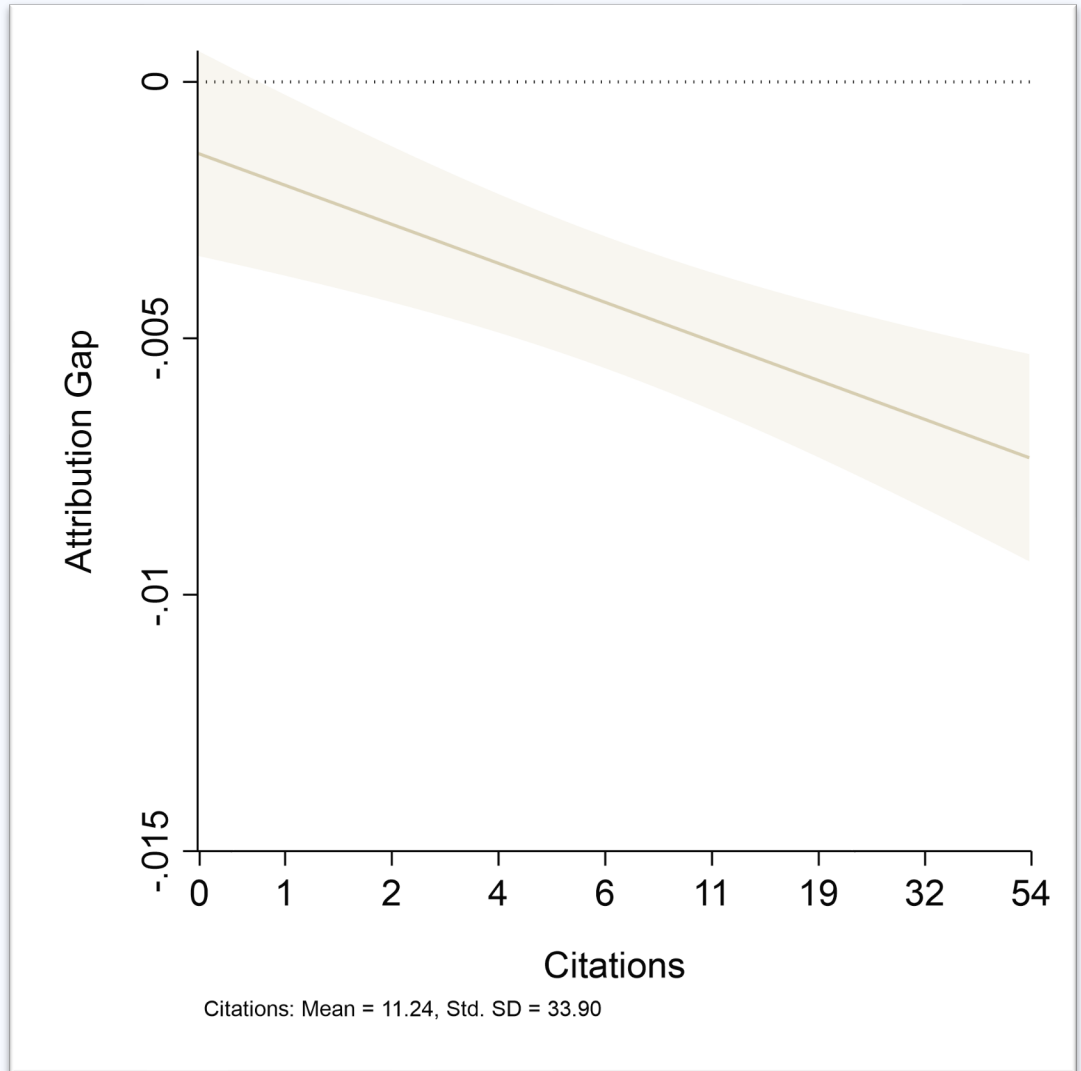


Articles



Patents

## Women's attribution declines with article impact



Context

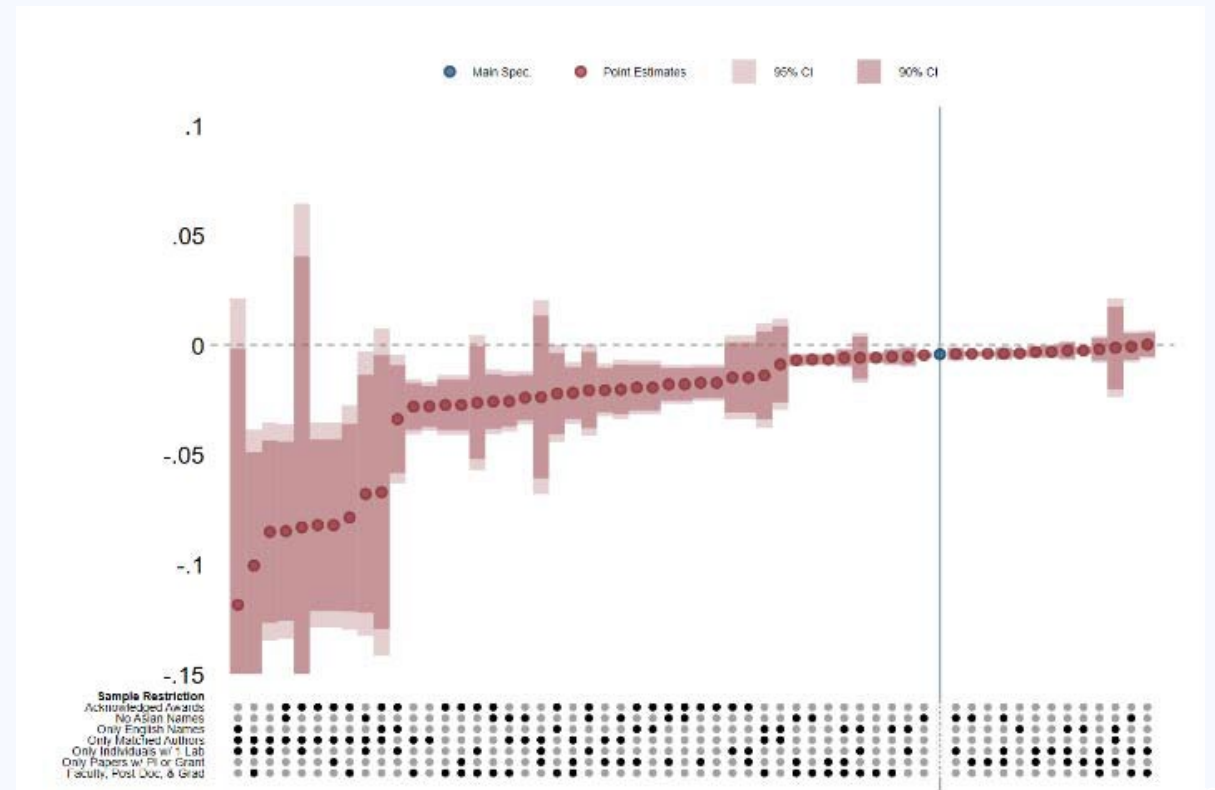
Data

Evidence

Policy

# The finding that women in research teams are significantly less likely to be credited with authorship is incredibly robust...

- Different gender imputations
- Different definitions of potential authors
- Different types of research output
- Timing of research output relative to employment
- Sample construction
- Logistic model
- Definition of time working in labs



Connect  
ions and  
Credit:  
How  
Social

Context

Data

Evidence

Policy

## Survey + interviews to understand mechanisms, corroborate results, and hear voices

*Dear [embedded name],*

*Advancement in science relies on individuals being credited for scientific contributions. We are asking you to provide input into our research study on how scientific credit has been allocated in research teams in which you have worked.*

*If you agree to be in this study, we would like you to think about a paper that was produced by a research team in which you worked and answer a series of questions about that paper, followed by some more general questions. Information on the paper comes from public data sources, such as ORCID and the Web of Science. This work has been approved by the University of Pennsylvania IRB.*

*The survey will take fewer than 5 minutes.*

*We will send you the aggregate results when they are finalized. Your responses will, of course, be kept confidential.*

*If you have any questions about the survey, please contact Britta Glenmon at [bglennon@wharton.upenn.edu](mailto:bglennon@wharton.upenn.edu) or the IRB at 215-898-2614.*

Authors who appeared on at least one WoS article after 2014 and who had an e-mail address

