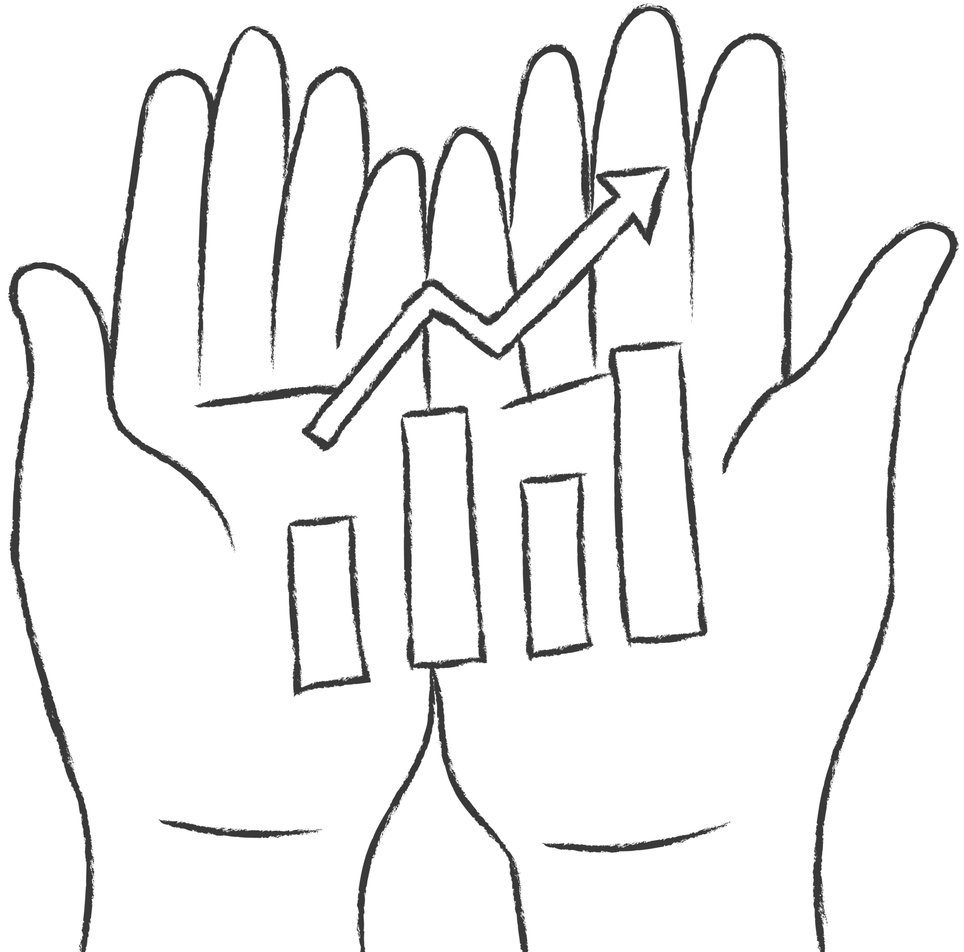
Reproducibility & data accessibility in science



