

USING TOOLS TO MAKE SCIENCE MORE OPEN

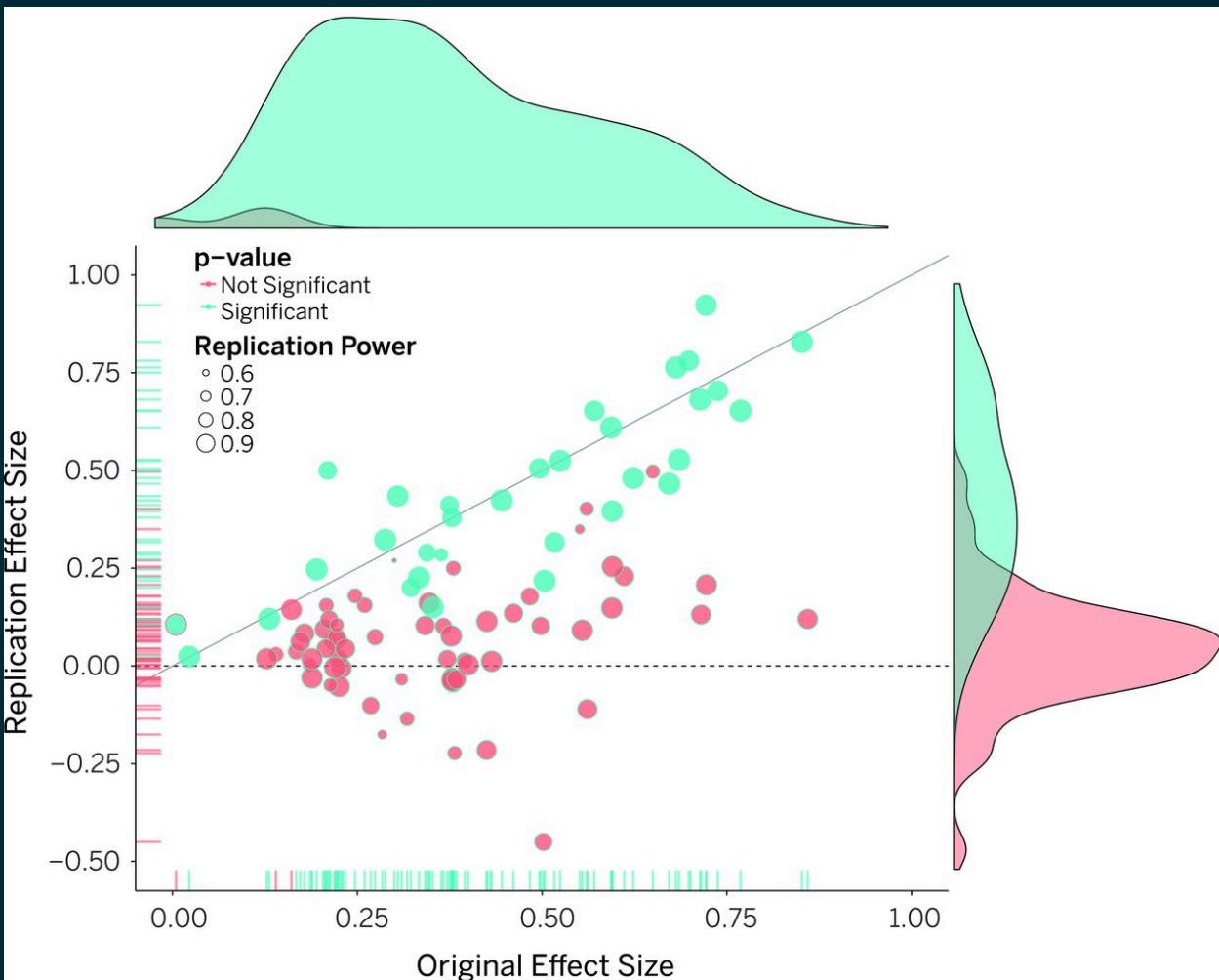
Tim Vantilborgh

INTRODUCTION

A COUPLE OF QUESTIONS

A COUPLE OF QUESTIONS

REPLICATION CRISIS IN PSYCHOLOGY



OPEN SCIENCE COLLABORATION (2015)

