Simulation-based Inference (SBI)

Analysis, Design, and Operations in Science Experiments

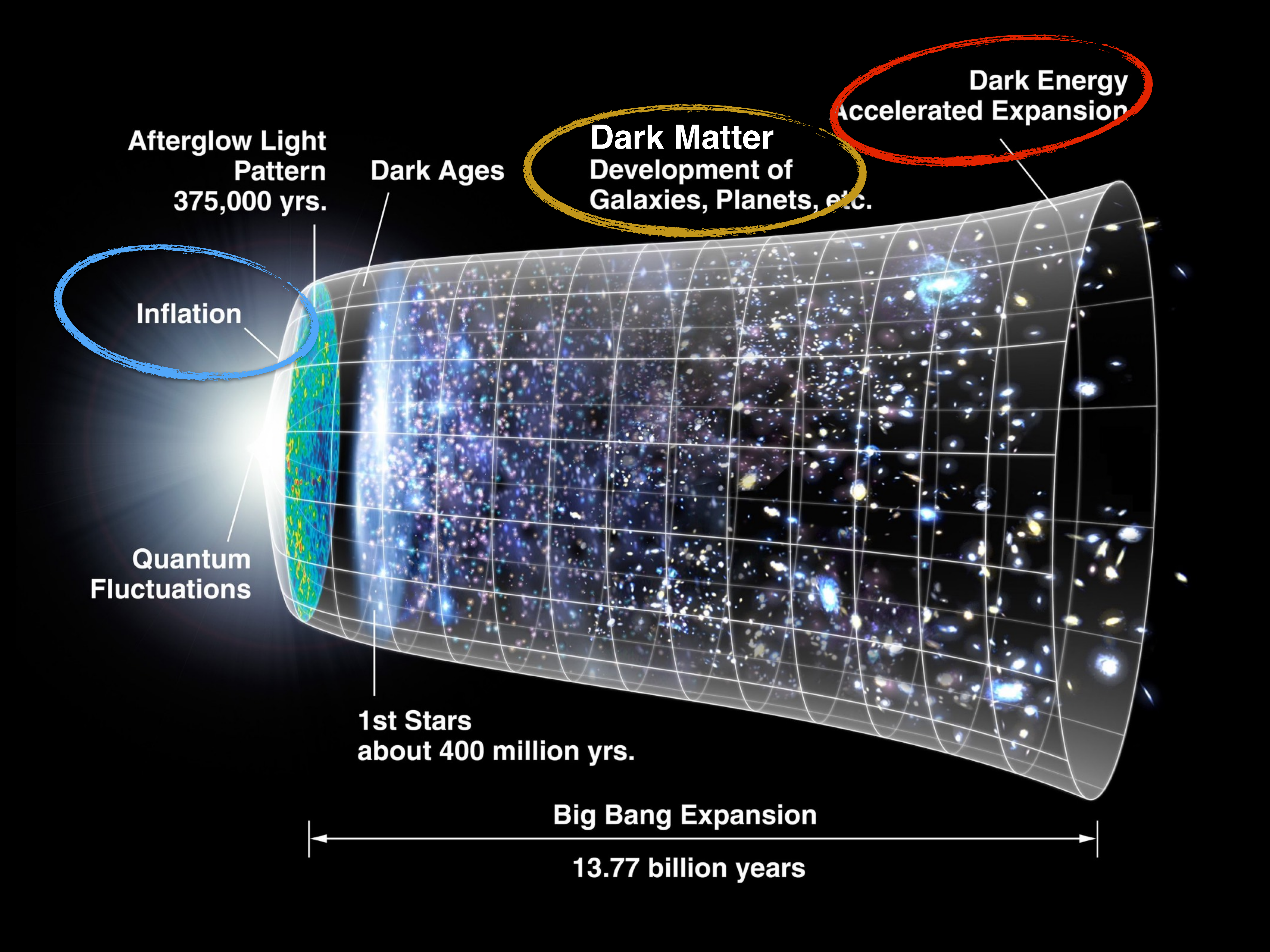
Brian Nord
Deep Skies Lab
Fermilab
University of Chicago

2025 May 16
CMU STAMPS Seminar

Today's questions

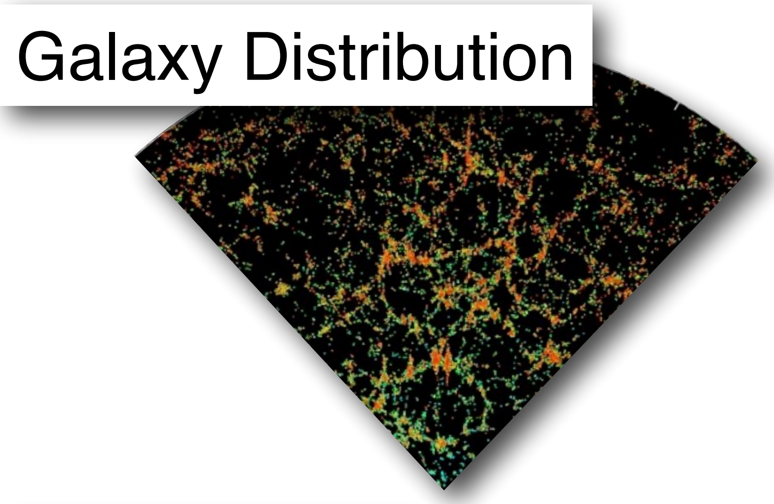
- What is the state of the field for SBI in data analysis?
- What is the future of SBI and other AI methods for designing and operating science experiments?

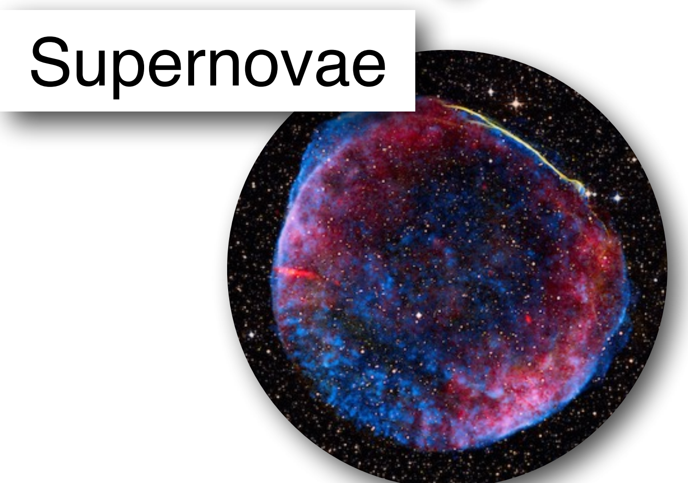
All the stars are closer

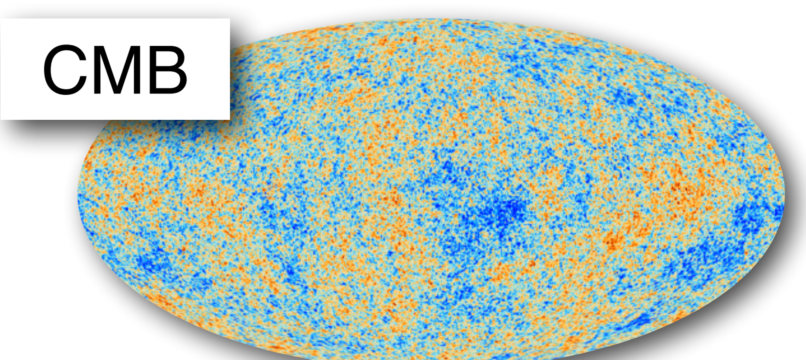


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
375,000 yrs.
Dark Ages
Dark Matter
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.77 billion years

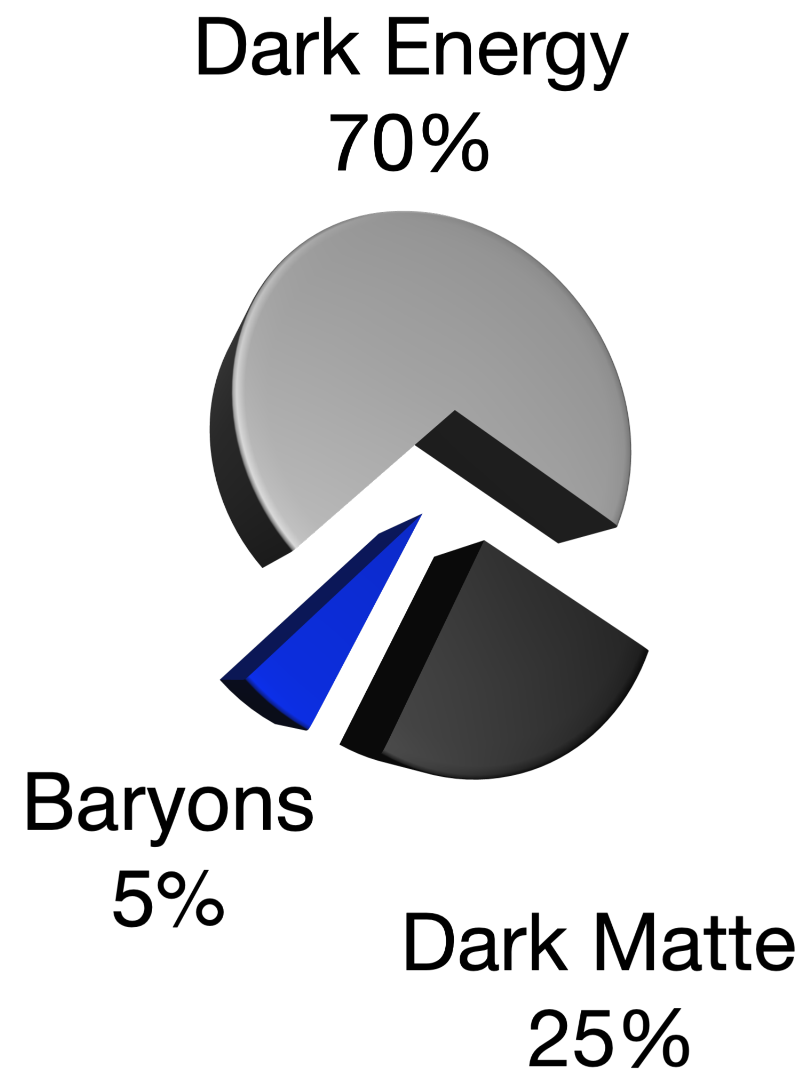
Path to the Modern Cosmological Paradigm





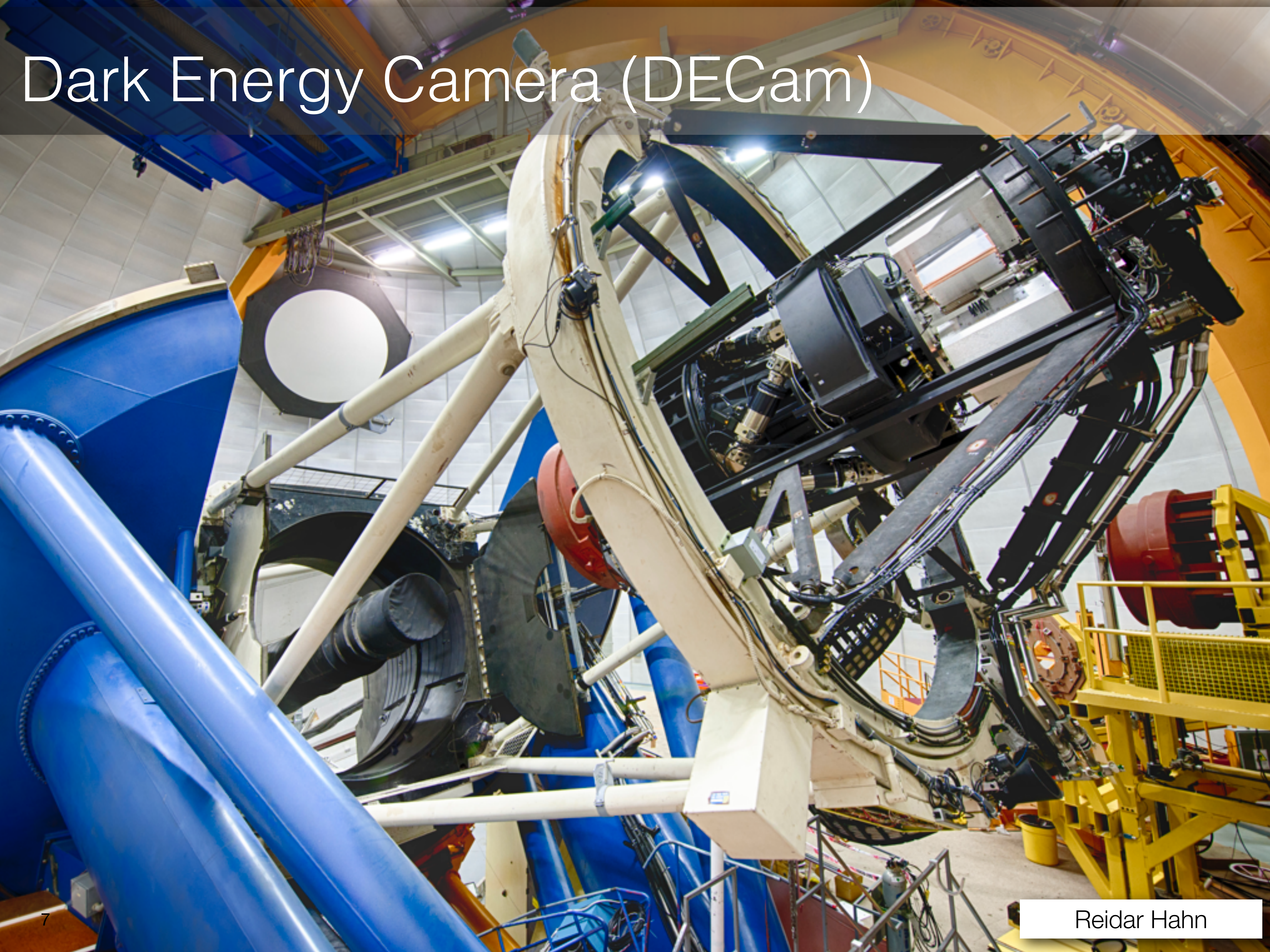


Structure Growth
+
Expansion
+
Geometry



Dark Energy Survey (DES)

Dark Energy Camera (DECam)

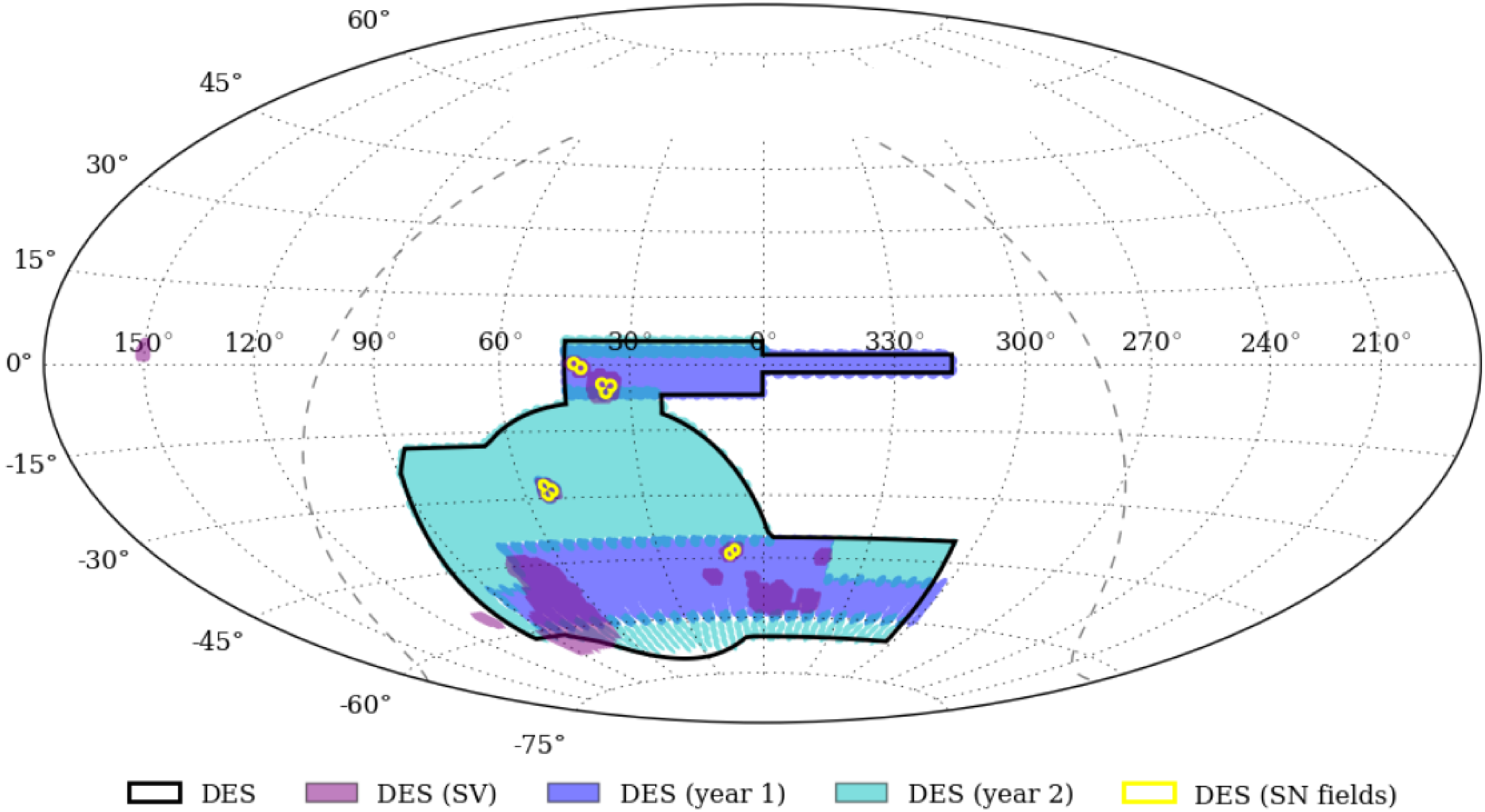


Reidar Hahn

DES Data

This is one of 72 CCDs in the DECam focal plane.

