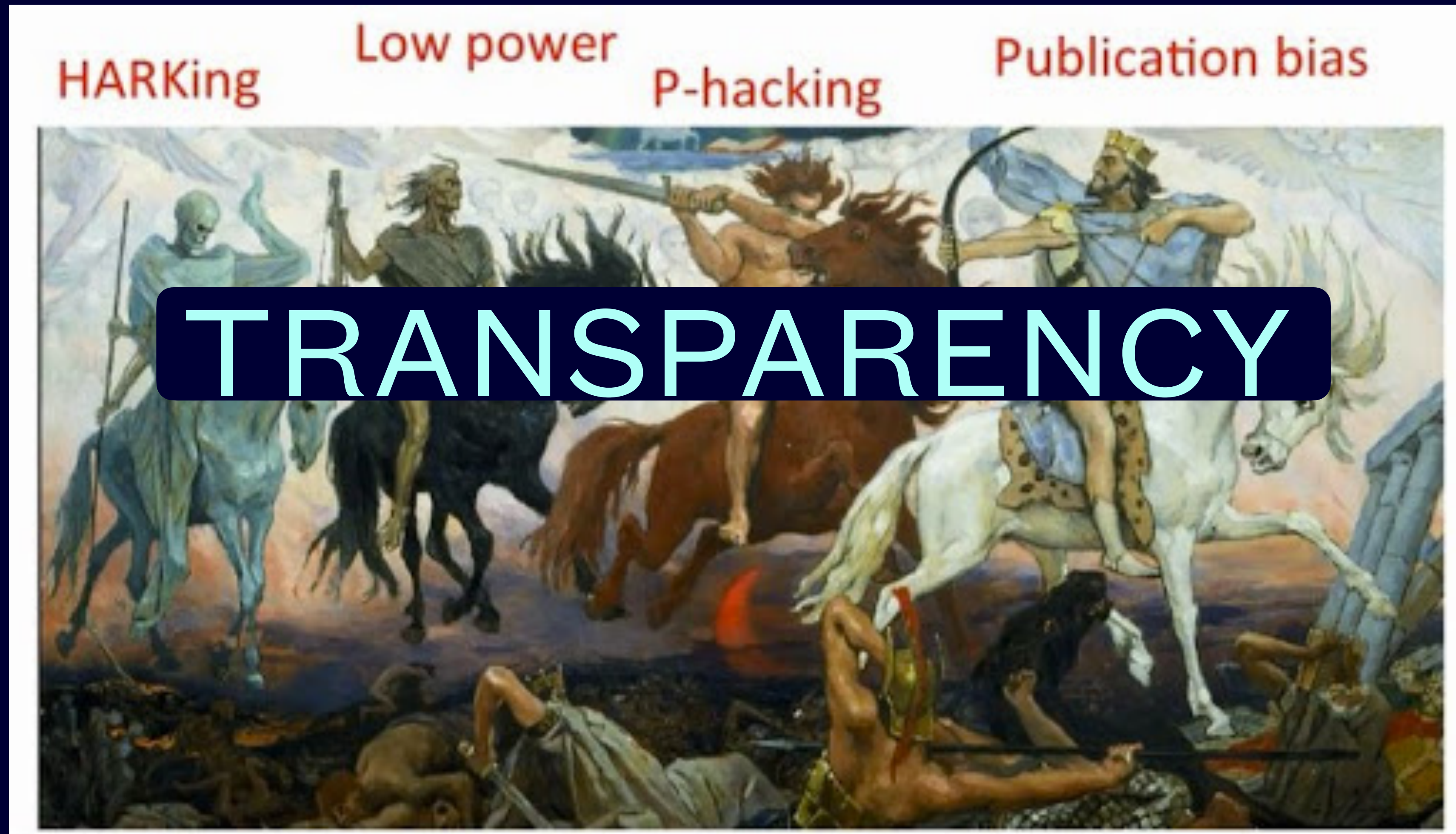


SInCLab meeting 24 Feb 21

Open science : practical solutions

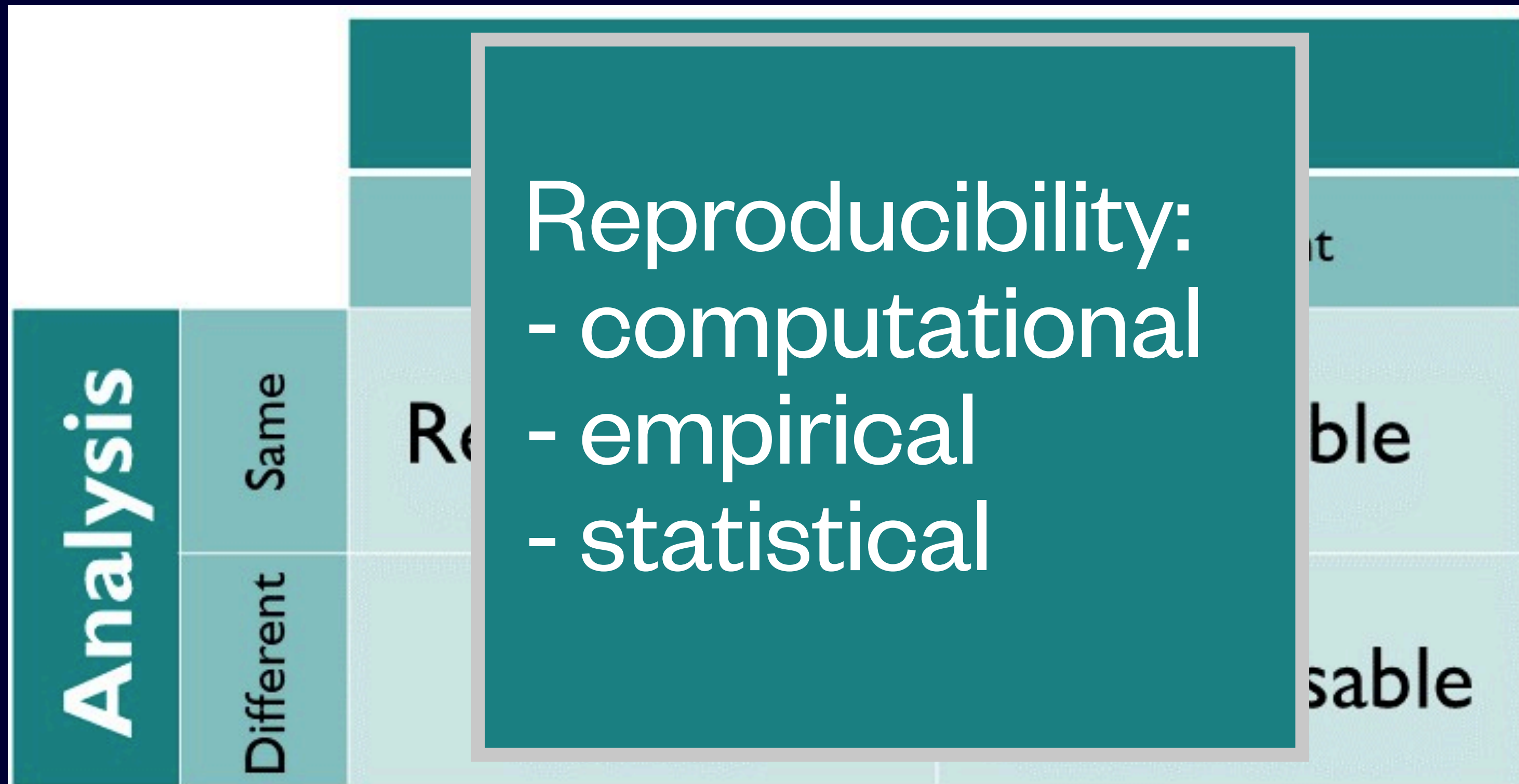
NIKA ADAMIAN

The problem



Definitions

Open science is “the process of making the content and process of producing evidence and claims transparent and accessible to others” (Munafò et al., 2017)

