LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

How to build a **strong** scientific field

Lessons from other disciplines and an outline for structural changes

Presented at the „Social Cognition 2.0 – What Makes Good Social Cognition Research?“ Meeting, July 2017, at the Social Cognition Center Cologne

PD Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität München

COS CENTER FOR OPEN SCIENCE

RESEARCH TRANSPARENCY

BITSS

www.nicebread.de
www.researchtransparency.org
@nicebread303

John Platt (1964): Strong Inference

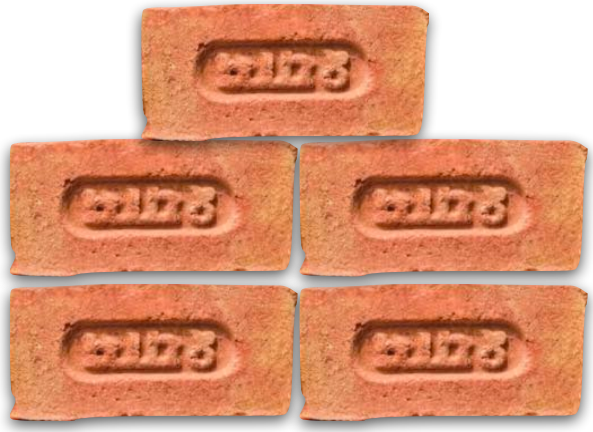
- „Scientists these days tend to keep up a polite fiction that all science is equal. […]“
- „But I think anyone who looks at the matter closely will agree that some fields of science are moving forward very much faster than others, perhaps by an order of magnitude“
- „These rapidly moving fields are fields where a particular method of doing scientific research is systematically used and taught, an accumulative method of inductive inference that is so effective that I think it should be given the name of **strong inference**."

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353. doi:10.1126/science.146.3642.347

John Platt (1964): Strong Inference

1. „Devising alternative hypotheses;
2. Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
3. Carrying out the experiment so as to get a clean result;
4. Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain; and so on.“

- „It is like climbing a tree.“



Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353. doi:10.1126/science.146.3642.347

My assumptions …

- Not all scientific fields have the same quality of the scientific process
- Some fields (more) reliably produce robust and valid knowledge, while other fields produce weak evidence and non-reproducible findings to a larger extent
- Differences in standards, quality control and incentives can explain these differences to some extent
  - Of course many other factors influence the strength of inference; such as the complexity of the object of research

