Open Science and Reproducibility

Alexandra Sarafoglou

Eric-Jan Wagenmakers

@OSCAmsterdam

opensciecne-amsterdam.com

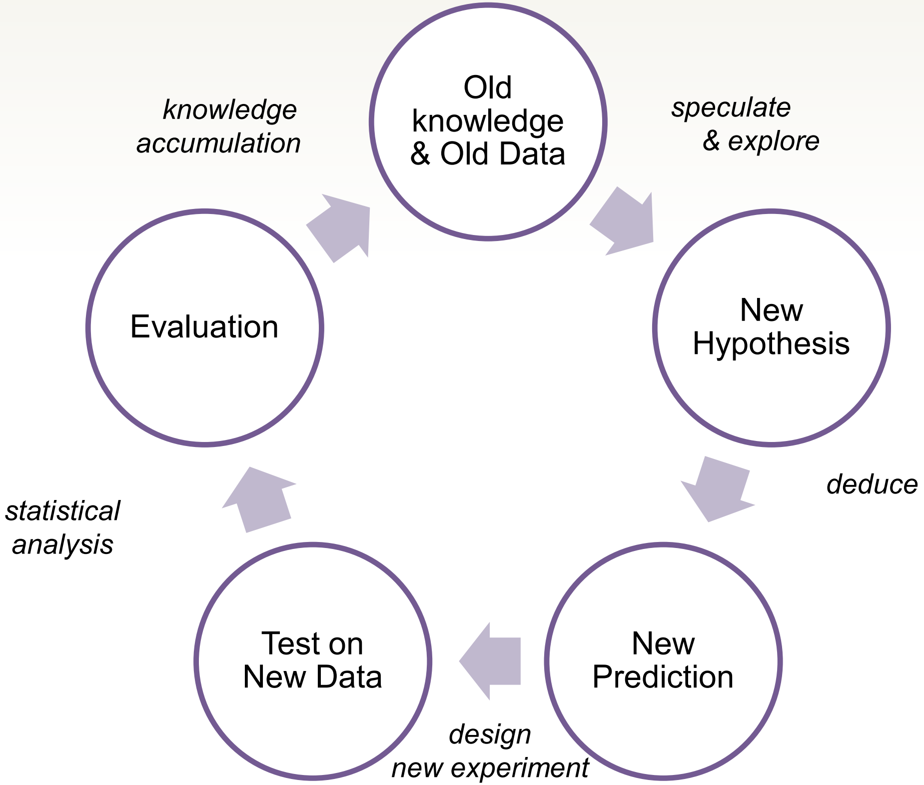
Overview

- What researchers want
- What the field gets
- How to uncover hidden uncertainty

Methodology 101

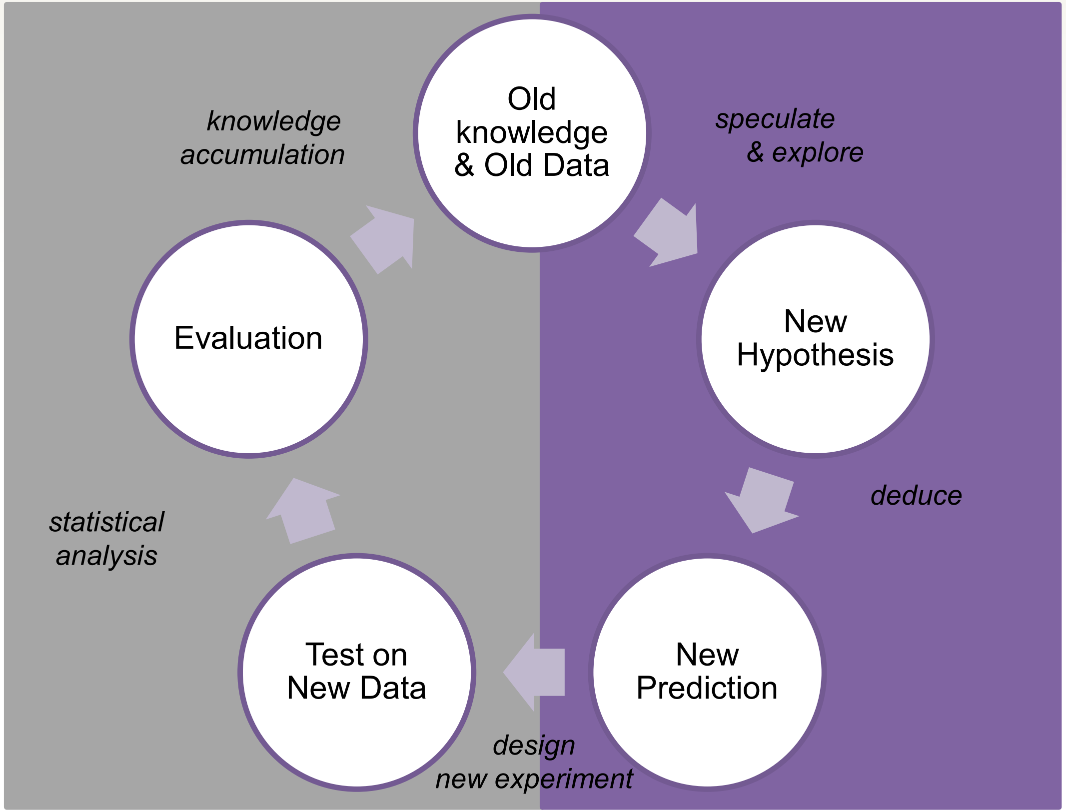
- There is a conceptual distinction between **hypothesis-generating** and **hypothesis-testing** research (De Groot 1956/2014; Reichenbach, 1938)
- When the data inspire a hypothesis, you cannot use the same data to test this hypothesis

Methodology 101



Methodology 101

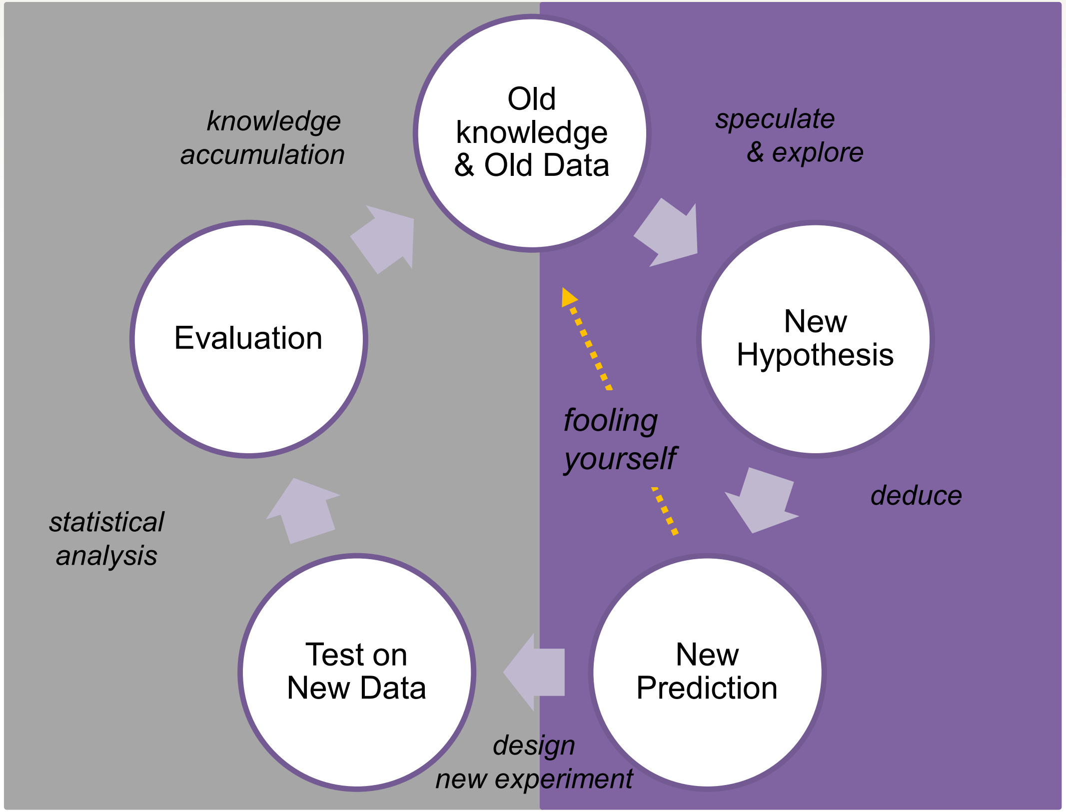
The Statistical Context of Justification
–
Confirmatory Research



The Creative Context of Discovery
–
Exploratory Research

Methodology 101

The Statistical Context of Justification
–
Confirmatory Research



The Creative Context of Discovery
–
Exploratory Research

Main Dilemma

- 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.