DES Footprint



- <u>Total area</u>: 5000 sq. deg. (~1/8 of the full sky)

Evolution of Optical Survey Experiments

SDSS I-II
(2000-08)

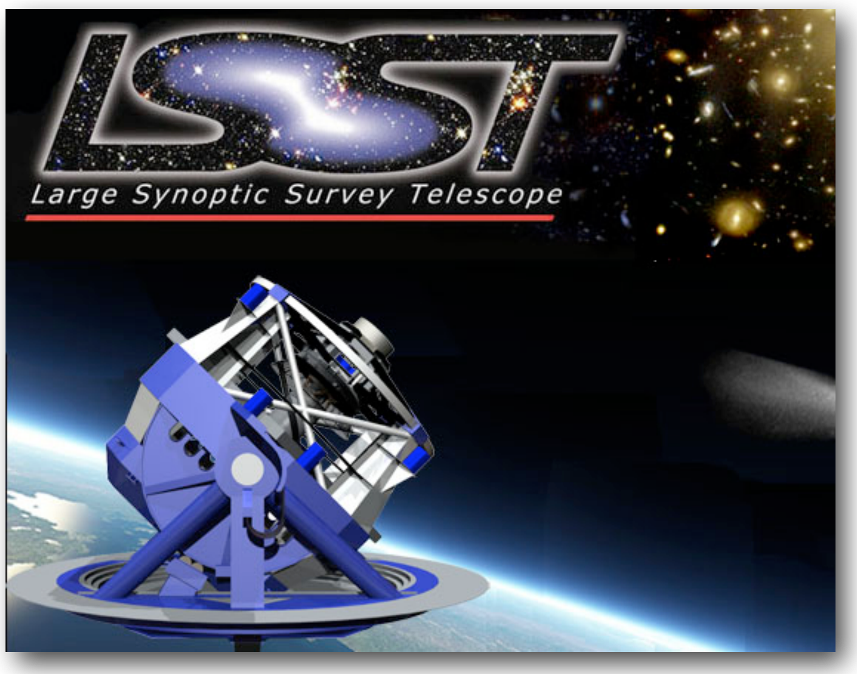
100M Galaxies
10k sq. deg.
0.2 Terabyte/Night

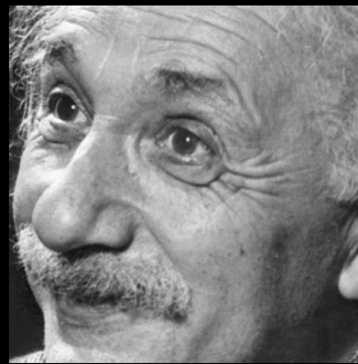
DES
(2013-19)

100M Galaxies
5k sq. deg.
1 Terabyte/Night

LSST
(2025-35)

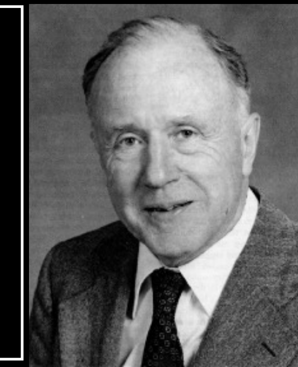
1000M Galaxies
20k sq. deg.
20 Terabyte/Night

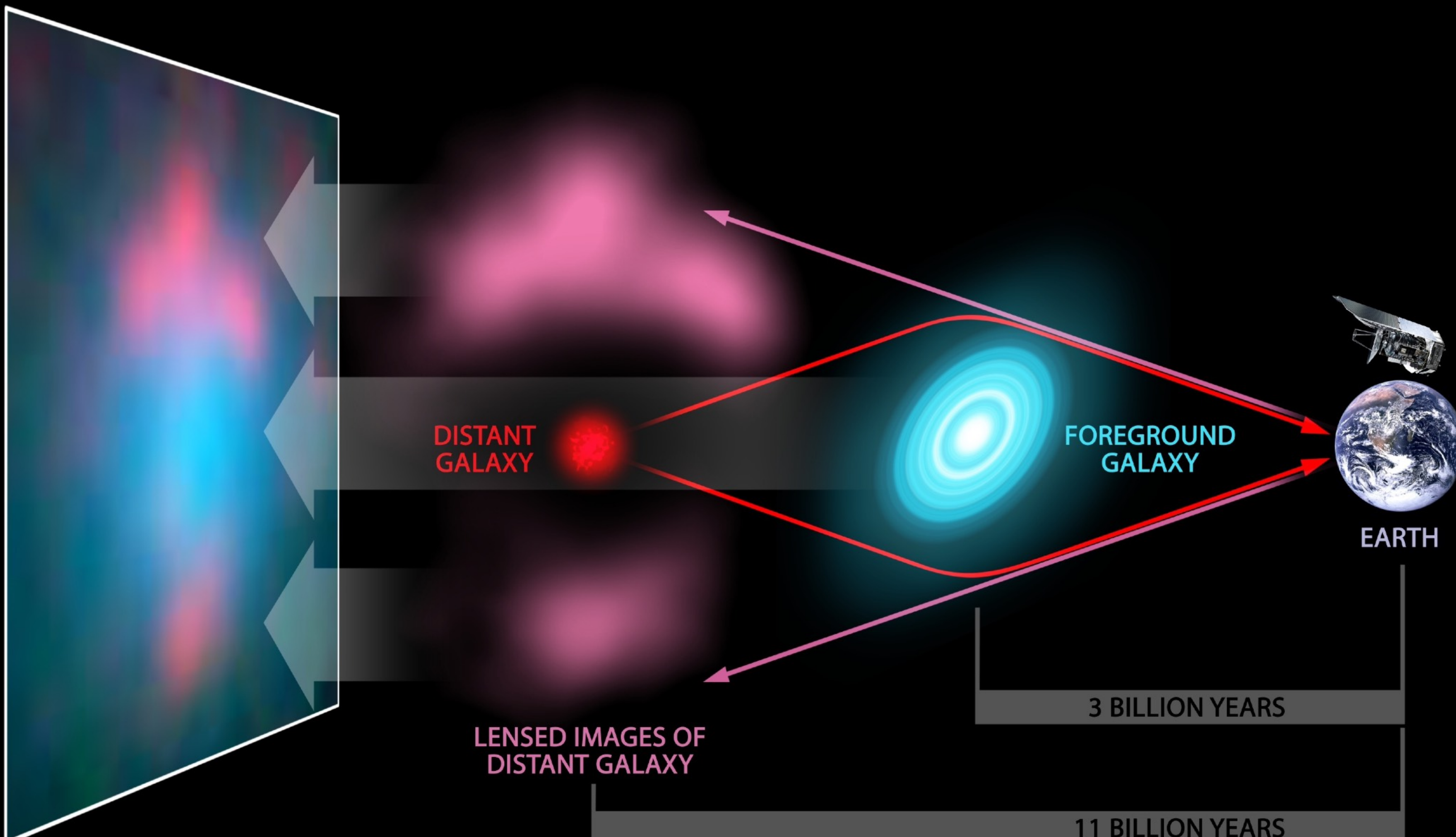




"Energy tells space how to curve, and space tells energy how to move."

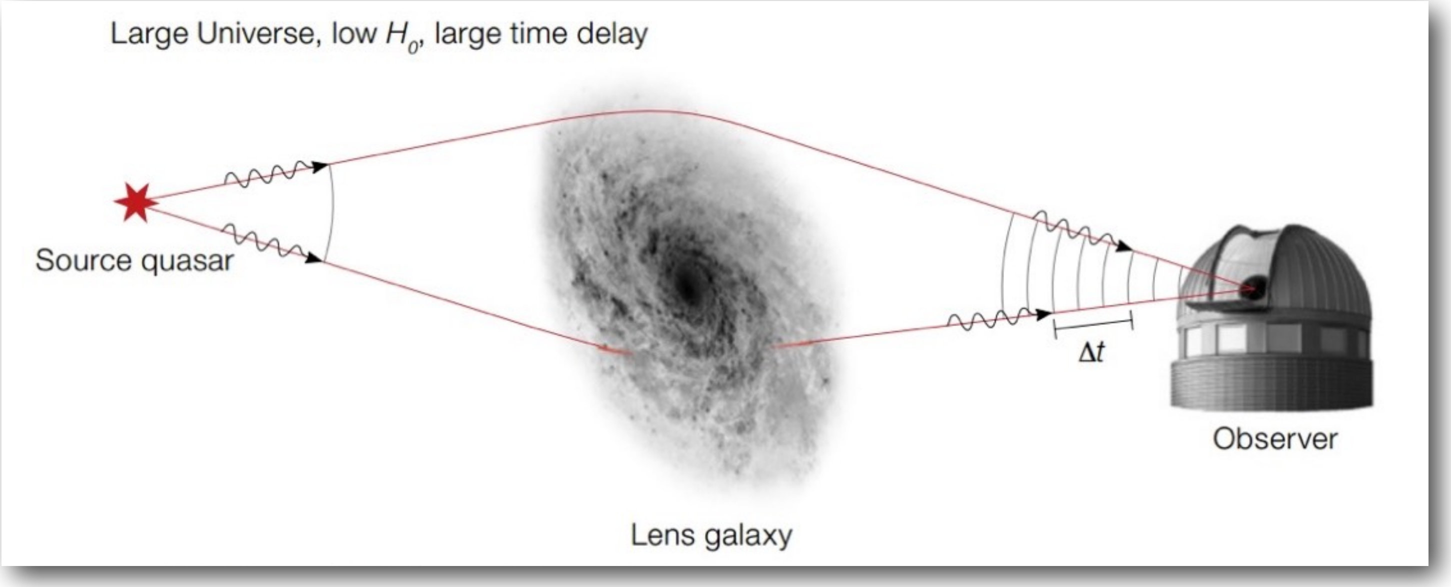
—John Wheeler





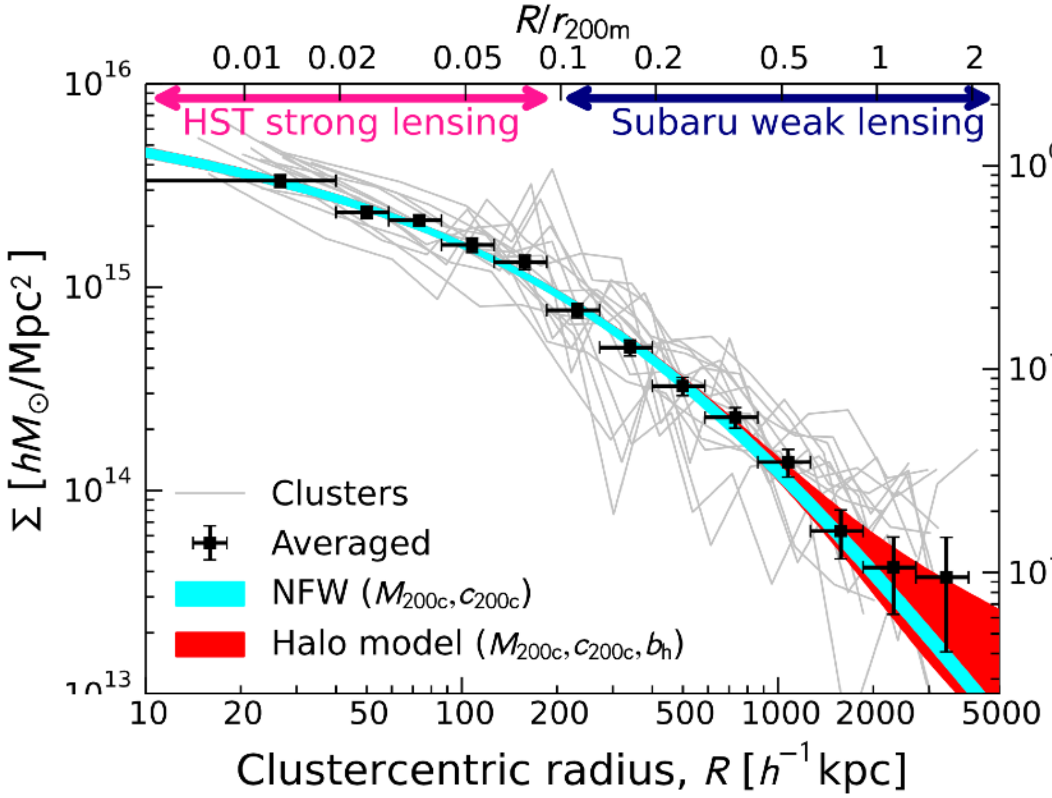
Cosmology with Strong Lensing

- Lensed quasars and supernovae
- ***The time delay*** between different light paths is proportional to the $\boldsymbol{H_0}$



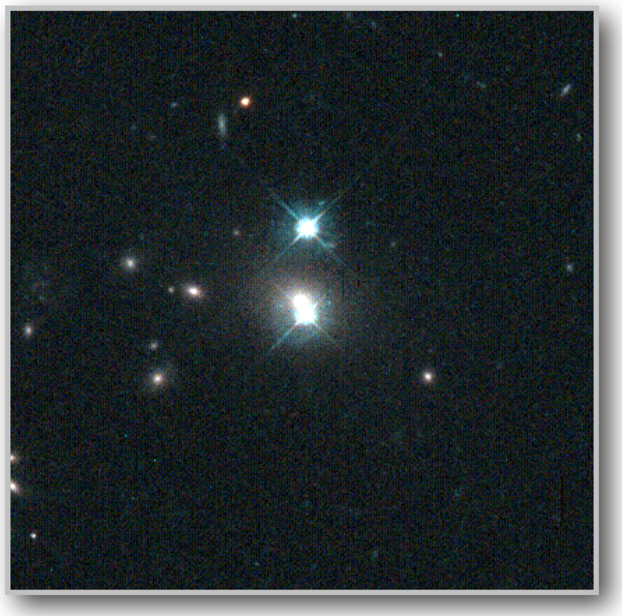
- Double-source plane systems
- Ratio of **distance ratio**s constraints **dark energy**

