
1: Intro into reproducibility

Reproducibility, Open Software and Open Data, Reproducible Scripts

Kosmas Hench

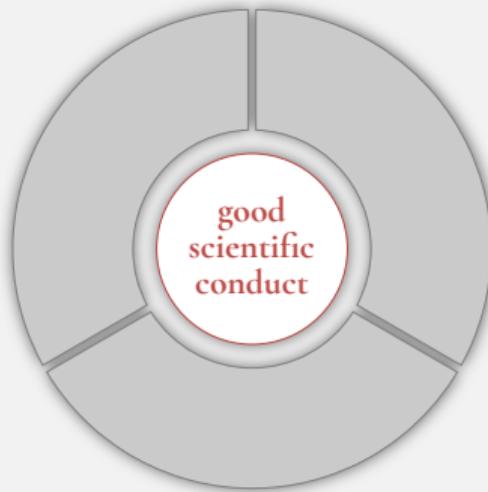


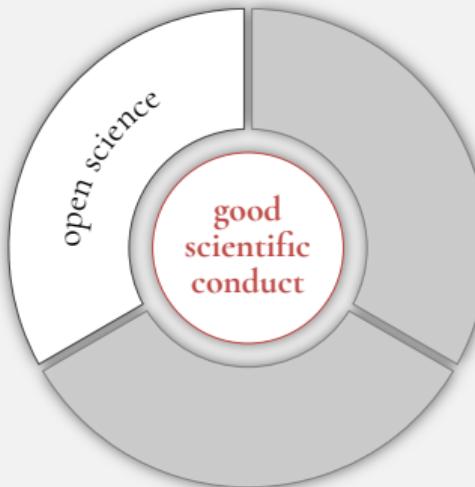
2025-05-26



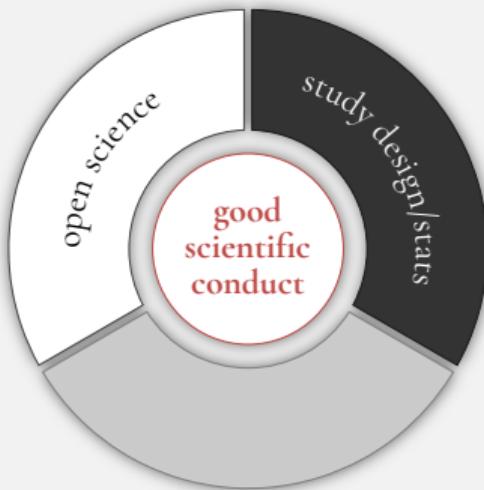
Good Scientific Conduct

Image by Pop & Zebra

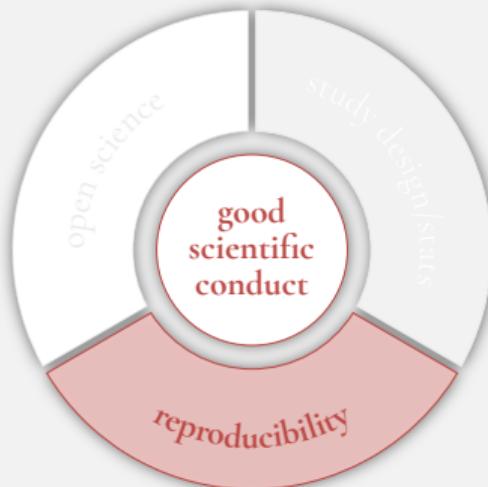


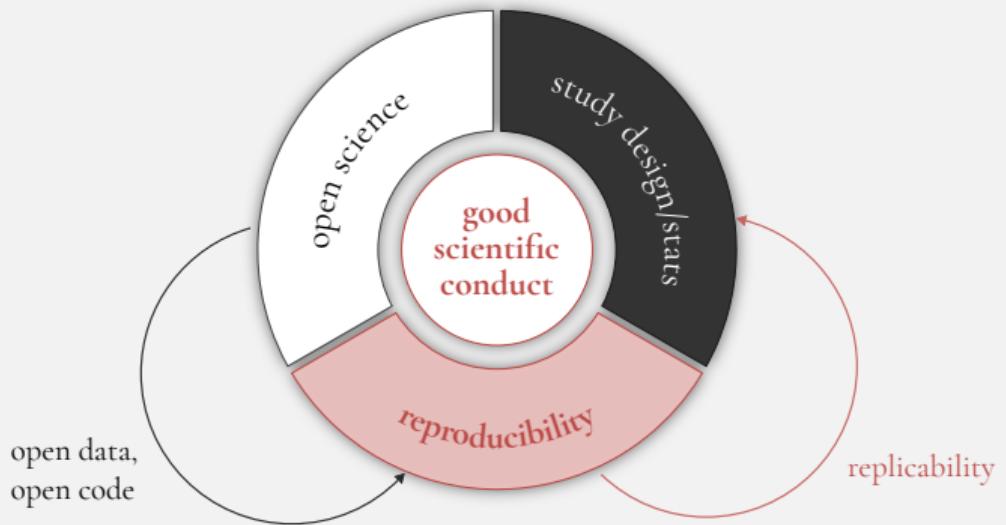


Garbage in, garbage out!



Same garbage in, same garbage out!