Main Dilemma

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

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

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

X

The first principle is that you must not fool yourself—and you are the easiest person to fool

Richard Feynman

Main Dilemma

- 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.

A Perfect Storm

- 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, as it violates an implicit social contract.

What Researchers Want

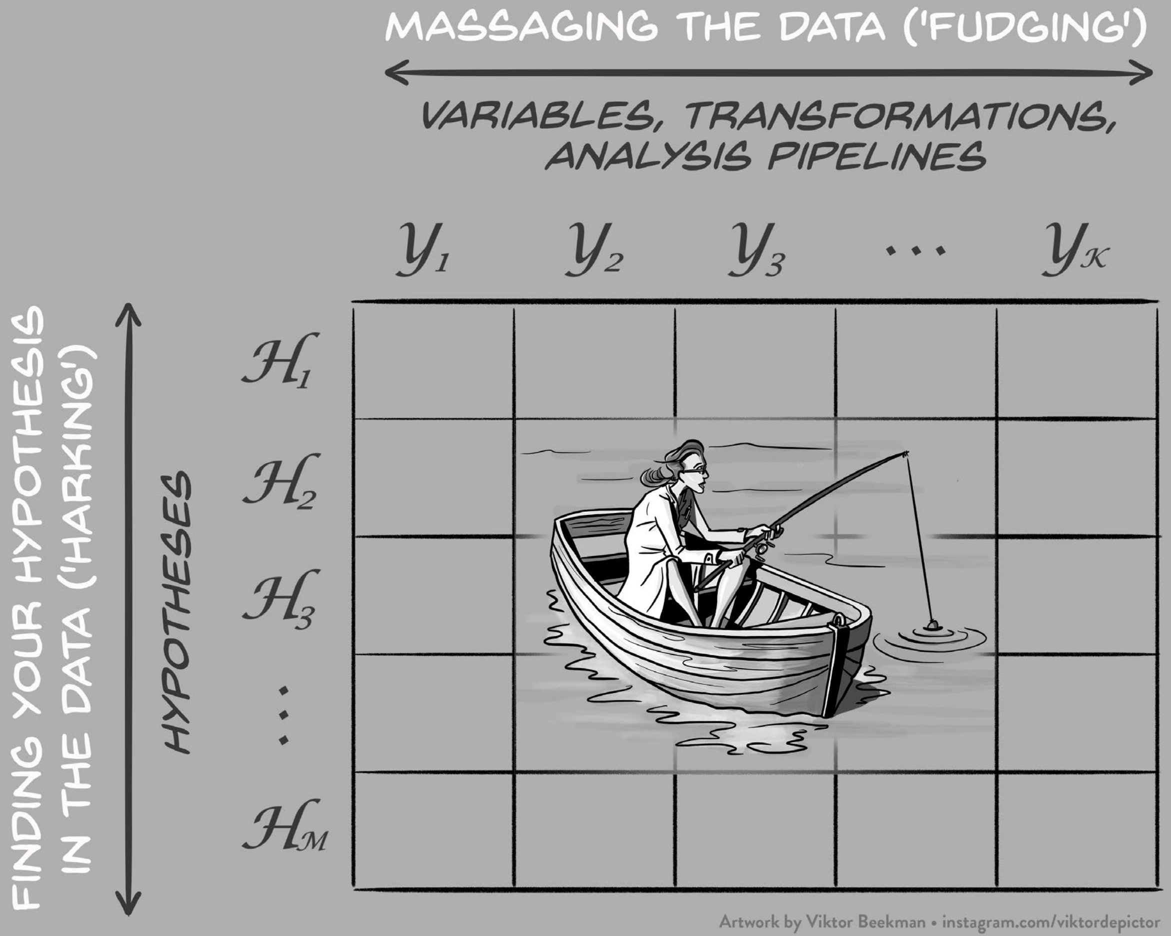
To discover the truth, but also:

- To present compelling data that leave no room for doubt or dissent
- To develop a coherent theoretical framework
- To publish papers that make interesting claims

What The Field Gets

Fruits of Perverse Incentives and Uncertainty-Allergy:

- Publication bias
- Fudging
- HARKing


MASSAGING THE DATA ('FUDGING')
VARIABLES, TRANSFORMATIONS, ANALYSIS PIPELINES
y_1 y_2 y_3 ... y_K
FINDING YOUR HYPOTHESIS IN THE DATA ('HARKING')
HYPOTHESES
$\mathcal{H}_1$ $\mathcal{H}_2$ $\mathcal{H}_3$ ... $\mathcal{H}_M$
Artwork by Viktor Beekman • instagram.com/viktordepictor

What The Field Gets

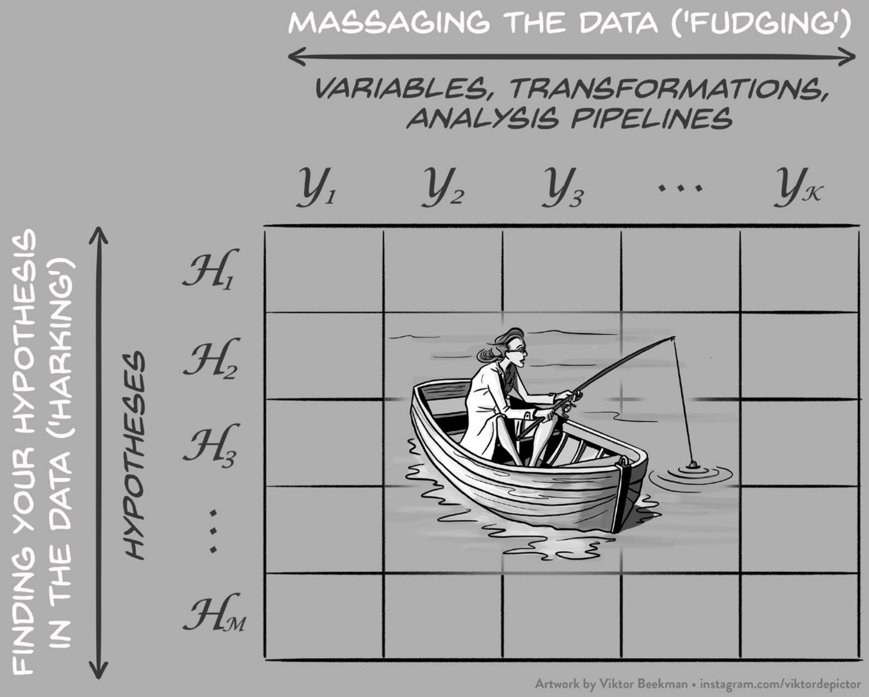
Table 1. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Simmons, Nelson, Simonsohn (2011)

Overconfident Claims.

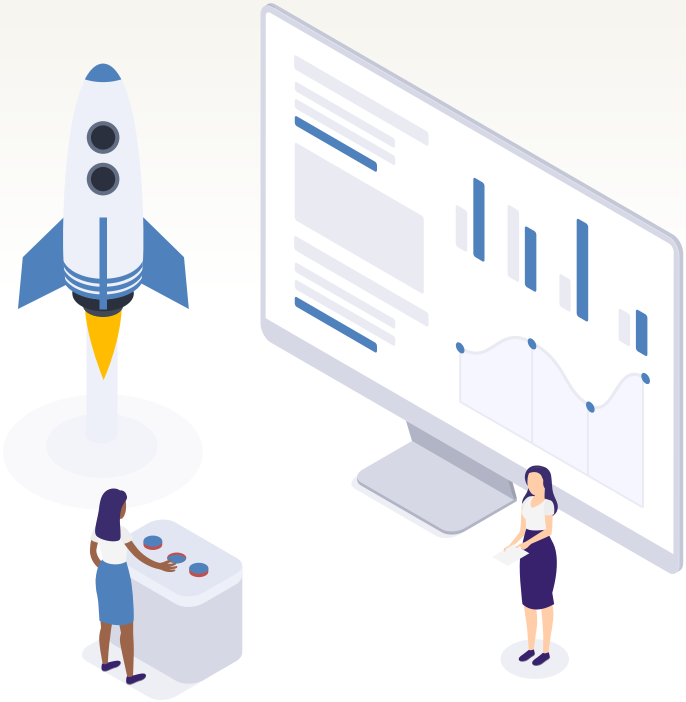
Spurious Results.