Rachel Warnock

10.04.2024

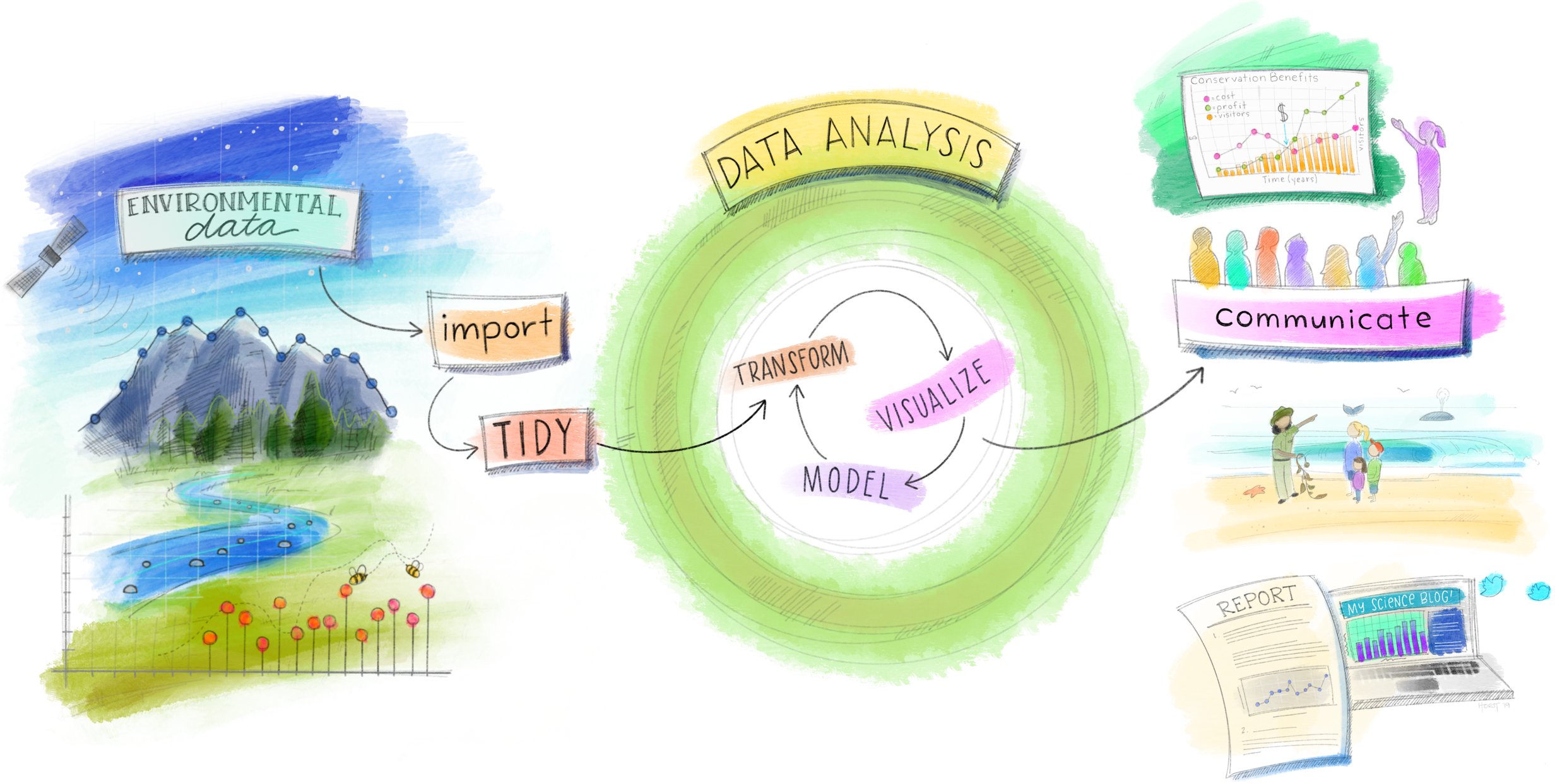
Reproducibility

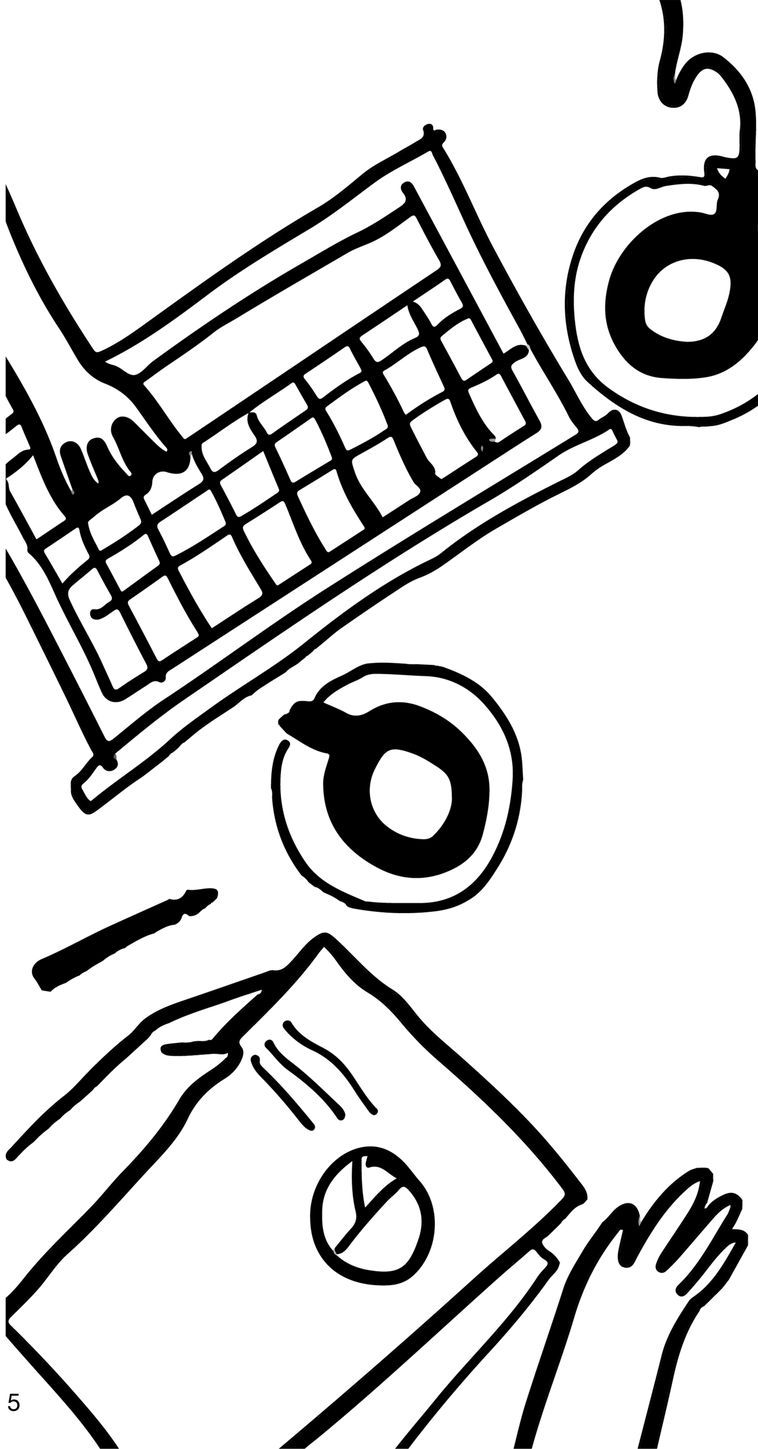
Requires that scientific papers provide sufficiently clear instructions to allow successful **repetition** and **extension** of analyses.

Replicability vs. reproducibility

replicability — Obtaining consistent results across studies aimed at answering the same scientific question.

reproducibility — Obtaining consistent computational results using the same input data, computational steps, methods, code and analytical conditions.


ENVIRONMENTAL data
import
TIDY
DATA ANALYSIS
TRANSFORM
VISUALIZE
MODEL
Conservation Benefits
cost
profit
visitors
$
Time (years)
visitors
Communicate
REPORT
My Science BLOG!

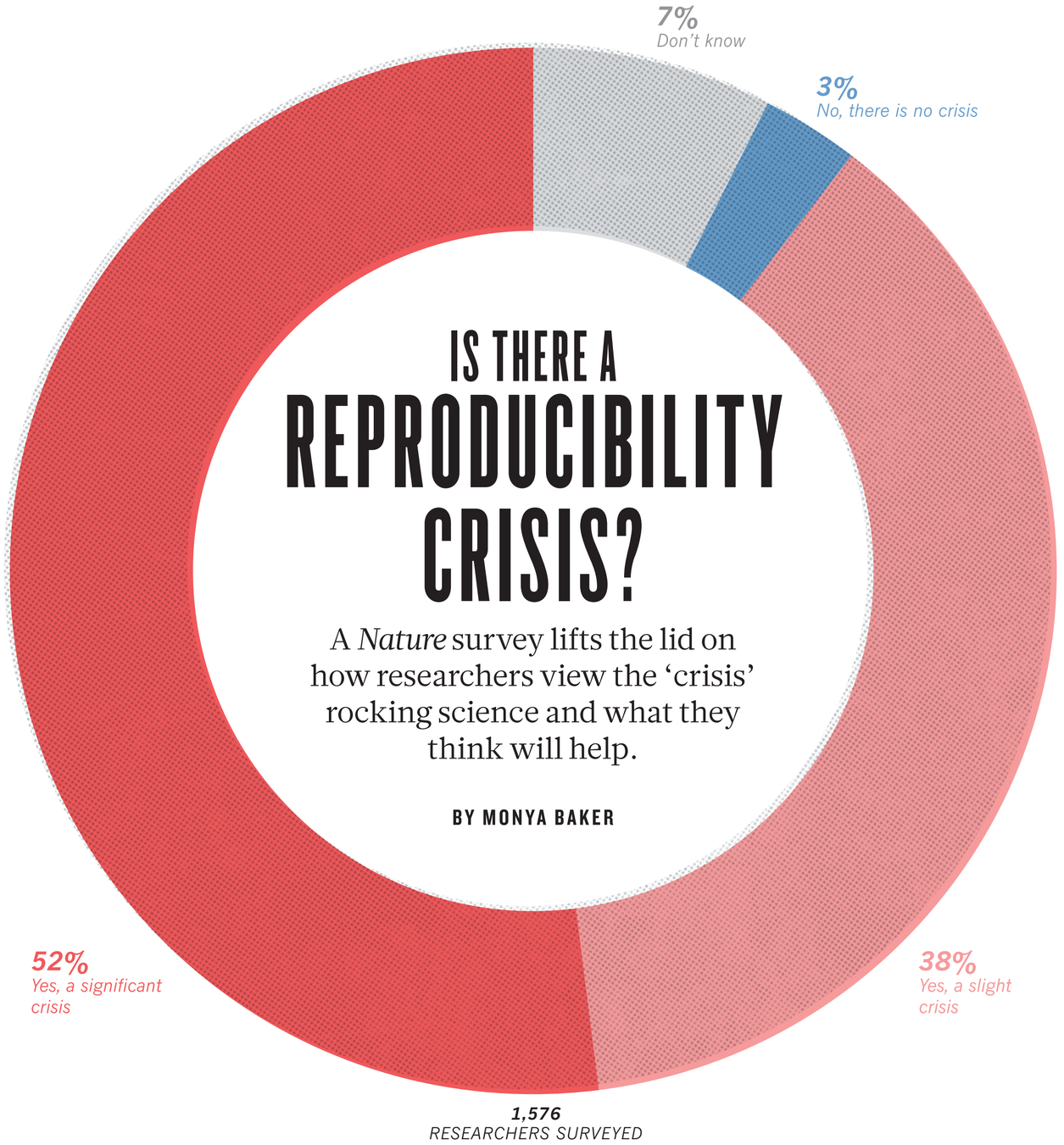


We’re often doing complicated things — with more data, complex tools and increased interdisciplinary research (across multiple research teams).

But lessons in reproducibility are valuable even when you're working alone or *not* doing something hugely complex!

There are many reasons to benefit from the steps taken to ensure reproducibility.