THE FOUR HORSEMEN AND THE REPLICATION CRISIS

HARKing

Low power

P-hacking

Publication bias



https://en.wikipedia.org/wiki/Four_Horsemen_of_the_Apocalypse#/media/File:Apocalypse_vasnetsov.jpg

A POSSIBLE SOLUTION: OPEN SCIENCE PRACTICES

PRACTICAL INFORMATION

GOALS WORKSHOP

- Stimulate awareness of importance open science practices
- Increase confidence that you are able to preregister your own research, create reproducible research, and share research materials

OVERVIEW WORKSHOP

1. Brief introduction to open science practices
2. Preregistration: An illustration
3. Reproducible research: A brief workshop
4. Sharing research materials: An illustration

ILLUSTRATIVE EXAMPLE

- Imagine that I would like to study the effectiveness of this workshop
- I could set up an experiment with 2 conditions (workshop group versus control group)
- I could measure attitude towards open science practices at the start and the end of the workshop in both conditions
- Hypothesis: there is a stronger increase in positive attitude towards open science practices in the workshop condition, compared to the control condition

PRACTICAL ASPECTS

- Ideally, you have a laptop with an internet connection and R + Rstudio installed
- All materials can be downloaded from:

BRIEF INTRODUCTION TO OPEN SCIENCE PRACTICES

WHY DO WE NEED OPEN SCIENCE?

Example Eric-Jan Wagenmakers:

- Dr. X has a favorite theory that she has worked on and published about previously.
- Dr. X designs an experiment to test a prediction from her theory.
- Dr. X collects the data, a painstaking and costly process. Part of her career and those of her students ride on the outcome.

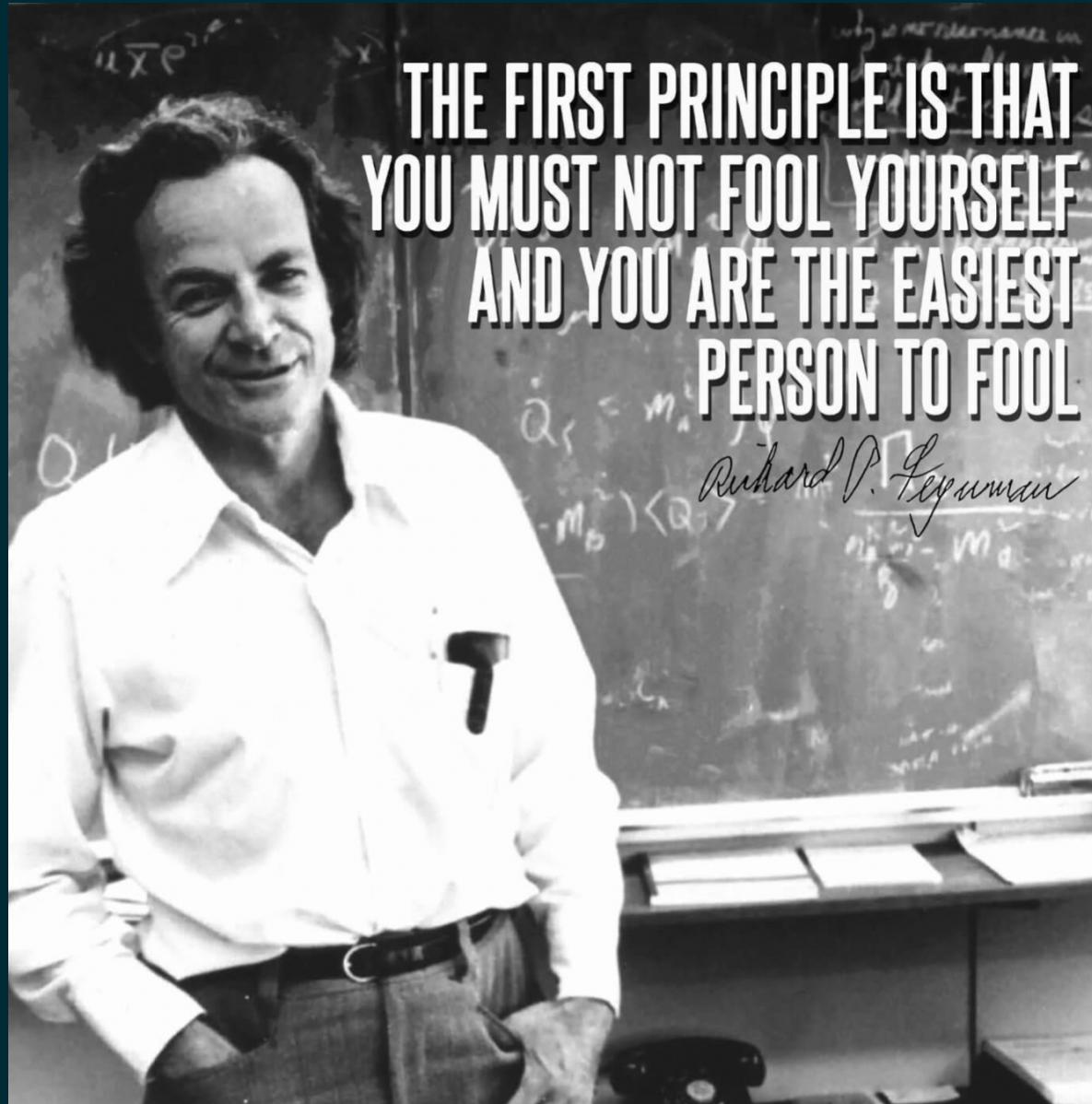
WHY DO WE NEED OPEN SCIENCE?