1

How To Uncover Hidden Uncertainty

Open Science Tools

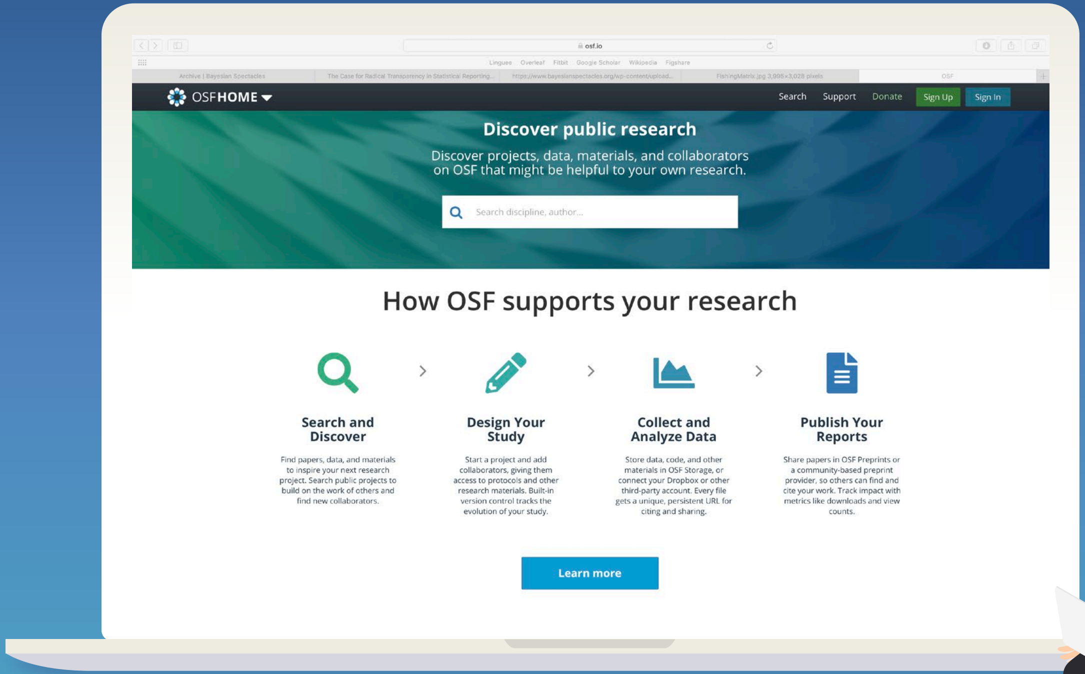


Method 1: Preregistration of Analysis Plans

- Strict separation between exploratory and confirmatory research
- Specify **hypotheses** and **all statistical analyses** before data collection

Open Science Framework (https://osf.io)

Preregistration is published with timestamp in a trusted online repository

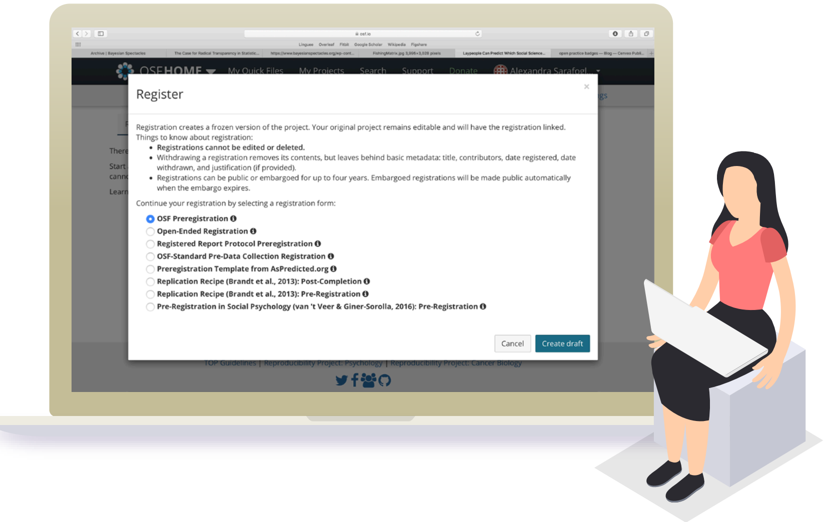


Method 1: Preregistration of Analysis Plans

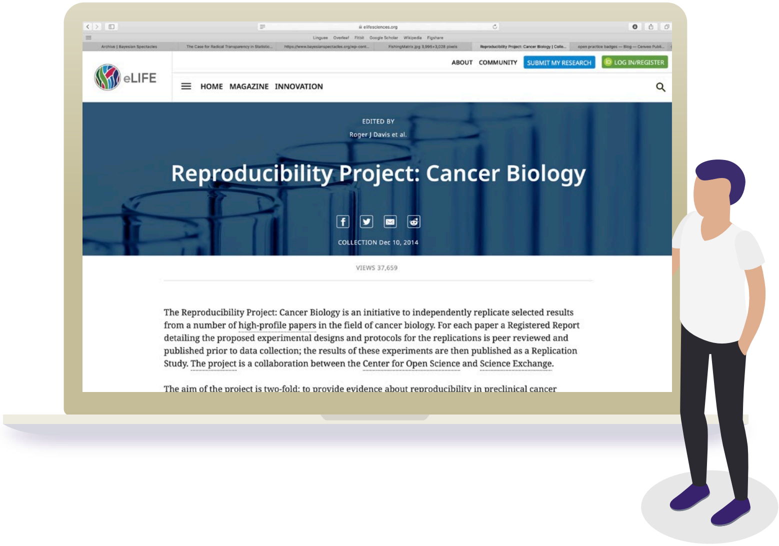
- After data collection, the preregistered analyses are conducted in an automated fashion
- Forces researchers to adhere to the empirical cycle
- Does not rule out exploratory analyses; just labels them as such
- Most efficient way to combat implicit and explicit forms of significance seeking