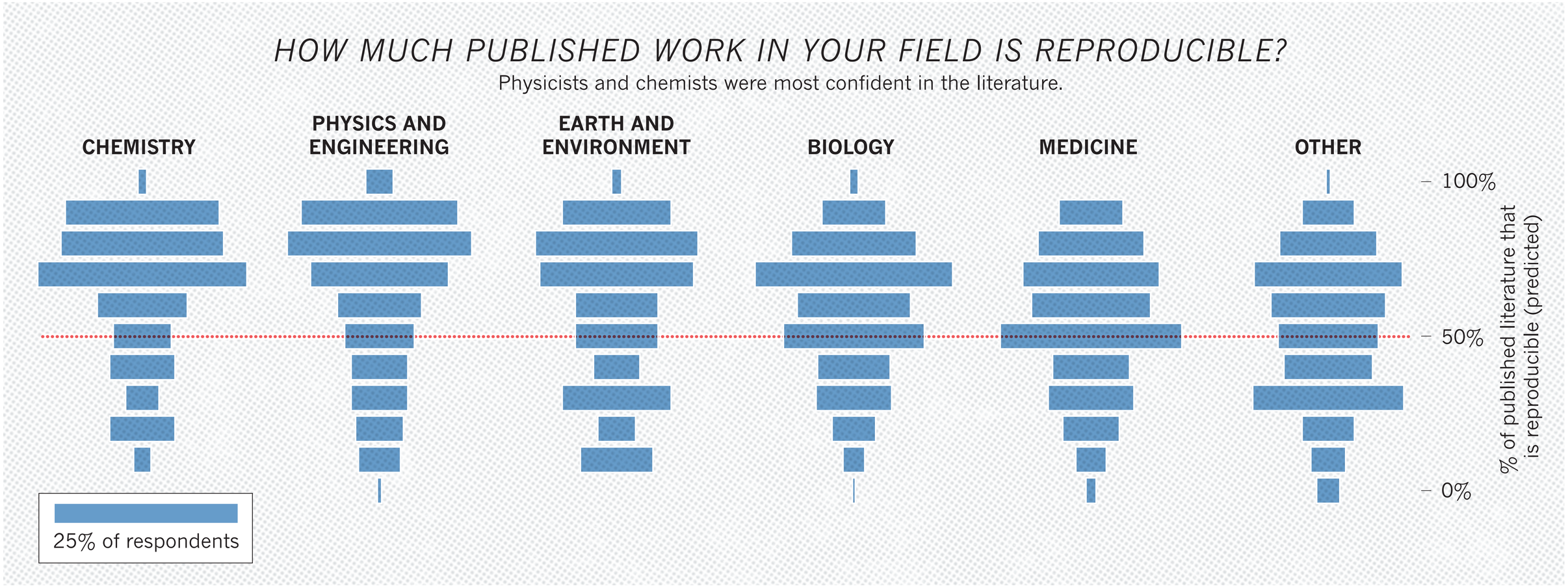
Research shows scientific papers often leave out details that are necessary or essential to reproduce the results.

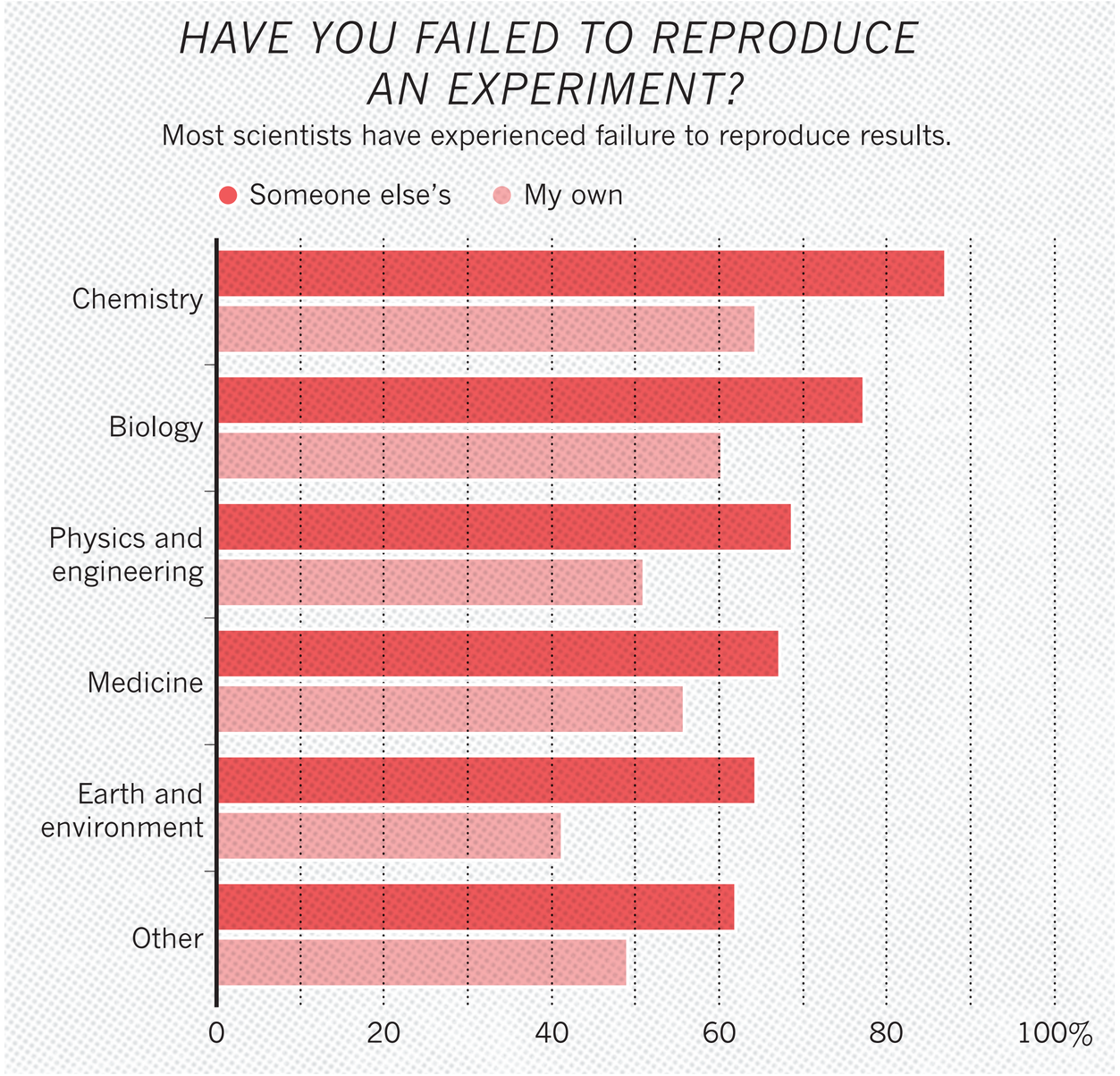


IS THERE A REPRODUCIBILITY CRISIS?

A *Nature* survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help.