- Now the data need to be analyzed.
- If $p < .05$, the experiment is deemed a success; if $p > .05$, it is deemed a failure.

WHY DO WE NEED OPEN SCIENCE?

Who is, without a shadow of a doubt, the most biased analyst in the entire galaxy, past, present, and future?

WHY DO WE NEED OPEN SCIENCE?



WHY DO WE NEED OPEN SCIENCE?

- So the world's most biased analyst, Dr. X, the easiest person to fool, proceeds to analyze the data.
- Dr. X can do this alone, without any oversight whatsoever. In most cases, the data and analysis code never leave the lab.

WHY DO WE NEED OPEN SCIENCE?

- Data are analyzed with no accountability, by the person who is easiest to fool, often with limited statistical training, who has every incentive imaginable to produce $p < .05$.
- When $p < .05$, the result is declared “significant” and any further doubt is frowned upon.

WHAT IS OPEN SCIENCE?

“Open Science” is an umbrella term used to refer to the concepts of openness, transparency, rigour, reproducibility, replicability, and accumulation of knowledge, which are considered fundamental features of science ¹

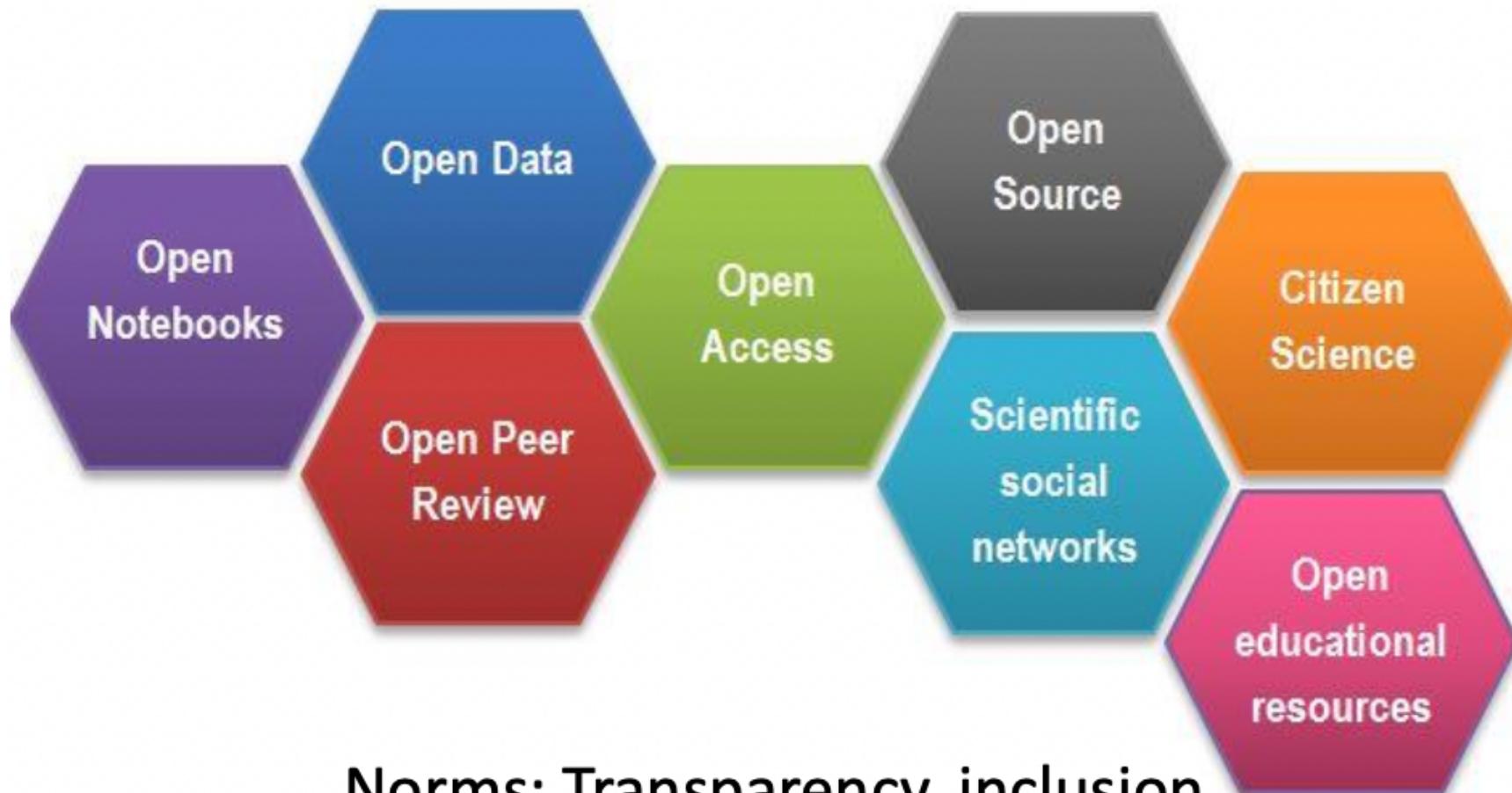
WHAT IS OPEN SCIENCE?

Crüwell et al. (2019):

- Open access
- Open data
- Preregistration
- Reproducible analyses
- Best practices in statistics
- Replications
- Teaching open science

WHAT IS OPEN SCIENCE?

