Jonathan Decker, Simon Steuernagel

Repeatability, Reproducibility & Replicability

The three R’s of science

- 2 Utah Scientists claimed 23.03.1989 to have achieved "Cold Fusion"

Browne, “Physicists Debunk Claim Of a New Kind of Fusion”

- 2 Utah Scientists claimed 23.03.1989 to have achieved "Cold Fusion"
- Refused to publish details needed for follow-up experiments
- Many scientists attempted to reproduce the results
- Essentially all failed

Browne, “Physicists Debunk Claim Of a New Kind of Fusion”

- 2 Utah Scientists claimed 23.03.1989 to have achieved "Cold Fusion"
- Refused to publish details needed for follow-up experiments
- Many scientists attempted to reproduce the results
- Essentially all failed
- Theories about what Utah scientists actually measured came up
- Utah Scientists never admitted any mistake
- Lots of funds and time invested to disprove claim

Browne, “Physicists Debunk Claim Of a New Kind of Fusion”

“Non-reproducible single occurrences are of no significance to science.”

— Karl Popper, The Logic of Scientific Discovery, 2002, p. 66

Popper and Jason, *The Logic of Scientific Discovery*

Outline

Three R's of (Data) Science

- Repeatability,
- Reproducibility,
- Replicability

Fidler and Wilcox, "Reproducibility of Scientific Results"

Three R's of (Data) Science

- Repeatability,
- Reproducibility,
- Replicability

- Do something again to get the same result

Fidler and Wilcox, "Reproducibility of Scientific Results"

Three R's of (Data) Science

- Repeatability,
- Reproducibility,
- Replicability

- Do something again to get the same result
- Terms can be used interchangeably

Fidler and Wilcox, "Reproducibility of Scientific Results"

Reproducibility

- Exact definitions vary between scientific fields

Fidler and Wilcox, “Reproducibility of Scientific Results”

Reproducibility

- Exact definitions vary between scientific fields
- Direct replication:
  - Do the exact same

Fidler and Wilcox, “Reproducibility of Scientific Results”

Reproducibility

- Exact definitions vary between scientific fields
- Direct replication:
  - Do the exact same
- Conceptual replication:
  - Alter certain features

Fidler and Wilcox, “Reproducibility of Scientific Results”

Reproducibility

- Exact definitions vary between scientific fields
- Direct replication:
  - Do the exact same
- Conceptual replication:
  - Alter certain features
- Confirm results & Advance/Generalize hypothesis

Fidler and Wilcox, “Reproducibility of Scientific Results”

Outline

A crisis?

- A variety of scientific fields are said to be affected by a *replication crisis*
- Remember our session on statistics?
  - Discussed a variety of (not) reproducible results

A known problem

- Already in 2005: “Why most published research findings are false”
- In 2014: Already discussion on “rescuing the replication crisis”
- Computer Science as a field is affected too

Schooler, “Metascience could rescue the ‘replication crisis’”;
Ioannidis, “Why most published research findings are false”

ML/AI Reproducibility Crisis

- ML/AI is also affected
- Obvious Example: Huge Neural Networks (NNs)
  - Repeating training of such an NN is hard
  - Requires access to specialized hardware
  - and/or thousands of dollars

https://www.stateof.ai/

ML/AI Reproducibility Crisis

- ML/AI is also affected
- Obvious Example: Huge Neural Networks (NNs)
  - Repeating training of such an NN is hard
  - Requires access to specialized hardware
  - and/or thousands of dollars
- 2020: 15% of papers publish code
  2021: 26% of papers publish code
  in other words: 74% of papers don't

https://www.stateof.ai/

ML Reproducibility Challenge

- People are trying to tackle this problem
- Straightforward approach: Reproduce papers!

ML Reproducibility Challenge

- People are trying to tackle this problem
- Straightforward approach: Reproduce papers!
- “ML Reproducibility Challenge”
  - Started already in 2018
  - Reproduce papers from top ML conferences
  - Student courses
  - ’21 Edition: https://paperswithcode.com/rc2021

Paperswithcode

- https://paperswithcode.com/
- Goal: Link papers with implementations
  - Official and unofficial
- Track research across tasks or used data sets
- Provide easy-access to code, benchmarks, external references, ...

- If your research is related to ML/AI, you should check it out!