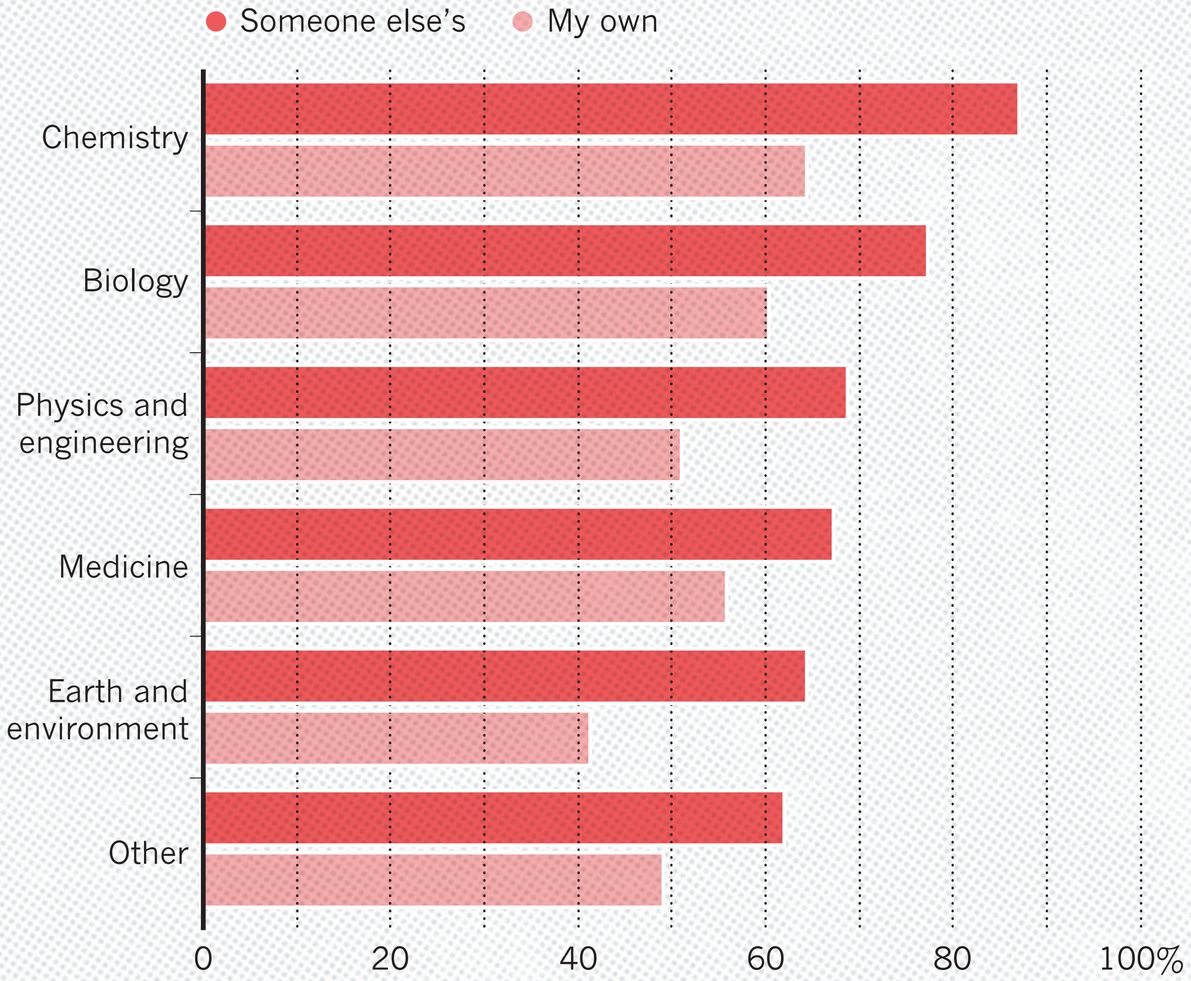
BY MONYA BAKER


HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?
Physicists and chemists were most confident in the literature.
CHEMISTRY
PHYSICS AND ENGINEERING
EARTH AND ENVIRONMENT
BIOLOGY
MEDICINE
OTHER
100%
50%
0%
% of published literature that is reproducible (predicted)
25% of respondents


HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%

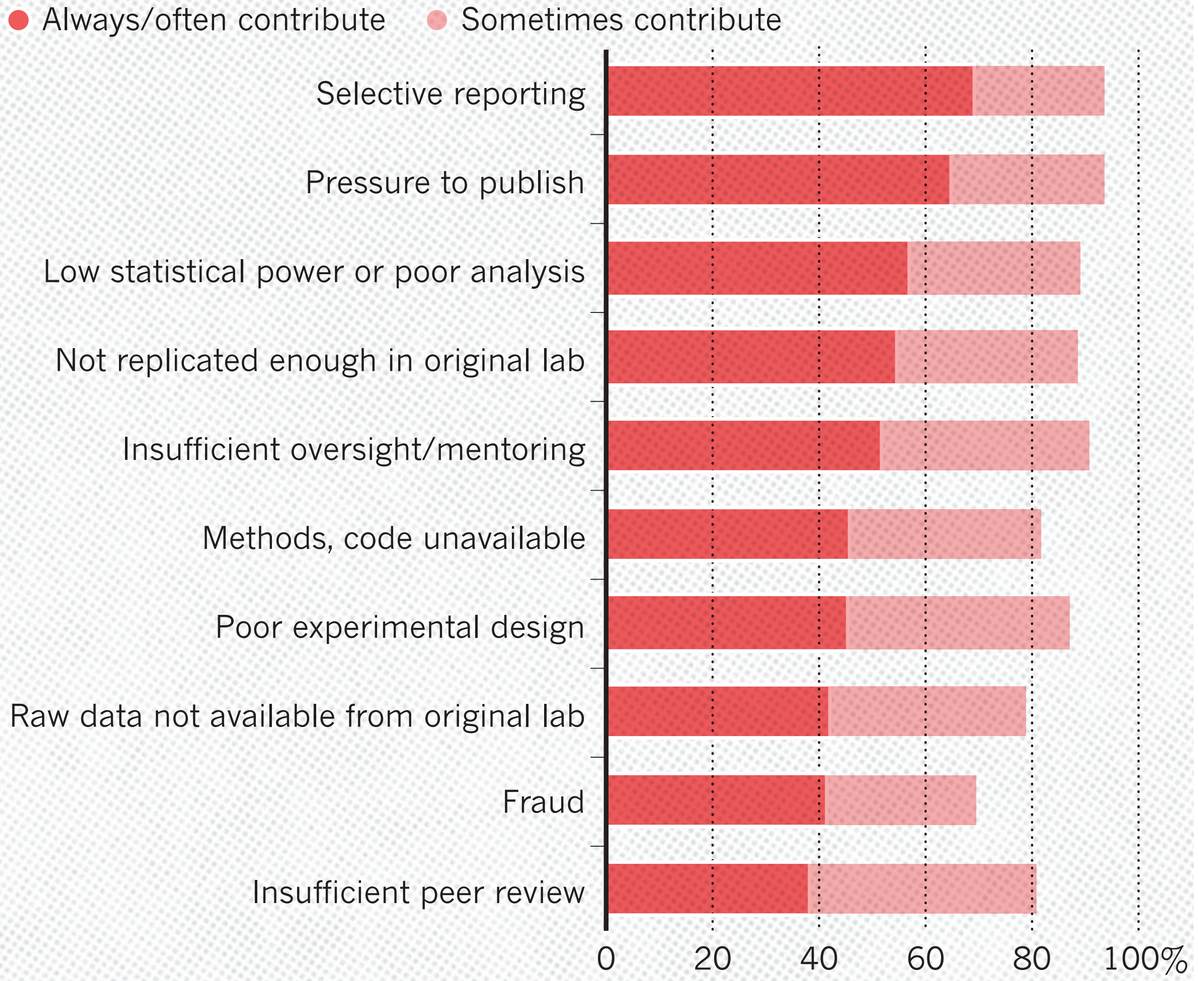
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



Retractions in the scientific literature: is the incidence of research fraud increasing?

