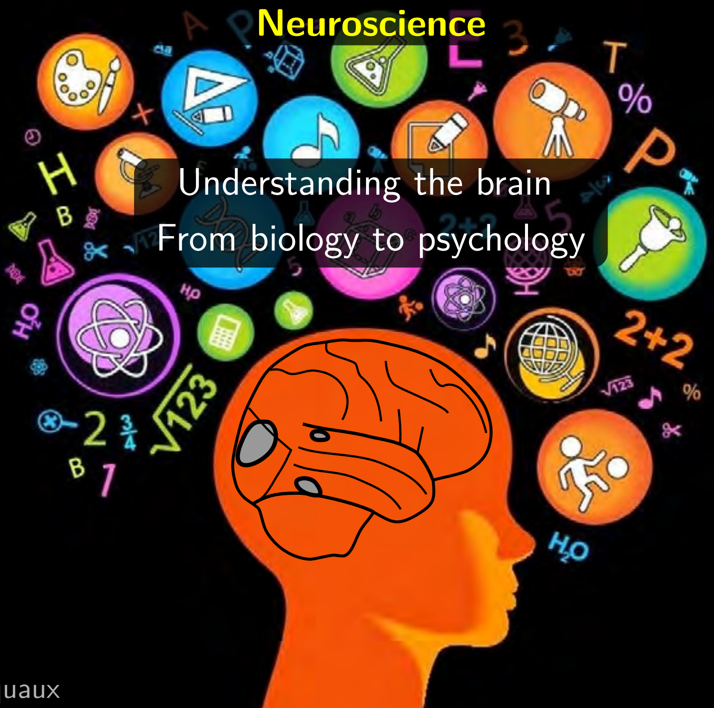
Computational and statistical reproducibility in data-intensive neuroscience