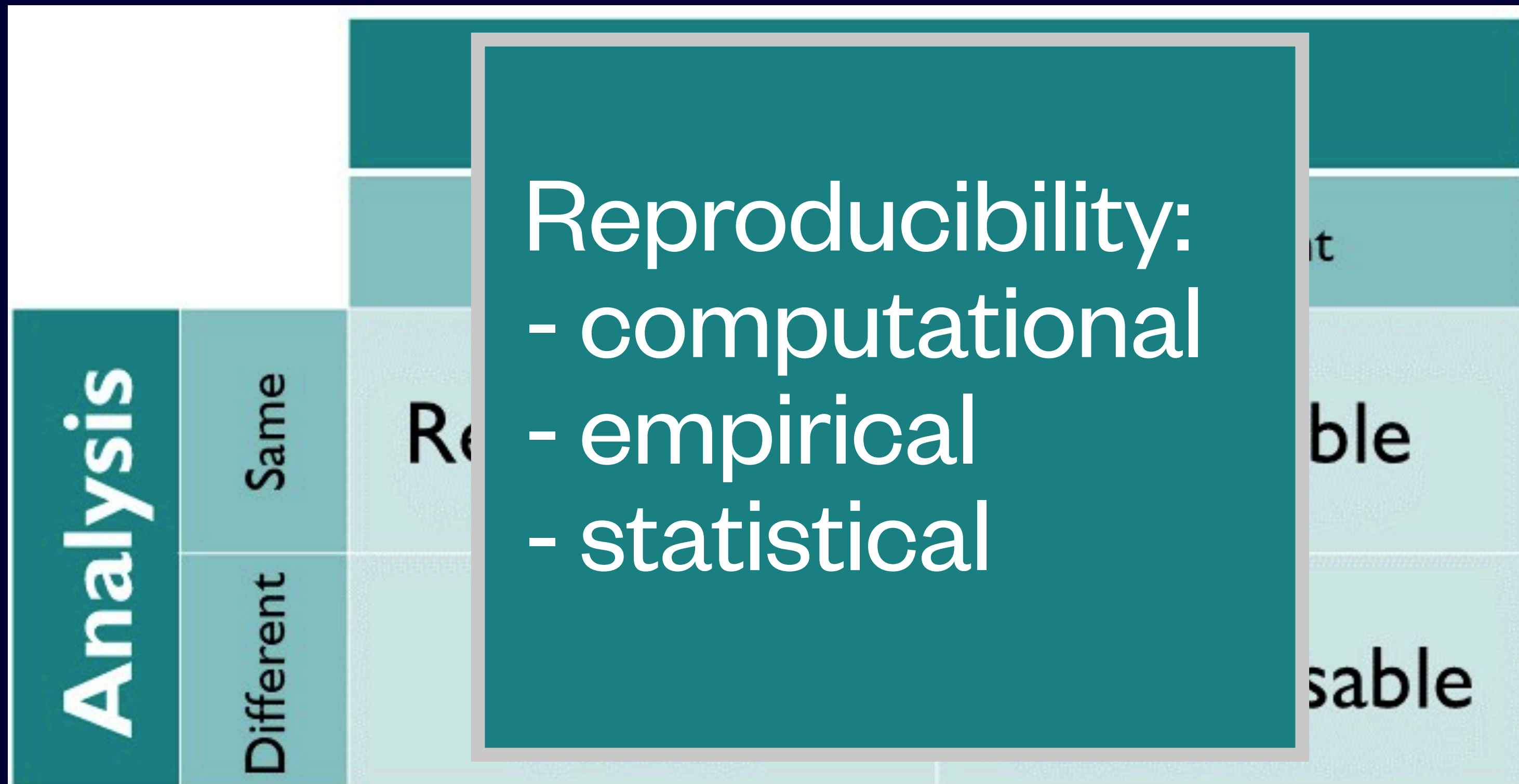


Open Science Practices:

1. Open Data
2. Open Materials / Code
3. Preregistration
4. Reproducible Analyses
5. Replications
6. Improvement of statistics and methodology
7. Open Access

Definitions

Open science is “the process of making the content and process of producing evidence and claims transparent and accessible to others” (Munafò et al., 2017)



Open Science Practices:

- 1. Open Data**
- 2. Open Materials / Code**
- 3. Preregistration**
4. Reproducible Analyses
5. Replications
6. Improvement of statistics and methodology
7. Open Access

Sharing



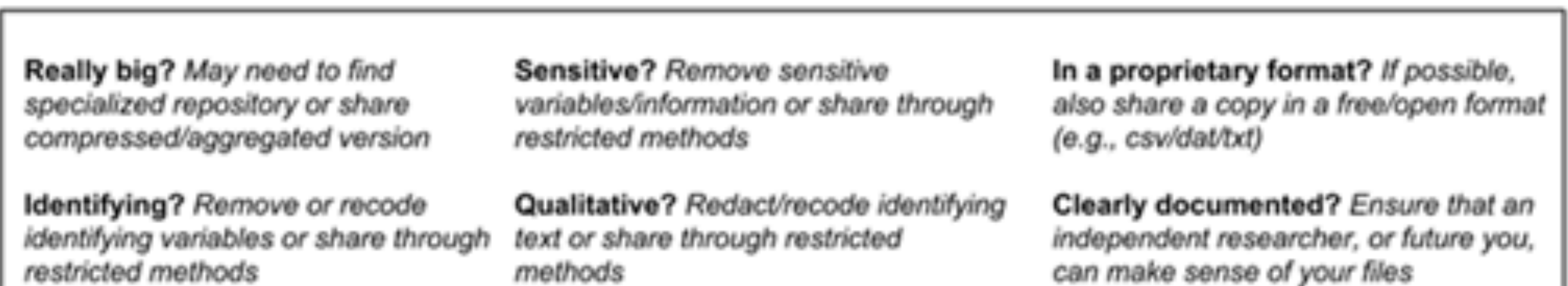
Sharing

Klein et al. (2018). A practical guide for transparency in psychological science.