Semantics

original study repeatable



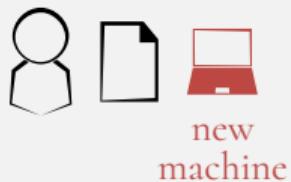
same results

Semantics

original study



Runnable



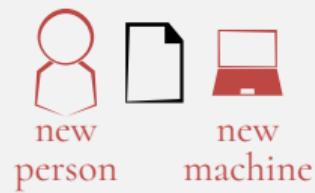
same results

Semantics

original study

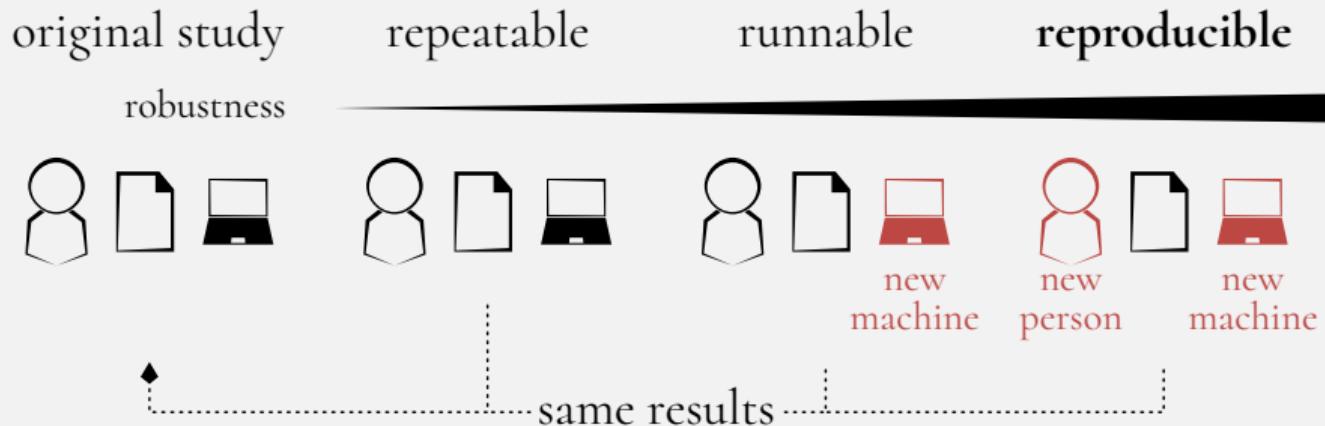


reproducible

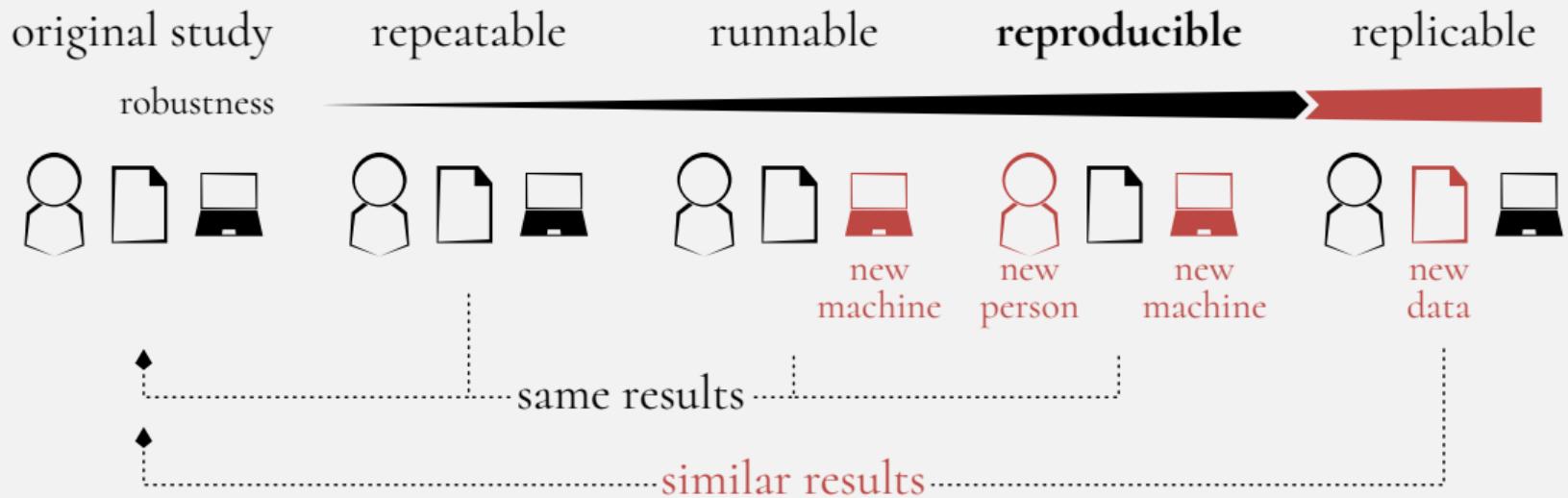


same results

Semantics

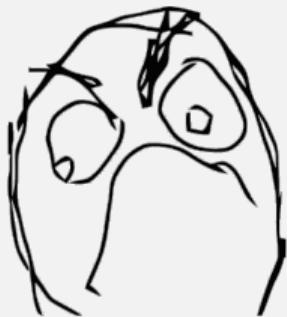


Semantics

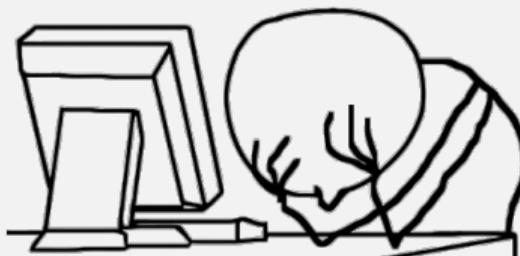


perils for reproducibility

different operating
systems



proprietary software



breaking changes
in software



reproducibility
by design

reproducibility by design

documentation

resource availability

structure

reproducibility by design

documentation

- commenting code
- markdown

resource availability

structure

reproducibility by design

documentation

- commenting code
- markdown

resource availability

- open code
- open data
- open software

structure

reproducibility by design

documentation

- commenting code
- markdown

resource availability

- open code
- open data
- open software

structure

repository types (DOI / accession nr)

code: zenodo (not github!)