Gaël Varoquaux

Inria

NeuroSpin

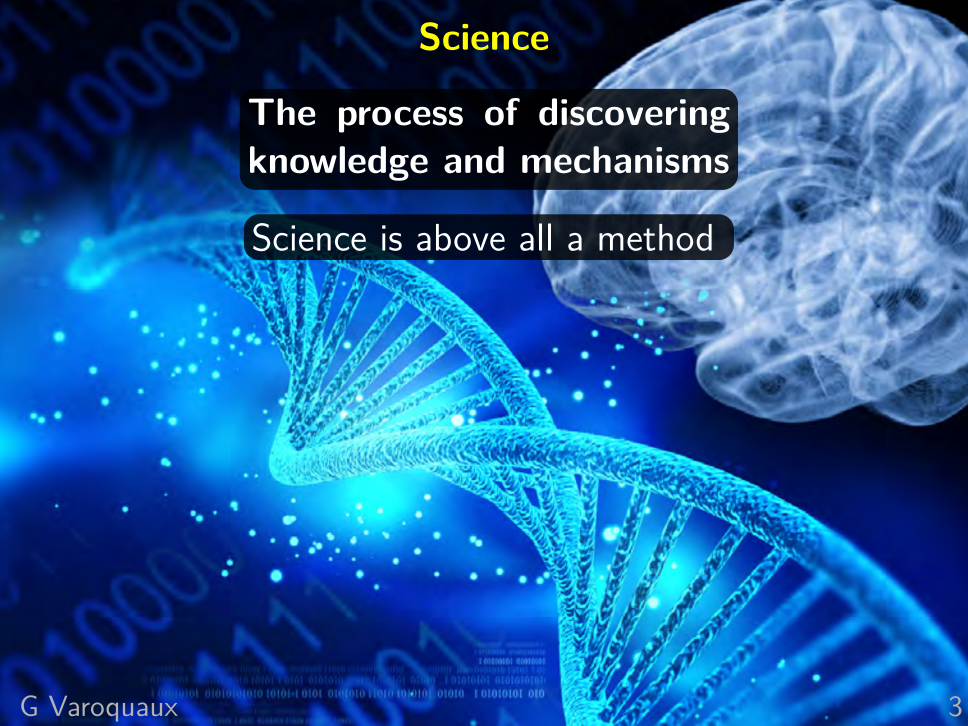
Neuroscience



Understanding the brain

From biology to psychology

Science



The process of discovering knowledge and mechanisms

Science is above all a method

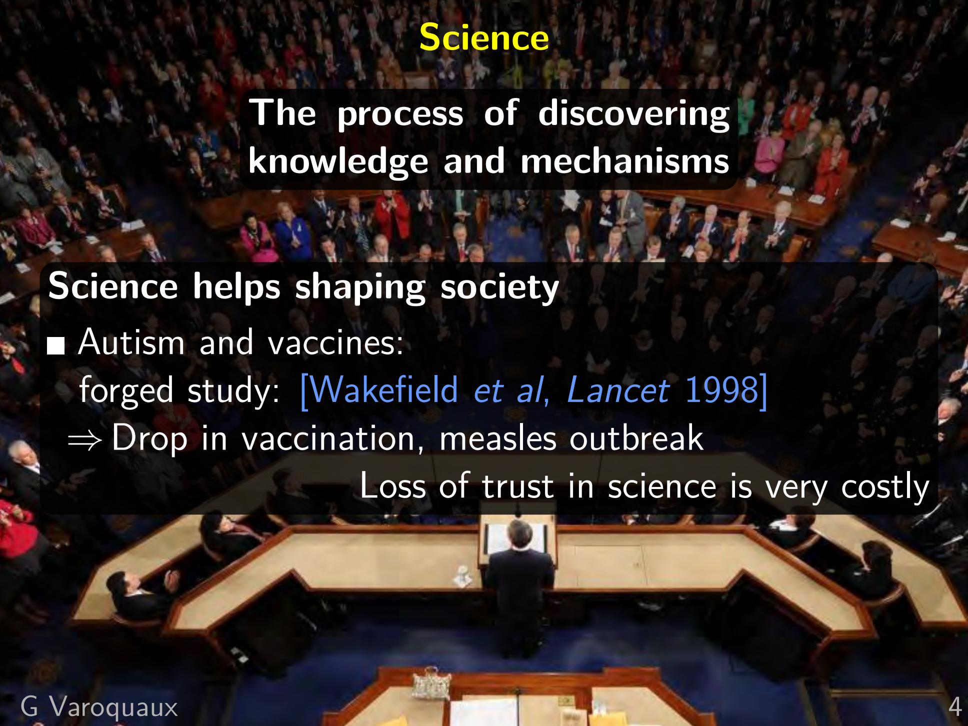
Science

The process of discovering knowledge and mechanisms

Science is above all a method

"Science is not a political construct or a belief system. Scientific progress depends on openness, transparency, and the free flow of ideas and people."

— Dr. Rush Holt, CEO of AAAS, testimony to the House Committee on Science, Space, and Technology, Feb 8, 2017



Science

The process of discovering knowledge and mechanisms

Science helps shaping society

- Autism and vaccines:
  forged study: [Wakefield *et al*, *Lancet* 1998]

$\Rightarrow$ Drop in vaccination, measles outbreak

Loss of trust in science is very costly

Science in the age of data:

good process = reproducibility

better: generalization & reuse

1. **Statistical reproducibility**

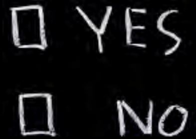
2. **Computational reproducibility**

1 Statistical reproducibility

Brain theories will just emerge from data

Brain theories will just emerge from data

Biology, neuroscience, psychology...
rely on working hypotheses

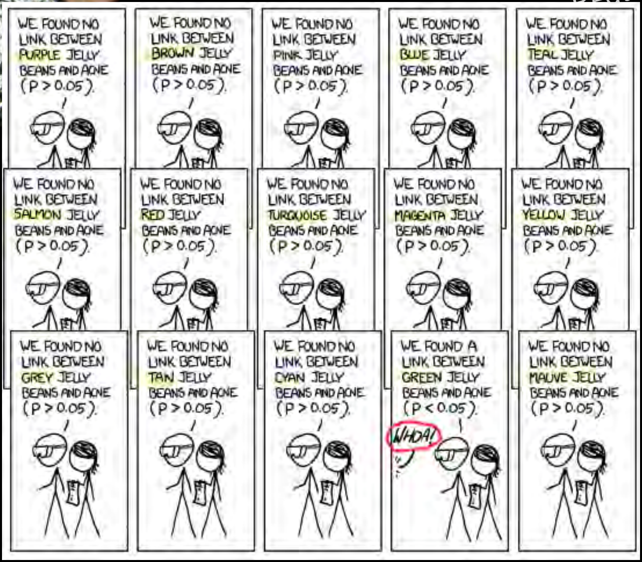


Why the tyranny of the *Working hypothesis*?

Reductionist

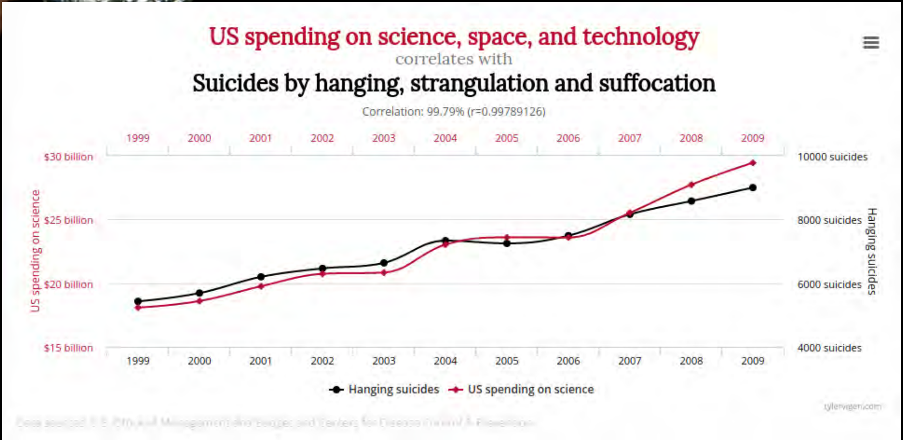
Why the tyranny of the *Working hypothesis*?

Reductionist



Why the tyranny of the *Working hypothesis*?

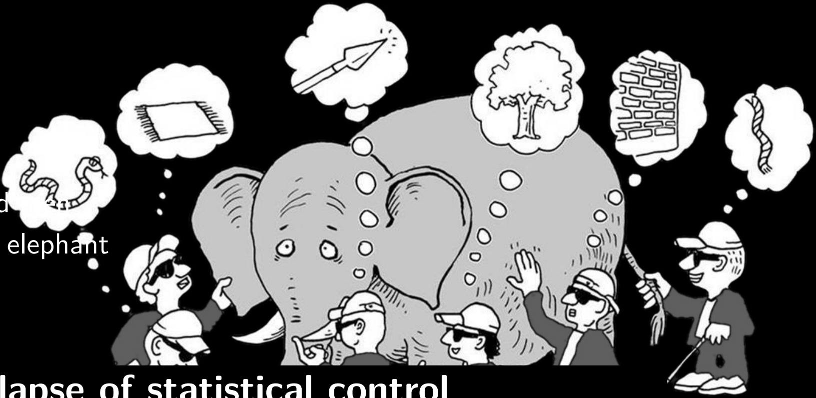
Reductionist



http://www.tylervigen.com/spurious-correlations

Reductionist approach: $\Rightarrow$ fragmentation

Six blind men
and the elephant



Collapse of statistical control

- Analytic variability $\Rightarrow$ uncontrolled variance
- Publication incentives $\Rightarrow$ selection bias