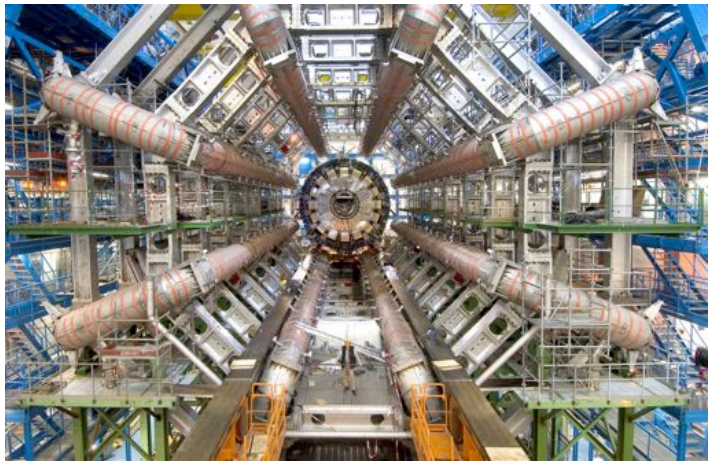
A look at other disciplines

https://home.cern/about/experiments

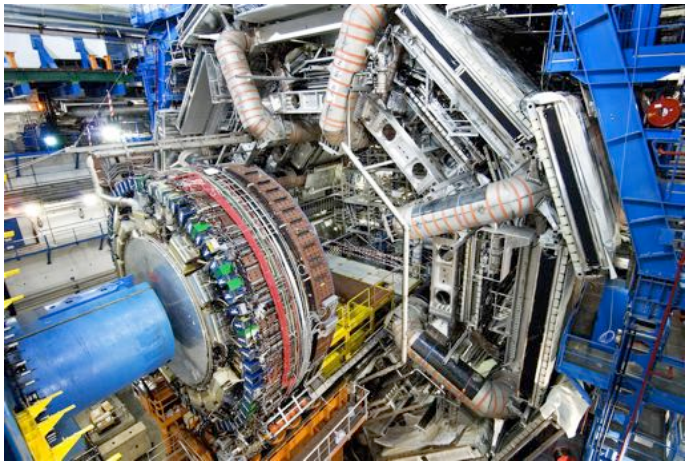
Physics: CERN

Built-in independent replication

ATLAS, 550 Mio €



CMS, 350 Mio €



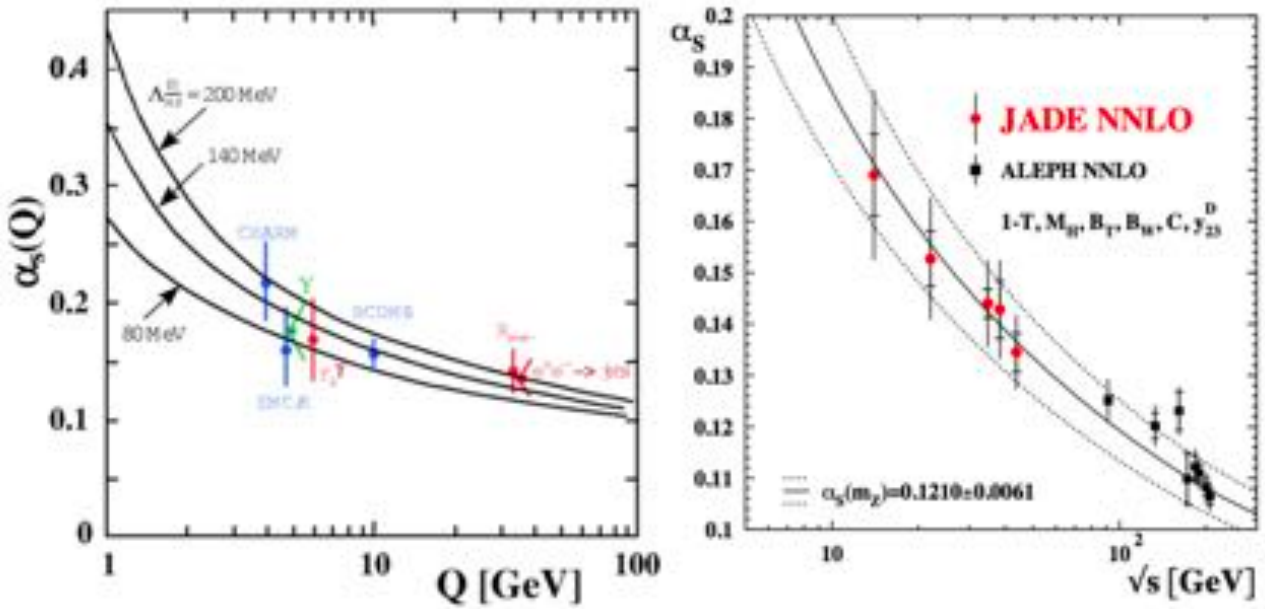
The biggest of these experiments, ATLAS and CMS, use general-purpose detectors to investigate the largest range of physics possible. Having two independently designed detectors is vital for cross-confirmation of any new discoveries made. ALICE and

➙ Independent replication is built into the system. A „discovery“ is only declared as a discovery when it has been independently replicated

Physics: Open Data

Why? Uniqueness of data

- JADE experiment (1979–1986) on PETRA accelerator at DESY
- JADE data still cover unique e^+e^- energy range in 2017
- JADE data being re-analysed even ~35 years later!