data: dryad, pangaea, figshare

DNA: ENA, SRA, GenBank

reproducibility by design

documentation

- commenting code
- markdown

resource availability

- open code
- open data
- open software

structure

- folder structure
- code organization
- rstudio projects
- workflow manager
(snakemake, nextflow)

reproducibility by design

documentation

- commenting code
- markdown

resource availability

- open code
- open data
- open software

structure

- folder structure
- code organization
- rstudio projects
- workflow manager
(snakemake, nextflow)

working environment

- version numbers
- conda environments
- containerization

reproducibility by design

documentation

- commenting code
- markdown

resource availability

- open code
- open data
- open software

structure

- folder structure
- code organization
- rstudio projects
- workflow manager
(snakemake, nextflow)

repository types (DOI / accession nr)

code: zenodo (not github!)

data: dryad, pangaea, figshare

DNA: ENA, SRA, GenBank

working environment

- version numbers
- conda environments
- containerization



Reproducible Scripts

Image by Glenn Carstens Peters

Design



Image by Maxence Pira

Execution

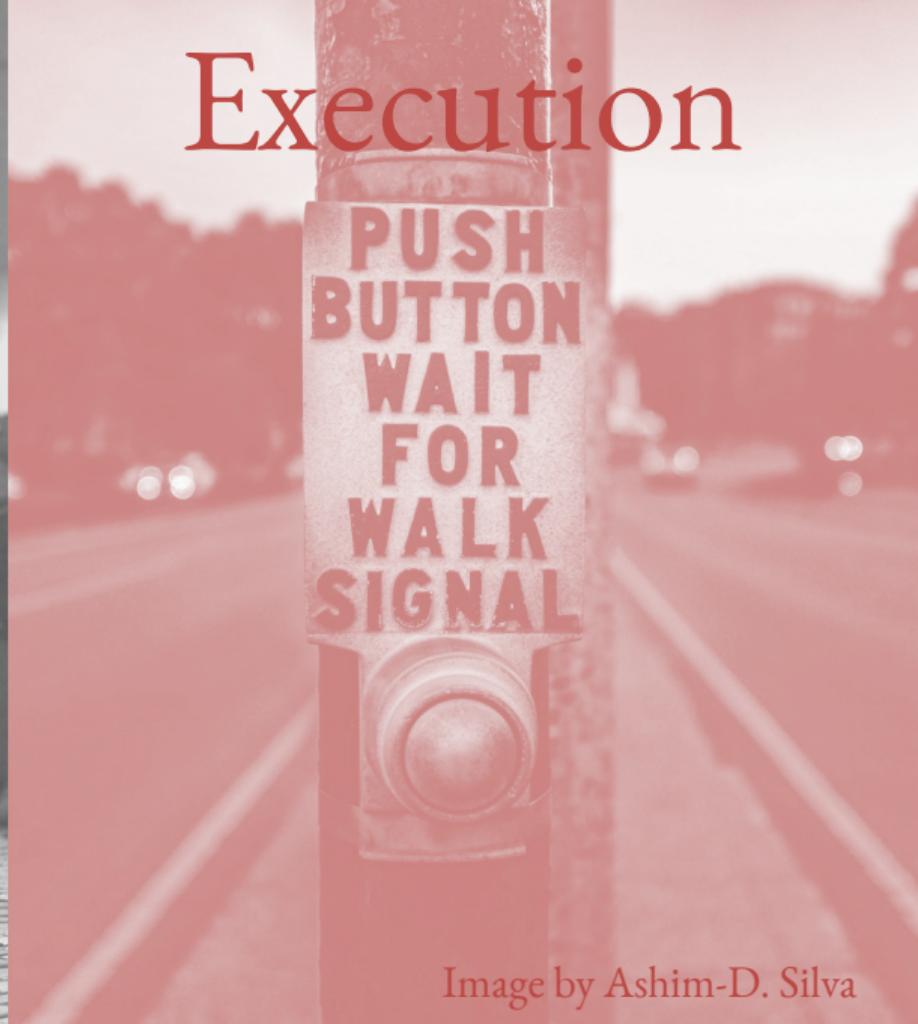
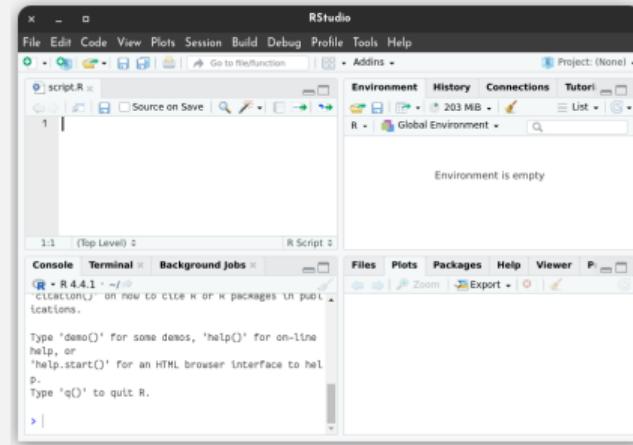
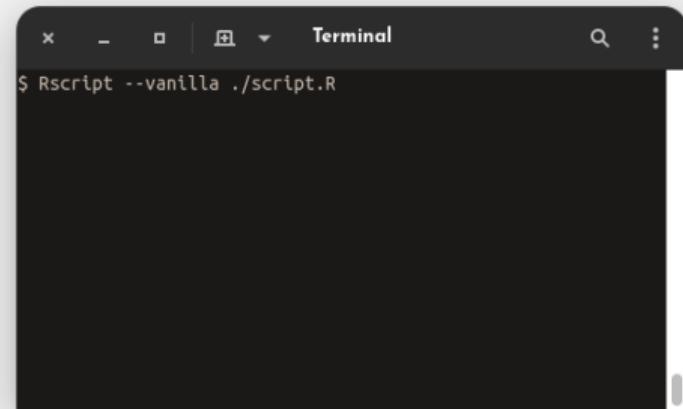