Method 1: Preregistration of Analysis Plans

- Online templates

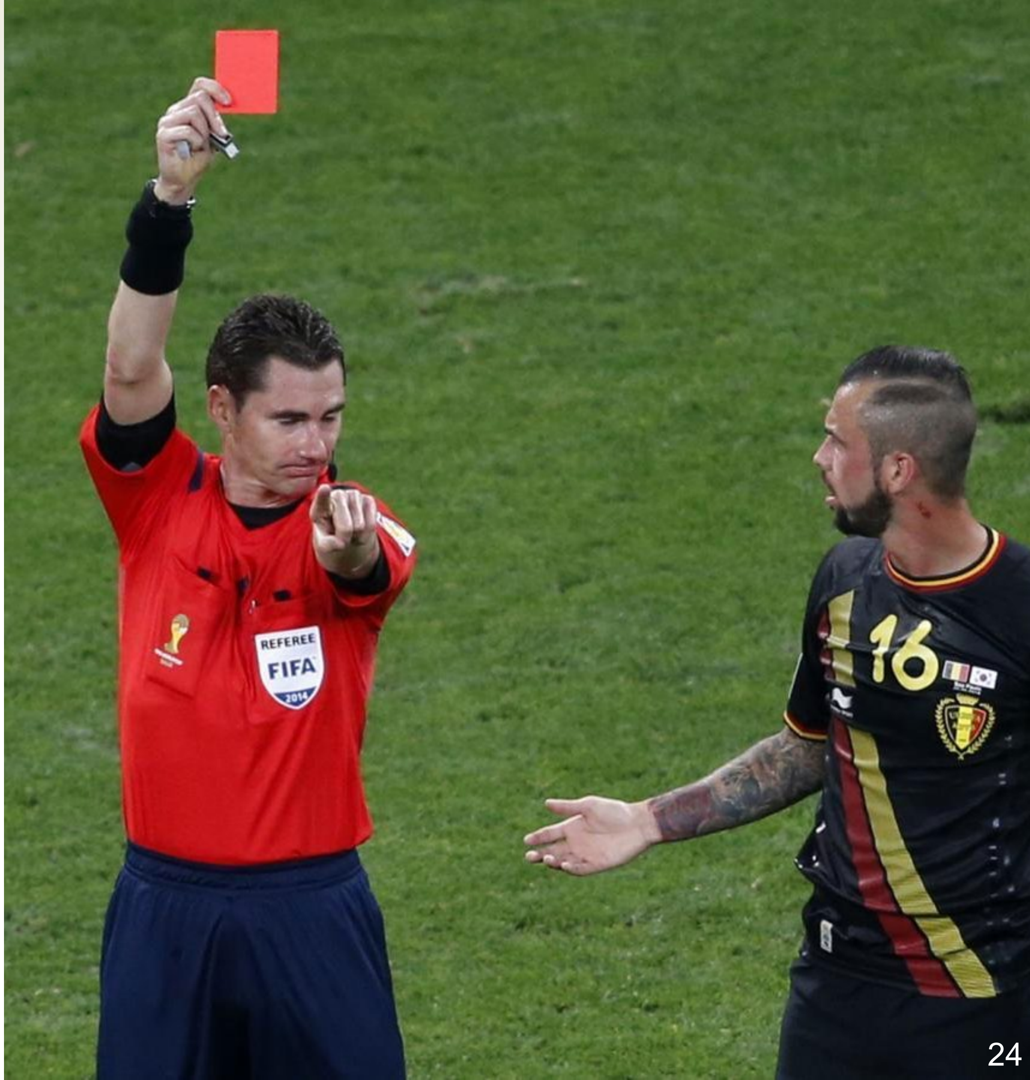


- Published Preregistrations



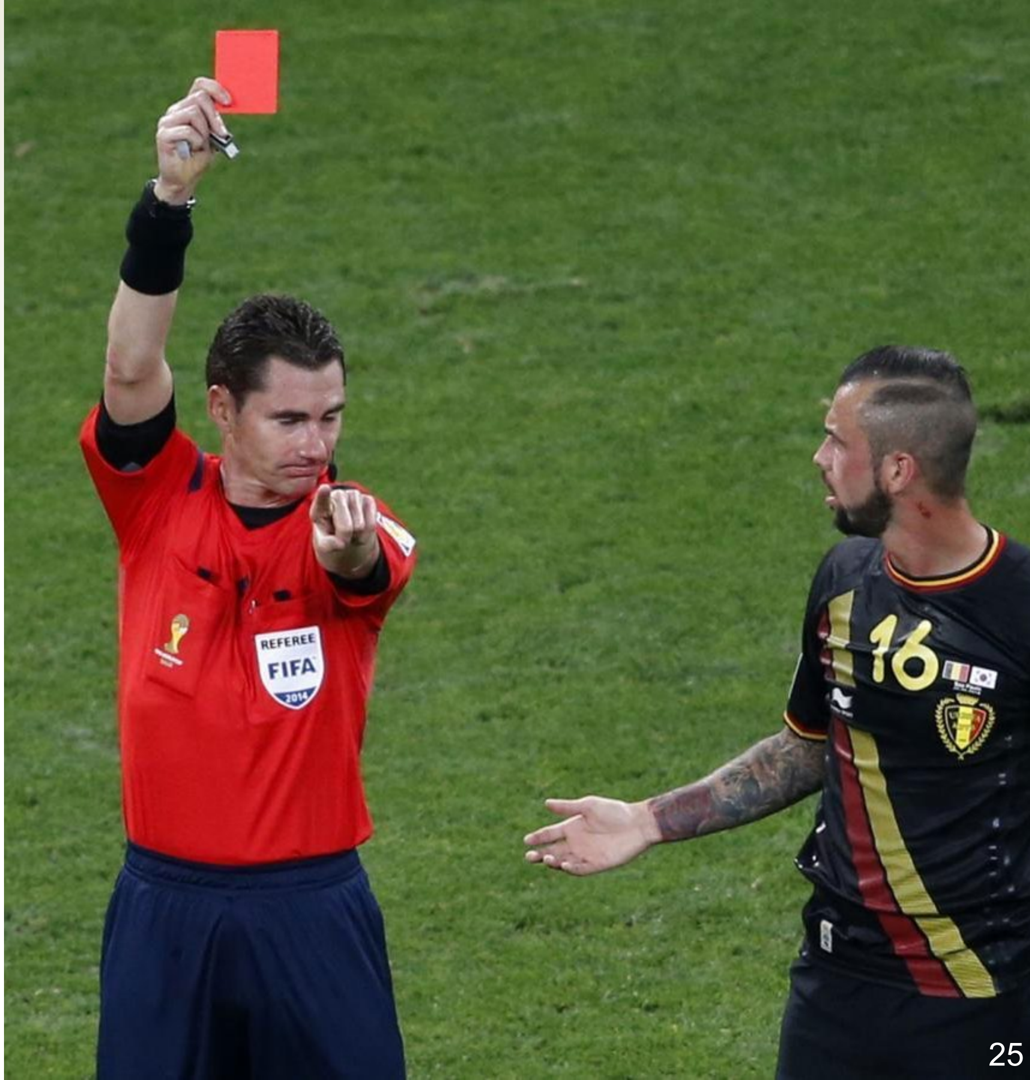
How to write a good preregistration

"Are soccer referees more likely to give red cards to players with dark skin than to players with light skin?"

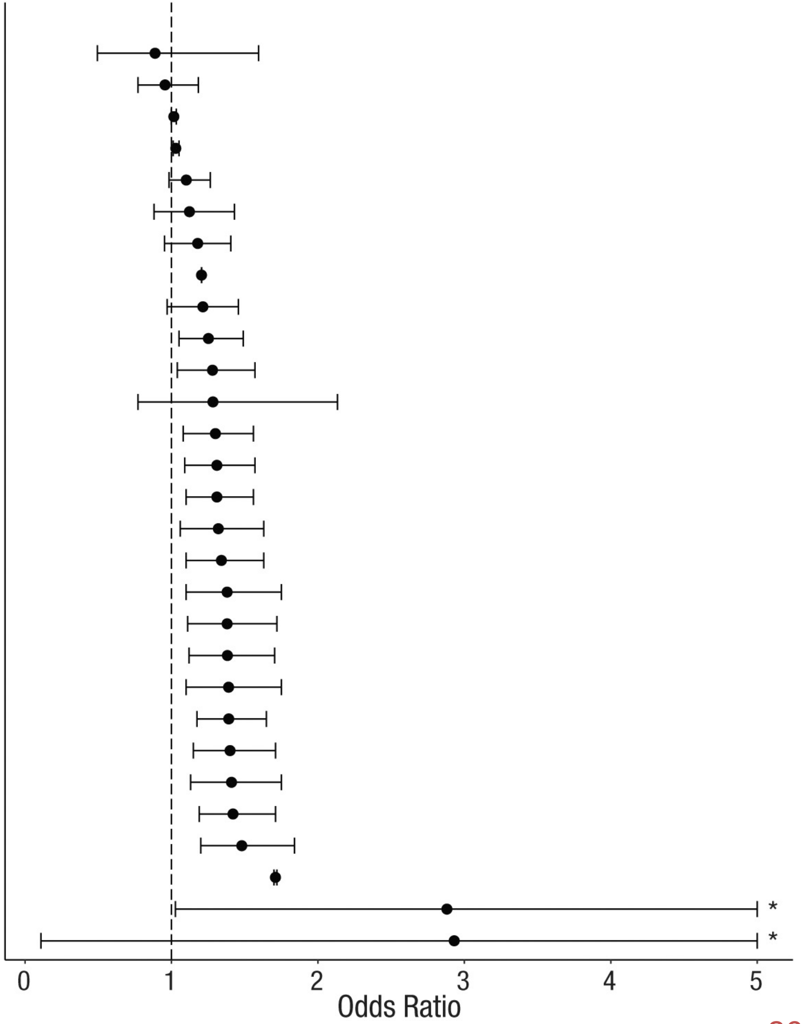


How to write a good preregistration

- How to measure skin color?
- Control for referees skin color?
- Is each red card decision independent?
- What about other ethnicities?