Transparency



Norms: Transparency, inclusion

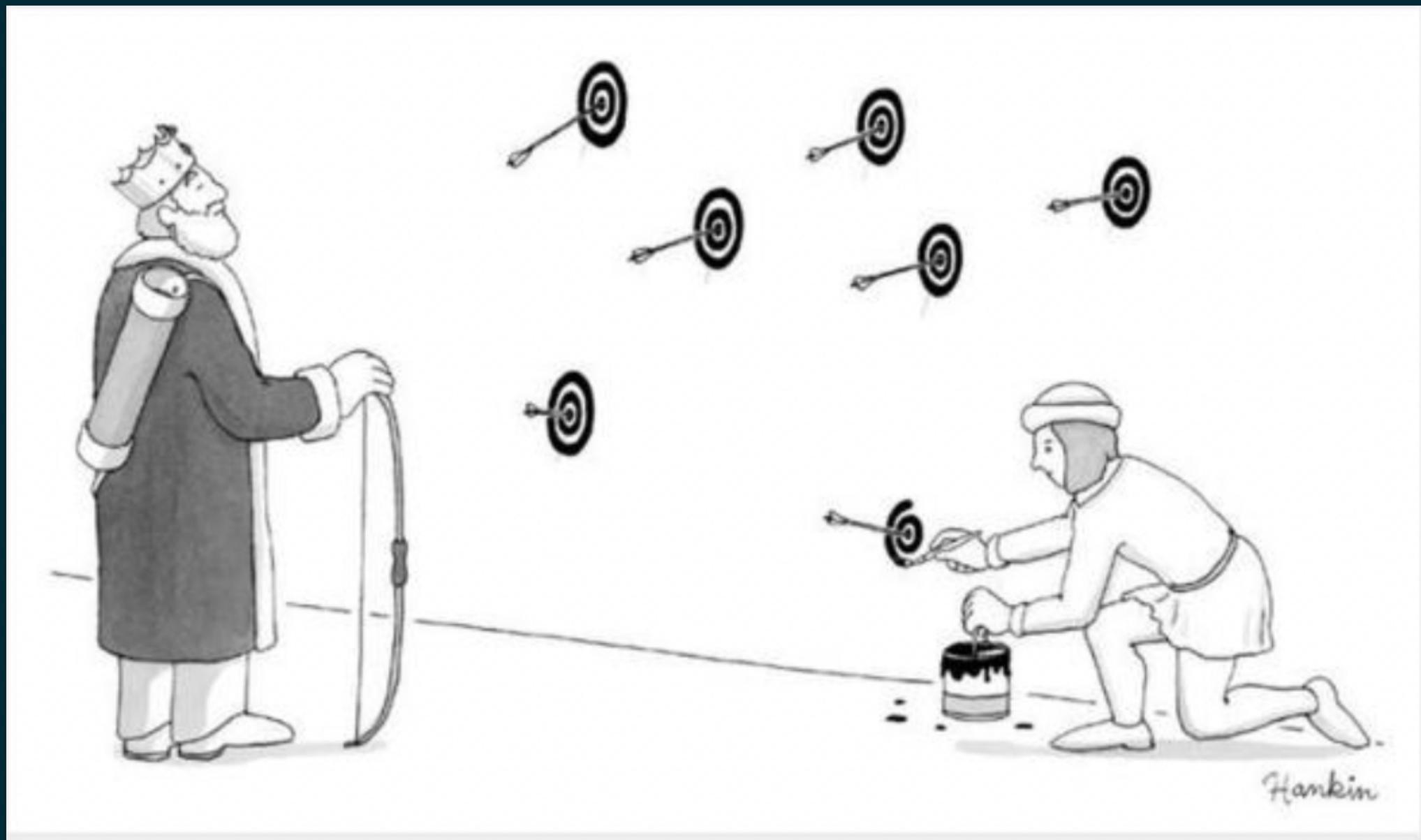
Anti-norms: Secrecy, elitism

WHAT IS PREREGISTRATION?

“...you’re simply specifying your research plan in advance of your study and submitting it to a registry.” ([Science](#) n.d.)

“The specification of a research design, hypotheses, and analysis plan prior to observing the outcomes of a study” ([Lindsay 2018](#))

WHAT IS PREREGISTRATION?



WHAT IS PREREGISTRATION?

- Distinguishes exploratory from confirmatory findings
- Prevents researchers from fooling themselves
- Does not rule out exploratory findings; just labels them accordingly

EXPLORATORY VS CONFIRMATORY RESEARCH

WHY PREREGISTER?

- Makes the research process more transparent
- Makes you thoroughly think through your data collection and analysis before starting
- Allows to clearly distinguish between a priori and post-hoc decisions (confirmatory and exploratory analyses)
- Strict documentation of studies
- Allows others to reproduce your results (with code and data sharing)
- Makes it harder to fool yourself and others

WHAT IS REPRODUCIBLE RESEARCH?

Research is reproducible when others can reproduce the results of a scientific study given only the original data, code, and documentation ([Alston and Rick 2021](#))

WHAT ARE BENEFITS OF REPRODUCIBLE RESEARCH?

Alston and Rick (2021):

- A log of all steps taken in the research process
- Easy to update analyses
- Easy to reuse materials and save time
- Signals rigor, transparency, and trustworthiness
- Increases citation rate
- Allows others to learn from your work
- Simplifies follow-up research
- Prevents mistakes from compounding over time

WHAT ARE CHALLENGES OF REPRODUCIBLE RESEARCH?

Alston and Rick (2021):

- Learning curve
- Complexity
- Technological change
- Fear of human error
- Intellectual property rights

WHAT ARE OPEN MATERIALS?

Sharing relevant research material (data, code, questionnaire, stimuli, lab notebook, ...) and/or research output (presentations, manuscripts) with the broader public.

WHAT ARE BENEFITS OF OPEN MATERIALS?