Outline

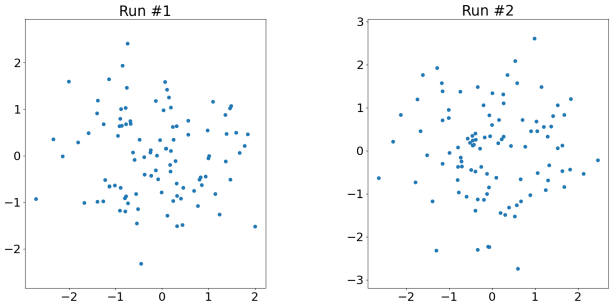
RNG Seeding

Practical Example:

- Problem: Generating random numbers is, well, random!
- How can you repeat randomized experiments?
  $\longrightarrow$ Seeding!

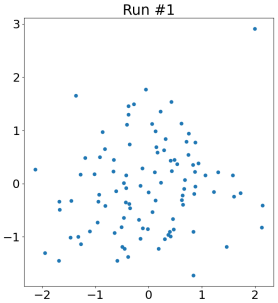
- Here: Repeat random number drawing in python

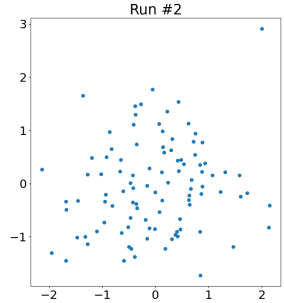
Python Example: (lack of) RNG Seeding




```
1        np.random.multivariate_normal(mean, covariance, n_samples)
```

Python Example: RNG Seeding






```
1        SEED = 42
2        RNG = np.random.default_rng(SEED)
3        data = RNG.multivariate_normal(mean, covariance, n_samples)
```

Benefits of Reproducibility

Why **you** should care

Alston and Rick, “A Beginner’s Guide to Conducting Reproducible Research”

Personal Advantages from Reproducibility

- Help remember the **how** and **why**

Alston and Rick, “A Beginner’s Guide to Conducting Reproducible Research”

Personal Advantages from Reproducibility

- Help remember the **how** and **why**
- Quickly modify figures
- Quickly modify experiments

Alston and Rick, “A Beginner’s Guide to Conducting Reproducible Research”

Personal Advantages from Reproducibility

- Help remember the **how** and **why**
- Quickly modify figures
- Quickly modify experiments
- Indicate rigorous work to fellow researchers
  $\longrightarrow$ Easier peer reviewing

Alston and Rick, “A Beginner’s Guide to Conducting Reproducible Research”

Personal Advantages from Reproducibility

- Help remember the **how** and **why**
- Quickly modify figures
- Quickly modify experiments
- Indicate rigorous work to fellow researchers
  $\longrightarrow$ Easier peer reviewing
- Reproducible research increases paper citation rates

Alston and Rick, "A Beginner's Guide to Conducting Reproducible Research"

Advantages for the Community

- You are part of a research community
  - Within your research niche...
  - ...and outside of it
- This community ($\rightarrow$ and you!) benefits from replicable research

Advantages for the Community

- Researchers can learn from each others' work
- Easier follow-up studies, and building upon previous work
- Prevent mistakes by allowing easier analysis

Alston and Rick, "A Beginner's Guide to Conducting Reproducible Research"

Outline

Dilemma

No luck

I have run an experiment, but the results did not work out. I am disappointed because I had carefully designed all the manipulations and stimuli, and the previous experiment(s) that I ran for the same project had worked out. I am now writing the paper. What do I do?

A I fully report the failed experiment as one of the main studies in the paper and speculate about the potential reasons behind the unsuccessful results in the discussion section.

B I mention the unsuccessful experiment in one sentence and ask the interested readers to contact me for more details.

C I do not mention the unsuccessful experiment anywhere.

D I leave out the unsuccessful experiment from the paper, but mention it in the cover letter to the editor and suggest it can be included if so desired.

Your experience

What are your experiences?

Your experience

What are your experiences?

- Have you been contacted by someone unable to reproduce your work?
- Papers *without* code?
- Any results you could not reproduce?
  - Maybe even got contradicting (or different) results?
- . . .

Please share!