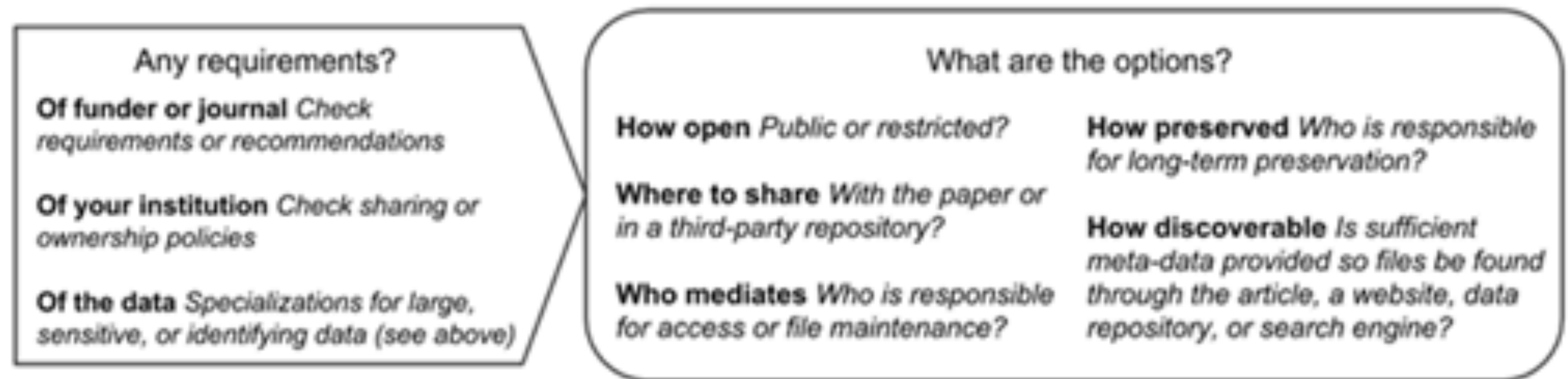
1. Scope your sharing



2. Assess your research products: Are your files...



3. Decide how to share



SOP including sharing

View

Wiki Version: (Current) Brian A. Nosek: 2014-01-07 20:29:42+00:00 UTC ▼

Checklist for Research Workflow

1. ___ Once idea and design is elaborated, evaluate the project's feasibility, potential impact, and perceived likelihood of success. Confirm allocation of resources after comparing this project with alternative resource investments
2. ___ Complete [research implementation checklist](#) to initiate research, document the research plan, and share with collaborators
3. ___ Once data collection is complete, post raw data files to OSF project
4. ___ Once initial data analysis is complete, post analysis scripts, codebook, and cleaned data files to OSF project
5. ___ Once final data analysis is complete, post analysis scripts, and updated codebook and cleaned files to OSF project
6. ___ Add a short, narrative summary of the project findings to OSF project, register project
7. ___ If project is private, decide whether and when to make project or project components public
8. ___ Decide next step: project dead, project paused, investigation continuing, or writing report
 - a. If investigation continuing, restart this checklist
 - b. If writing report, move on to [writing report checklist](#)

Open Data

FAIR principles

1. Findable

Data have sufficient metadata, and a unique and persistent identifier

2. Accessible

Data is deposited in a trusted repository

3. Interoperable

Data uses a broadly applicable language or format

4. Reusable

Data is well-described, anonymised, and under a clear reuse license

Open Data

Considerations for sharing small(ish)-scale data

1. Good metadata — description of data and variables

Example REAMDE: Confidence Project (<https://osf.io/72uxf/>)

2. Non-proprietary format (csv/dat/txt)

3. Choice of repository:

- OSF
- Zenodo
- Figshare

4. Digital Object Identifier (DOI)

5. License (CC0 or CC-BY; <https://chooser-beta.creativecommons.org/>)

6. If sensitive — consider synthetic data (*synthpop*; Quintana, 2020)

Specialised neuroscience solutions

A free and
MEG

BIDS-EEG

dataset_description.json	▶ description of dataset in JSON format
README	▶ readme file for users
participants.tsv	▶ participants table files in tab delimited format
participants.json	▶ description of participants table file columns (above)
sourcedata	▶ original raw data if converted to a supported format
stimuli	▶ original stimuli (sound files and images)
sub-001	▶ anonymized subject 1 folder
eeg	
sub-001_ses-01_task-meditation_channels.tsv	▶ file describing channels
sub-001_ses-01_task-meditation_eeg.bdf	▶ raw EEG data file (not all raw formats are possible)
sub-001_ses-01_task-meditation_eeg.json	▶ amplifier and recording information
sub-001_ses-01_task-meditation_events.tsv	▶ events in tabular format

sub-001_ses-01_task-meditation_events.tsv — Edited					
onset	duration	trial_type	response_time	sample	value
71.3867187500	n/a	stimulus	n/a	18275	128
75.7304687500	n/a	response	n/a	19387	2
79.7734375000	n/a	response	n/a	20422	2
125.6093750000	n/a	stimulus	n/a	32156	128
179.8007812500	n/a	stimulus	n/a	46029	128
183.0976562500	n/a	response	n/a	46873	2
185.6328125000	n/a	response	n/a	47522	4
284.8593750000	n/a	stimulus	n/a	72924	128
287.7578125000	n/a	response	n/a	73666	2
289.9609375000	n/a	response	n/a	74230	2

BIDS specification:

<https://bids-specification.readthedocs.io/en/latest/>

This example:


<https://openneuro.org/datasets/ds001787>

<https://github.com/scn/bids-matlab-tools/wiki>

Open Materials / Code

Levels of transparency

Analysis code

- 
- Script that generates figures
 - Script with figures and statistical tests / models
 - Full historical transparency through Git (very easy to integrate with RStudio)
 - Manuscript prepared in RMarkdown

Share as much as you are comfortable with, and a little bit more

Special case: stimulation scripts

Why special?