R Grant Steen

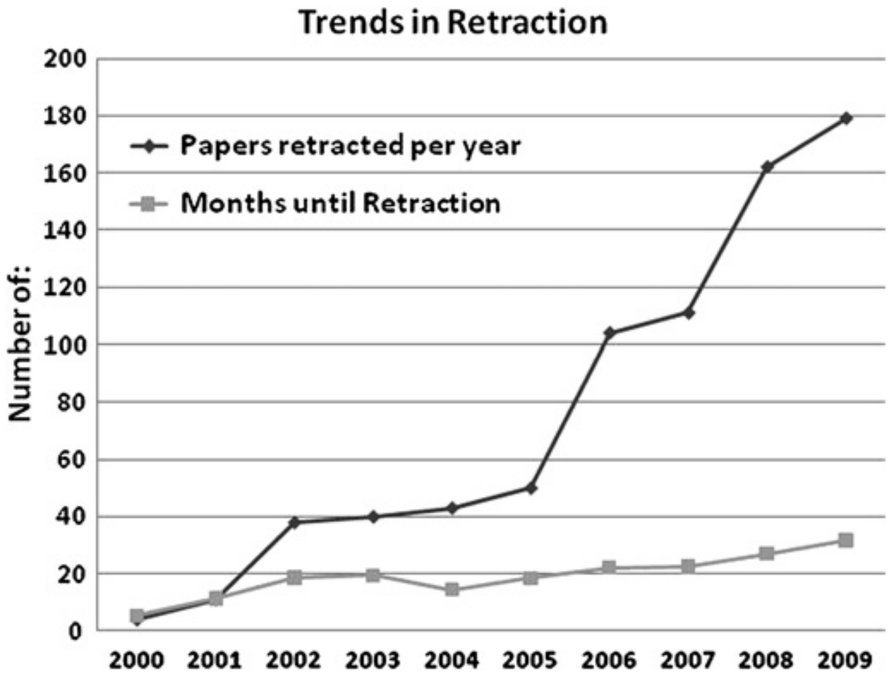


Figure 1 Trends in paper retraction showing the number of papers retracted by year, including retractions for all reasons. The average time until retraction is also shown.

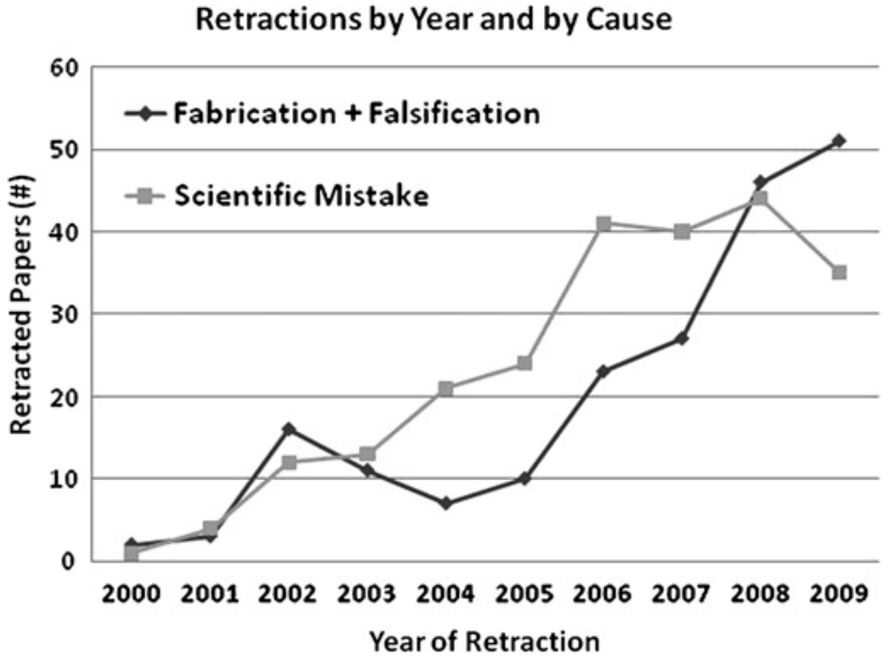


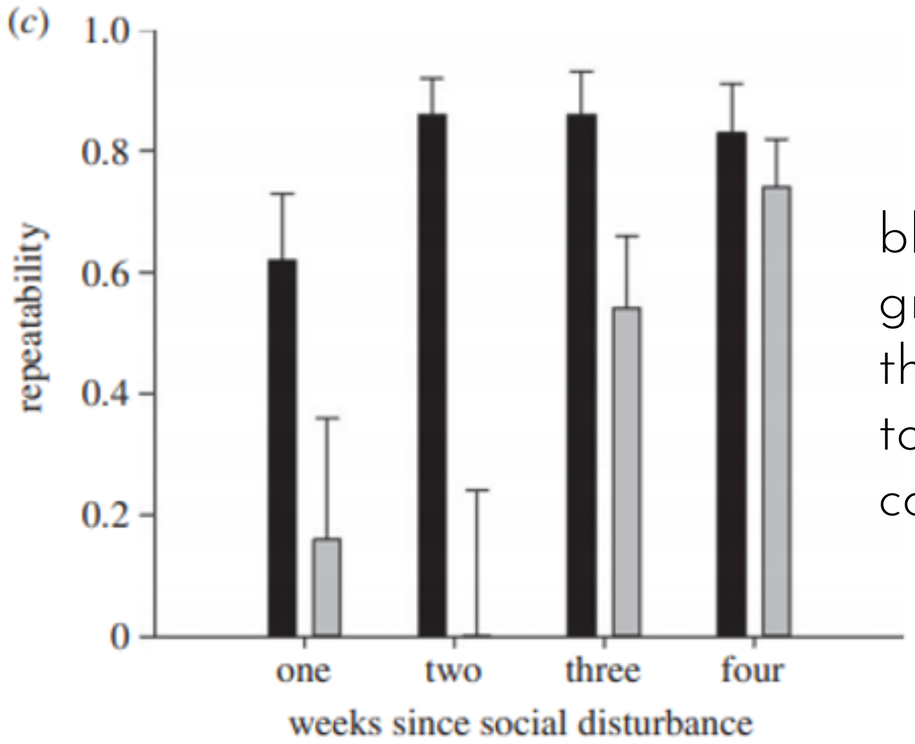
Figure 2 Number of papers retracted by year, including only 197 papers retracted for fraud (fabrication or falsification) and 234 papers retracted for scientific mistake. This figure therefore excludes all papers retracted for non-scientific error (eg, plagiarism, duplicate publication).

Spider biologist denies suspicions of widespread data fraud in his animal personality research

Behavioral ecologists are in turmoil as dozens of research papers involving an expert on social spiders draw scrutiny

31 JAN 2020 • BY ELIZABETH PENNISI

Science



black= control
grey = spiders
that have moved
to another
colony

Downloaded from https://royalsocietypublishing.org/ on 23 February 2022

PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Laskowski KL, Pruitt JN. 2014 Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider. *Proc. R. Soc. B* **281**: 20133166. http://dx.doi.org/10.1098/rspb.2013.3166

Received: 3 December 2013
Accepted: 28 February 2014

Subject Areas:
behaviour

Keywords:
social niche specialization, familiarity, variance partitioning, repeatability, *Stegodyphus mimosarum*

Author for correspondence:
Kate L. Laskowski
e-mail: klaskowski@igb-berlin.de

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2013.3166 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider

Kate L. Laskowski[1] and Jonathan N. Pruitt[2]

[1]Department of Biology and Ecology of Fishes, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, Berlin 12587, Germany
[2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

While there are now a number of theoretical models predicting how consistent individual differences in behaviour may be generated and maintained, so far, there are few empirical tests. The social niche specialization hypothesis predicts that repeated social interactions among individuals may generate among-individual differences and reinforce within-individual consistency through positive feedback mechanisms. Here we test this hypothesis using groups of the social spider *Stegodyphus mimosarum* that differ in their level of familiarity. In support of the social niche specialization hypothesis, individuals in groups of spiders that were more familiar with each other showed greater repeatable among-individual variation in behaviour. Additionally, individuals that were more familiar with each other exhibited lower within-individual variation in behaviour, providing one of the first examples of how the social environment can influence behavioural consistency. Our study demonstrates the potential for the social environment to generate and reinforce consistent individual differences in behaviour and provides a potentially general mechanism to explain this type of behavioural variation in animals with stable social groups.