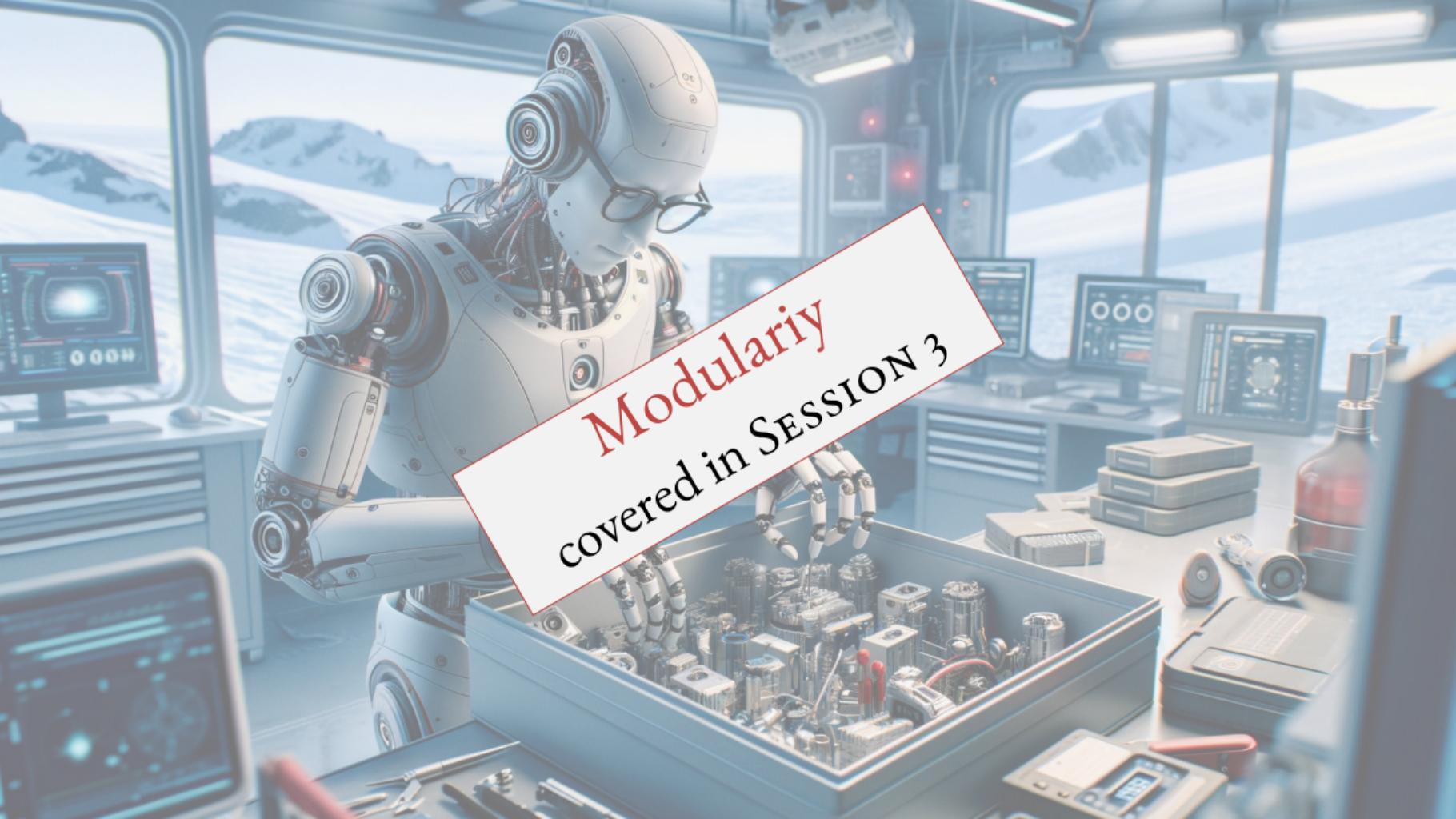
Image by Ashim-D. Silva

Design



Execution



A white and silver humanoid robot with large, expressive eyes and a complex mechanical structure is shown from the waist up, looking down at a workbench. It is positioned in a high-tech laboratory or workshop with multiple computer monitors displaying data in the background. In the foreground, a red-bordered white banner with a diagonal gradient from light blue to red contains the text.

Modularity
covered in SESSION 3

Random Numbers

→ set your seed

sampling / subsetting
MCMC runs
simulation
parameter optimization
plot jittering
...



```
> sessionInfo()
```

Version Numbers

```
R version 4.4.1 (2024-06-14)
Platform: x86_64-pc-linux-gnu
Running under: Debian GNU/Linux 12 (bookworm)

Matrix products: default
BLAS:    /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
LAPACK:  /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p0.3.21.so; LAPACK version 3.11.0

locale:
[1] LC_CTYPE=en_US.UTF-8     LC_NUMERIC=C      LC_TIME=en_US.UTF-8     LC_COLLATE=en_US.UTF-8

time zone: Europe/Berlin
tzcode source: system (glibc)

attached base packages:
[1] stats      graphics   grDevices utils      datasets   methods    base

other attached packages:
[1] ape_5.8

loaded via a namespace (and not attached):
[1] compiler_4.4.1  parallel_4.4.1 Rcpp_1.0.12   nlme_3.1-165  grid_4.4.1   digest_0.6.35 lattice_0.22-6
```



Code Clarity

Documentation

why, not what
in moderation

Consistent Style

check out style guides
if possible, don't repeat yourself

DFG



Thank you for
your attention!

created using dalle3