Outline

Ensuring Reproducibility

- Describe experiment setup and methods
- Share raw data
  - Use common (future proof) format
  - Follow FAIR principles

Ensuring Reproducibility

- Describe experiment setup and methods
- Share raw data
  - Use common (future proof) format
  - Follow FAIR principles
- Make backups

FAIR Data

Findable, **A**ccessible, **I**nteroperable, **R**eusable

https://www.go-fair.org/fair-principles

FAIR Data

Findable, **A**ccessible, **I**nteroperable, **R**eusable
https://www.go-fair.org/fair-principles

- Use identifier
  https://www.doi.org

FAIR Data

Findable, **A**ccessible, **I**nteroperable, **R**eusable
https://www.go-fair.org/fair-principles

- Use identifier
  https://www.doi.org
- Provide metadata
  - When, where, what data
  - Instruments used
  - Author contact
  - Domain relevant
  - Machine-readable

FAIR Data

Findable, **A**ccessible, **I**nteroperable, **R**eusable
https://www.go-fair.org/fair-principles

- Use identifier
  https://www.doi.org
- Provide metadata
  - When, where, what data
  - Instruments used
  - Author contact
  - Domain relevant
  - Machine-readable
- Openly retrievable

Code

- Use Version control (e.g. git)
  - Tag version used for paper:

```
git tag -a "Paper Title Year" -m "Description"
```

Code

- Use Version control (e.g. git)
  - Tag version used for paper:

    ```
    git tag -a "Paper Title Year" -m "Description"
    ```
- Write good documentation
- Ensure portability

Code

- Use Version control (e.g. git)
  - Tag version used for paper:

    ```
    git tag -a "Paper Title Year" -m "Description"
    ```
- Write good documentation
- Ensure portability
- Use common programming languages
- Seed random numbers

Code

- Use Version control (e.g. git)
  - Tag version used for paper:
    ```
    git tag -a "Paper Title Year" -m "Description"
    ```
- Write good documentation
- Ensure portability
- Use common programming languages
- Seed random numbers
- Use Citation File Format (CFF)
  - Cite software like paper
    https://citation-file-format.github.io

Citation File Format

https://citation-file-format.github.io

- Add *CITATION.cff* file to repo
- Contains metadata
- DOI via Zenodo https://zenodo.org
- References to other work
- Separate sets of authors for paper and code
- Supported by GitHub
  - And Zotero

```yaml
cff-version: 1.2.0
title: ironik
message: "If you use this software, please cite it as below."
type: software
authors:
  - orcid: 'https://orcid.org/0000-0002-7384-7304'
    affiliation: Georg-August-Universität Göttingen
    email: jonathan.decker@uni-goettingen.de
    family-names: Decker
    given-names: Jonathan
identifiers:
  - type: doi
    value: 10.1109/5.771073
repository-code: 'https://gitlab.gwdg.de/jonathan.decker1/ironik'
license: MIT
version: 0.1.5
date-released: '2022-05-16'
```

Discussion: Reproducibility in Your Research

1. Describe experiment setup and methods
2. Share all results and raw data
3. Make backups
   - Use common (future proof) format
   - Follow FAIR principles
     - Use identifier
     - Provide metadata
     - Openly retrievable
4. Use Version control (e.g. git)
   - Tag version used for paper
5. Write good documentation
6. Ensure portability
7. Use common programming languages
8. Seed random numbers
9. Use Citation File Format (CFF)
   - Cite software like paper

Outline

Pressure on Scientists

- Publication bias
  - Journals discouraged replication studies
  - Only "novel", "original" and "groundbreaking" research
  - Replication study literature is rare

Fidler and Wilcox, “Reproducibility of Scientific Results”

Pressure on Scientists

- Publication bias
  - Journals discouraged replication studies
  - Only "novel", "original" and "groundbreaking" research
  - Replication study literature is rare
- Publish or Perish
  - Every study must be significant
  - Results are cherry-picked or p-hacked
  - Manipulate for statistical significance
  - Null hypothesis ignored

Fidler and Wilcox, “Reproducibility of Scientific Results”

Pressure on Scientists

- Publication bias
  - Journals discouraged replication studies
  - Only "novel", "original" and "groundbreaking" research
  - Replication study literature is rare
- Publish or Perish
  - Every study must be significant
  - Results are cherry-picked or p-hacked
  - Manipulate for statistical significance
  - Null hypothesis ignored
- Lack of incentives to share data
- Hypothesising After Results are Known (HARKing)
- Leads to questionable research practice and irreproducible results