- Cluster and galaxy profiles
- Profile **slopes** indicate amount and type of **dark matter**

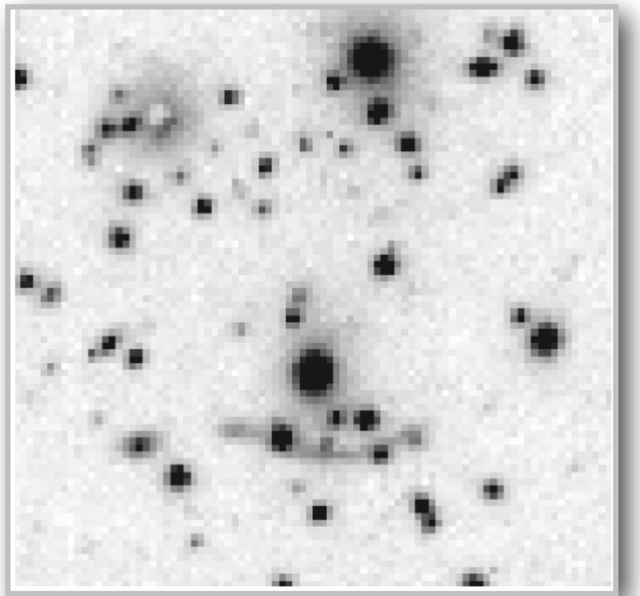


Milestones

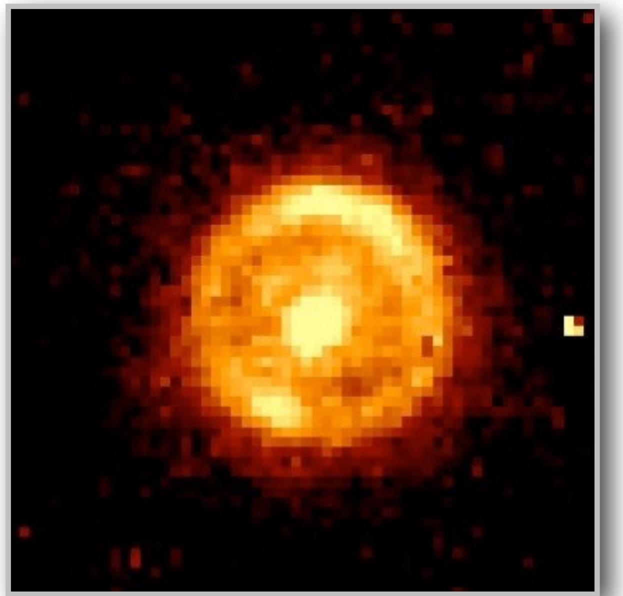
- **<u>1979: Quasar</u>**
  Twin Quasar SBS 0957+561
- **<u>1986: arcs</u>**
  Cluster Abell 370
- **<u>1998: Einstein Ring</u>**
  Galaxy JVAS B1938+666
- **<u>2014: Supernova</u>**
  Cluster MACS J1149.6+2223



- Walsh, Carswell, Weyman 1979



- Lynds & Petrosian 1986;
  Soucail+1987



- King+1998

- Kelly+2014

Too many to count: a paradigm shift

Lens type

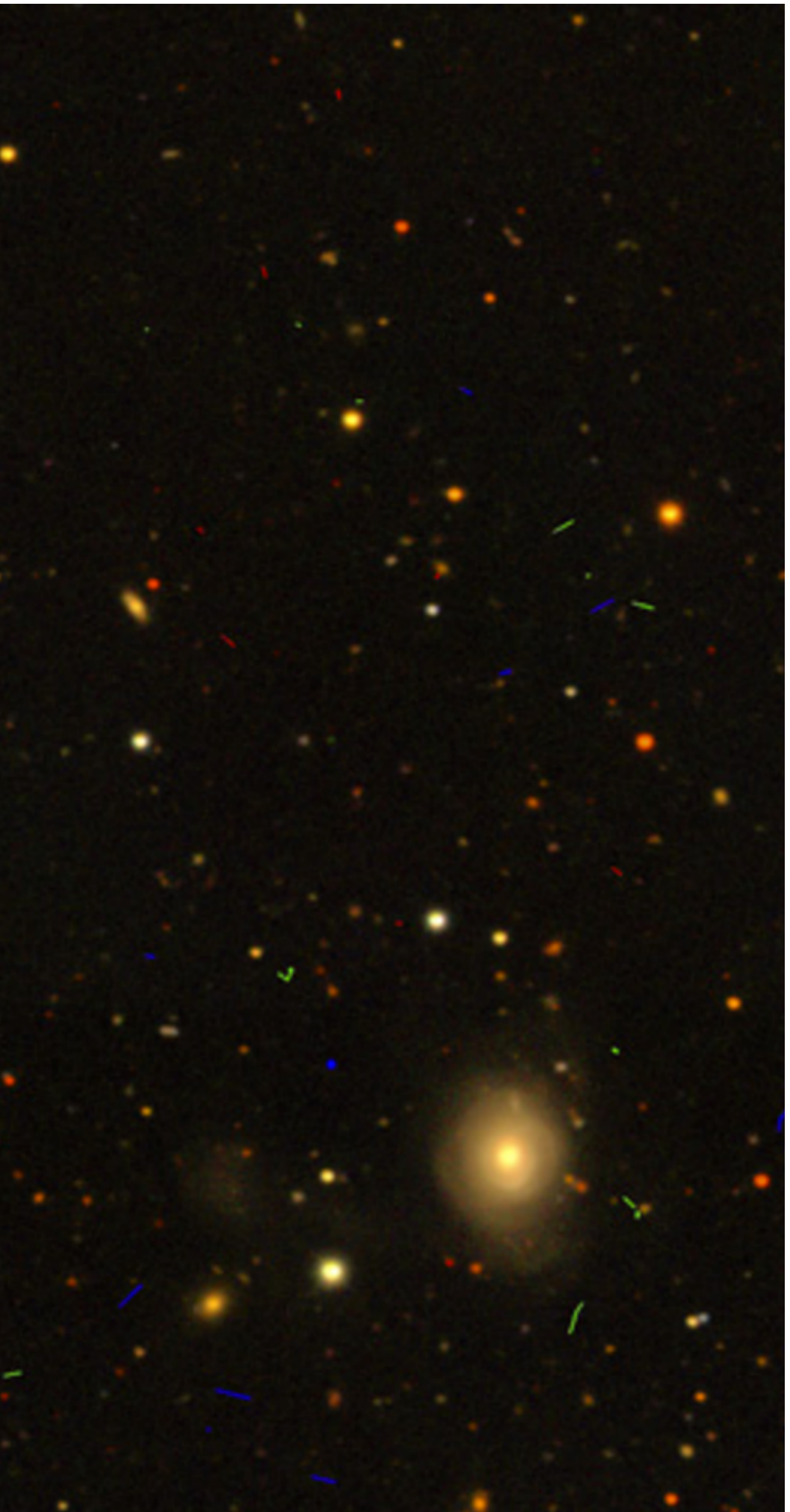
Survey

	Galaxy	Quasar	SNe
Today	**1000**	**<50**	**~2**
DES	**2,000**	**120**	**5**
LSST	**120,000**	**8,000**	**120**
Euclid	**170,000**	**-**	**-**

Nord+2016; Collett+2015; Gavazzi+2008; Oguri+Marshall, 2010

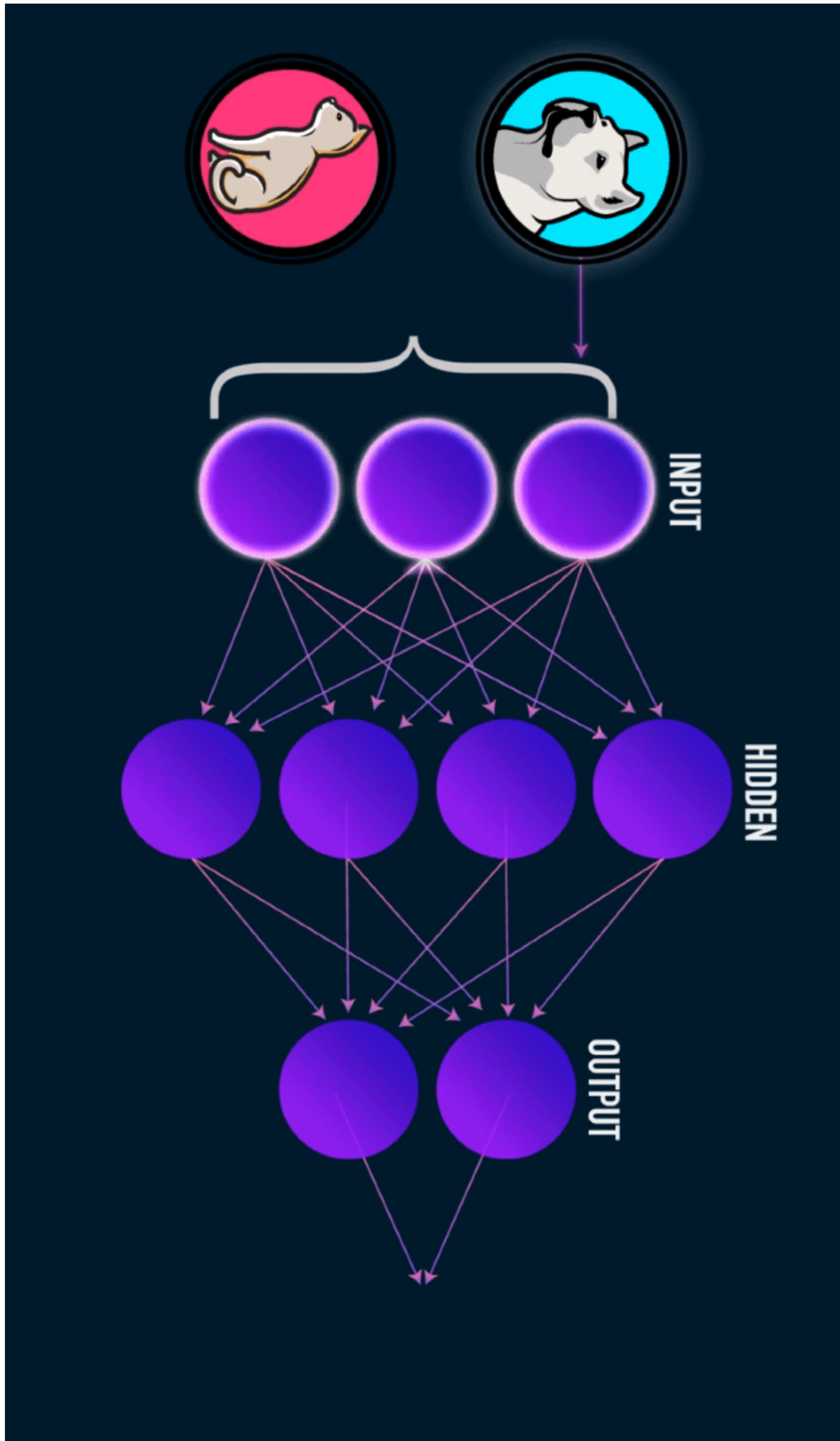
Evolving Size and Complexity of Experiments & Data

Data

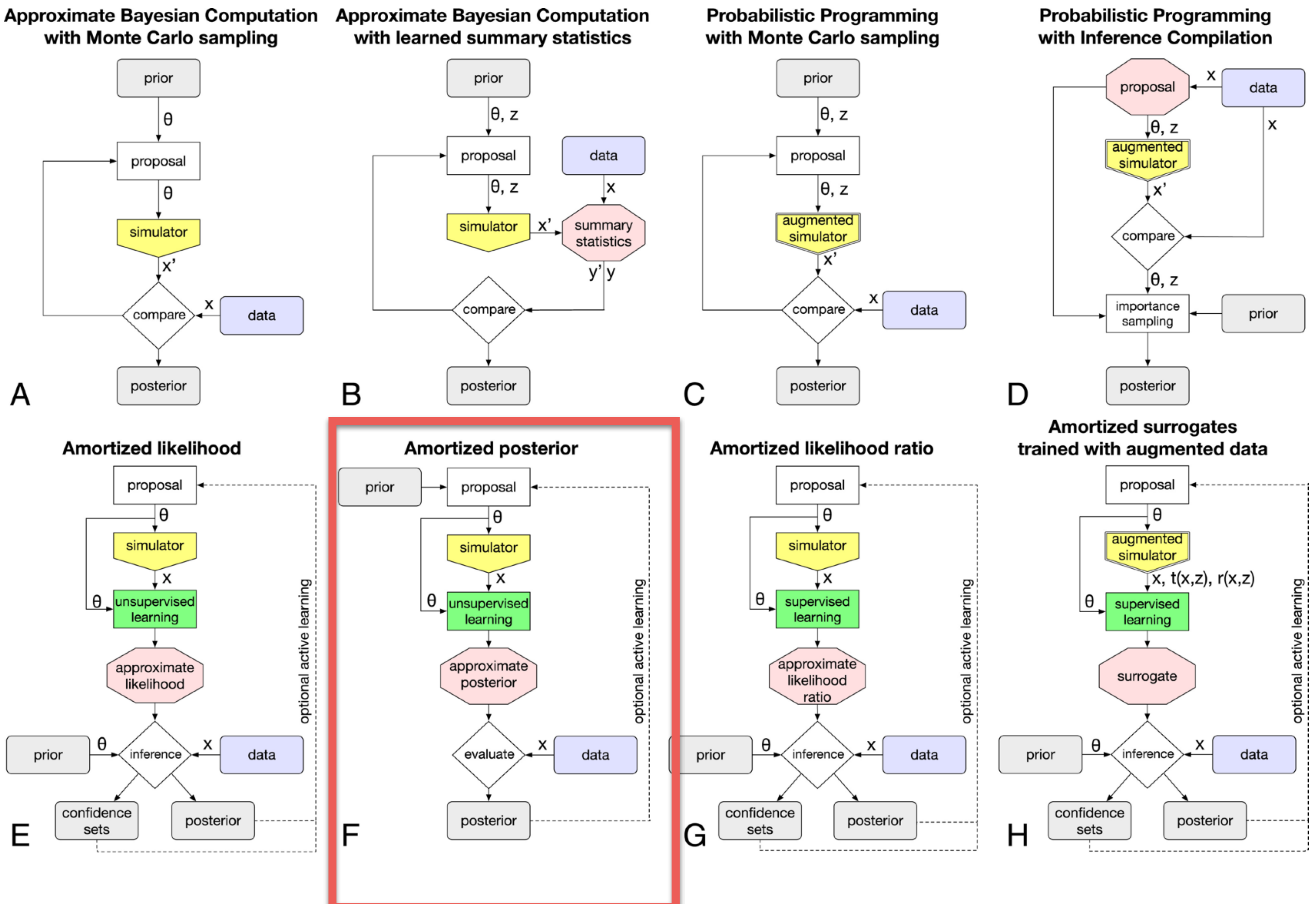


Instruments

Models



Modern Simulation-based Inference



- Cranmer, Brehmer, Louppe, 2020

Simulation-based Inference: Definition and Context

- General characteristics of SBI
  - **Prediction** of latent (physically meaningful) parameters.
  - **Amortization** and training with simulations
  - **Expressive densities** for comprehensive uncertainty quantification

- Family of SBI-like things
  - Anything that provides an aspect of a density
    — e.g., regression models with UQ, MC Dropout, Deep Ensembles.