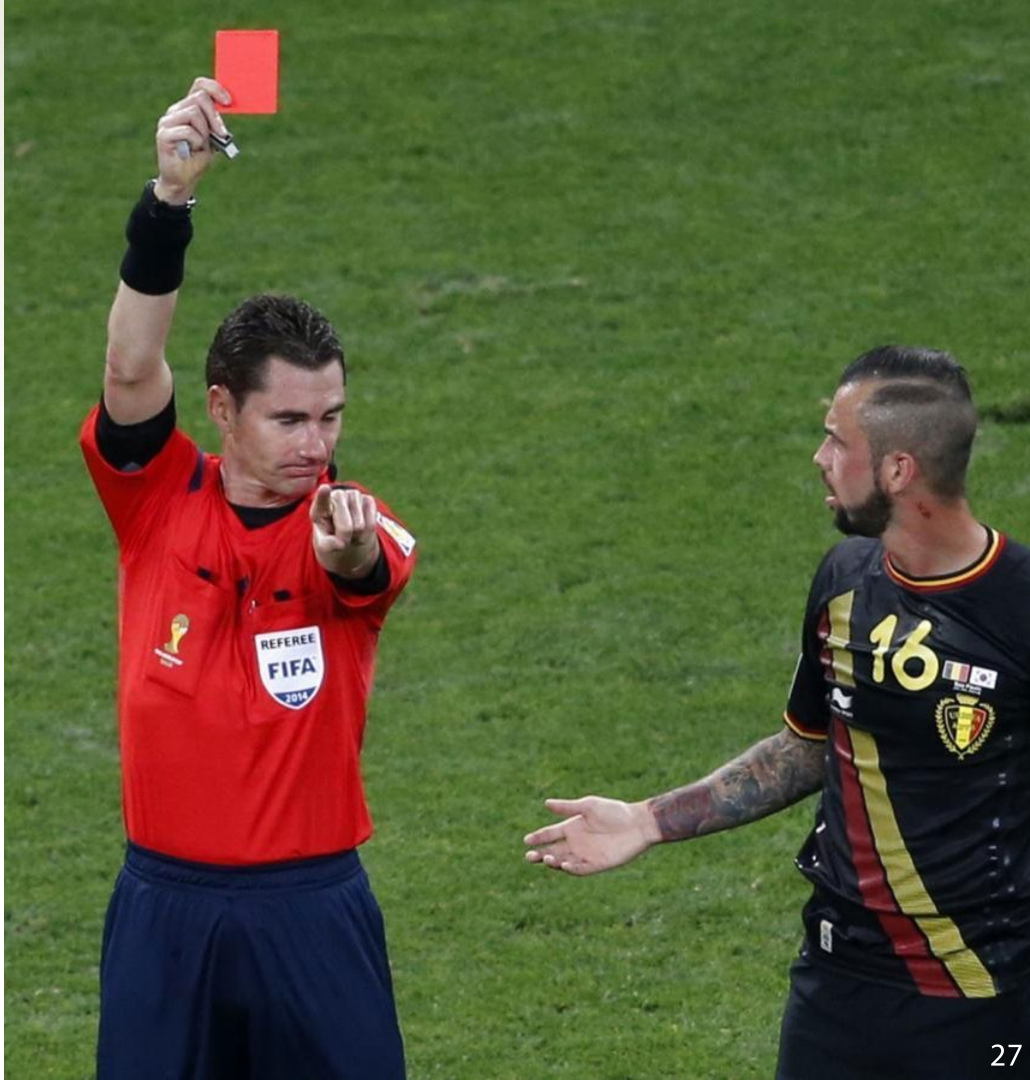
Team	Analytic Approach	Odds Ratio
12	Zero-Inflated Poisson Regression	0.89
17	Bayesian Logistic Regression	0.96
15	Hierarchical Log-Linear Modeling	1.02
10	Multilevel Regression and Logistic Regression	1.03
18	Hierarchical Bayes Model	1.10
31	Logistic Regression	1.12
1	OLS Regression With Robust Standard Errors, Logistic Regression	1.18
4	Spearman Correlation	1.21
14	WLS Regression With Clustered Standard Errors	1.21
11	Multiple Linear Regression	1.25
30	Clustered Robust Binomial Logistic Regression	1.28
6	Linear Probability Model	1.28
26	Hierarchical Generalized Linear Modeling With Poisson Sampling	1.30
3	Multilevel Logistic Regression Using Bayesian Inference	1.31
23	Mixed-Model Logistic Regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear Probability Model, Logistic Regression	1.34
5	Generalized Linear Mixed Models	1.38
24	Multilevel Logistic Regression	1.38
28	Mixed-Effects Logistic Regression	1.38
32	Generalized Linear Models for Binary Data	1.39
8	Negative Binomial Regression With a Log Link	1.39
20	Cross-Classified Multilevel Negative Binomial Model	1.40
13	Poisson Multilevel Modeling	1.41
25	Multilevel Logistic Binomial Regression	1.42
9	Generalized Linear Mixed-Effects Models With a Logit Link	1.48
7	Dirichlet-Process Bayesian Clustering	1.71
21	Tobit Regression	2.88
27	Poisson Regression	2.93



How to write a good preregistration

Create and Analyze Dummy Data

- Simulations
- Pilot Studies
- Existing Data



Method 2: Sensitivity Analyses

Examine sensitivity to modeling choices:

- Multiverse analysis
- Crowd sourcing

Ideally this is done by independent labs

Method 3: Blinded Analyses

Challenges of Preregistration: Unexpected features of the data

- Dutilh et al. (2017): Preregistration of an impossible analysis
- Reproducibility Project: Cancer Biology
  - Horrigan (2017): Spontaneous tumor regressions
  - Arid, Kandela & Mantis (2017): Unexpected early deaths in control group