Fidler and Wilcox, “Reproducibility of Scientific Results”

Experimenter's Regress

- Experiment B is replica of A
- What if results conflict, either:
  - Conflicts are genuine $\rightarrow$ Combined hypothesis
  - B was not proper replica of A (including mistakes in A)

Fidler and Wilcox, "Reproducibility of Scientific Results"

Experimenter's Regress

- Experiment B is replica of A
- What if results conflict, either:
  - Conflicts are genuine $\rightarrow$ Combined hypothesis
  - B was not proper replica of A (including mistakes in A)

- Well-conducted experiment gives correct results
- Correct results confirmed via well-conducted experiment

Fidler and Wilcox, "Reproducibility of Scientific Results"

Experimenter's Regress

- Experiment B is replica of A
- What if results conflict, either:
  - Conflicts are genuine $\rightarrow$ Combined hypothesis
  - B was not proper replica of A (including mistakes in A)
- Well-conducted experiment gives correct results
- Correct results confirmed via well-conducted experiment
- Disproving original study requires multiple failed replicas

Fidler and Wilcox, "Reproducibility of Scientific Results"

Changes Needed

- Incentivize data-sharing and reproducibility
- Train scientists in statistical methods and bias detection

Fidler and Wilcox, “Reproducibility of Scientific Results”

Changes Needed

- Incentivize data-sharing and reproducibility
- Train scientists in statistical methods and bias detection
- Pre-register scientific predictions
  `https://osf.io`
  `https://aspredicted.org`
- Update journal publishing practices
  - Replica studies
  - "Failed" and non-significant studies
  - Require data-sharing

Fidler and Wilcox, “Reproducibility of Scientific Results”

Changes Needed

- Incentivize data-sharing and reproducibility
- Train scientists in statistical methods and bias detection
- Pre-register scientific predictions
  `https://osf.io`
  `https://aspredicted.org`
- Update journal publishing practices
  - Replica studies
  - "Failed" and non-significant studies
  - Require data-sharing
- Peer reviews check for data required for replica
- Award studies that follow these practices

Fidler and Wilcox, “Reproducibility of Scientific Results”

What's Next

- Check your research
  - Could someone else continue your work given your documentation?

What's Next

- Check your research
  - Could someone else continue your work given your documentation?
  - What if your computer died, could you recover?

What's Next

- Check your research
  - Could someone else continue your work given your documentation?
  - What if your computer died, could you recover?
- Help others use your work!
  - Scientific progress is iterative

“Kohn’s Second Law: An experiment is reproducible until another laboratory tries to repeat it.”

— Alexander Kohn, Fortune or Failure, 1989

Kohn, *Fortune Or Failure: Missed Opportunities and Chance Discoveries*

References

Alston, Jesse M. and Jessica A. Rick. “A Beginner’s Guide to Conducting Reproducible Research”. In: *The Bulletin of the Ecological Society of America* 102.2 (2021), e01801. DOI: https://doi.org/10.1002/bes2.1801. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/bes2.1801. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/bes2.1801.

Browne, Malcolm. “Physicists Debunk Claim Of a New Kind of Fusion”. In: (May 3, 1989). URL: https://archive.nytimes.com/www.nytimes.com/library/national/science/050399sci-cold-fusion.html (visited on 06/16/2022).

Fidler, Fiona and John Wilcox. “Reproducibility of Scientific Results”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2021. Metaphysics Research Lab, Stanford University, 2021. URL: https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/ (visited on 06/15/2022).

Ioannidis, John PA. “Why most published research findings are false”. In: *PLoS medicine* 2.8 (2005), e124.

Kohn, A. *Fortune Or Failure: Missed Opportunities and Chance Discoveries*. Blackwell, 1989. ISBN: 9780631160878. URL: https://books.google.de/books?id=SNyNQgAACAAJ.

Popper, Karl Raimund and Gary James Jason. *The Logic of Scientific Discovery*. Psychology Press, 2002. 548 pp. ISBN: 978-0-415-27844-7.

Schooler, Jonathan W. “Metascience could rescue the ‘replication crisis’”. In: *Nature* 515.7525 (2014), pp. 9–9.