- Verification
- Reproducibility
- Improves meta-analyses
- Increases citations
- Improves fairness
- Speeds up scientific advancement

HURDLES FOR ADOPTING OPEN SCIENCE PRACTICES

- Lack of awareness
- Lack of training
- Fear
- Slow process is not ideal for ECRs
- Lack of incentives
- Not supported by the system

OPEN SCIENCE TOOLS

- Software (Rmarkdown, Quarto)
- Packages (worcs, papaja)
- Websites (OSF, aspredicted, github, zenodo)

PREREGISTRATION: AN ILLUSTRATION

WHERE TO PREREGISTER?

As predicted:

- Short: 9 questions
- All authors approve
- Does not become public until authors act to publish it

OSF:

- Large number of templates
- All participating authors can cancel within 48 hours
- Prereg can be kept private for up to four years but then gets published
- Withdrawing possible but leaves behind basic metadata
- Also allows to make scripts, materials, preprints public

OSF PREREGISTRATION TEMPLATES

Registration Forms and Templates

Registration Forms	Description	Templates
OSF Prereg* (citable from this preprint)	This is our standard, comprehensive, and general purpose preregistration form.	Google Doc , OSF Workflow , R Markdown by Frederik Aust , R Markdown by James Bartlett
Open-Ended Registration	Summary of registered work with a time-stamped snapshot of a research project. Use this one if you are registering a completed project with data or materials.	Word , GoogleDoc
Qualitative Preregistration* (Haven et al 2020)	Template for registering primarily qualitative work.	Word , GoogleDoc , FAQ
AsPredicted Preregistration* form here	Eight questions derived from content recommended by AsPredicted.org .	Word , GoogleDoc
OSF-Standard Pre-Data Collection Registration	State whether data have been collected or viewed and other pertinent comments. Use this one if your pre-analysis plan is uploaded on OSF as a doc	Word , GoogleDoc
Replication Recipe (Brandt et al., 2013): Pre-Registration*	Register a replication study with a series of questions regarding the original work.	Word , GoogleDoc
Replication Recipe (Brandt et al., 2013): Post-Completion	Register a replication study after it has been conducted with questions regarding the outcomes of the replication.	Word , GoogleDoc
Pre-Registration in Social Psychology (van 't Veer & Giner-Sorolla, 2016)*	Preregister a research study outlining the hypotheses, methods, and analysis plan	Word , GoogleDoc , OSF
Registered Report Protocol Preregistration	Register your protocol AFTER having been given "in-principle acceptance" from a Registered Report journal	Word , GoogleDoc , OSF Workflow
Secondary Data Preregistration*	For preregistering a research project that uses an existing dataset.	OSF Page , Example , FAQ

[^<https://osf.io/zab38/wiki/home/>]

LET'S PREREGISTER OUR STUDY

- We will use the standard OSF preregistration template
- Tip: download the template to your computer before filling out the online form

REPRODUCIBLE RESEARCH: A BRIEF WORKSHOP

PAPAJA

papaja is short for ‘preparing APA journal articles’ and is the name of this R package designed to create fully reproducible journal articles that seamlessly fuse statistical analyses, simulations, and prose

- papaja creates dynamic scientific reports
- papaja ensures that manuscript complies to APA rules
- Manual: http://frederikaust.com/papaja_man/