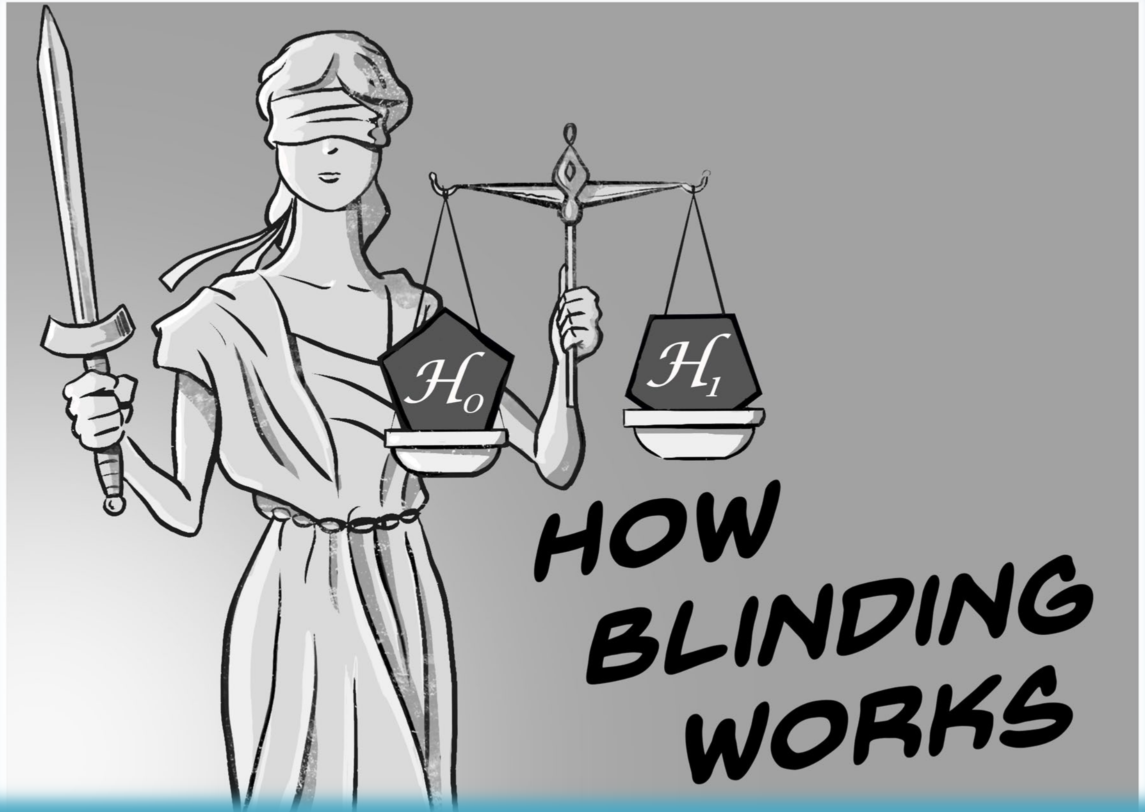
Method 3:
Blinded Analyses

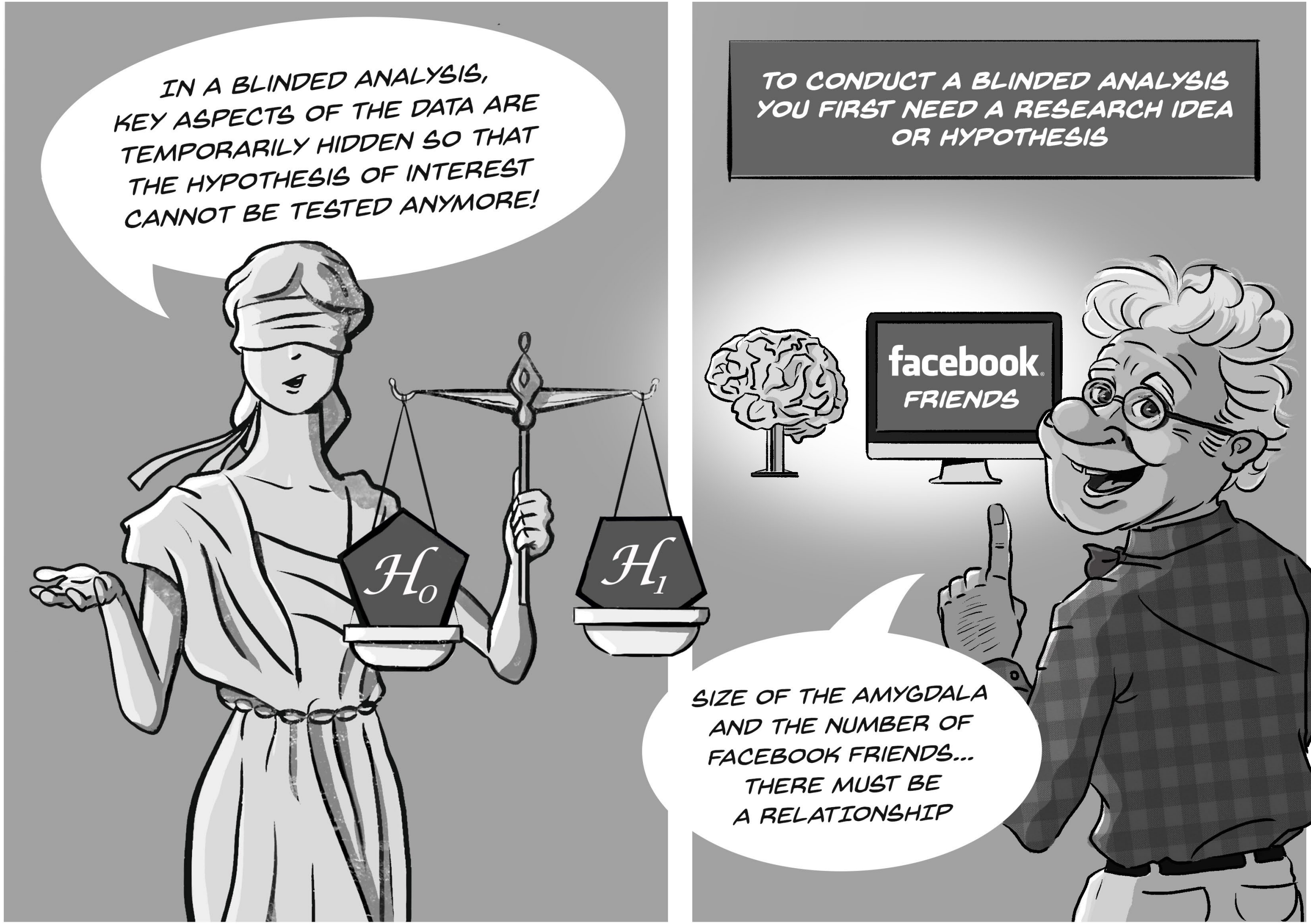
Create and Analyze Dummy Data

- Common Practice in (Astro-) physics
- Allows researchers to make data depend choices without introducing bias