- History of SBI-like things
  - 1990's: "Indirect Inference" — economics
  - 2000's: "Likelihood-free Inference" — physics and cosmology
  - 2010's: "Simulation-based Inference" — particle physics

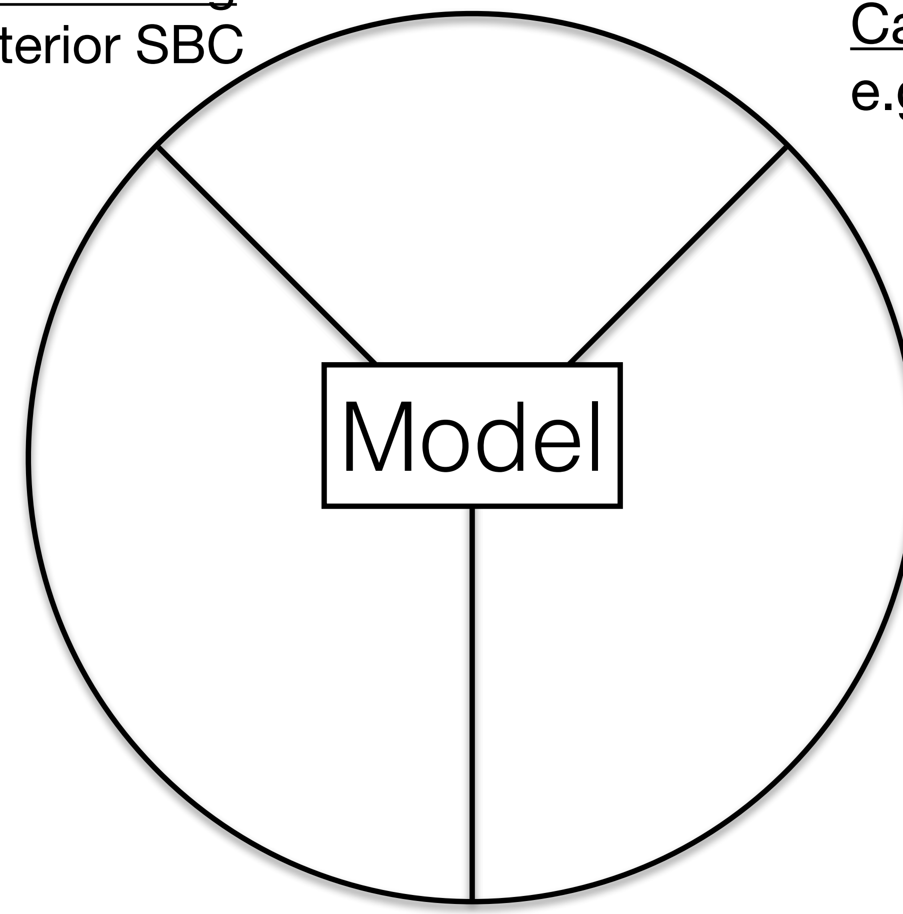
- Advancements that came with or enabled modern SBI
  - Neural density estimators
  - Differentiability
  - Simulation-based calibration (SBC)

Assessing Models: We need pressure points

- Common Modern Methods
  - Neural Posterior estimation (NPE)
  - Neural Likelihood estimation (NLE)
  - Neural Ratio Estimation (NRE)
- What's next
  - Extrapolation — e.g., for prediction out of distribution
  - Statistical guarantees —e.g., locally valid credible regions
  - Hierarchical structure
  - Comprehensive pressure points
  - More expressive densities

Forward Modeling
e.g. posterior SBC

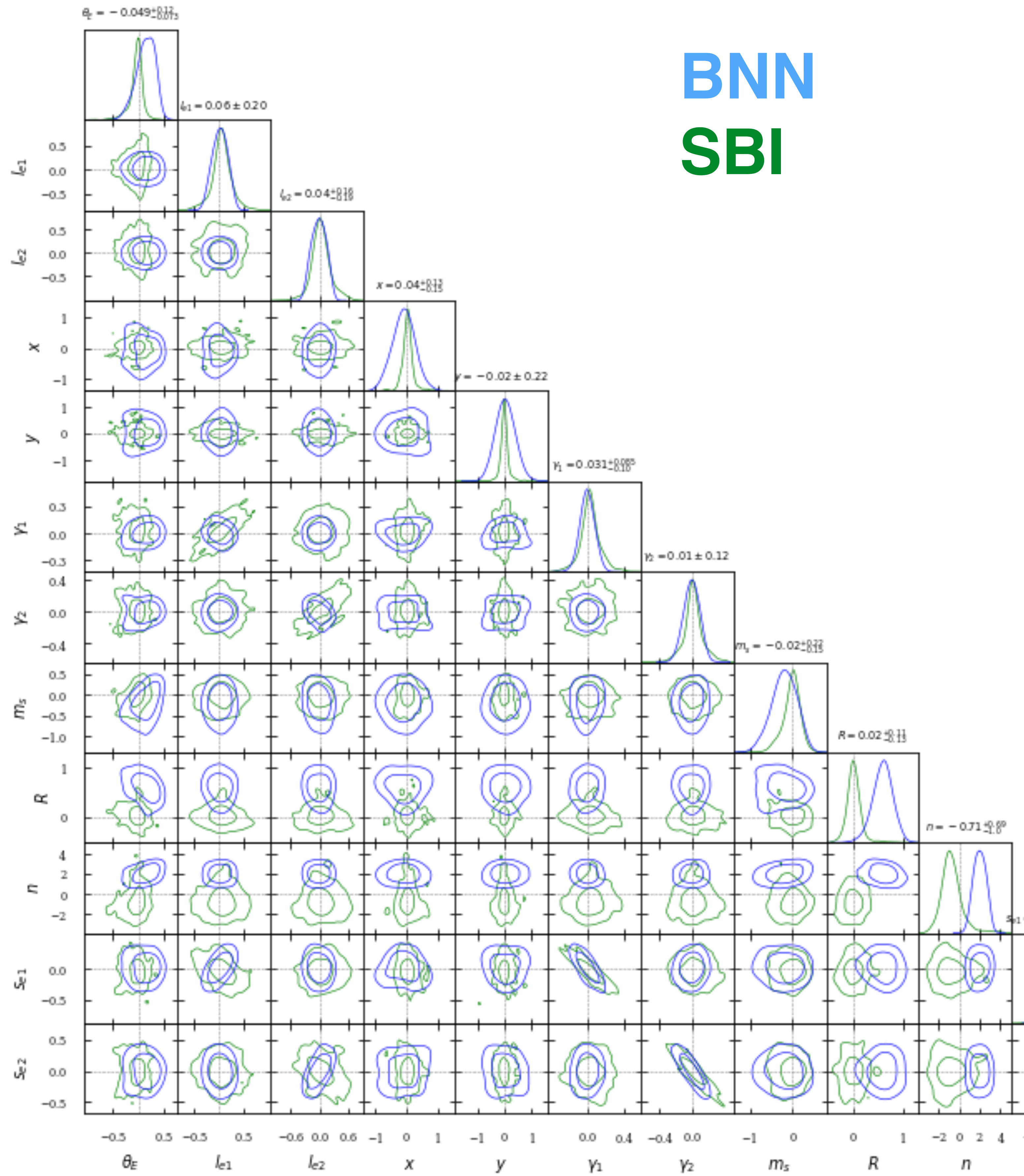
Simulation-based Calibration (SBC)
e.g., Talts et al, 2018



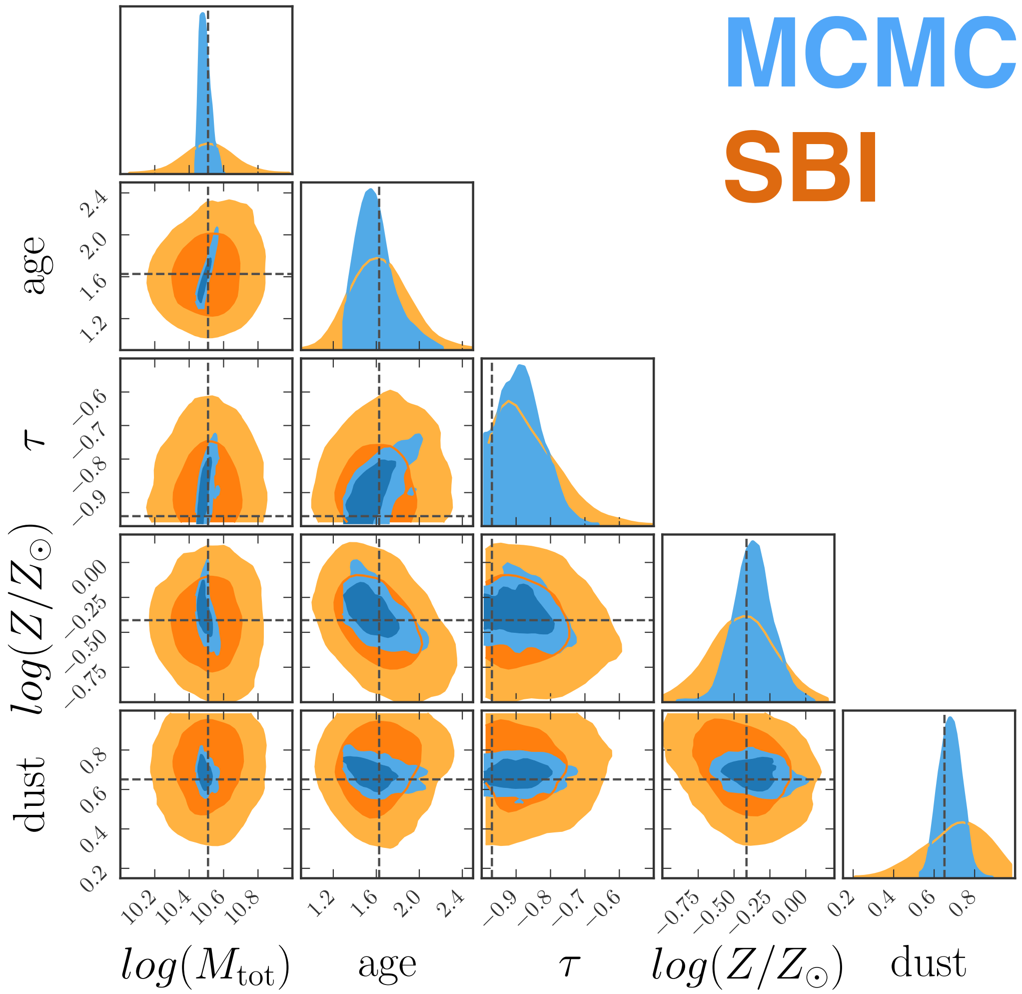
Statistical Guarantees
e.g., WALDO (Masserano+2023), **Hierarchical Conformal Regions with Validity (Trivedi and Nord, 2025)**

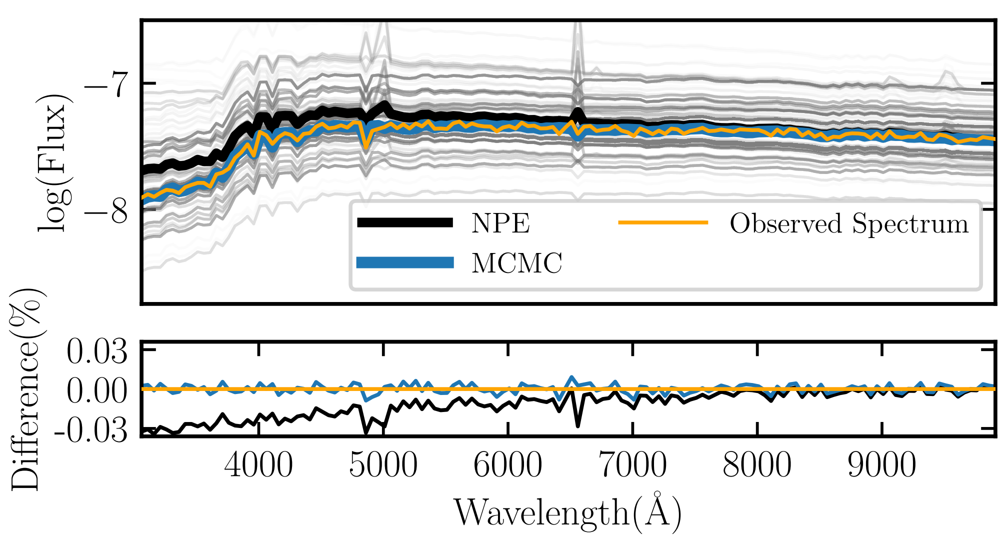
SBI and galaxy-galaxy lenses (Poh+2025, JCAP)

- **Simultaneously estimate posteriors** of multiple parameters — radius, magnification, profile, source ellipticity.
- SBI is **more accurate, precise, and stable** than BNNs
- SBI/BNN both **orders of magnitude faster** than traditional MCMC methods.
- Need comprehensive assessments to ensure model fidelity.



SBI and galaxy spectra (Khullar+2022, MLST)