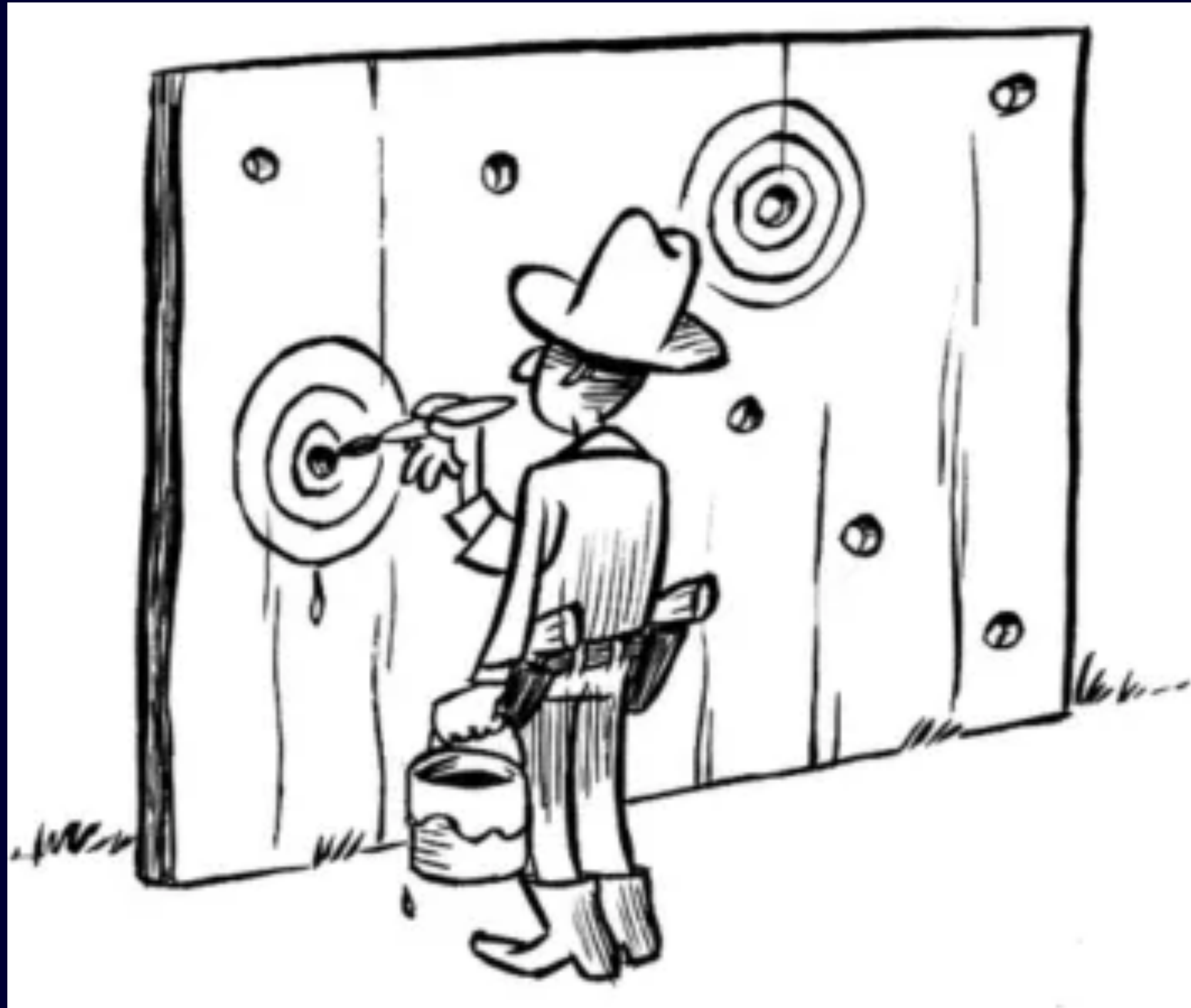
- Rely either on proprietary software, or on a special arrangement of open-source dependencies (or both)
- Incorporate functions written by different people in different eras
- Depend on hardware
- VERY messy

Solutions

- Become a Research Software Engineer
- Share anyway, disclose that this is demo only (e.g. MATLAB/Octave)
- Share anyway, describe your equipment and other info in the readme
- Share procedure via alternative means e.g. gif or video recording

So why not share?

Preregistration



Forms of preregistration

1. Unreviewed preregistration : description of your plans in a time-stamped uneditable archive
2. Reviewed preregistration (Registered Report): submit your plan to a journal before proceeding;
 - Supported by many journals
 - Most useful where reviewers can assess how informative the results will be
3. Registered Replication Reports
 - Including multi-site efforts such as ManyLabs

Tools and templates

- aspredicted.org
- OSF (templates: <https://osf.io/t6m9v/>, checklist <https://osf.io/93znh/>)
- journal policies: https://katie-drax.shinyapps.io/cos_registered_reports/

Journal	Attention, Perception & Psychophysics
Includes pre-study peer review	Yes
Offers provisional pre-study acceptance	Yes
Permanence of adoption	Indefinite
Offers RRs for novel studies	Yes
Offers RRs for replication studies	Yes
Offers RRs for meta-analysis	No
Offers RRs for analyses of existing data sets	No
Publishes Registered Reports only	No
Allows reporting of unregistered analyses	Yes
Includes post-study peer review	Yes
Allows inclusion of unregistered pilot studies	Yes
Requires public data deposition	Yes
Specifies structured criteria for editorial decisions	Yes
Requires submitted protocols to have prior ethical approval	Yes
Specifies minimum statistical power requirements	Yes
Will publish 'Withdrawn Registrations'	No
Publishes accepted protocols, in full or part, prior to study completion	No
Offers incremental (sequential) registration	No
Offers incremental addition of unregistered studies	No
Offers RRs for qualitative research	No
Requires deposition of protocol in public registry following Stage 1 acceptance	No