Artwork by Viktor Beekman
instagram.com/janovitsj

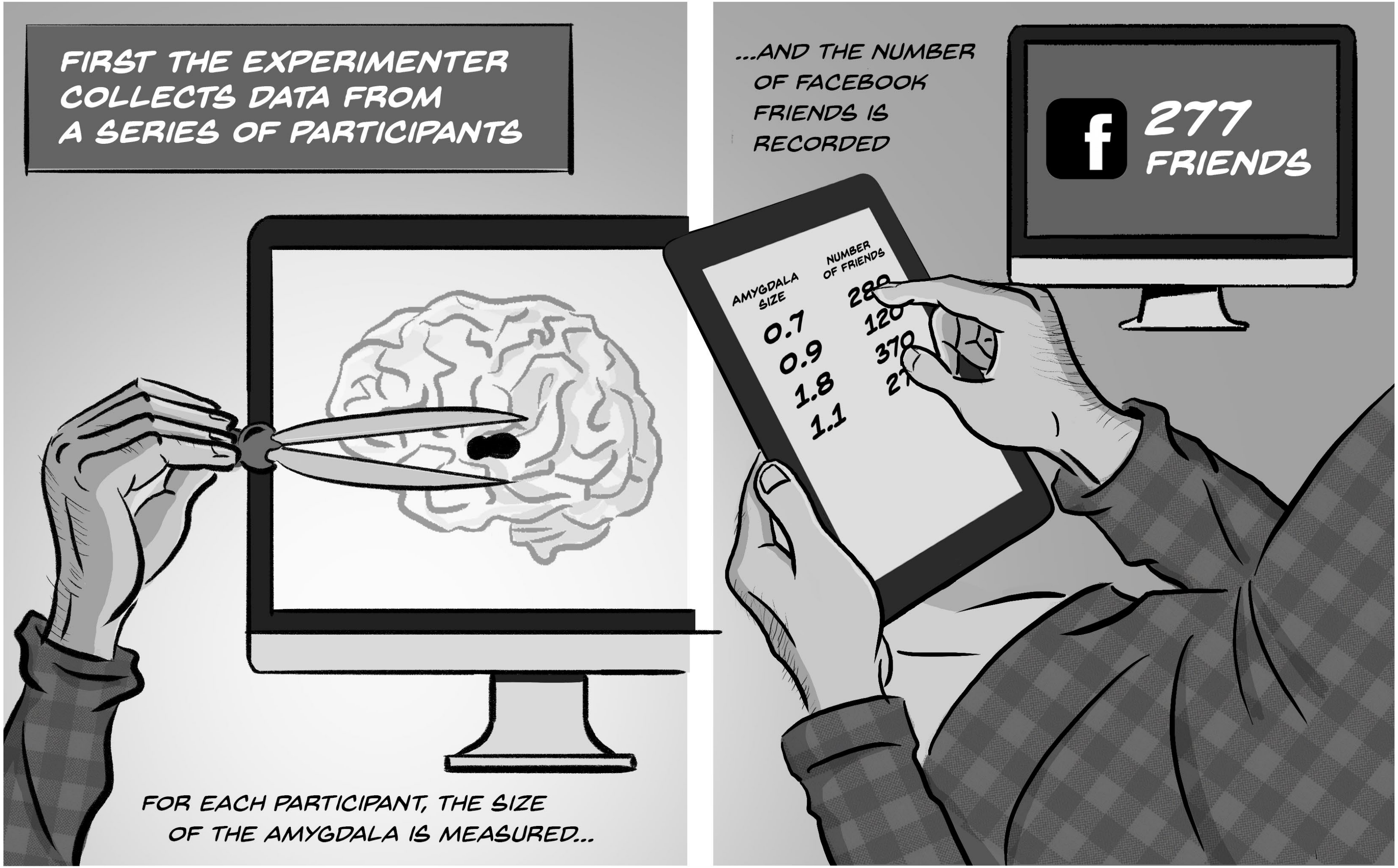
How Blinding Works

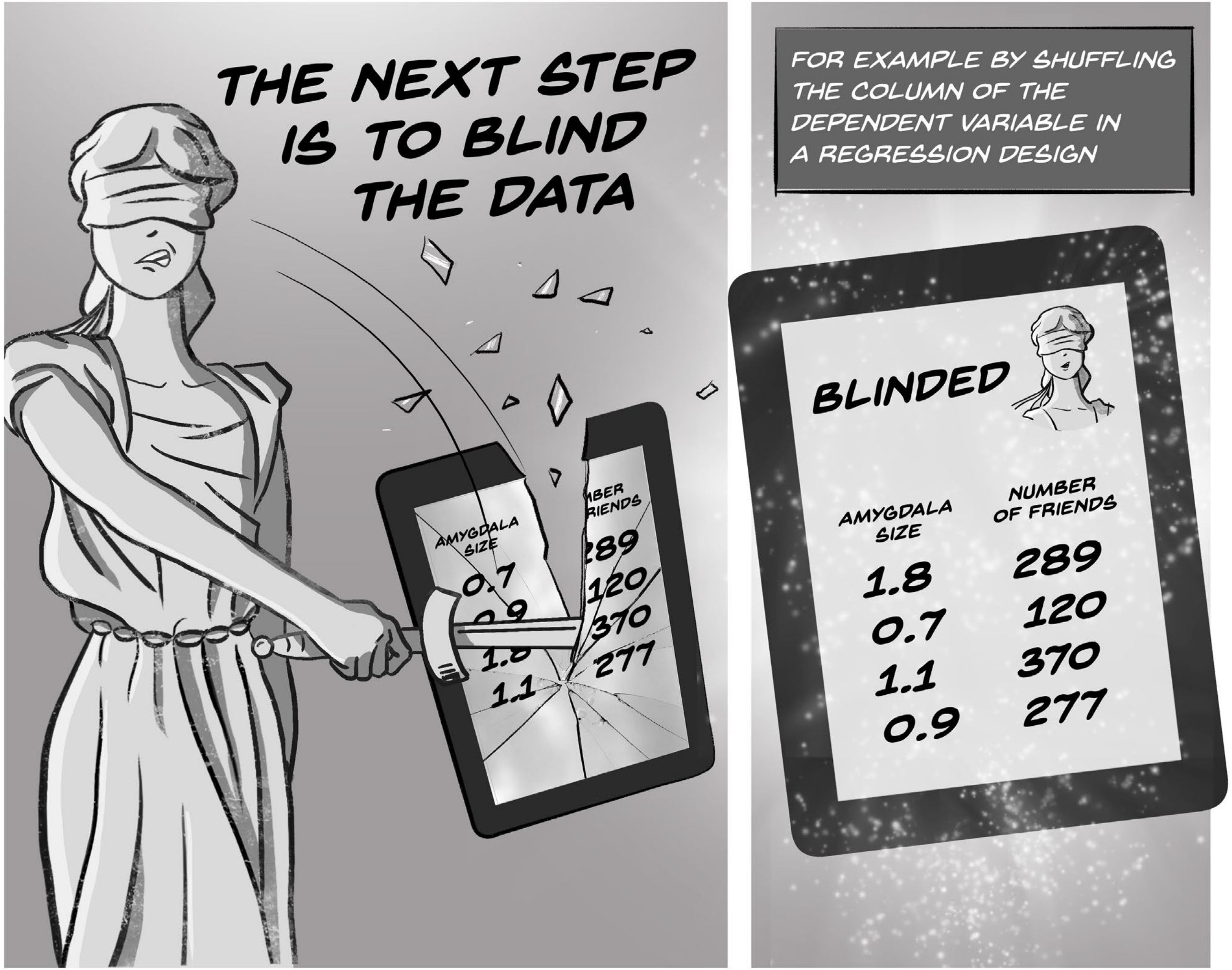




Artwork by Viktor Beekman - instagram.com/viktordepictor

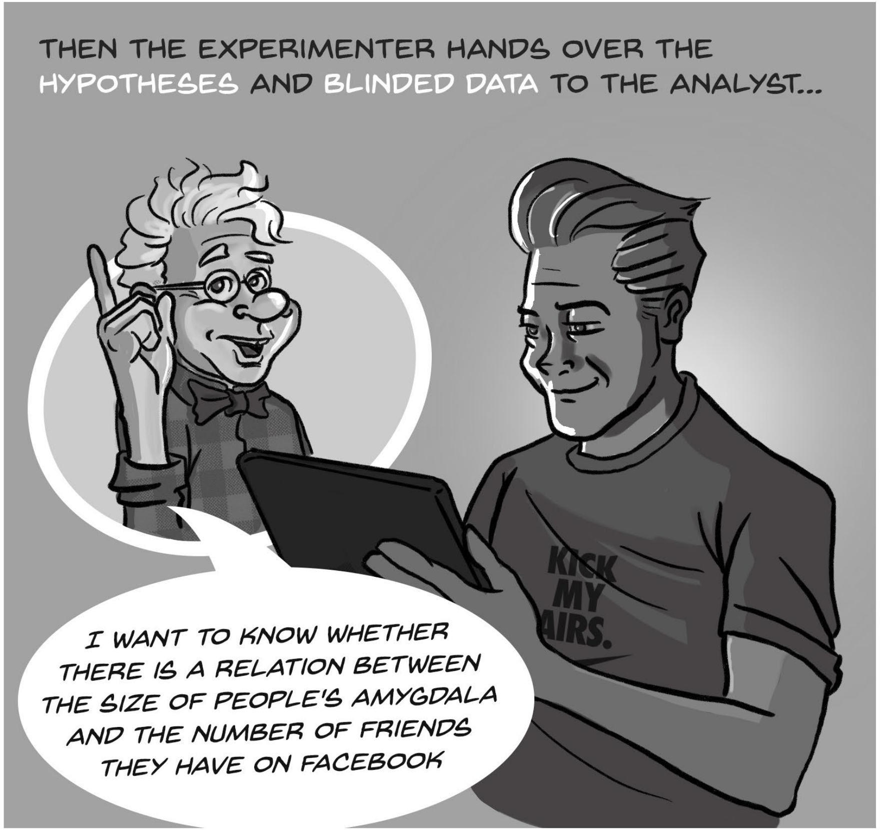
1/3


FIRST THE EXPERIMENTER COLLECTS DATA FROM A SERIES OF PARTICIPANTS
FOR EACH PARTICIPANT, THE SIZE OF THE AMYGDALA IS MEASURED...
...AND THE NUMBER OF FACEBOOK FRIENDS IS RECORDED
277 FRIENDS
AMYGDALA SIZE
NUMBER OF FRIENDS
0.7
0.9
1.8
1.1
120



Artwork by Viktor Beekman - instagram.com/viktordepictor

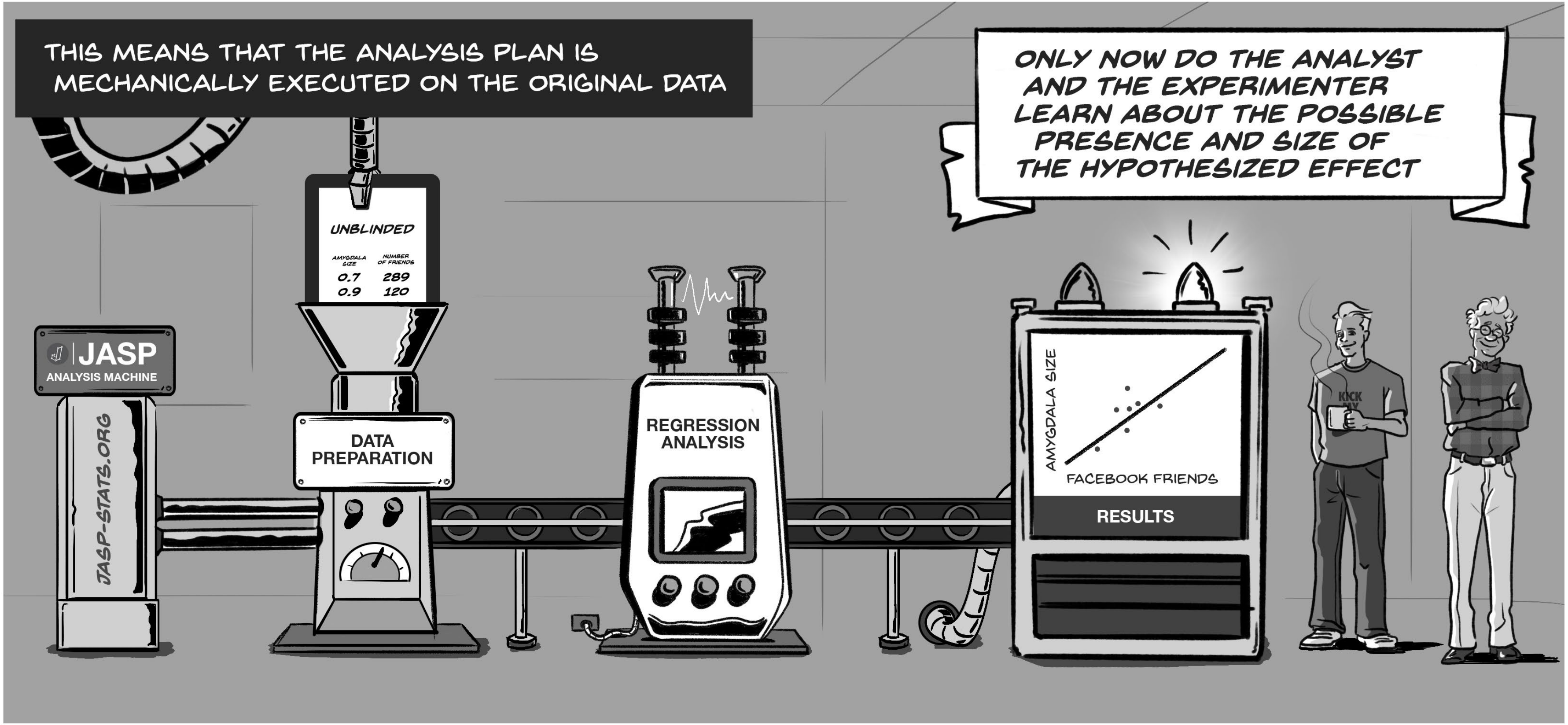
2/3


THEN THE EXPERIMENTER HANDS OVER THE HYPOTHESES AND BLINDED DATA TO THE ANALYST...
I WANT TO KNOW WHETHER THERE IS A RELATION BETWEEN THE SIZE OF PEOPLE'S AMYGDALA AND THE NUMBER OF FRIENDS THEY HAVE ON FACEBOOK
KICK MY AIRS.