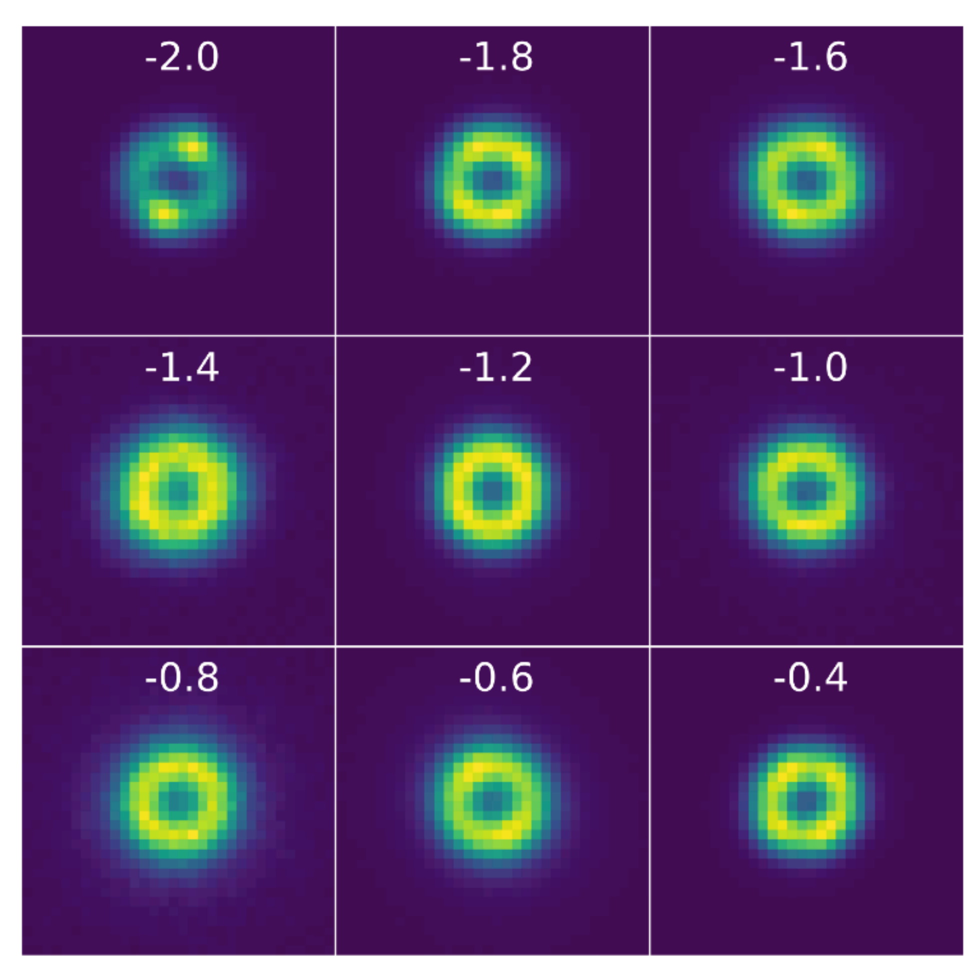


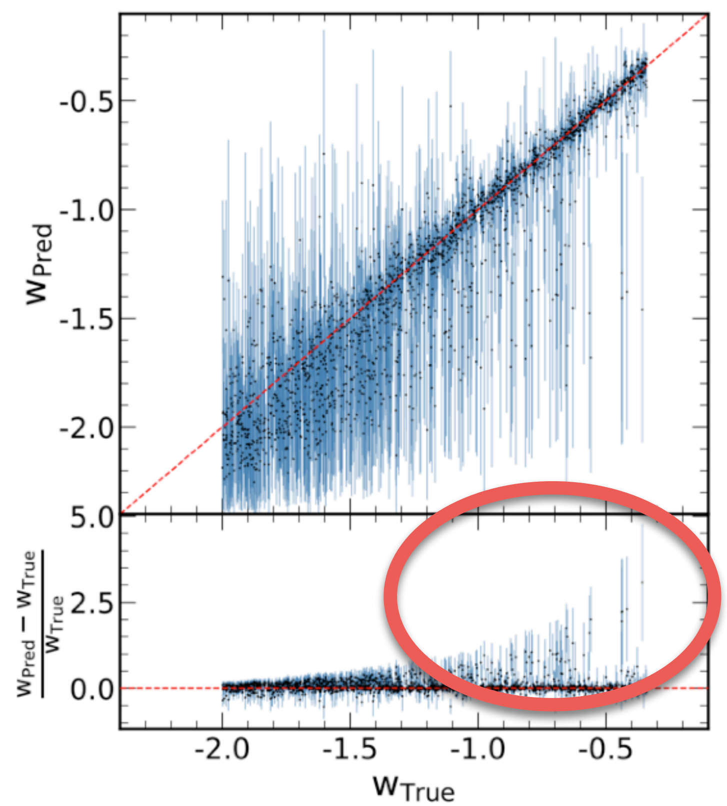


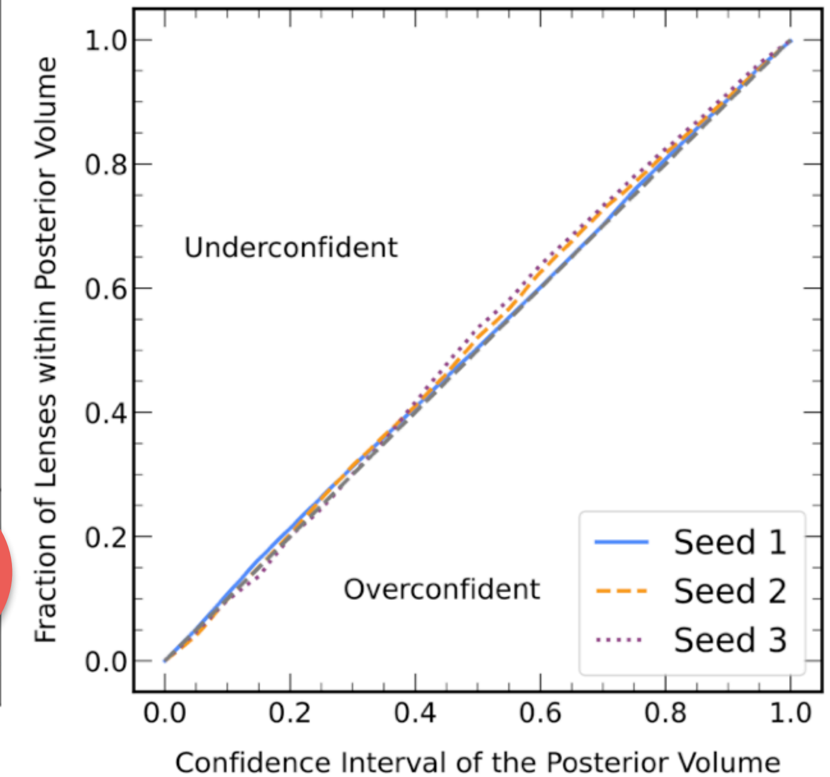
- Simultaneously estimate **posteriors of galaxy age, metallicity, dust, star formation**
- SBI performs is as **accurate** as MCMC, nearly as precise, but is **100,000x faster**.

Gourav Khullar

NRE for Population-Level Inference (Jarugula+2024)







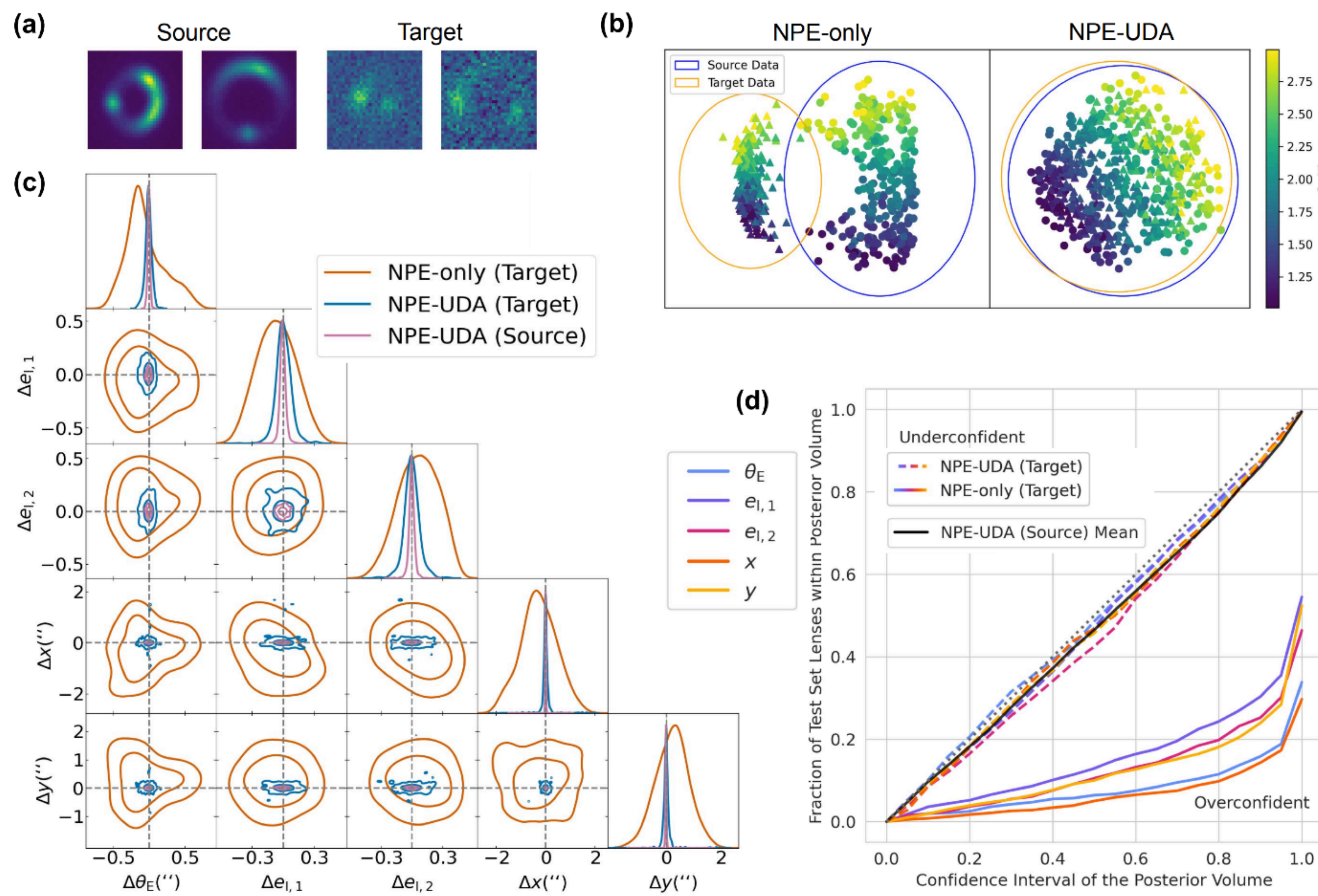
- Data: Einstein rings (future: arcs)
- Model: posterior is well-calibrated, but also bi-modal
- Next: two-parameter inference

NPE with Domain Adaptation (Swierc+2024)

- Source Domain: noiseless images
- Target Domain: Noisy images
- Without DA:
    - Latent spaces are disjoint
    - Model is overconfident on target data
    - Model is inaccurate
- With DA:
    - Latent spaces overlap
    - Model is calibrated and accurate

Examples

Latent space

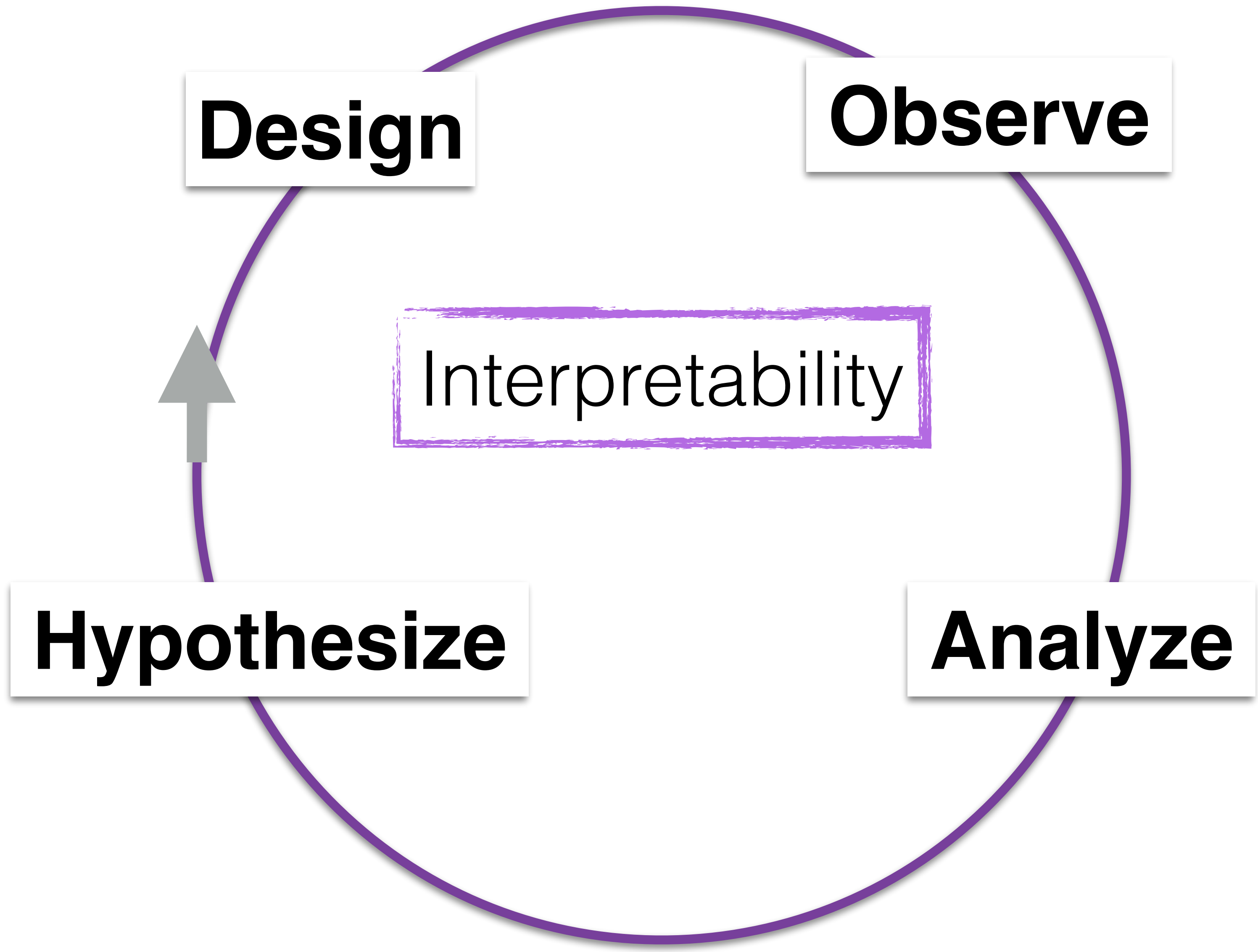


Posteriors

Coverage

Automating the Scientific Cycle

- <u>Cost of manual science</u>
  - instruments
  - personnel and human time
- <u>Next big experiments?</u>
  - Colliders
  - Telescopes
  - Fusion devices
- <u>DOE Community Direction</u>
  - AI Town Hall: Automated Cosmic Experiment
  - Future Scientific Methodologies Workshop
  - Self-driving facilities
- <u>AI Community Direction</u>
  - MODE Collaboration
  - Simulation Intelligence



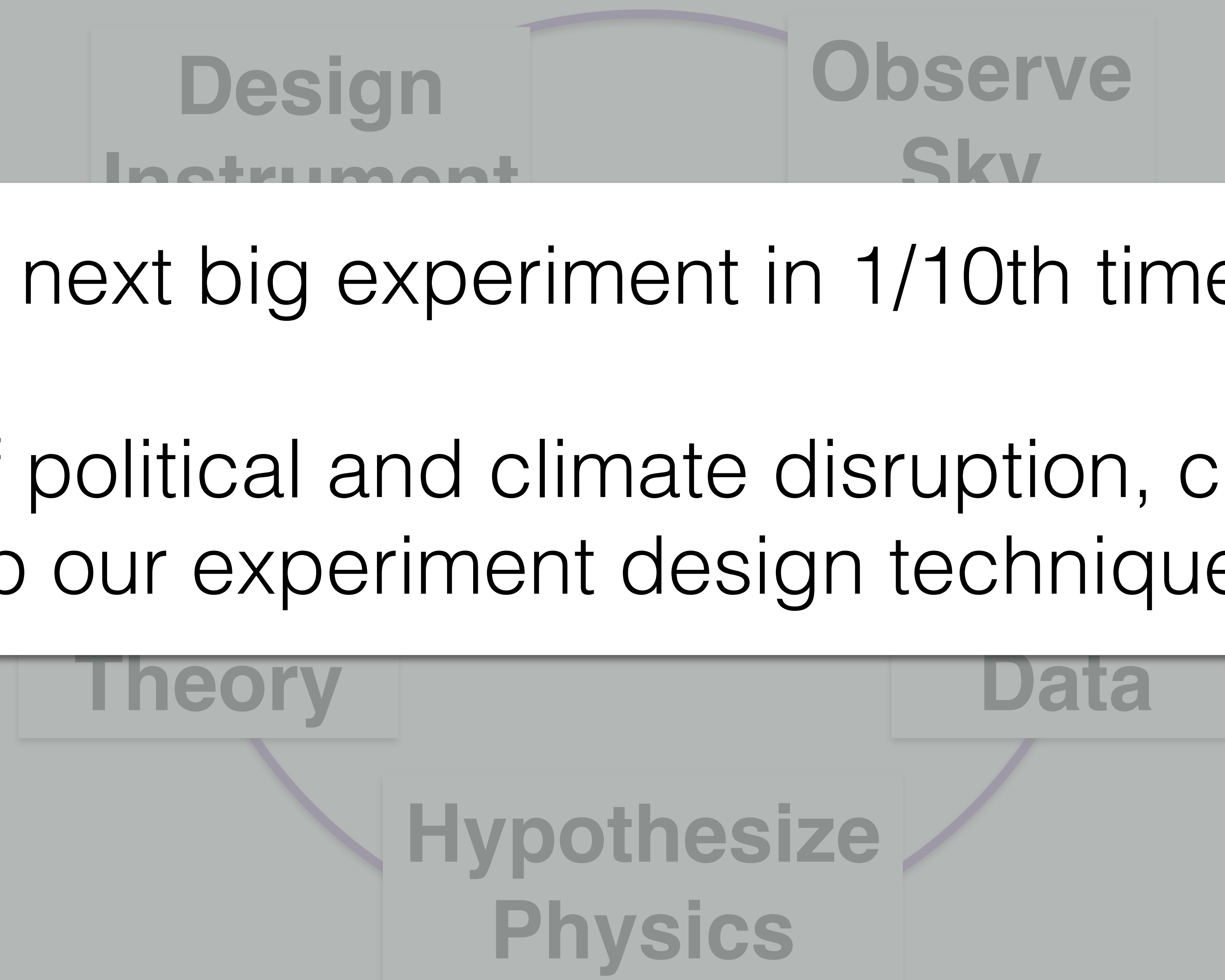
Automating the Scientific Cycle

- Cost of manual science
  - instruments
  - personnel

Imagine designing the next big experiment in 1/10th time.