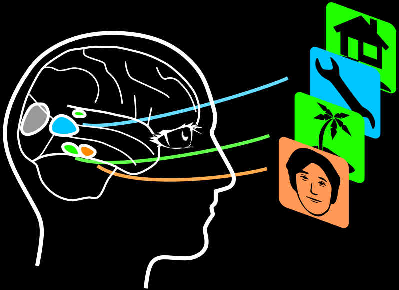
[Ioannidis 2008]

Generalization

can build broader theories

[Varoquaux and Poldrack 2019]

Paradigm 1: Seen



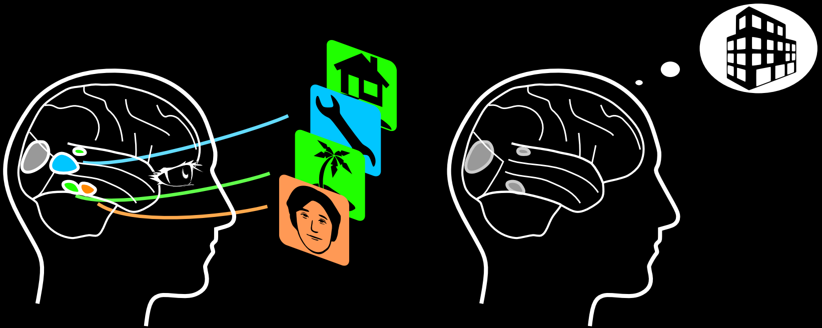
Generalization

can build broader theories

[Varoquaux and Poldrack 2019]

Paradigm 1: Seen

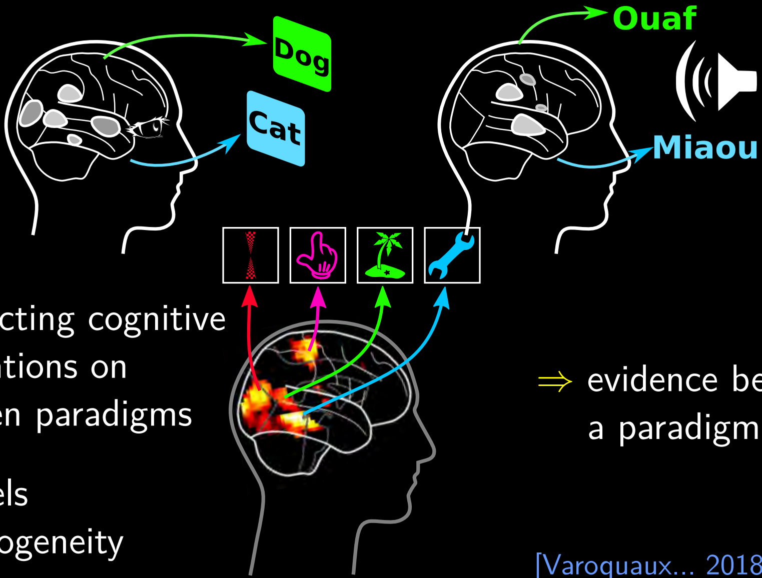
Paradigm 2: Imagined



"A theory is a good theory if it satisfies two requirements: It must **accurately describe a large class of observations** on the basis of a model that contains only a few arbitrary elements, and it must **make definite predictions about the results of future observations**."

Stephen Hawking, A Brief History of Time.