DPHEP https://arxiv.org/abs/1205.4667

@tiborsimko 4 / 30

https://home.cern/cern-people/opinion/2014/11/road-open-science
https://cds.cern.ch/record/2263664/files/cap-ilide-2017-04-04.pdf

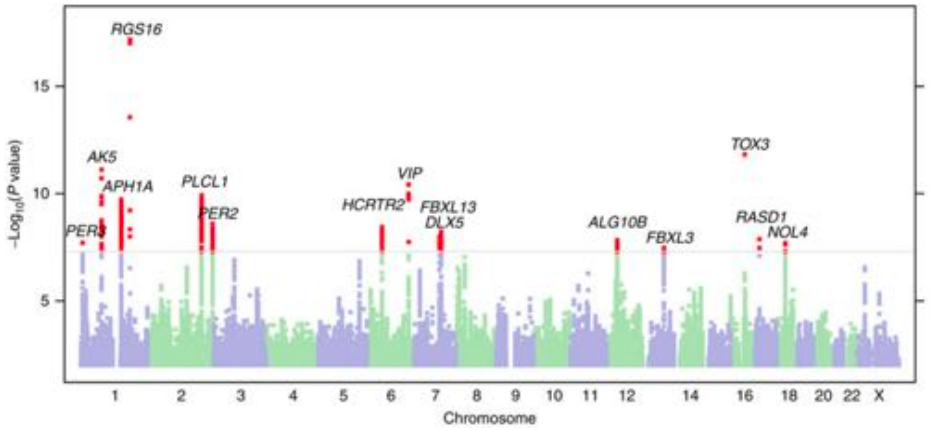
Chemistry

Danheiser is the editor-in-chief of the unconventional journal *Organic Syntheses* that has verified the experiments of all the papers it has published since it launched in 1921. The journal does this by having the research replicated by independent chemists before publishing them – a practice that is almost unheard of in chemistry or any other research field (the exception being a few brief instances in history). All experiments are checked for reproducibility in the lab of one of the journal's board of editors, often by graduate students and postdoctoral researchers working under the supervision of the *Organic Syntheses* editor.

Between 2010 and 2016, the journal rejected 7.5% of submissions due to irreproducibility of yield or selectivity, Danheiser notes. 'Most chemists would consider that to be frightening,' he adds, as papers in conventional journals are therefore less likely to be reproducible.

https://www.chemistryworld.com/news/taking-on-chemistrys-reproducibility-problem/3006991.article

Genome-wide associations studies

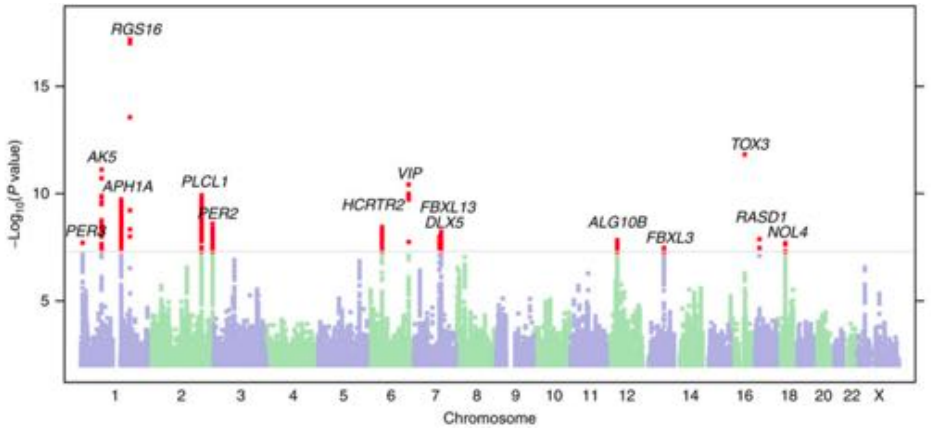


- „prior to 2005, the field was largely a scientific wasteland scattered with the embarrassing and wretched corpses of unreplicated genetic association studies“
  (Daniel MacArthur, 2009)

© KC Green

https://www.wired.com/2009/04/genome-wide-association-studies-failure-or-success/

Genome-wide associations studies



- „prior to 2005, the field was largely a scientific wasteland scattered with the embarrassing and wretched corpses of unreplicated genetic association studies“
  (Daniel MacArthur, 2009)
- But after that …

✓ new statistical standard of evidence: $p < .0000005$

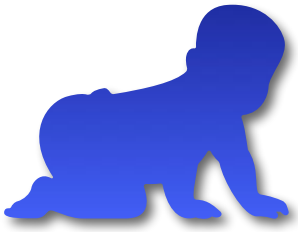
✓ independent replication is standard

✓ all raw data are shared openly (e.g., European Genome-phenome Archive)

https://www.wired.com/2009/04/genome-wide-association-studies-failure-or-success/

https://osf.io/rpw6d/

Developmental Psychology: ManyBabies project



- Developmental psychology has notoriously low sample sizes, low power, and huge publication bias: „We simply cannot get large samples with baby studies!“
- But then …

✓ ManyBabies 1 project (spearheaded by Michael Frank): *Infant Directed Speech Preference*

✓ Decide as a community: What are the most relevant, most pressing, theoretically most important research questions?