Given the imminence of political and climate disruption, can we afford to not level up our experiment design techniques?

- DOE Community Direction
  - AI Town Hall: Automated Cosmic Experiment
  - Future Scientific Methodologies Workshop

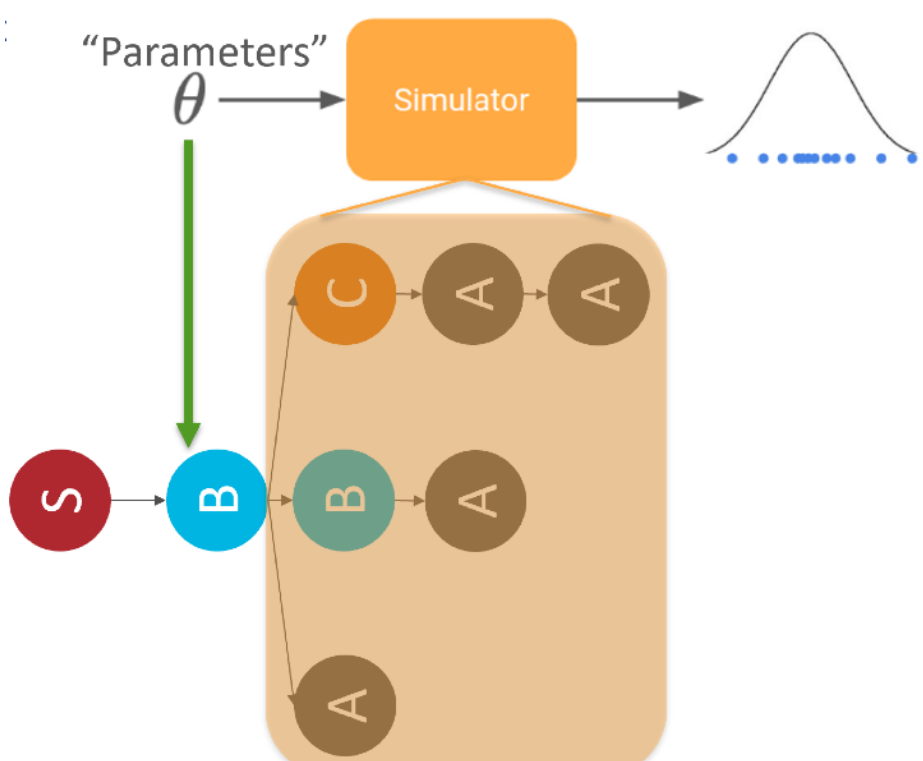


Auto-Optics Modeling (Cohen + Nord 2025, in prep.)

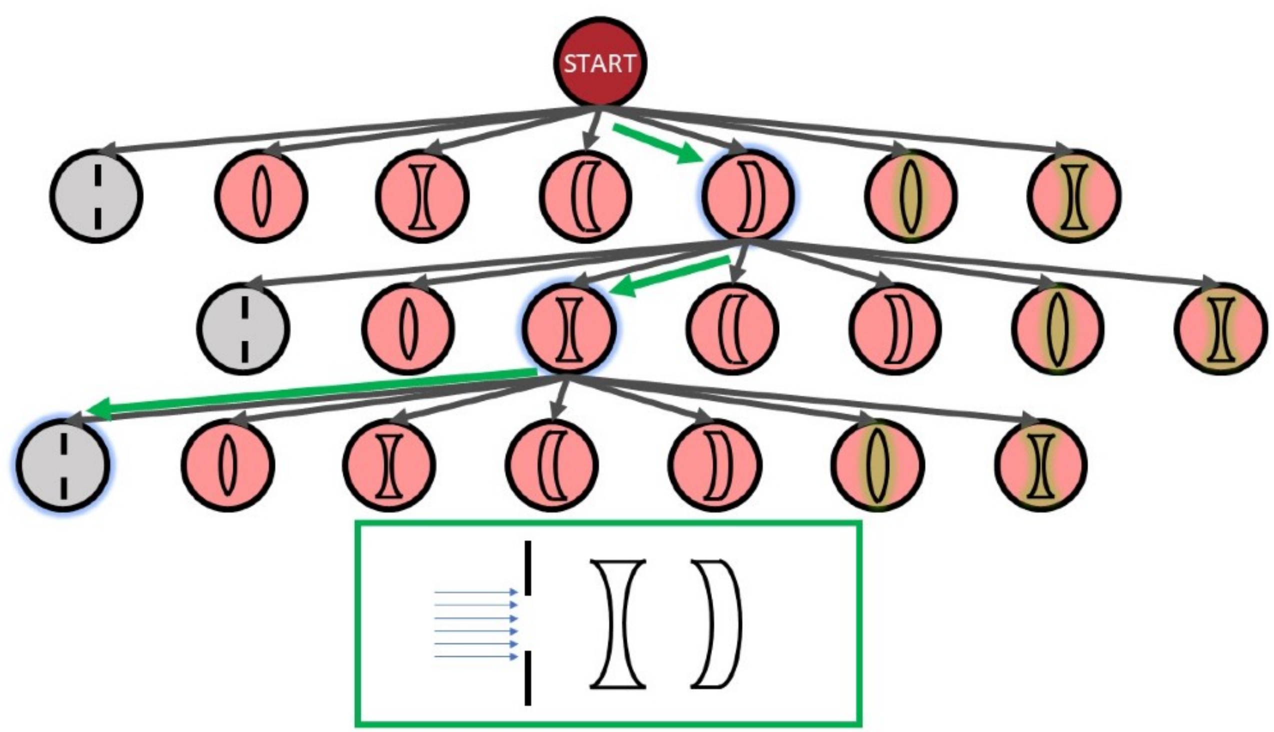
- Problem: optimize **telescope optics** — a discrete+continuous space
- Solution: **Tree Search** + **Simulation-based Inference**
  - Produces sets of optics configurations with **probabilities**.
  - **Competitive** with existing algorithms, but also **explainable**
  - Also works for symbolic regression.

Benjamin Cohen

Optimization Loop



Resulting Optics Tree



DES Observing

Self-Driving Telescopes

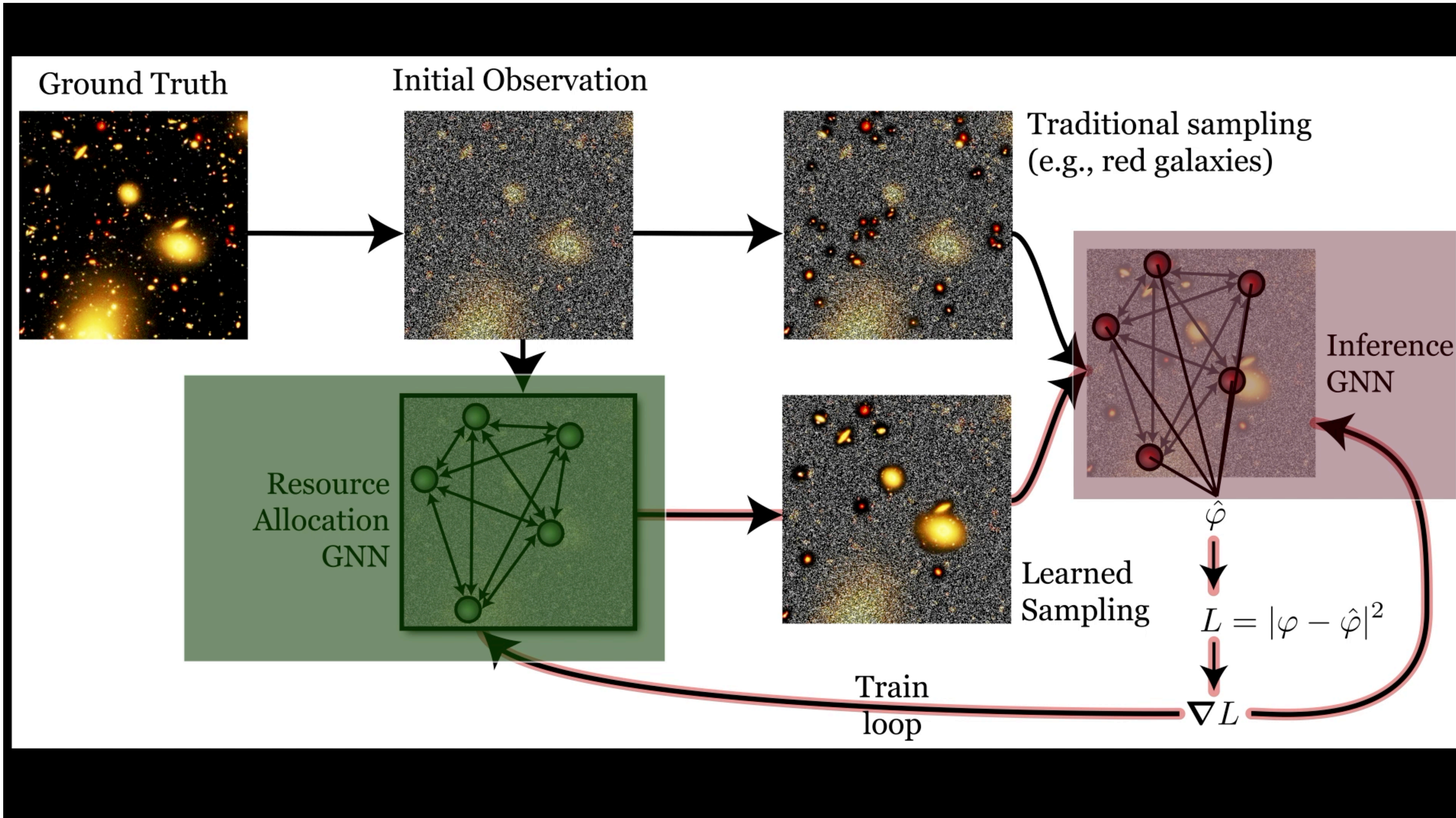
- Problem: **competing metrics** of success for different cosmic probes; time is a limited resource.

- Supervised Solution: **Reinforcement Learning** (Voetberg, Zhou, Neilsen+ 2022, in prep.)

- Unsupervised Solution **Graph Neural Networks** (Cranmer, Melchior, Nord, 2022, in prep.)

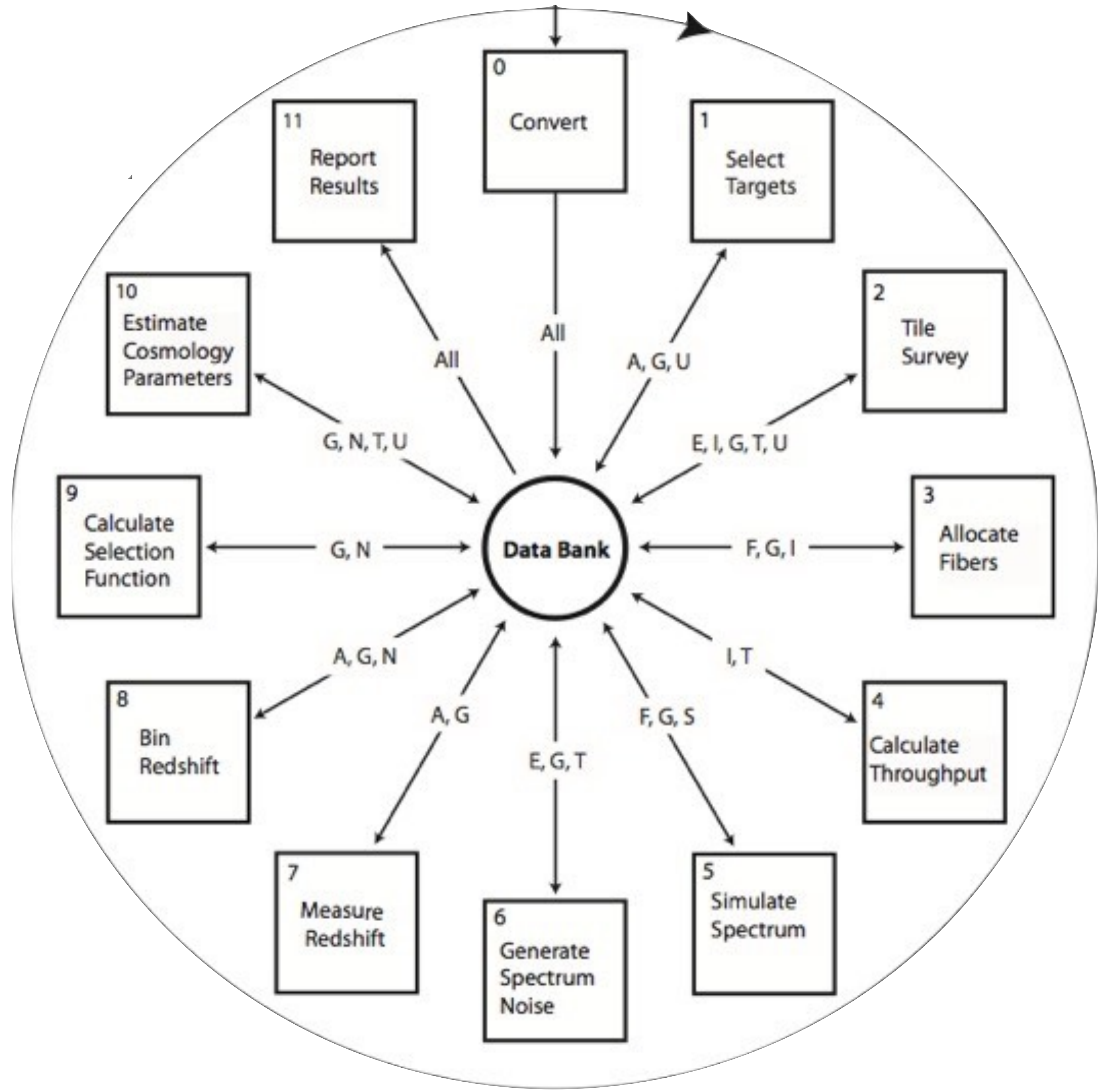
Unsupervised Resource Allocation with GraphNets

Cranmer, Melchior, Nord, 2021



Digital Twins: End-to-End Simulations of Experiments

- **SPOKeS**: Spectroscopic Ken Simulation (Nord+2016)
- Start with galaxy data, simulate every aspect of survey and compute cosmological constraints