Across tasks: atlasing cognition



Predicting cognitive operations on unseen paradigms

Models heterogeneity

$\Rightarrow$ evidence beyond a paradigm

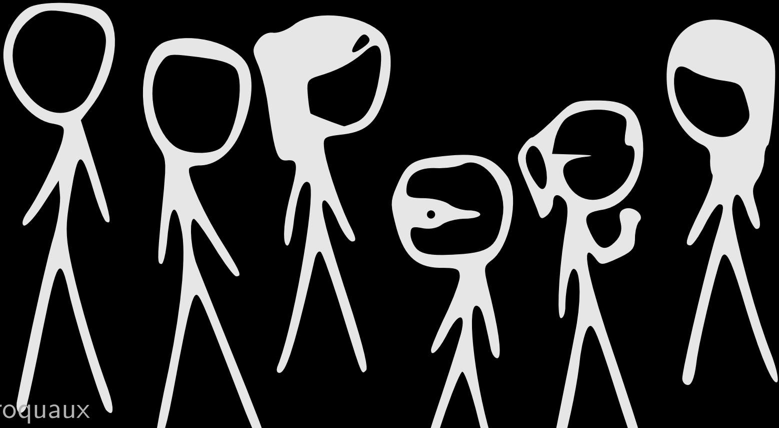
[Varoquaux... 2018]

Across subjects: biomarkers

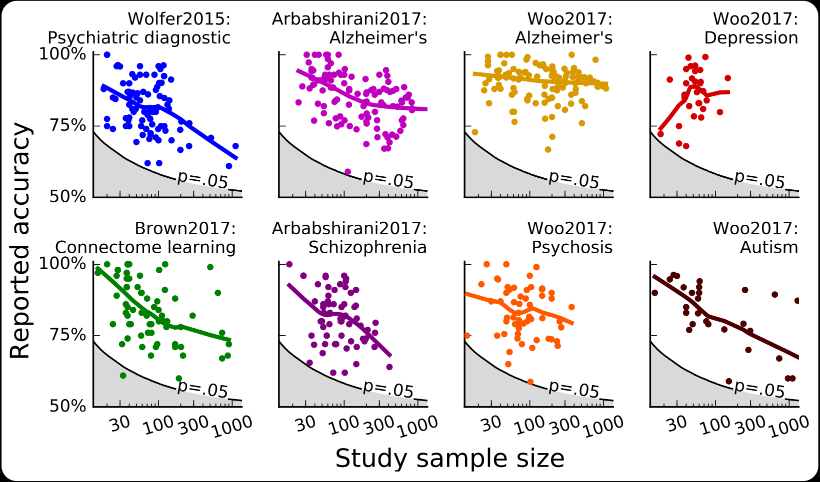
Predicting autism status to new sites [Abraham... 2017]

Many samples overcome heterogeneity

strong generalization



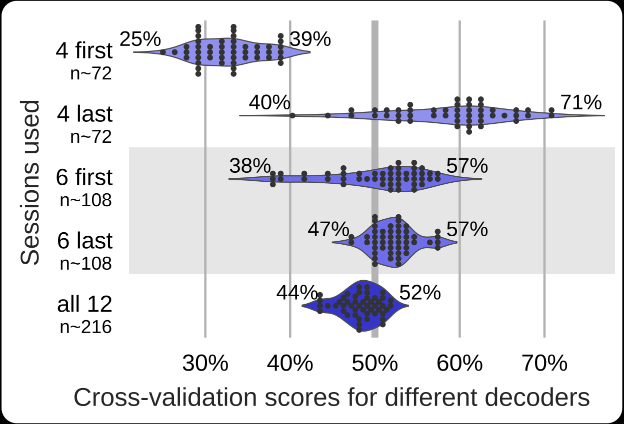
In the literature, effect sizes decrease with sample sizes



[Varoquaux 2017]

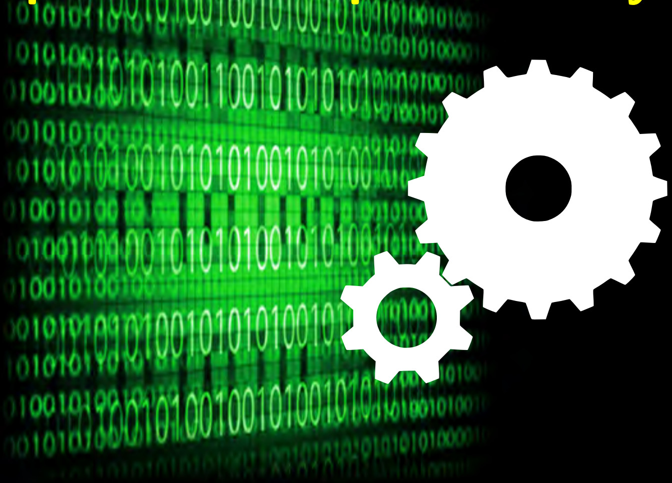
1 Cross-validation failure: not enough data