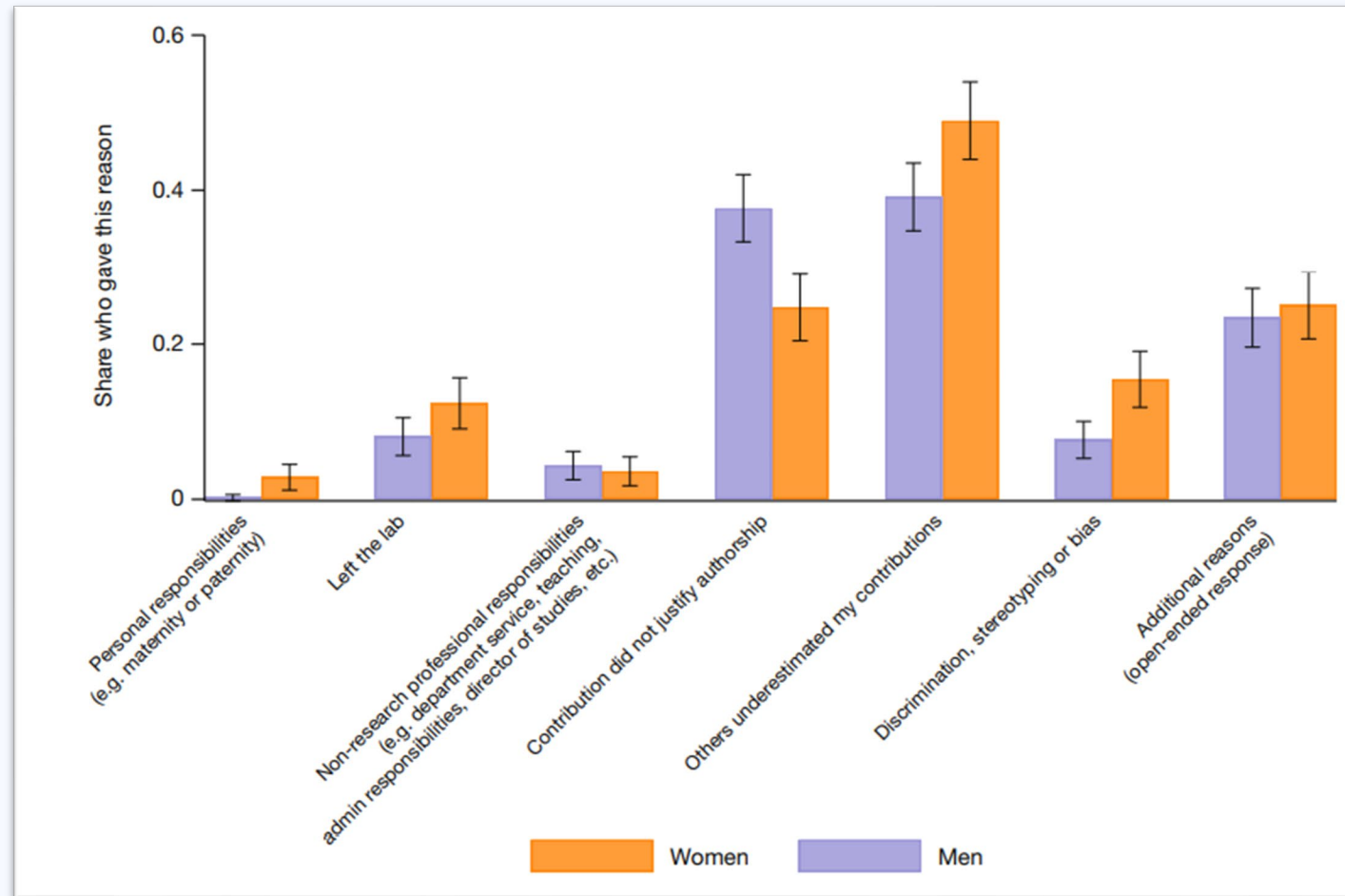
Asked about:

- Their roles on named papers
- Whether they had been left off of other papers and why

Of the 2,660 responses, 43% of women and 38% of men report being excluded as an author ( $p=.015$ )