1. Introduction

A fundamental goal in the field of the animal personality literature is to understand the mechanisms responsible for generating and maintaining consistent individual differences in behaviour. While there are now a number of well-developed theoretical models predicting potential causal mechanisms (reviewed in [1–3]), thus far empirical tests of these predictions are extraordinarily few (but see [4–6]). In order to fully explain the presence of consistent individual differences in behaviour, or personalities, a potential mechanism would need to address the two key aspects of personality: among-individual variation in behaviour and within-individual consistency over time. A recent hypothesis, termed the social niche specialization hypothesis, predicts that the social environment may play a key role in both promoting individual differentiation and individual consistency [7,8].

The social niche specialization hypothesis predicts that when a group of individuals interacts repeatedly, it can be beneficial for them to develop 'social niches' [7,8]. These social niches provide a way to reduce competition among individuals and increase individual pay-offs. First, by differentiating their behaviour from each other, individuals can reduce direct competition with group mates [6,9]. This type of niche specialization has been well established in ecology, where the presence of competing conspecifics can generate diet specializations among individuals [10,11], and the social niche specialization hypothesis expands this concept beyond the foraging context. The presence of among-individual variation in behaviour can increase colony productivity [12] and individual reproduction

© 2014 The Author(s) Published by the Royal Society. All rights reserved.

Retracted

A	B	C				pre-treatment					Post treatment				
Source.colony	Expt.colony	Indv.ID	Weeks	Treatment	Individual	Bold1	Bold2	Bold3	Bold4	Bold5	Bold6	Bold7	Bold8	Bold9	Bold10
UPP-A8	UPP-A8_3	M_3_35	4	Mixed	M_4_10	297.91	356.56	600	160.66	600	160.66	600	119.83	153.53	241.34
UPP-A8	UPP-A8_3	M_3_35	4	Mixed	M_4_11	600	600	600	77.56	313.32	477.56	313.32	384.26	339.12	357.55
KAL-AA1	KAL-AA1_4_cor	C_139	4	Mixed	M_4_12	600	600	600	335.09	212.22	335.09	412.22	414.03	415.46	399.41
KAL-1	KAL-1_3	M_3_3	4	Mixed	M_4_13	103.5	442.25	241.34	339.63	119.72	139.63	119.72	210.73	113.47	111.31
KAL-AA1	KAL-AA1_3_cor	C_99	4	Mixed	M_4_14	600	600	600	178.41	68.57	463.77	441.23	420.89	369.81	402.63
UPP-A8	UPP-A8_2	M_2_33	4	Mixed	M_4_15	561.71	600	600	257.62	124.72	257.62	124.72	600	600	600
KAL-B3	KAL-B3_4_cont	C_144	4	Mixed	M_4_16	600	281.46	600	562.01	183.25	562.01	183.25	600	84.53	600
KAL-AA1	KAL-AA1_4_cor	C_140	4	Mixed	M_4_17	600	600	259.94	600	452.7	492.23	452.7	499.45	594.72	458.71
UPP-A8	UPP-A8_3	M_3_38	4	Mixed	M_4_18	600	170.13	58.75	435.60	28.22	35.21	28.22	83.75	600	12.21
KAL-B3	KAL-B3_4_cont	C_144	4	Mixed	M_4_19	600	90.56	43.48	600	276.63	58.75	233.51	358.66	502.37	63.77
KAL-B3	KAL-B3_4_cont	C_145	4	Mixed	M_4_20	600	277.53	63.77	600	165.58	600	165.58	46.35	334.00	600
KAL-AA1	KAL-AA1_3_cor	C_97	4	Mixed	M_4_21	600	600	107.37	349.00	327.22	349.00	327.22	600	408.28	288.32
UPP-A8	UPP-A8_3	M_3_36	4	Mixed	M_4_22	264.59	600	583.78	600	209.16	482.35	409.16	484.53	495.62	317.35
N10-C3	N10-C3_3	M_3_18	4	Mixed	M_4_23	600	600	475.82	600	419.38	600	600	600	600	600
UPP-B77	UPP-B77_2	M_2_39	4	Mixed	M_4_24	35.72	137.59	100.26	113.47	312.17	13.54	19.5	53.97	66.21	58.75
N10-C3	N10-C3_1_cont	C_22	4	Mixed	M_4_25	403.03	234.66	554.88	31.72	76.66	31.72	76.66	42.13	39.53	54.23
N10-C3	N10-C3_2_cont	C_64	4	Mixed	M_4_26	83.00	170.56	600	518.31	600	518.31	600	305.85	498.31	554.88
UPP-A7	UPP-A7_2_cont	C_74	4	Mixed	M_4_27	600	600	305.85	153.53	358.66	453.53	358.66	452.65	358.94	359.94
KAL-AA1	KAL-AA1_1_cor	C_7	4	Mixed	M_4_28	600	600	600	91.91	590.41	591.91	590.41	500.21	454.23	443.37
KAL-1	KAL-1_1_contro	C_3	4	Mixed	M_4_29	35.06	114.27	404.35	600	600	600	155.42	119.25	600	598.62
KAL-1	KAL-1_2	M_2_1	4	Mixed	M_4_30	252.47	600	600	262.83	194.62	262.83	194.62	24.56	530.35	171.75
KAL-B3	KAL-B3_1	M_1_11	4	Mixed	M_4_31	600	600	143.37	404.35	233.00	404.35	233.00	61.2	252.72	114.28
KAL-B3	KAL-B3_4_cont	C_144	4	Mixed	M_4_32	110.53	600	600	600	600	495.56	577.45	585.56	497.56	583.78
UPP-A8	UPP-A8_1	M_1_33	4	Mixed	M_4_33	581.03	600	600	600	600	600	600	600	600	600
UPP-A10	UPP-A10_3	M_3_24	4	Mixed	M_4_34	23.06	600	183.28	552.25	181.06	552.25	181.06	114.04	188.09	214.94
KAL-AA1	KAL-AA1_2_cor	C_53	4	Mixed	M_4_35	600	337.74	171.75	184.65	452.09	184.65	452.09	107.28	518.31	600
UPP-A8	UPP-A8_4	M_4_34	4	Mixed	M_4_36	600	600	598.62	237.40	91.79	237.40	91.79	600	383.22	199.38
UPP-A7	UPP-A7_3_cont	C_121	4	Mixed	M_4_37	600	449.44	600	285.47	262.85	285.47	262.85	600	600	143.37
UPP-A10	UPP-A10_1_cor	C_27	4	Mixed	M_4_38	251.06	600	600	600	198.57	512.23	398.57	477.00	482.09	482.34
KAL-AA1	KAL-AA1_1	M_1_10	4	Mixed	M_4_39	42.21	600	96.00	600	191.54	600	232.1	600	600	600
UPP-A10	UPP-A10_1	M_1_22	4	Mixed	M_4_40	471.34	600	146.53	201.83	260.09	201.83	260.09	471.75	72.19	100.26
UPP-A10	UPP-A10_4_cor	C_159	4	Mixed	M_4_41	600	403.25	317.35	298.59	191.54	298.59	191.54	600	344.09	600
N10-C3	N10-C3_2	M_2_20	4	Mixed	M_4_42	600	84.53	600	600	348.88	600	348.88	600	441.94	182.09
KAL-B3	KAL-B3_4_cont	C_146	4	Mixed	M_4_43	600	600	600	94.72	570.93	594.72	570.93	499.87	525.48	541.02
KAL-1	KAL-1_2_contr	C_48													
KAL-B3	KAL-B3_4_cont	C_145													

More on this story

Blog by coauthor Kate Klaskowski
What to do when you don’t trust your data anymore

Blog by American Naturalist editor Dan Bolnick
17 months

Note this story is more about replicability and scientific integrity than reproducibility, but the two are linked. Reproducibility supports replicability.