Analytic variability strikes back



[Varoquaux 2017]

2 Computational reproducibility



2 Trust and productivity: reproducible research

"if it's not open and verifiable by others, it's not science, or engineering, or whatever it is you call what we do"

— V. Stodden, *The scientific method in practice*

Computational reproducibility:

- Automate everything
- Control the environment

Reproducibility

Rerun and come to the same conclusion

An argument for copying all scripts for each paper

Better: a tag in git

2 Libraries enable reproducible science

Reproducibility

Rerun and come to the same conclusion

An argument for copying all scripts for each paper

Better: a tag in git

Copying scripts scales very poorly:

- Accumulation of half-dead code $\Rightarrow$ cognitive overload
- No consolidation across studies
- Each study has different variants of different bugs

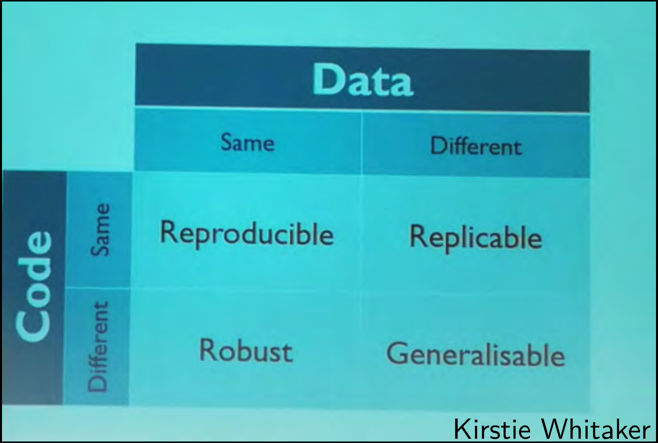
Garden of forking code

Reproducibility

Rerun and come to the same conclusion

An argument for copying all scripts

for each paper   Better: a tag in git



Kirstie Whitaker

2 Libraries enable reproducible science

Reproducibility

Rerun and come to the same conclusion

An argument for copying all scripts for each paper

Frozen food

Reusability

Apply the approach to a new problem

Being able to understand, modify, run in new settings

Generalization is the true scientific test

Reusable computational science = **libraries**

scikit-learn

Make research in machine-learning models and algorithm useable to people who do not understand them

nilearn

Make it easy to answer neuroimaging problems with them

2 Building foundations of neuroimaging with computers

scikit-learn

Make research in machine-learning models and algorithm useable to people who do not understand them

Challenges:
- **Variety of that space**
- **Statistical concepts $\gg$ coding concepts**

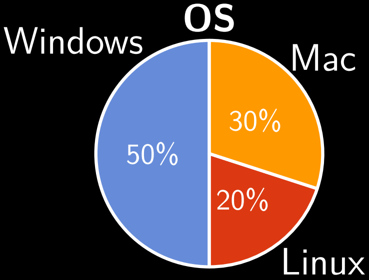
nilearn

Make it easy to answer neuroimaging problems with them

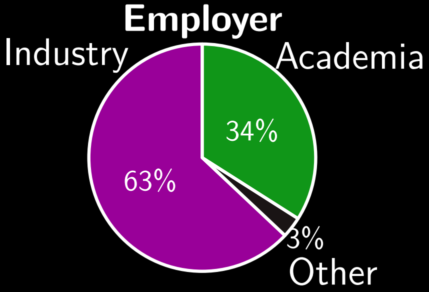
Challenges:
- **Onboarding technology-adverse users**