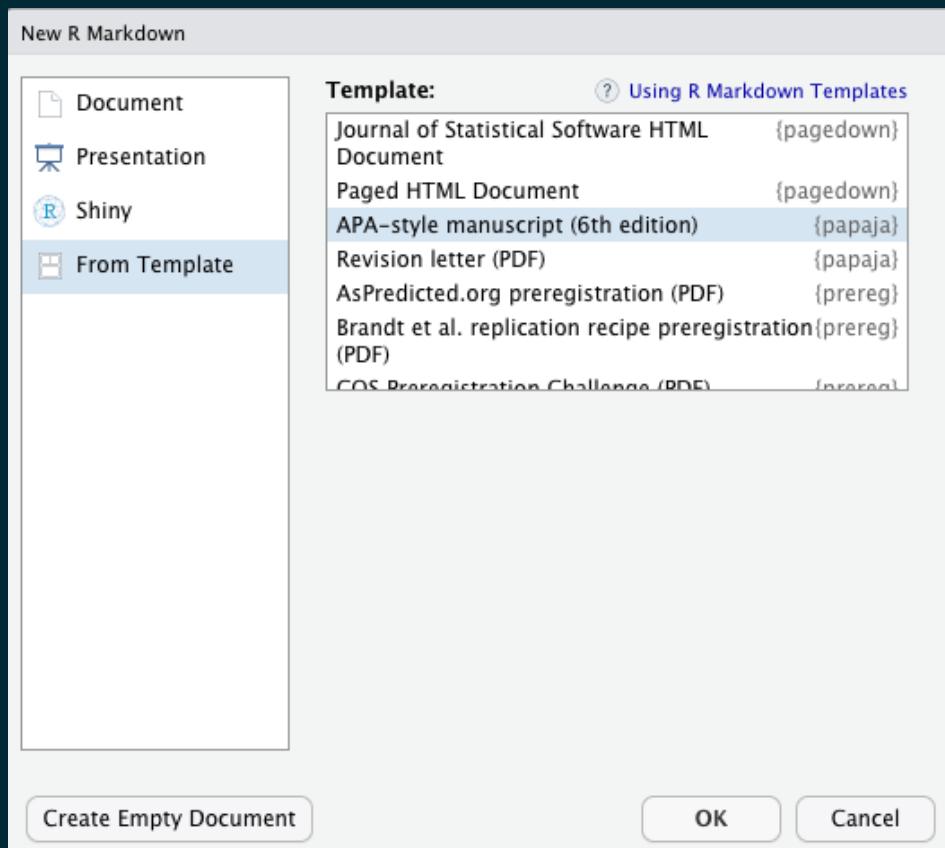

```
# Install devtools package if necessary
if(!"devtools" %in% rownames(installed.packages()))
  install.packages("devtools")

# Install the stable development verions from GitHub
devtools::install_github("crsh/papaja")

# Install the latest development snapshot from GitHub
devtools::install_github("crsh/papaja@devel")
```

CREATING A DOCUMENT

- Go to File > New file > R markdown
- Click on “From Template”



CREATING A DOCUMENT

- Click on the “knit” button to create pdf, html, or word version of rmarkdown file
- When you open a new document from template, it already contains some placeholder information

STRUCTURE OF DOCUMENT

- YAML front matter
- Manuscript metadata
- Rendering options
- Body
- Uses markdown for layout
(<https://www.markdownguide.org/cheat-sheet/>)
- Can contain R code chunks

CITATIONS

- Tip: Rmarkdown and papaja work really well in combination with zotero
- Create a .bib file with collection of references
- Add .bib file to YAML front matter
- Tip: the “citr” package provides an add-in to search and insert citations

CITATIONS

Citation type	Syntax	Rendered citation
Citation within parentheses	[@james_1890]	(James, 1890)
Multiple citations	[@james_1890; @bem_2011]	(Bem, 2011; James, 1890)
In-text citations	@james_1890	James (1890)
Year only	[–@bem_2011]	(2011)

LET'S ADD A DATASET TO OUR MANUSCRIPT

```
48 ````{r setup, include = FALSE}
49 # Load libraries
50 library("papaja")
51 library("tidyverse")
52 library("report")
53 library("kableExtra")
54 library("afex")
55
56 # Set up references to R scripts in a bib file
57 r_refs("r-references.bib")
58
59 # Load dataset
60 df = read_csv2("220908_simulated-data.csv")
61 ````
```

NUMERICAL VALUES

```
82 ## Participants
83
84 Our simulated sample consists of `r length(df$id)` participants.
85 `r printnum(sum(df$condition %in% "work"), numerals=FALSE, capitalize=TRUE)` participants were randomly assigned to the
86 workshop condition, while the remaining `r printnum(sum(df$condition %in% "contr"))` participants were randomly assigned
87 to the control condition.
88 `r printnum(sum(df$gender %in% "male"), numerals=FALSE, capitalize=TRUE)` participants were male, while `r
89 printnum(sum(df$gender %in% "female"))` participants were female.
90 On average, participants were `r printnum(mean(df$age))` years old ($SD = `r sd(df$age)`$ years).
```

TABLES

```
115 ~ ````{r descriptive}
116 df %>%
117   select(pretest, posttest, conscientiousness, condition) %>%
118   report_sample(
119     group_by = "condition"
120   ) %>%
121   apa_table(
122     ,
123     caption = "Descriptives of key variables by condition.",
124     escape=TRUE
125   )
126 ~ ````
```