Publication pressure threatens the integrity of palaeontological research

Nussaïbah B. Raja (nussaibah.raja.schoob@fau.de, @mauritiantales) & Emma M. Dunne (dunne.emma.m@gmail.com, @emmadnn)

This manuscript has been submitted for publication in **The Geological Curator** and is undergoing peer-review. Please note that the manuscript has not been formally accepted for publication and is as such a **non-peer reviewed preprint**. Subsequent versions of this manuscript may have slightly different content. If accepted, the final version of this manuscript will be available via the Peer-review Publication link on this web-page. Feel free to contact us if you have any queries or comments; we welcome any feedback that you may have.

Benefits of reproducibility to the community

Saves time, effort, and money.

Increases transparency, helps others find mistakes.

Increases scientific integrity.

It makes science more accessible and equitable.

Many funding agencies / journals have requirements.

Benefits of reproducibility to you

At an absolute minimum you should be able to reproduce the results yourself!

Part of ensuring reproducibility is doing future you a favour.

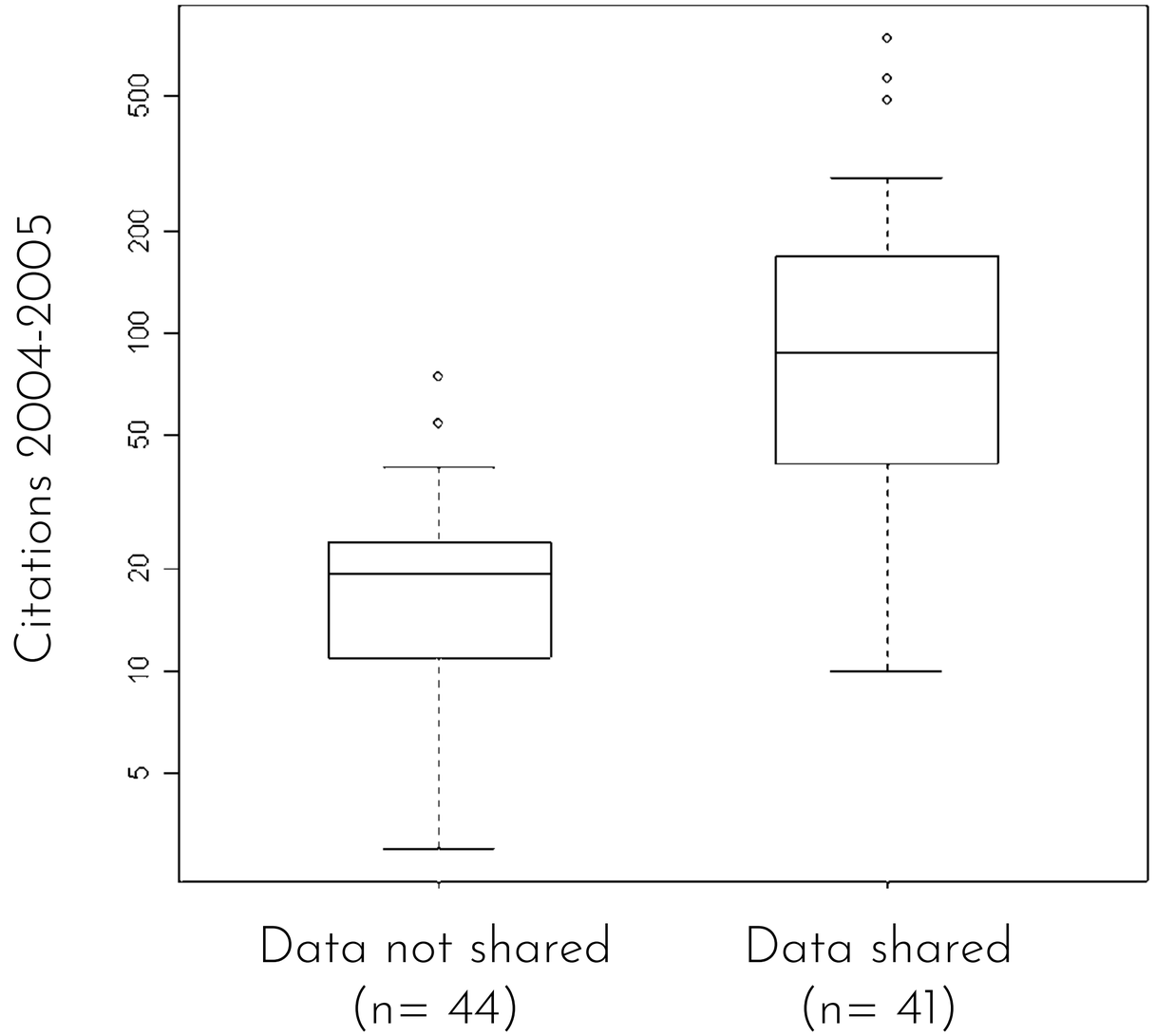
If you can routinely reuse data and code with ease, you can save a lot of time.

OPEN ACCESS Freely available online

PLoS one

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma



How can we establish practices that increase transparency and reproducibility?

Data (and code) accessibility and documentation

Data (and code) accessibility

Fourth group exercise

In you working groups, establish the following and report back:

- Can you access the raw data associated with your study?
- Can you access the code used to perform the statistical analysis?

What to do when you can't access the data?

Email the authors? Email the editor?

Can you extract the data from a PDF?

- tabulizer package in R
- tesseract software

Reproducibility

Something to think about:

Is enough information provided for you to reproduce the analysis?

Guidelines for reproducible computational research

Sandve et al. 2013. Plos Comp Bio.

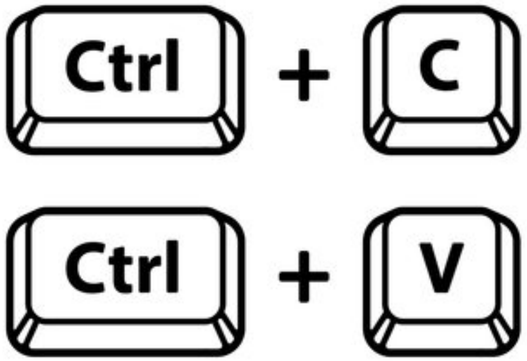
At an absolute minimum you should be able to reproduce the results yourself.

Rule 1: for every result, keep track of how it was produced

- all parameter choices
- pre- and post-processing
- any manual steps

How exactly were specimens measured?