Scikit-learn, an impact

½ million returning users



12 000 citations



[1] monthly users reading the docs

Factor 10 in time investment

- Corner cases in algorithm (numerical stability)
- Multiple platforms and library versions
- Documentation
- Making it simpler (and get less educated users)
- User and developer support ($\sim$ 100 mails/day)

An impact on science and society

Factor 10 in time investment

Technical + scientific tradeoffs

- Ease of install/ease of use rather than speed
- Focus on "old science"

- Software good practice mandatory:
  - Automated testing
  - Version control

- Open source to grow a community

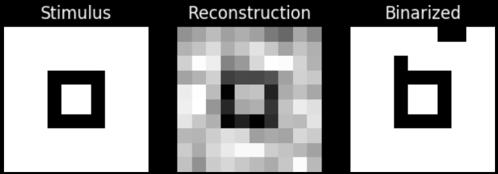
2 Reusable science

scikit-learn is the new machine-learning textbook

nilearn is the new neuroimaging review article

Experiments reproduced at each commit

eg: brain reading



nilearn.github.io/auto_examples/02_decoding/plot_miyawaki_reconstruction.html

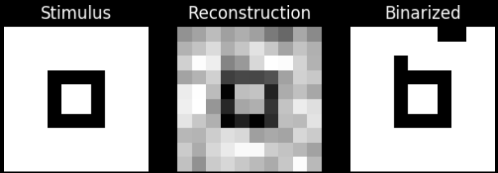
2 Reusable science

scikit-learn is the new machine-learning textbook

nilearn is the new neuroimaging review article

Experiments reproduced
at each commit

eg: brain reading



nilearn.github.io/auto_examples/02_decoding/plot_miyawaki_reconstruction.html

Resource intensive Continuous integration:

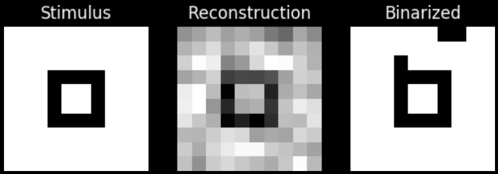
- Data ⇒ Fight for good open data
- Computation ⇒ Find good algorithms and tradeoffs

Forces us to distill the literature (as a review)

scikit-learn is the new machine-learning textbook

nilearn is the new neuroimaging review article

Experiments reproduced
at each commit
eg: brain reading



nilearn.github.io/auto_examples/02_decoding/plot_miyawaki_reconstruction.html

Package development consolidates science and moves it outside the lab

scikit-learn is the new machine-learning textbook

nilearn is the new neuroimaging review article

Experiments reproduced
at each commit

eg: brain reading

`scipy-lectures`: **living book for Python in science**

Package development consolidates science and moves it outside the lab

Statistical and computational reproducibility

Statistical

- Variability in question & methods makes control hard
- Aim for generalization:

Broader theories in the face of heterogeneity

[Varoquaux and Poldrack 2019]

Computational

- Libraries enable reproducibility
- Aim for reusability

Easy of reuse and reproducibility fosters innovation

@GaelVaroquaux

Open data is essential

References I

A. Abraham, M. P. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 147:736–745, 2017.

J. P. Ioannidis. Why most discovered true associations are inflated. *Epidemiology*, 19(5):640–648, 2008.

G. Varoquaux. Cross-validation failure: small sample sizes lead to large error bars. *NeuroImage*, 2017.

G. Varoquaux and R. A. Poldrack. Predictive models avoid excessive reductionism in cognitive neuroimaging. *Current Opinion in Neurobiology*, 2019.

G. Varoquaux, Y. Schwartz, R. A. Poldrack, B. Gauthier, D. Bzdok, J. Poline, and B. Thirion. Atlases of cognition with large-scale brain mapping. *PLOS Comp Bio*, 2018.