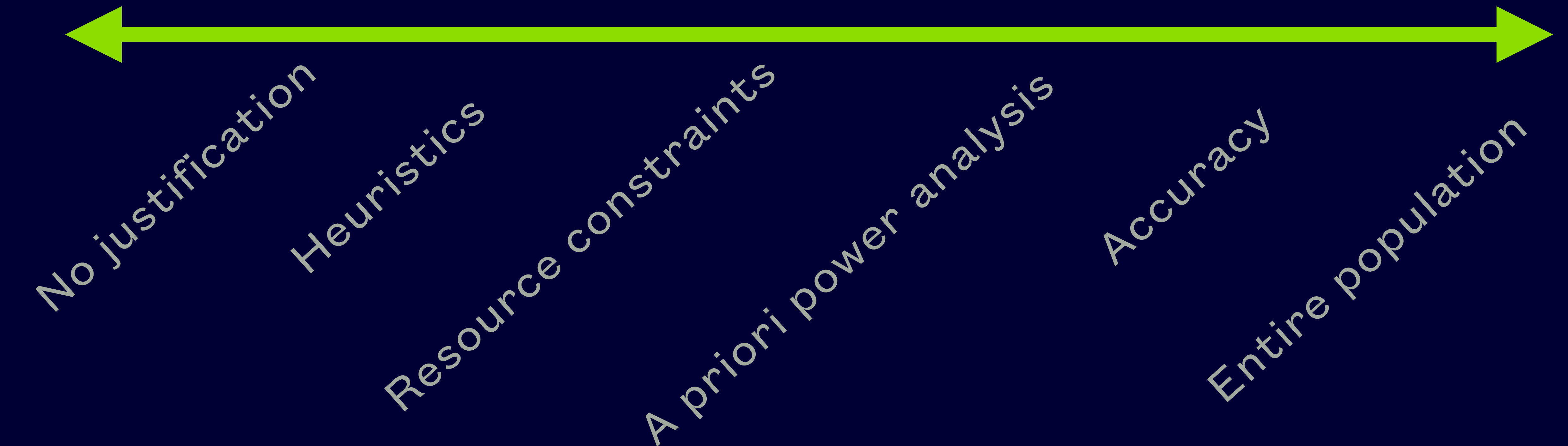
Registered Reports

1. Delineate confirmatory hypotheses: motivation, test, interpretation. Define discriminatory data patterns.
2. Demonstrate sufficient statistical power
3. Ensure reproducibility and replicability: higher reporting standards

Sample size justification

Very imprecise

Very precise



Sample size justification

What effect sizes are interesting?

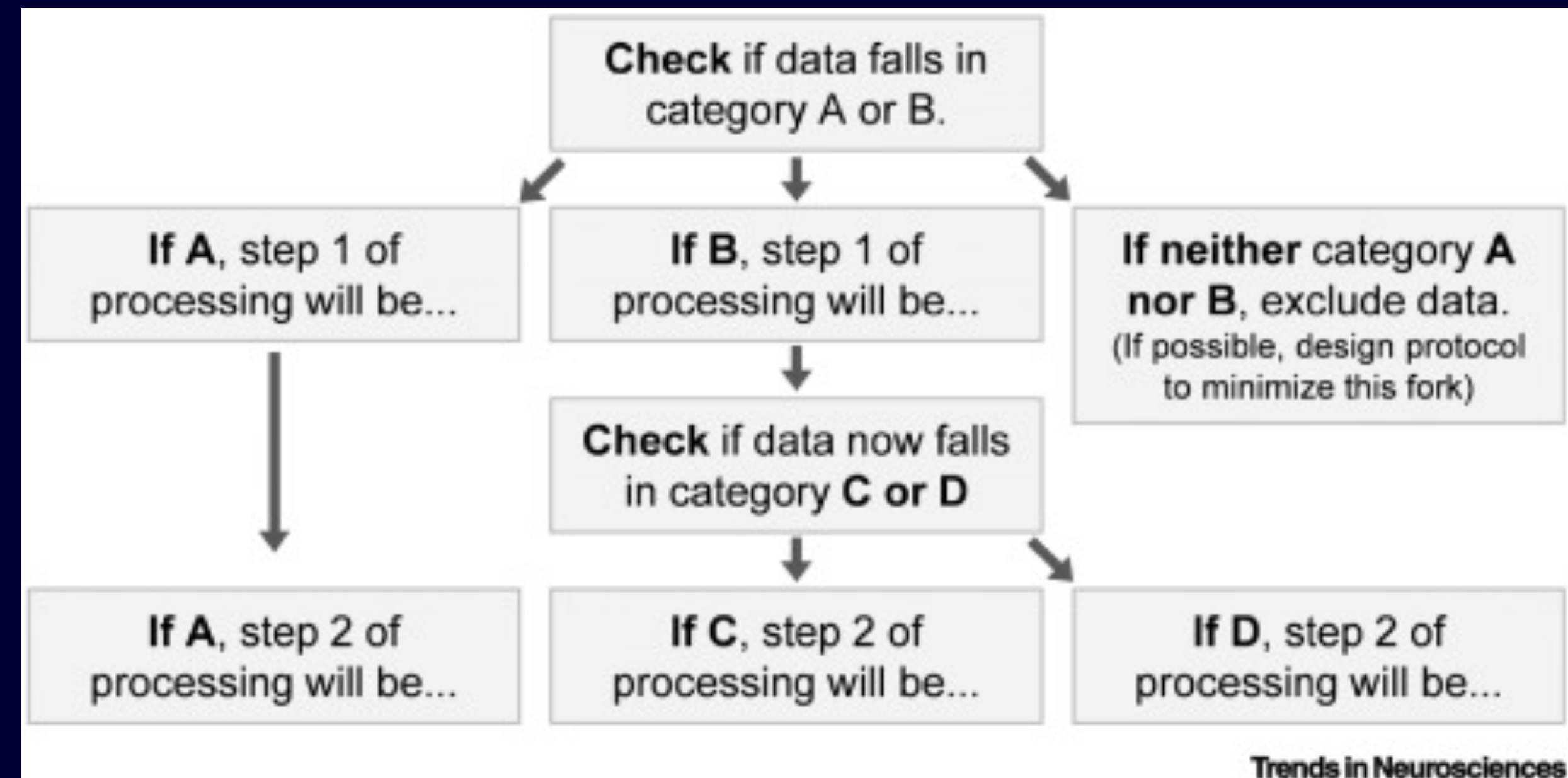
- smallest effect size of interest
- minimal statistically detectable effect
- estimate from a meta-analysis (watch out for bias and heterogeneity)
- width of confidence interval
- effect size from a previous or pilot study (if large enough)