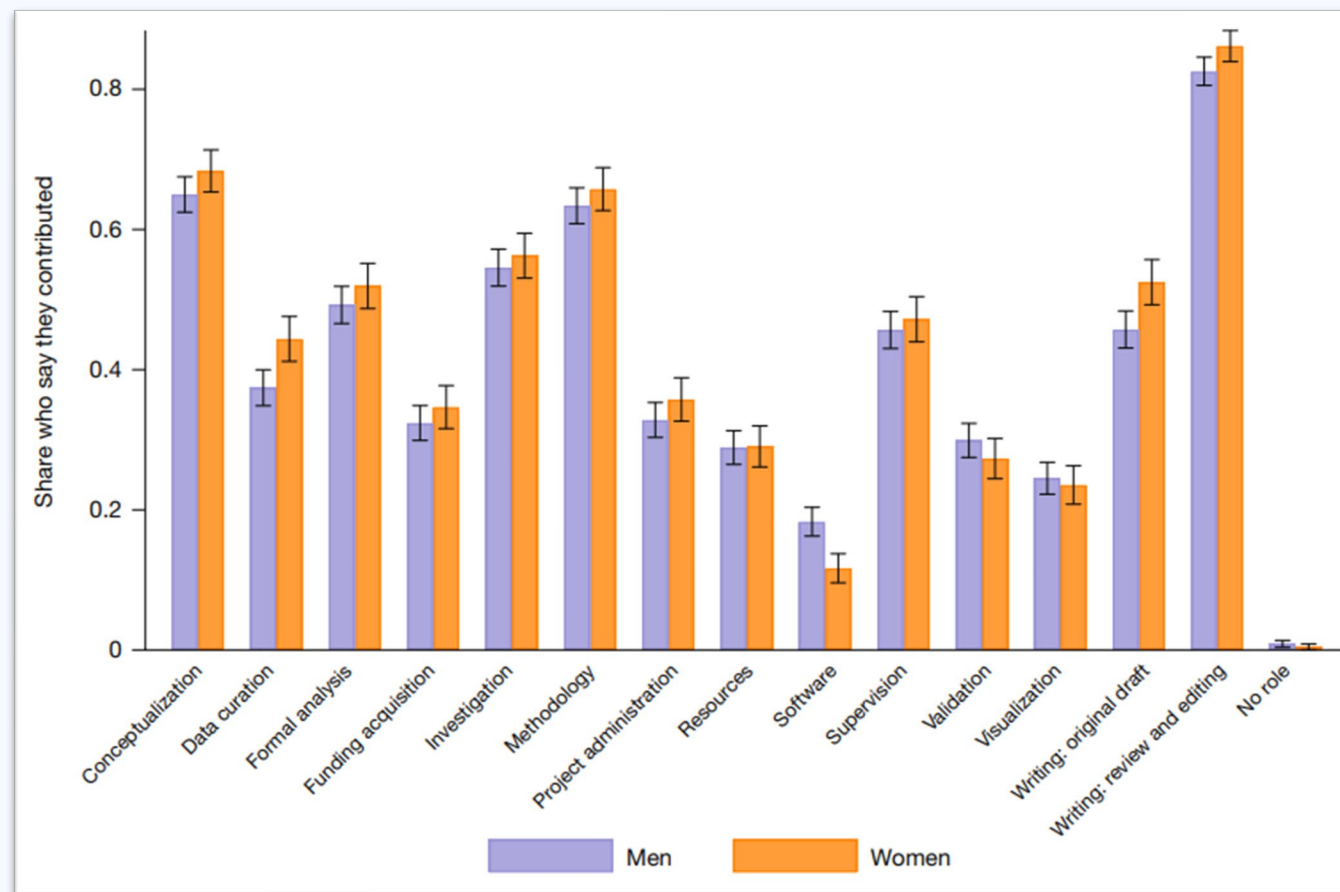
Optional open-ended responses + willingness to be interviewed

## Women are more likely to report that their contributions were underestimated or that there was discrimination





# Women report making more contributions than men on authored papers



( $\mu_{Women}=6.34$ ;  $\mu_{Men}=6.11$ ;  $p=.09$ )

## **Narrative responses: Women don't ask...but asking can also backfire**

“I did not push to be listed as an author.” (F)

“Senior authors shamed me in front of group for asking for recognition (trying not to be a female-doormat stereotype backfires pretty much every time I have tried...)” (F)

“I was scared to make too much of a fuss” (F)

“when one of them tried to express her feelings, she heard she was too sensitive. From then on, we all just "accepted" the situation.” (F)

“I have also had to fight to have my name included on papers in which I have heavily contributed. As a result of this fight, I have not been excluded from papers; but, this has created rifts in the workplace. I have had multiple individuals refuse to work with me following such discussions.” (F)

## Narrative responses: Power dynamics and personal preferences

“Publications were used as reward and punishment.” (F)

“This is not that uncommon an occurrence. It is largely related to bias and favouritism in the lab management.” (F)

“The problem of adding/omitting the correct authors to a paper is not uncommon. It is used to disadvantage or advantage someone for his/her career. It is used by seniors/PIs just because they have more power over other colleagues” (F)

“As a young RA I did all the data analyses and a lot of the writing for a paper but was working for a arrogant older male.” (F)

“I have seen friends in this situation, where some undergrads and grad students were "overlooked", while other "contributors" higher up in the academic hierarchy were included in papers despite not having contributed at all.” (M)