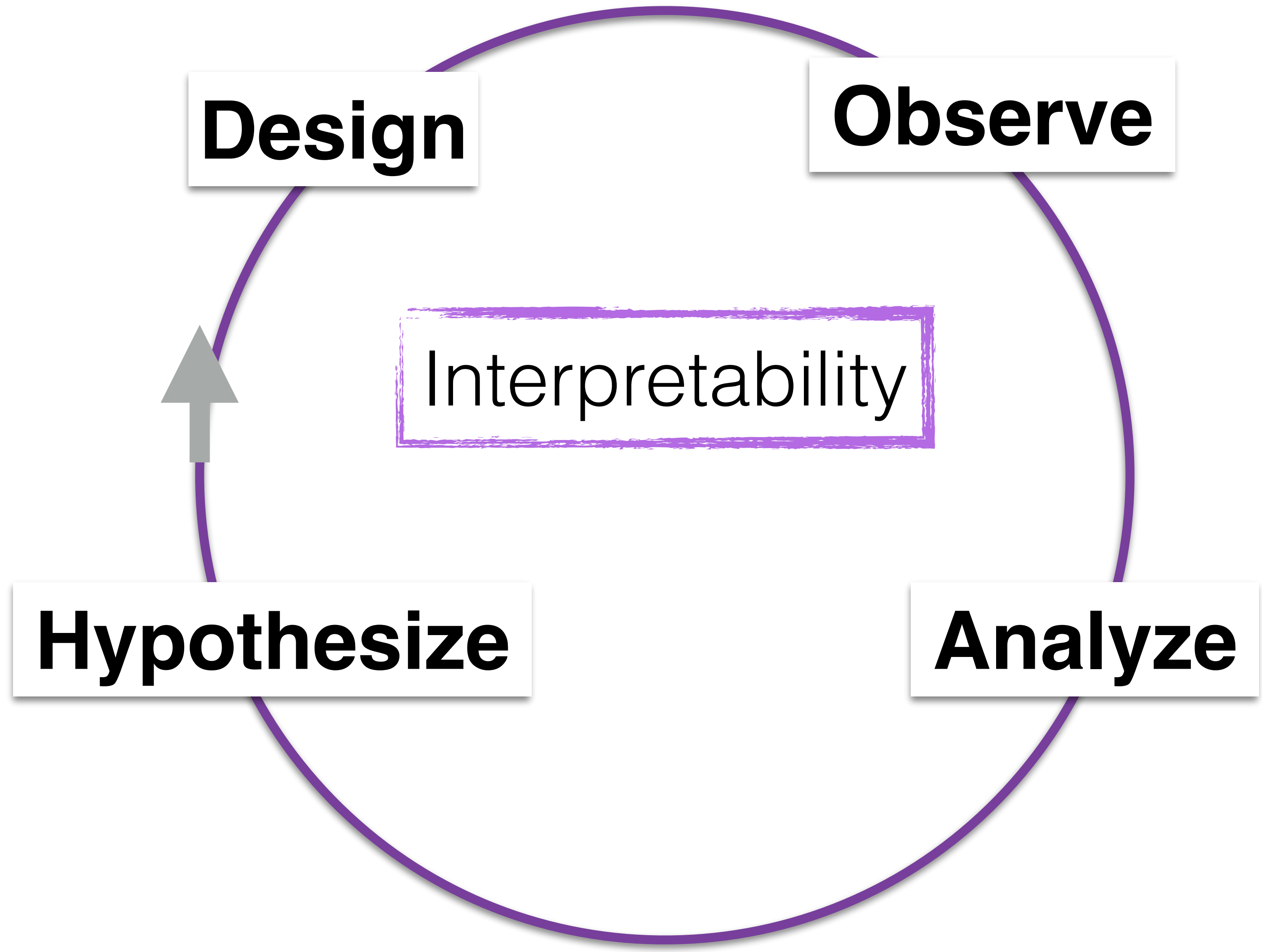


Automating the Scientific Cycle

- SBI in data analysis is well-tested but there remain outstanding issues:
  - Extrapolation — e.g., for prediction out of distribution
  - Statistical guarantees —e.g., locally valid credible regions
  - Hierarchical structures
  - Comprehensive pressure points
  - More expressive densities
  - Benchmarks

- It behooves the community to align on our jargon and methods for model assessment.

- Applying SBI to other elements of the scientific cycle is on the horizon.



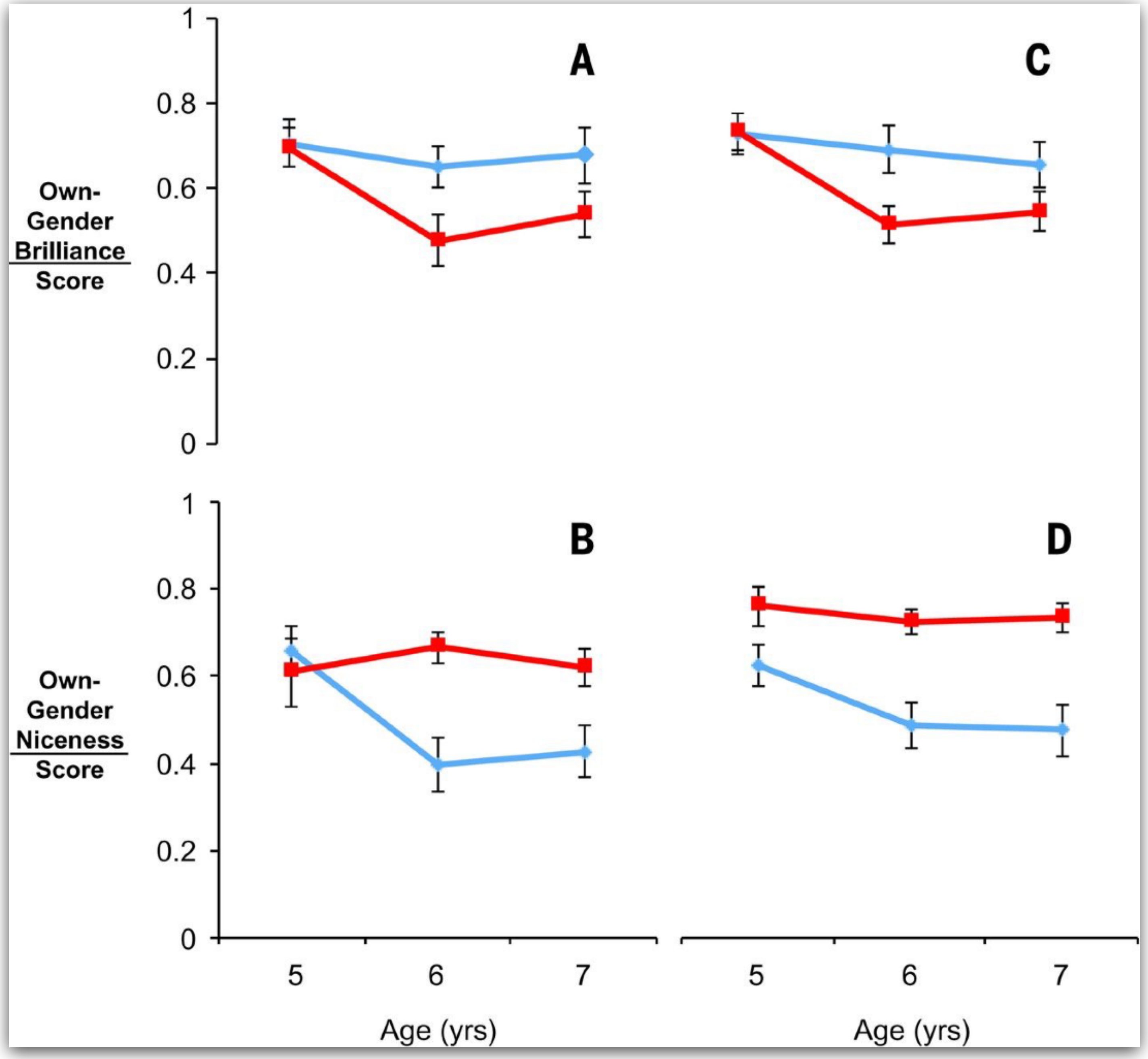
Extras

Childhood

- Pre-defined gender stereotypes affect people for a long time.

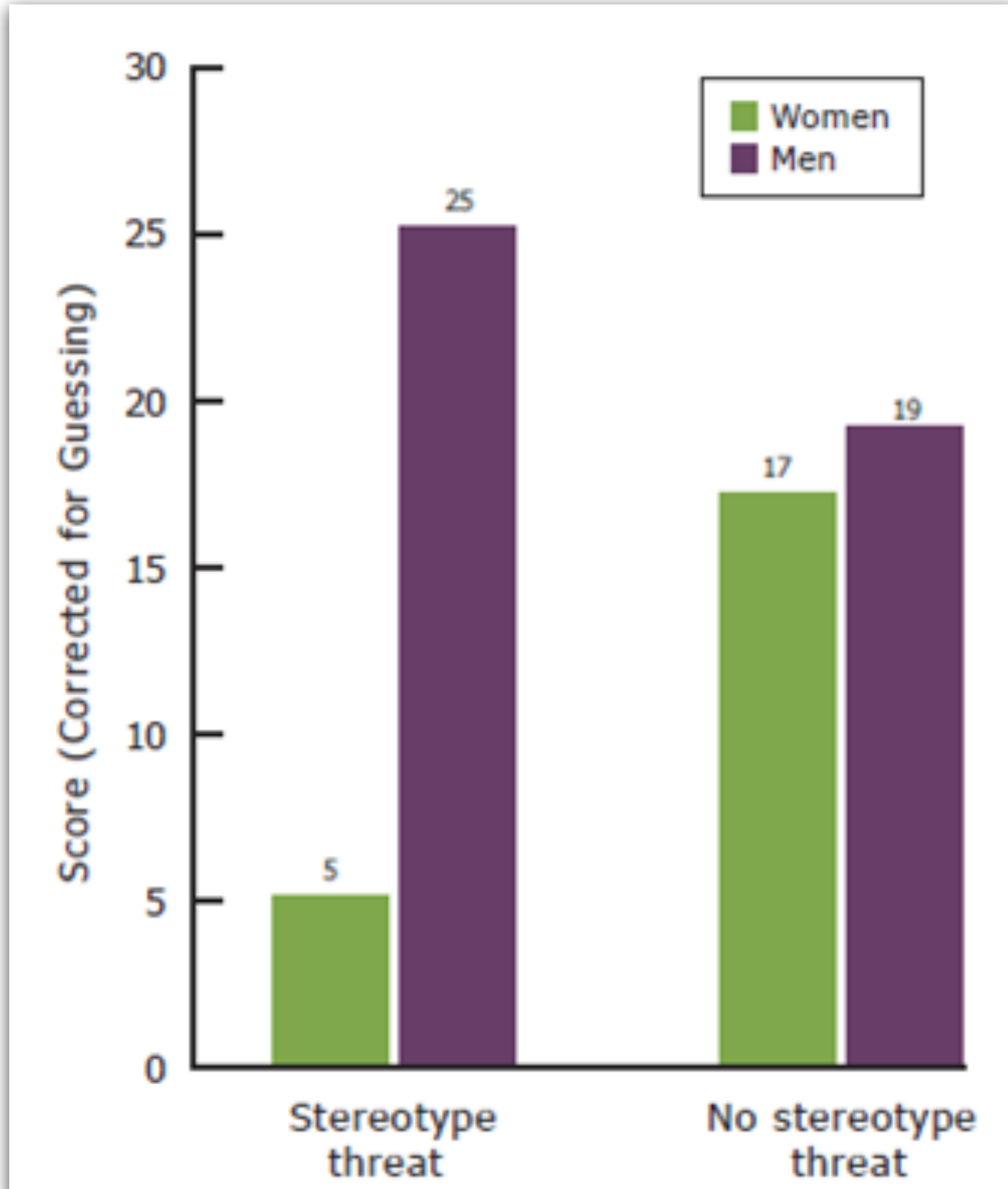
Childhood: Bias emerges early

- <u>Experiment containing four studies :</u>
  - N = 400 children
  - Mostly middle class backgrounds and 75% White
  - Children are read a story about brilliant people (without identifying gender). Then, they are asked to select what they think is the brilliant protagonists gender.
  - Ask children at each age, 5, 6, 7
- <u>Results:</u>
  "By age 6, girls were prepared to lump more boys into the 'really, really smart' category and to steer themselves away from games intended for the 'really, really smart.'"
- This is a good example of *internalized bias*
- Additional experiments and studies:
  - Farenga & Joyce 1999; Ambady+2001; Lavy & Sand, 2015; Buck+2008; Nguyen & Ryan 2008



Bian+2017 (*Science*)

Education: test-taking and stereotype threat

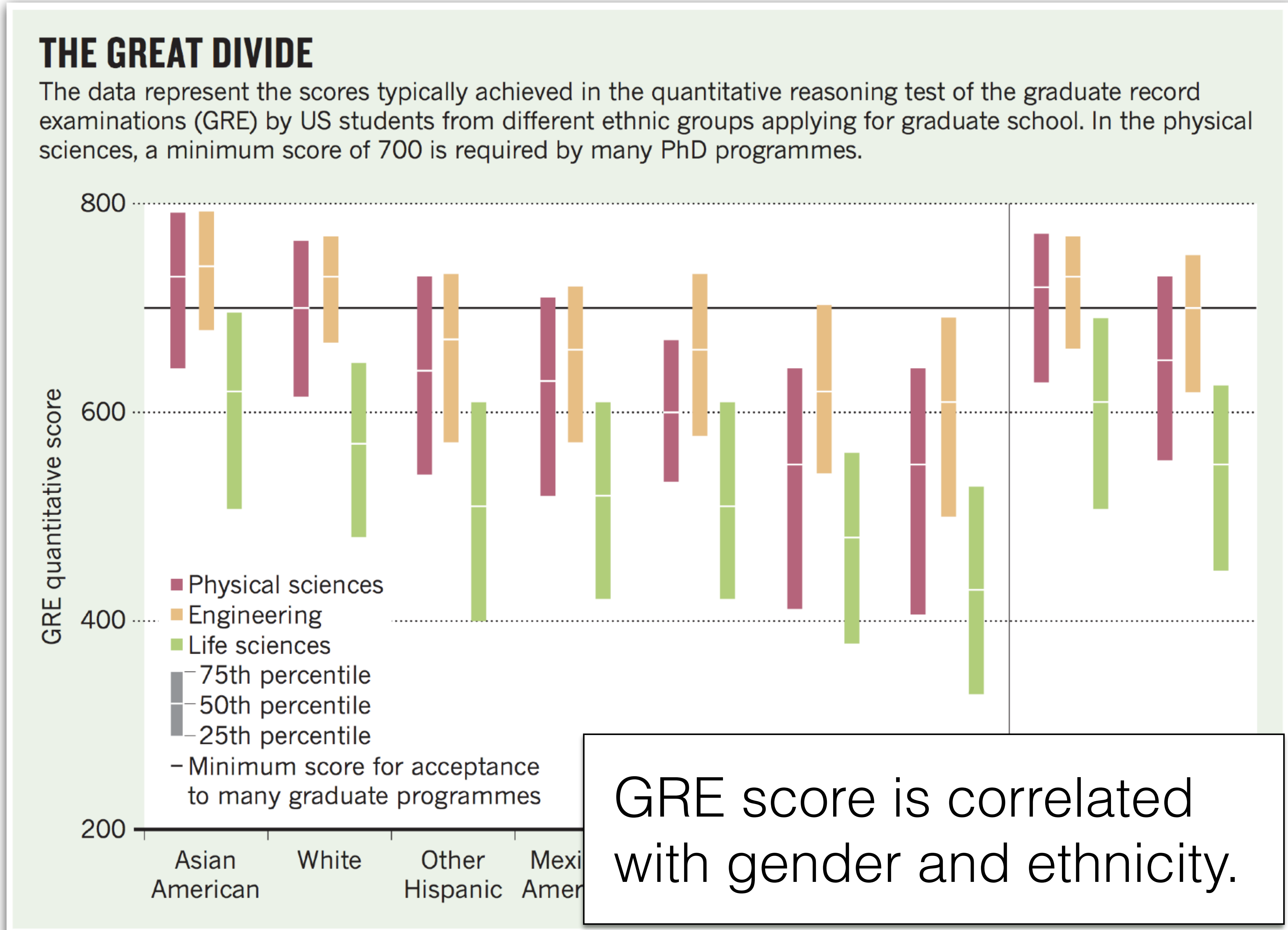


visualization of data from Spencer+1999 (by J. Schmelz)

- Study of ~60 people, about half men/women
- Split into two groups, and each given a different test
    - Test 1: discusses potential gender disparities in math tests like this one.
    - Test 2: does not
- The threat of playing to one's stereotype impedes performance.
- *Stereotype threat* discovered in studies of Black students (Katz, Roberts, & Robinson, 1965)