RESULTS FROM STATISTICAL TESTS

```
137 ````{r, anova-example}
138 m1 = aov_ez(
139   data=df,
140   dv="delta",
141   id="id",
142   between=c("condition", "gender")
143 )
144
145 m1_apa = apa_print(m1)
146
147 apa_table(
148   m1_apa$table
149   , caption = "A really beautiful ANOVA table."
150   , note = "Note that the column names contain beautiful mathematical copy: This is because the table has variable
151   labels."
152 )
152 ^
153
154 Condition(`r m1_apa$full$condition`) affected change in positive attitude to open science practices.
155 Gender was not related to change in positive attitude to open science practices, `r m1_apa$full$gender`.
156 There was no significant interaction effect between condition and gender, `r m1_apa$full$condition_gender`.
```

RESULTS FROM STATISTICAL TESTS

```
167 ### Regression
168
169 ```{r, regression-example, results="asis"}
170 m2 = lm(delta ~ condition + conscientiousness, data=df)
171 print(report(m2))
172 ```
173
174 ```{r, regression-table}
175 apa_m2 = apa_print(m2)
176 apa_table(apa_m2$table,
177             caption = "Results from regression model.",
178             escape=TRUE)
179 ```
180
```

FIGURES

```
157  
158 • ````{r, beeplot}  
159 apa_beeplot(  
160   data = df  
161   , id = "id"  
162   , dv = "delta"  
163   , factors = c("condition", "gender")  
164 )  
165 • ````  
166
```

LIMITATIONS

- Collaboration can be difficult when co-authors do not use rmarkdown
- Conversion to pdf works better than conversion to word
- Long-term computational reproducibility can be an issue as packages and R versions are updated. Docker can be a solution to this problem.

PERSONAL NOTE

- There are situations where a word document is easier than an Rmarkdown file
- It takes a while to get used to working with papaja
- Some experience with R is definitely needed!
- papaja can save you a ton of time when revising analyses!
- It is much easier for me to trace all the steps in a research process

SHARING RESEARCH MATERIALS: AN ILLUSTRATION

AVAILABLE PLATFORMS

- Many repositories exist (<https://www.re3data.org>)
- “A data repository is an online platform that is used to deposit completed datasets with the purpose to publish, share and/or preserve them. A data repository is database infrastructure that compiles, manages and gives access to data and associated metadata and documentation.”
- However, repositories can be used for more than just data
- Common repositories in our field include Open Science Framework (<https://osf.io>) and Zenodo (<https://zenodo.org>)

LET'S SHARE SOME MATERIALS ABOUT OUR STUDY

- Create project in OSF

CONCERNS ABOUT SHARING DATA

- Discussion: what concerns do you have about sharing data?

CONCLUDING THOUGHTS

ADOPTING OPEN SCIENCE TOOLS: A WORTHY INVESTMENT?

- Open science practices can help us strengthen the credibility and replicability of IO psychology research
- We need to acknowledge that adopting these practices takes time
- As an ECR, learning these practices and tools is a valuable skill
- As a senior researcher, we should not feel threatened by these practices and tools
- Widespread adaptation requires systemic changes
- We need to emphasize quality over quantity in research

THANK YOU FOR YOUR ATTENTION

BIBLIOGRAPHY

- Alston, Jesse M., and Jessica A. Rick. 2021. “A Beginner’s Guide to Conducting Reproducible Research.” *The Bulletin of the Ecological Society of America* 102 (2): e01801. <https://doi.org/10.1002/bes2.1801>.
- Crüwell, Sophia, Johnny van Doorn, Alexander Etz, Matthew Makel, Hannah Moshontz, Jesse Niebaum, Amy Orben, Sam Parsons, and Michael Schulte-Mecklenbeck. 2019. “8 Easy Steps to Open Science: An Annotated Reading List.” ArXiv. <https://doi.org/10.31234/osf.io/cfzyx>.
- Lindsay, Brian A. Nosek and D. Stephen. 2018. “Preregistration Becoming the Norm in Psychological Science.” *APS Observer* 31 (February).
<https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science>.
- OPEN SCIENCE COLLABORATION. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716.
<https://doi.org/10.1126/science.aac4716>.
- Science, Center for Open. n.d. “Preregistration.” Accessed September 2, 2022.
<https://www.cos.io/initiatives/prereg>.