ASTR 8500 · Professional Development & Proposal Writing · UVA

Does the Reproducibility Crisis Affect Astronomy?

Spoiler: Yes — but it's complicated (and kind of fascinating)

Sidhant Kumar Suar
Spring 2026

Some Definitions and an Interesting Statistic!!!

- Replicability: different data + same/similar method → same conclusion
- Reproducibility: same data + same/similar analysis → same result
- Consistency: different data + different method → same conclusion
- Root causes: p-hacking, publication bias, small samples, closed data
- **Does it affect astronomy?** 🔭

>70%

of researchers have tried and failed to reproduce another scientist's results

Note: This is across different fields

Nature survey, 1,576 researchers

THE RISK SPECTRUM

Astronomy: The Tension Zone

← Higher Replicability Risk

Lower Replicability Risk →

Psychology & Epidemiology

Complex systems

Flexible analyses

❌ Lack of a strong underlying theory

Astronomy

Strong underlying theory

Open data archives

No controlled experiments

⚠ Flexibility hides in data pipelines

Particle Physics

Strong underlying theory

“Repeatable” controlled experiments

✅ Gold standard for replicability

"We have the theory of physics — but our laboratory is the Universe"

The Measurement Gap

Early Universe (CMB / Planck)

67

km/s/Mpc

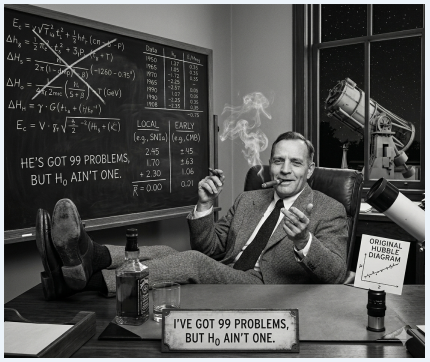
≠

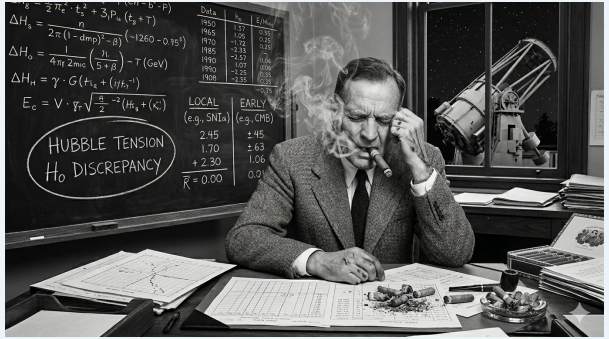
>5σ discrepancy

Local Universe (Cepheids / SN Ia)

73.0

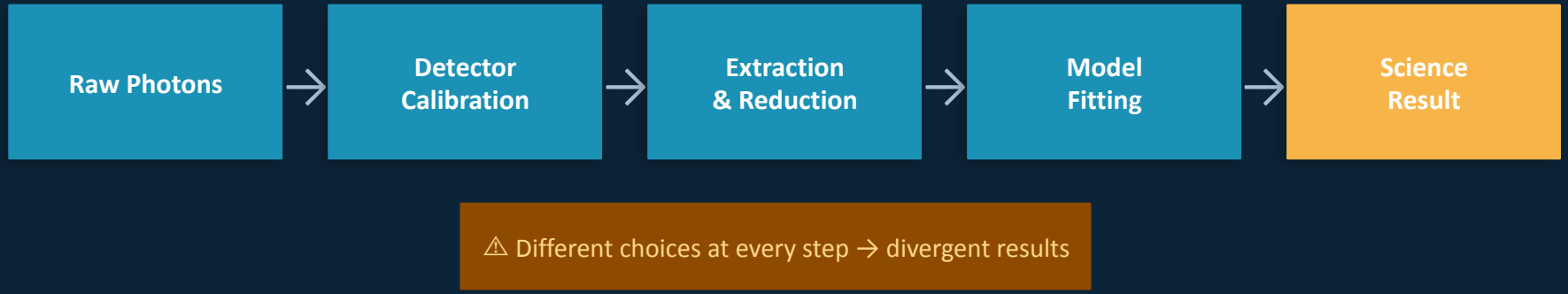
km/s/Mpc





CASE STUDY #2 — DATA PIPELINES

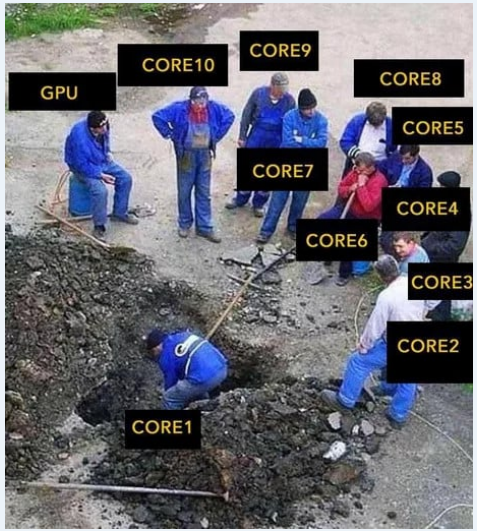
The Pipeline Problem: A Reproducibility Crisis in Exoplanet Science



- Exoplanet atmosphere spectra from JWST: multiple independent teams get measurably different absorption features from identical raw data
- Stellar limb-darkening choices, background subtraction, and systematics treatment all create “researcher degrees of freedom”

Numerical Inconsistencies

- Algorithm choices in your code can lead to different results on different computing clusters because of the difference in underlying hardware architecture. Even running the same code on the same machine could give spurious results if the algorithm chosen leads to unexpected truncation errors or has random seeds, and so on.



SILVER LININGS

The Astronomy Community Is Fighting Back

📂 Open Data Archives

MAST, VizieR, NED, Zenodo — decades of tradition of sharing raw and calibrated data

💻 Open-Source Pipelines

AstroPy, Eureka!, exoTEDRF — community-maintained, peer-reviewed, version-controlled

📓 Reproducible Notebooks

Jupyter notebooks + code-as-figures becoming standard practice for JWST papers

🔭 Multi-Pipeline Validation

Running multiple independent reduction pipelines on same data to identify systematics

🤝 Large Collaborations

SDSS, DES, DESI, CMB-S4 — blind analyses and internal review built into the structure

Practical Takeaways for Grad Students & Young Researchers

1. Version-control everything — Git + GitHub for your code. Tag versions used in each paper. Your future self will thank you.
2. Document your pipeline choices — Every threshold, every sigma cut, every prior — write it down. If you can't explain it, neither can a reviewer.
3. Release your data & code — Zenodo, GitHub, and ADS all let you cite your code. It's fast becoming a requirement for journals.
4. Be skeptical of your own results — If something is surprisingly clean or significant, that's a yellow flag. Check your systematics harder.
5. Value replication in proposals — Explicitly framing your work as confirming/extending prior results is scientifically valuable — say so.

REFERENCES

1. https://libguides.williams.edu/c.php?g=1476327&p=10995208
2. https://simplystatistics.org/posts/2016-08-24-replication-crisis/
3. Baker, M., 2016, Nature, 533, 7604, 452. doi:10.1038/533452a
4. Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLOS Medicine 2(8): e124. https://doi.org/10.1371/journal.pmed.0020124
5. https://www.templeton.org/news/the-biggest-crisis-in-astronomy
6. https://www.desi.lbl.gov/2025/03/10/closing-our-eyes-to-truly-see-blinding-in-desis-analysis-of-its-cosmological-measurements/