✓ Do multi-lab collaborations to achieve this goal

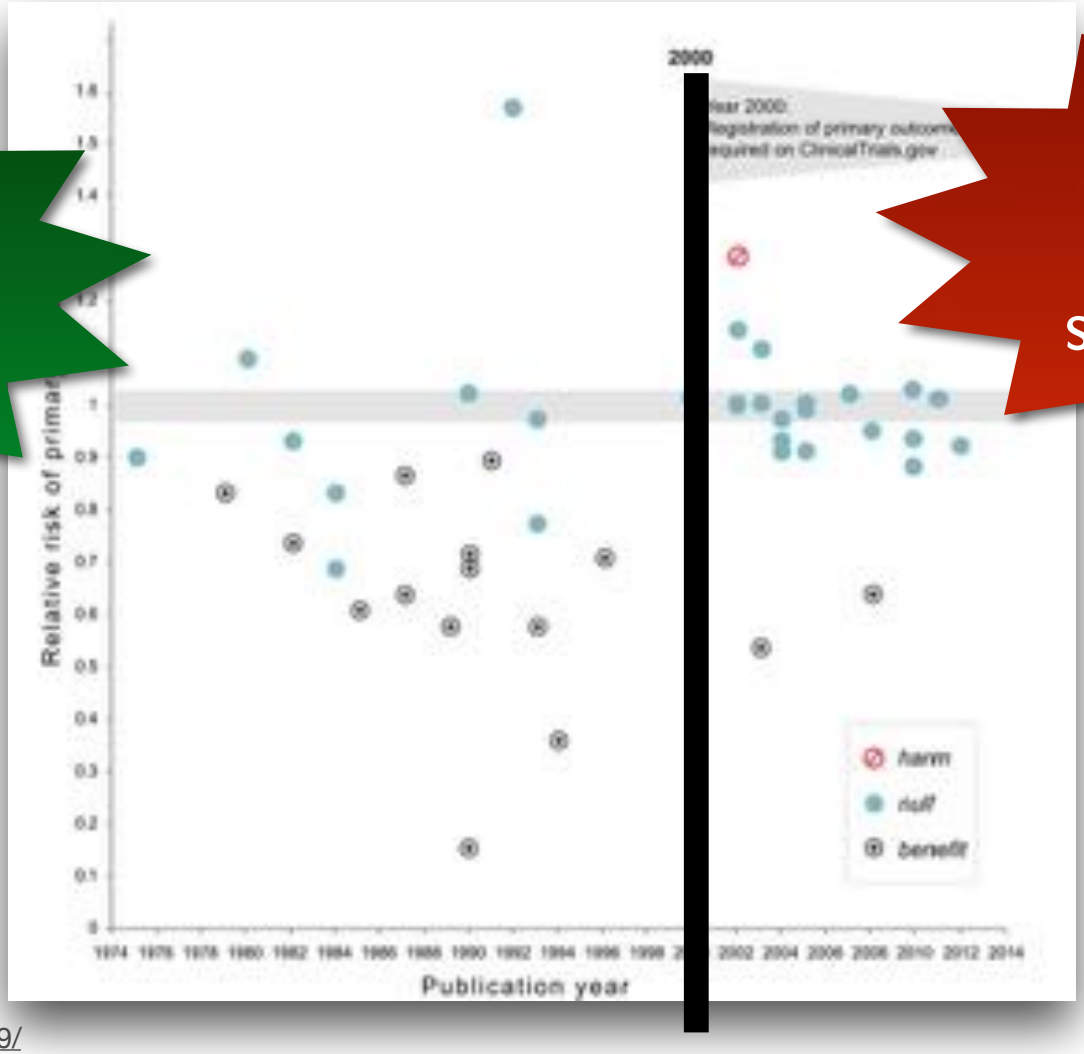
✓ First project: A registered replication report (RRR) with guaranteed publication; preregistered + open data