Rule 2: avoid manual data manipulation steps



Rule 3: record (or even better archive) the exact versions of all software used

Rule 4: use version control

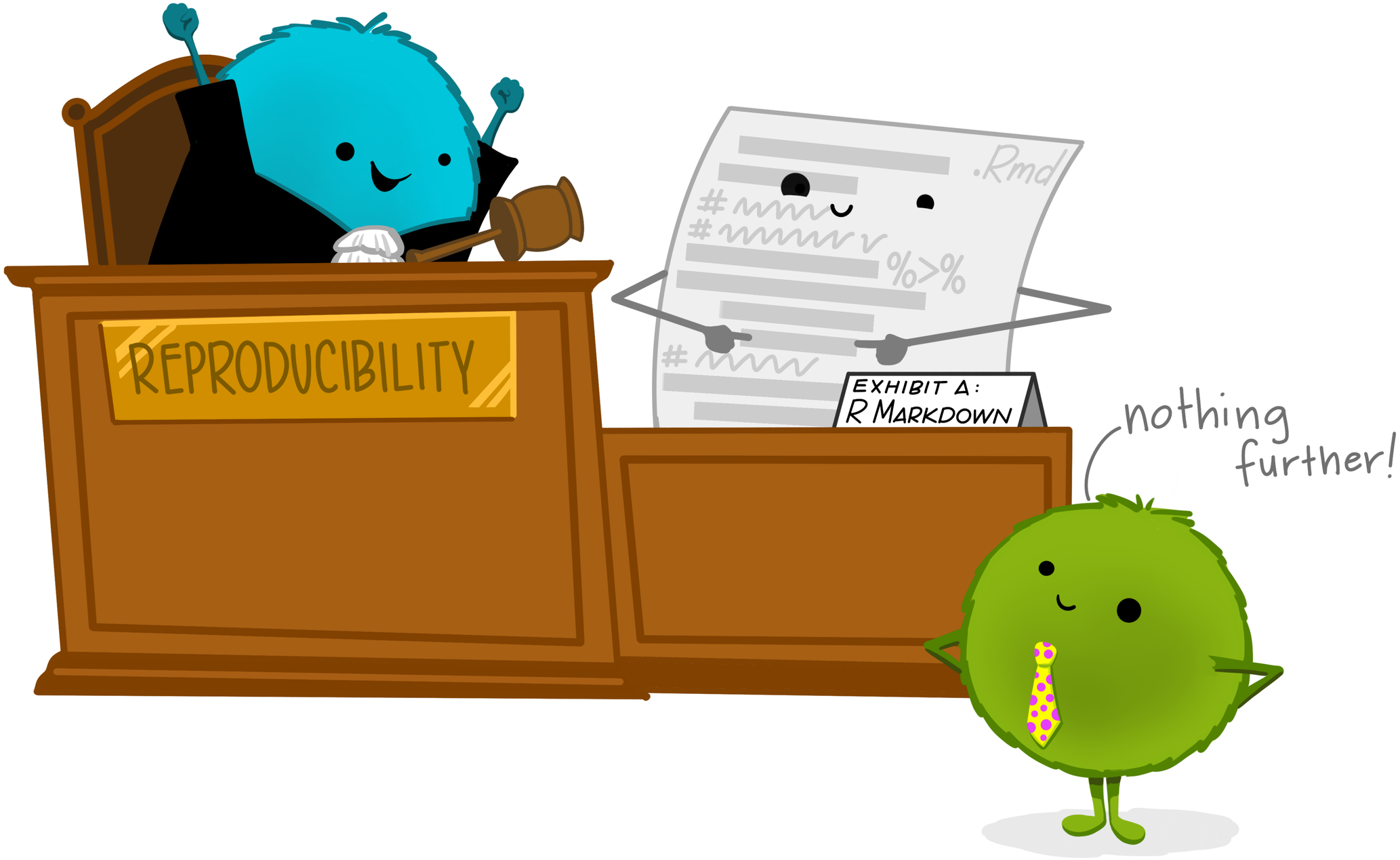
Rule 5: record all intermediate results and standardise formats

Rule 6: for analyses that include randomness, take note of the underlying seeds

quick demo!

Rule 7: always store raw data (and code) used to produce plots

Rule 8: connect textual results to underlying results



@allison_horst

Rule 9: provide public access to all scripts and results

- supplementary material
- code repositories, e.g., github
- dryad, zenodo

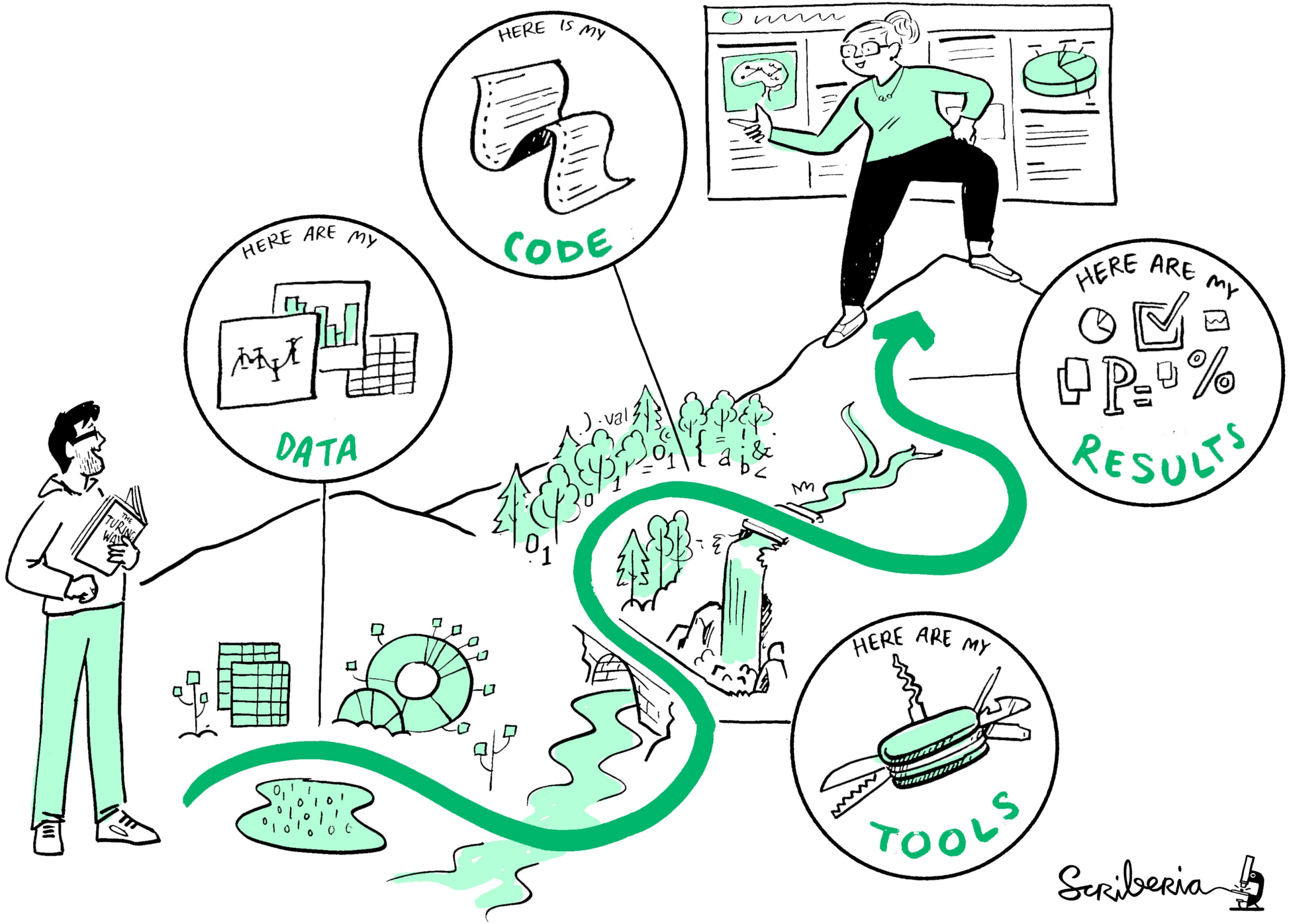
Accept that there is a **short-term** trade off between maximising reproducibility and getting results out there, but the long-term benefits are worth the time investment.

Further reading

A protocol for data exploration to avoid common statistical problems

Reproducible Research in R: A Tutorial on how to Do the Same Thing More Than Once

Guide for Reproducible Research


HERE IS MY
CODE
HERE ARE MY
DATA
HERE ARE MY
RESULTS
HERE ARE MY
TOOLS
Scriberia

Reproducibility workflow

Fifth group exercise

In you working groups:

- What would a reproducible work flow look like for your study?
- What are all the necessary steps you would need to record?


TEA
REX
LoofandTimmy.com