Power analysis

- G*power does NOT support factorial designs. Power beyond one-way ANOVA should be assessed via simulation (http://shiny.ieis.tue.nl/anova_power/) but correlation matrices are an issue for within-subject designs
- Power analysis should be reported
- Power analysis should be done for (as close as possible to) the exact test that you will be using (number of parameters matters)
- Multiple comparisons issue should be considered
- Power can be increased without increasing sample (more precise estimation through optimal design, one-sided tests, sequential analyses...)

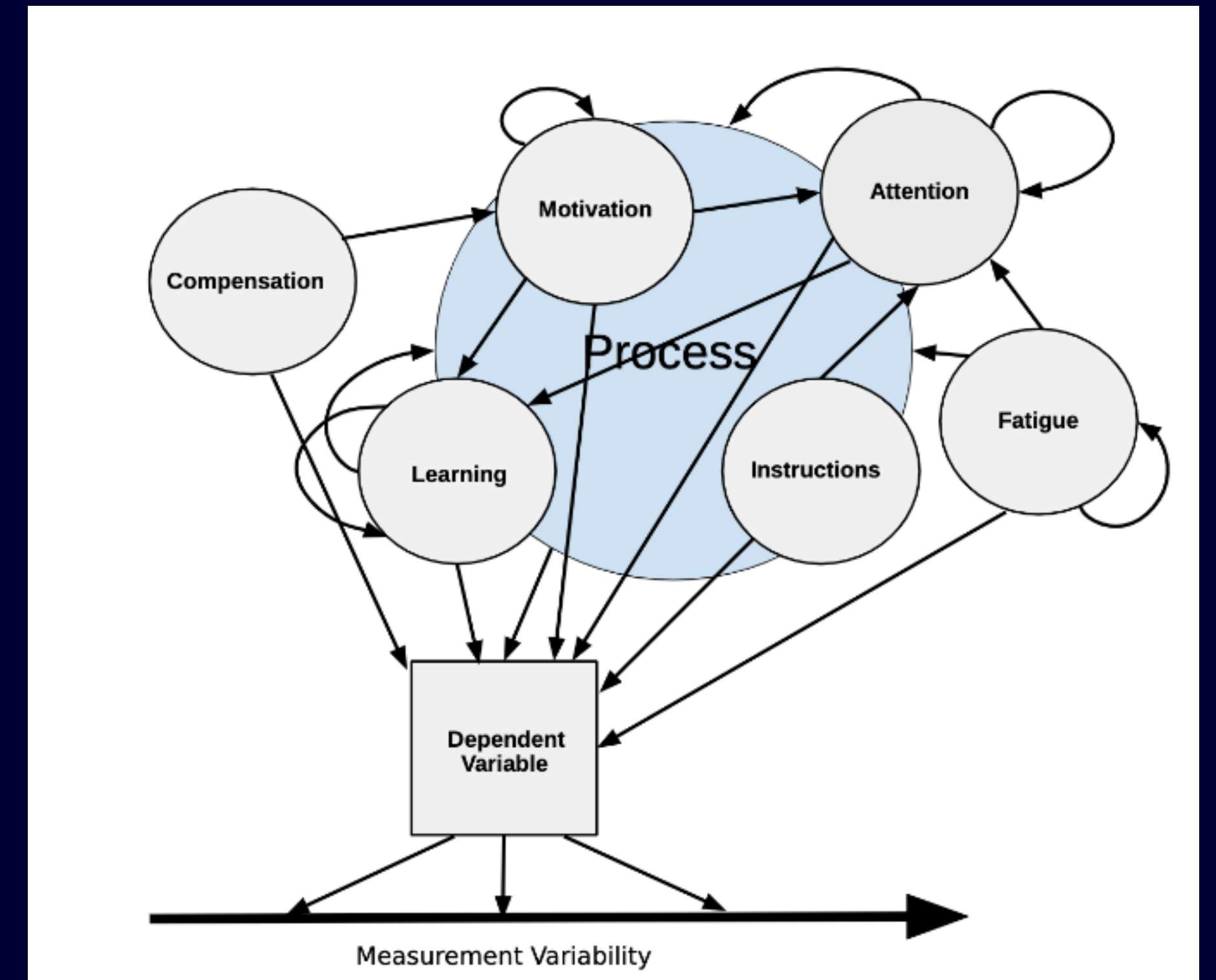
Challenges for preregistration

- Data collection does not go as planned
- Data does not look as expected
- Preexisting data
- Large and/or longitudinal datasets
- Sequential experiments
- Modeling