“The person who changed the order of the names was in a power position and I could not say anything.” (F)

“Academia is so competitive to gain tenure, grants etc that some people will take any opportunity to limit career progression of colleagues and collaborators and limiting authorship is one of the passive means to achieving this outcome” (M)

“Most of my fellow academics (especially women, and most especially women of color) have been harmed by faculty who decide to award authorship to other lab members who did not do the work. Sometimes authorship is removed because of lab favoritism; sometimes authorship is removed as a retaliatory move when the trainee reports harassment or bullying.” (F)

“In cases where the corresponding author didn't 'like' a lab member, he would not give authorship.” (F)

Context

Data

Evidence

Policy

## Narrative responses: Lack of clear criteria

"My experience is that the definition of intellectual contribution is sometimes not clear in certain projects and some PIs may not have fair criteria." (M)

"I have found through 30 years and three continents that often women's contributions are considered "being nice and helpful" and not acknowledged as scientific contribution" (F)

"The culture of inclusivity in listing people as contributors varies considerably within my field. Some researchers consider a long author list as a dilution, or want to avoid author-ordering discussions." (M)

"We don't have institutional guidelines to help determine authorship...Authorship here is directly linked to economic benefits and it is a means to control the income of research staff. This is slowly improving but it remains unfair" (M)

"A murky situation arises in collaborative team-based sampling collection and archiving where the junior folks do a huge share of sample collection, labwork, management/organization, etc of biological materials that later become resources." (M)

"I believe that most of the time there is a lack of clear "rules" regarding who should be included in a paper. Each research group makes its own "rules" and they usually change to one manuscript to another, especially when there are collaborations involved." (M)

"Some PI's I have published with have an overly-rigorous metric: "If someone didn't directly make a specific figure, no authorship". Others are even more obnoxious: "No undergrads or techs on papers - regardless of their level of contribution." Other interdisciplinary papers (mostly with a clinical component) only list the "main" clinical contact as an author, not the residents or fellows who actually did the work." (M)

# Consequences for research and innovation

Discouragement of early career researchers

Noisy productivity signals

Other groups

- Minorities
- Immigrants
- non-native speakers
- First-gen
- “quiet voices” in general

# What can be done in science?

Science is now big science

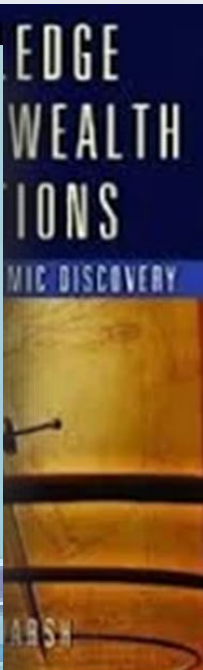
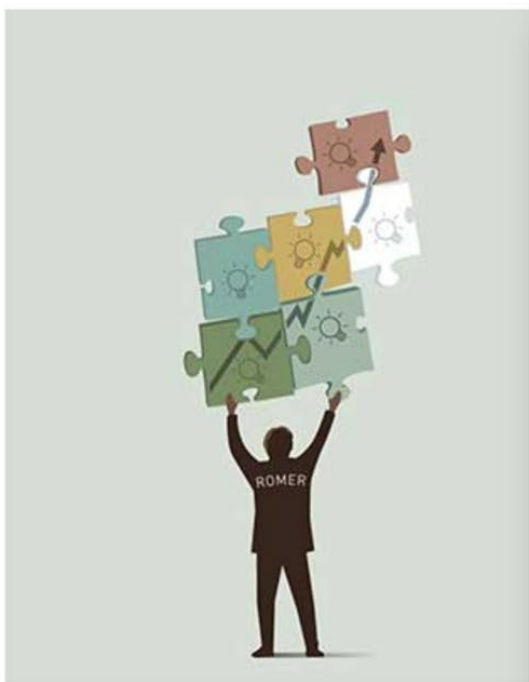
- Training in grad school
- Training for PIs
- Communicate as PI
- Communicate as team member
- Project CREDIT

Better measurement

- The effect of policies instituted by the research institutions on science quality
- Student retention, placements and career trajectories
- Business startups

# Broader implications

Low  
New  
New  
New  
.....



Romer's research shows how the accumulation of ideas sustains long-run economic growth.

Context

Data

Evidence

Policy

# Questions?

Julia Lane

[Julia.lane@nyu.edu](mailto:Julia.lane@nyu.edu)

<https://www.linkedin.com/in/julia-lane/>