Education: GRE's

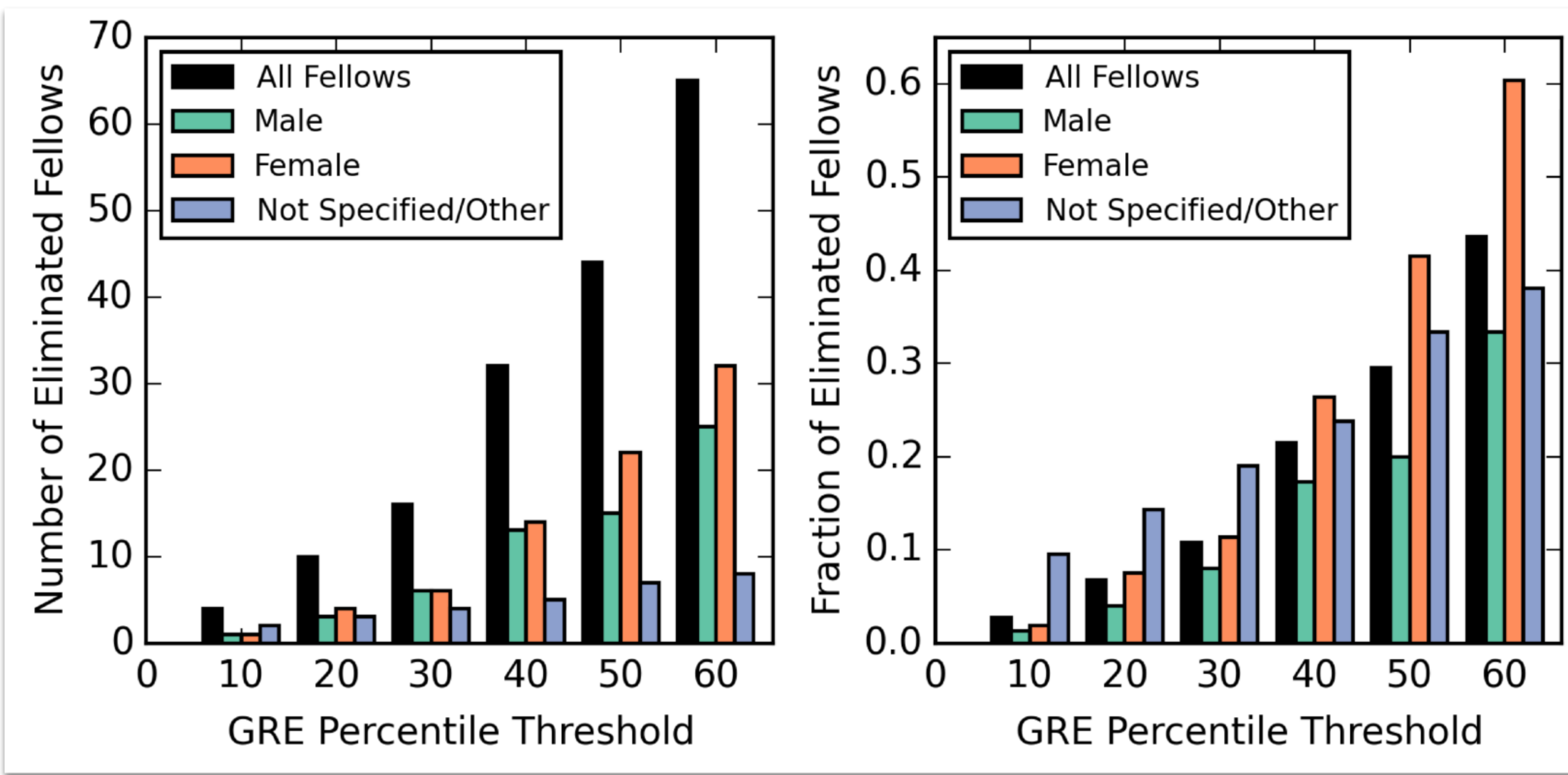
Miller & Stassun, 2014



GRE score is correlated with gender and ethnicity.

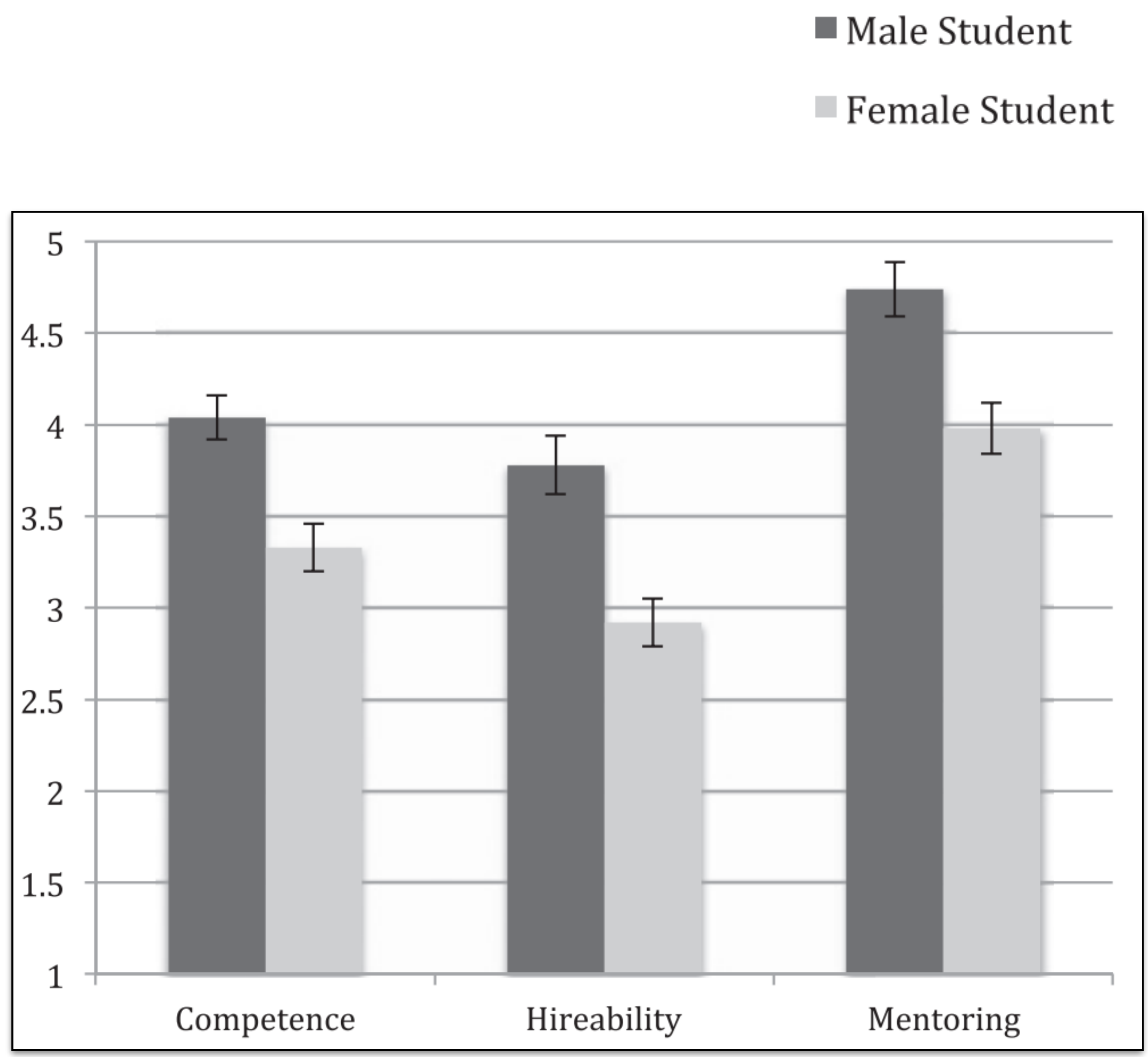
Education: GRE's

- There's a strong correlation between status as a URM and 'success' on the GRE.
- But, there is a very weak correlation between 'success' on the GRE and 'success' in science research and academia.
- Large fractions of prize fellows would be eliminated with strict PGRE percentile thresholds in admission. (Levesque, Bezanson, Tremblay, 2015)



Applications: what's in a name?

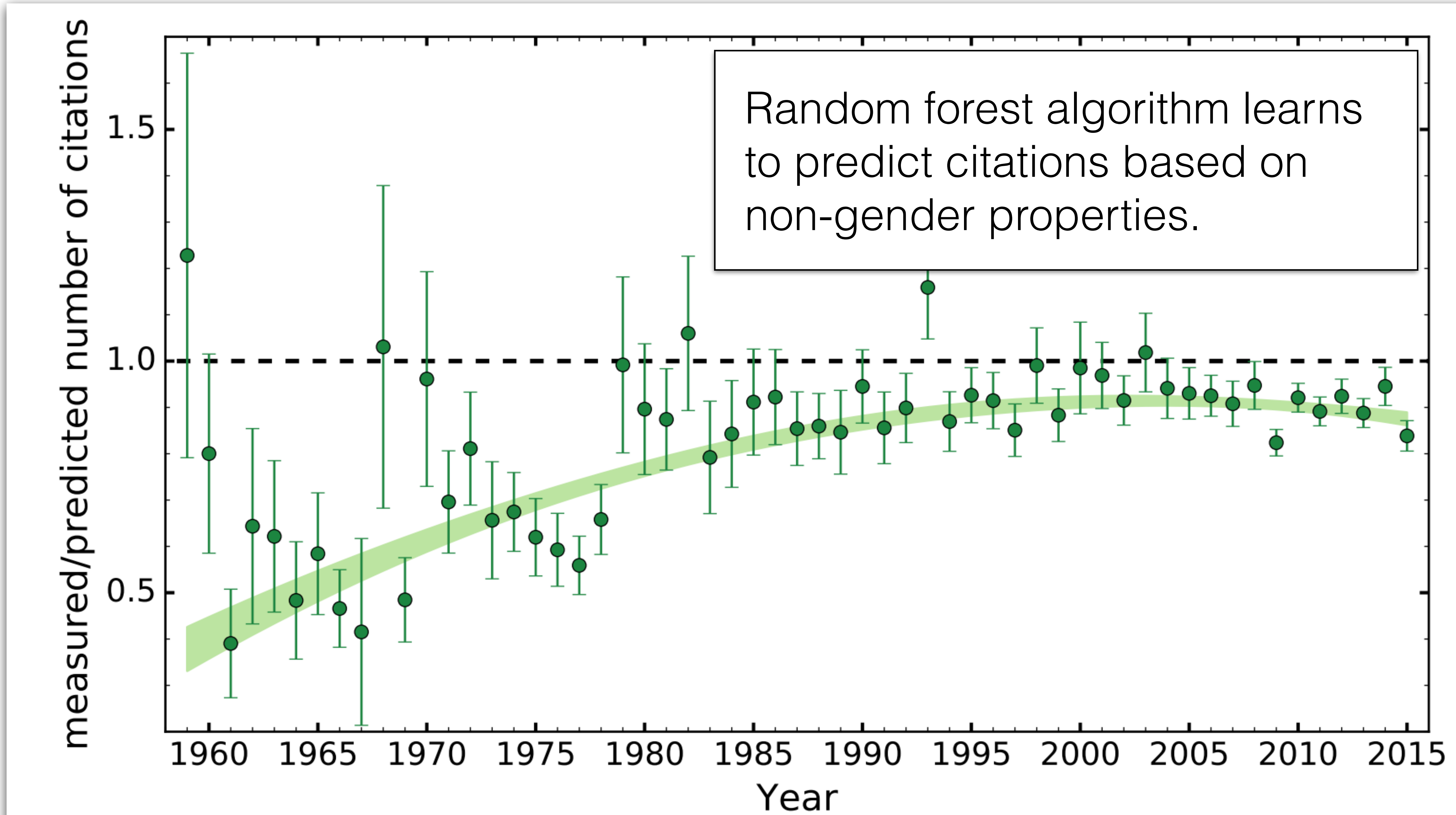
- 127 faculty at research-intensive institutions
- applications randomly assigned gender
- all differences in rating are statistically significant
- similar studies have been done for race, but can be difficult to control for socio-economic status.



Moss-Racusin+2012

Bias in citations

Caplar, Tacchella, & Birrer 2016



Sample of 200,000 astro papers demonstrates bias in citations.

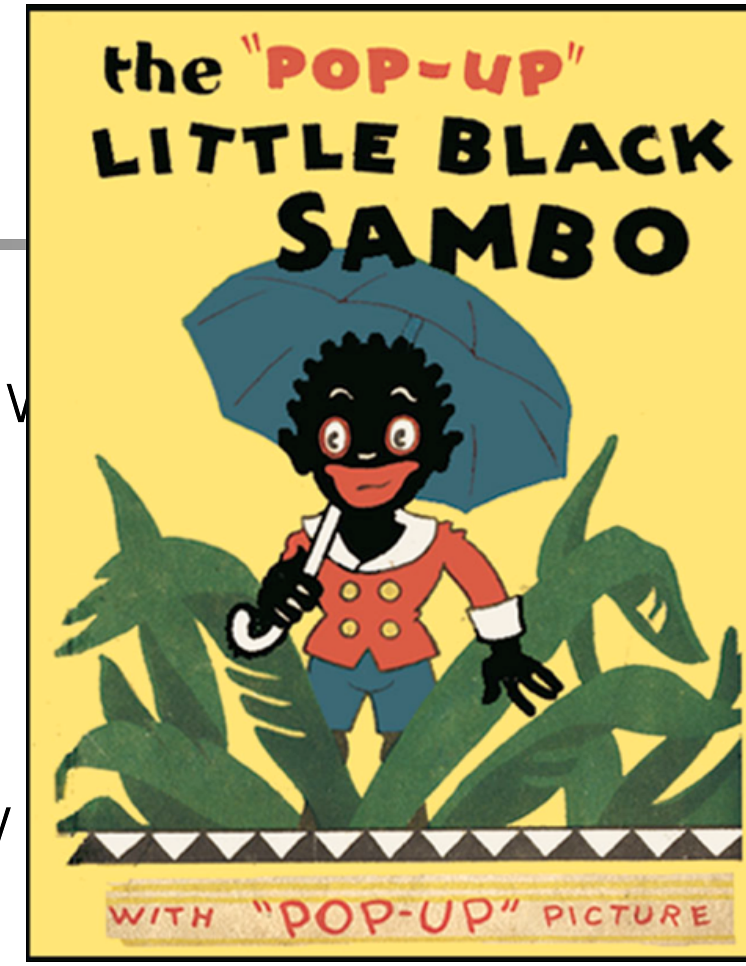
Work Environments

- Sexual harassment and assault are real and prevalent in academic and research environments.
- <u>Geoff Marcy, UC Berkeley</u>
  Serial harassment for > 10 years; was not fired, resigned of own volition
- <u>Christian Ott, Caltech</u>
  Fell in love with student and then fired her, discussed the issue with other student.
- <u>Timothy Slayter, Wyoming/Arizona:</u>
  "She would teach better if she did not wear underwear."
- <u>Many instances at other universities, other fields:</u>
  e.g., U. Chicago, AMNH, Anthropology, Biology, Philosophy


Geoff Marcy


Timothy Slayter

Personal Experiences



- I experience racism and the other side of privilege at w
  ~ 1/week
- Examples:
  - "Do you know the story of Sambo?"
    ~70-year-old White senior scientist discussing how
    collaborate.
  - "You are safe here. You shouldn't worry."
    ~50-year-old White senior scientist at work (gaslighting).
  - "You're not Black enough."
    ~90-year-old White woman at a political event.
  - "I guess you know about the mean streets."
    ~25-year-old junior scientists from outside U.S. at a science meeting.
  - "And you're mulatto, ya know!"
    ~35-year-old White woman at a pub in my city of residence.

We are our institutions

- History sets the context.
- We make decisions now.
- We set up the future.