What preregistration does not help with

- Making sure your hypotheses are grounded in theory
- Making sure your analytical choices are optimal given the situation
- Protecting from biases outside of data treatment



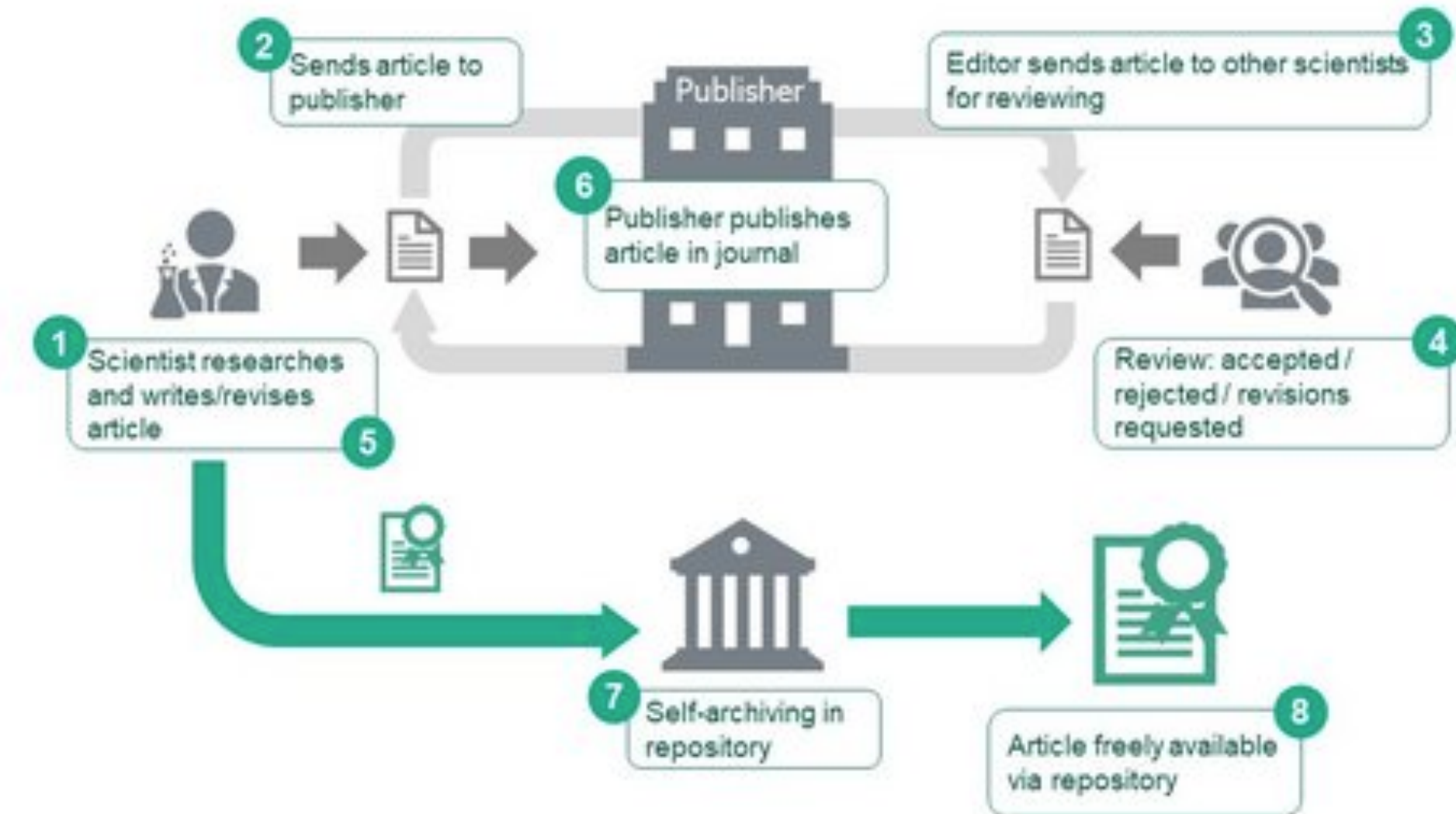
T. Van Zandt (Psychonomics 2018)

Will you preregister?

P.S. Open Access

- SHERPA/RoMEO (<http://www.sherpa.ac.uk/romeo/index.php>)
- DOI your important figures before publishing them

Scholarly Publishing: Green Open Access (Post-Prints)



Source: Based on Oberländer, Anja (2020): Open Access – Es ist nicht alles Gold, was glänzt. In: Open Science. Von Daten zu Publikationen. Zenodo. <http://doi.org/10.5281/zenodo.4018594>



Thank you!