...AND THE ANALYST DECIDES ON AN ANALYSIS PLAN.
WHICH VARIABLES SHOULD I INCLUDE?
HOW SHOULD I TRANSFORM MY DATA?
REGRESSION OR CORRELATION?
KICK MY AIRS.


WHEN THE ANALYST IS HAPPY, HE RETURNS THE ANALYSIS PLAN TO THE EXPERIMENTER


NOW THE BLIND IS LIFTED
$\mathcal{H}_0$
$\mathcal{H}_1$



Artwork by Viktor Beekman - instagram.com/viktordepictor

3/3

# Facebook Friends	Amygdala Size	Gender	Age	Overal Brain volume
190	0.655	male	28	1238
232	0.680	male	35	1386
84	0.647	female	18	1176
311	0.698	male	20	1172
138	0.699	female	27	1185
356	0.732	female	34	970

Blind Data by Shuffling Rows (Regression Designs or Correlational Data)

# Facebook Friends	Amygdala Size	Gender	Age	Overall Brain volume
84	0.655	male	28	1238
356	0.680	male	35	1386
190	0.647	female	18	1176
232	0.698	male	20	1172
138	0.699	female	27	1185
311	0.732	female	34	970

Blind Data by Shuffling Rows (Regression Designs or Correlational Data)

Gender

Education	male	female
high	13.7 (SD 1)	14.5 (SD 1.2)
low	11.8 (SD 2)	9.46 (SD 2.5)

Blind Data by Hiding the Labels (ANOVA Designs)

Gender

Education	?	?
?	13.7 (SD 1)	14.5 (SD 1.2)
?	11.8 (SD 2)	9.46 (SD 2.5)

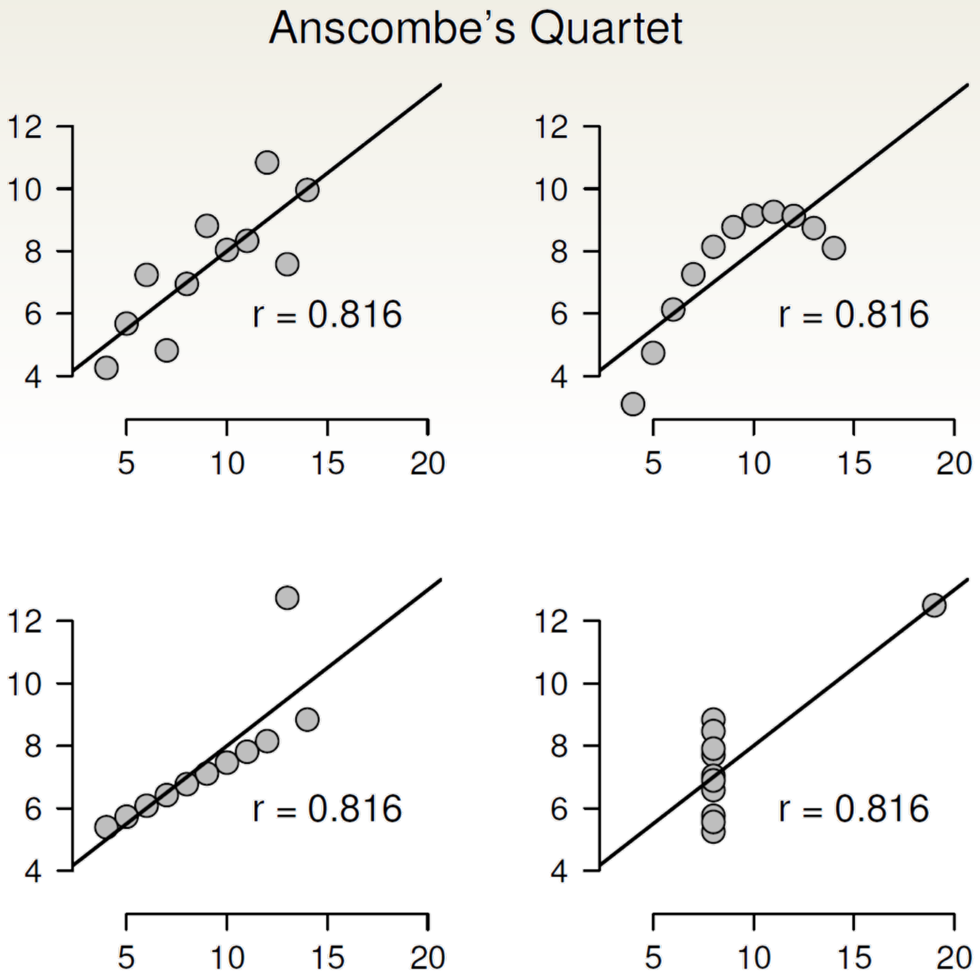
Blind Data by Hiding the Labels (ANOVA Designs)

Method 4: Share the Data

- Facilitates re-analysis, verification, and metaanalysis
- In review process, allows reviewers to propose and carry out informative alternative analyses

Method 4: Share the Data

Plot your data!



Method 5: Share Experiences

OSCA

Open Science Community Amsterdam

- @OSCAmsterdam
- openscience-amsterdam.com

Concluding Comments

- More transparency is needed
- Transparency means mental hygiene: the scientific equivalent of brushing your teeth, or washing your hands after visiting the restroom
- This requires a change in culture

Concluding Comments

- Journals and funders starting to demand mental hygiene
- Mental hygiene can also be rewarded. For instance, journals could prefer to publish preregistered studies, or studies that share their data, materials, and code

Concluding Comments

Social and Behavioral Sciences

- Transparency Checklist
- https://eltedecisionlab.shinyapps.io/TransparencyChecklist/

Title: A Consensus-Based Transparency Checklist for Social and Behavioral Researchers

Authors:

B. Aczel[1*], B. Szaszi[1], A. Sarafoglou[2], Z. Kekecs[1], Š. Kucharský[2], D. Benjamin[3], C. D. Chambers[4], A. Fisher[2], A. Gelman[5], M. A. Gernsbacher[6], J. P. Ioannidis[7], E. Johnson[5], K. Jonas[8], S. Kousta[9], S. O. Lilienfeld[10,11], D. S. Lindsay[12], C. C Morey[4], M. Monafò[13], B. R. Newell[14], H. Pashler[15], D. R. Shanks[16], D. J. Simons[17], J. M. Wicherts[18], D. Albarracin[17], N. D. Anderson[19], J. Antonakis[20], H. Arkes[21], M. D. Back[22], G. C. Banks[23], C. Beevers[24], A. A. Bennett[25], W. Bleidorn[26], T. W. Boyer[27], C. Cacciari[28], A. S. Carter[29], J. Cesario[30], C. Clifton[31], R.M. Conroy[33], M. Cortese[34], F. Cosci[35], N. Cowan[36], J. Crawford[37], E. A. Crone[38], J. Curtin[6], R. Engle[39], S. Farrell[40], P. Fearon[16], M. Fichman[41], W. Frankenhuis[42], A. M. Freund[43], M. G. Gaskell[44], R. Giner-Sorolla[45], D. P. Green[5], R. L. Greene[46], L. L. Harlow[47], F. Hoces de la Guardia[48], D. Isaacowitz[49], J. Kolodner[50], D. Lieberman[51], G. D. Logan[52], W. B. Mendes[53], L. Moersdorf[43], B. Nyhan[54], J. Pollack[55], C. Sullivan[56], S. Vazire[26], E-J. Wagenmakers[2]

Concluding Comments

- Student projects are**ideal** to try out and learn about Open Science Practices!

References

Slides were (mostly) created by Eric-Jan Wagenmakers

https://www.bayesianspectacles.org/wp-content/uploads/2018/01/RadicalTransparency.pdf

https://www.youtube.com/watch?v=LxRHkk2MBWw&t=35s

False-Positive Psychology

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.

Many Analyst Study

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337-356.

Literature on Sensitivity Analyses

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702-712.

Literature on Blinding

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: hide results to seek the truth. *Nature News*, *526*(7572), 187.

Dutilh, G., Sarafoglou, A., & Wagenmakers, E.-J. (in press). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese* Preprint available on PsyArXiv: https://psyarxiv.com/d79r8