✓ The next RRR on Theory of Mind is on its way

Pharmacology / Drug development

Drugs or dietary supplements for the treatment or prevention of cardiovascular disease

no prereg:
57%
success rate!



prereg:
8%
success rate...

http://chrisblattman.com/2016/03/01/13719/

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLoS ONE*, *10*(8), e0132382–12. http://doi.org/10.1371/journal.pone.0132382

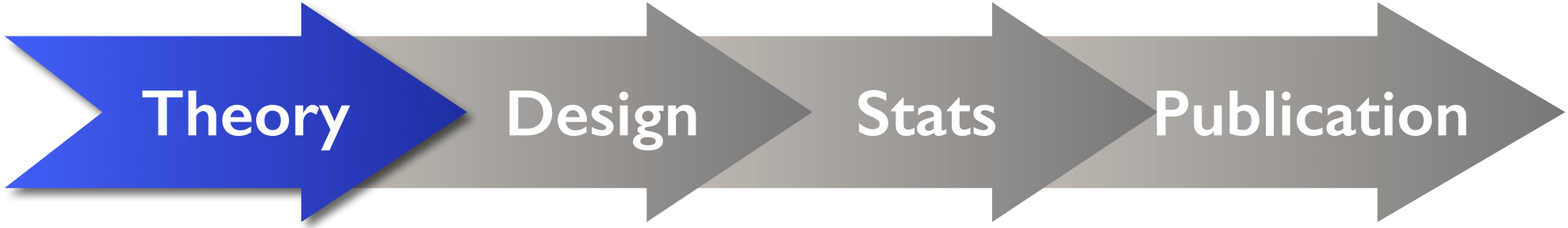
An engineer's perspective

- A project's success is judged by whether you have a working product (i.e., a product that does its intended purpose, reliably and reproducibly)
- Empirical efficacy and accountability: What is the external gold standard that psychology's result are judged against? (Morey, 2016)
- Would you bet 20% of your next grant funding on the existence of your effect?

A vision for a **strong scientific field**

Disclaimers:

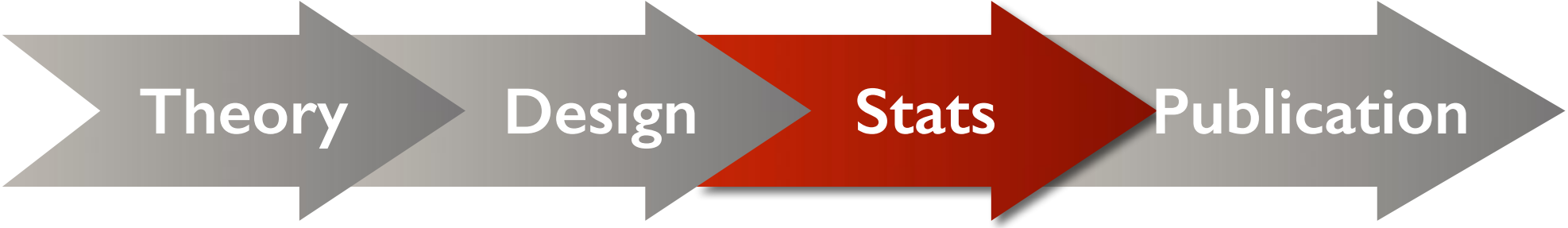
1. For the sake of discussion, the following is deliberately painted in black/white. Reality probably is more nuanced.
2. I do not claim that this is in any way complete. Probably I missed important points; so please see this as a starting point for a discussion.
3. The best practice examples from the other fields do not mean that their whole field is uniformly perfect, and other fields are entirely rotten. They provide examples what can be done.



A *weak* **scientific field …**	A ***strong*** **scientific field …**
has *X* pet theories of *X* researchers; construct inflation, no cumulation	has some consensus about the most pressing and theoretically important research questions, builds cumulatively on them*
Has imprecise and highly flexible theories that are nearly unfalsifiable	strives for falsifiable theories with precise predictions
Original authors use ad hoc adjustments („hidden moderators“) to defend their theories ➔ protective belt/degenerate research program	Original authors define conditions of falsification, and define theoretically plausible moderators *a priori*

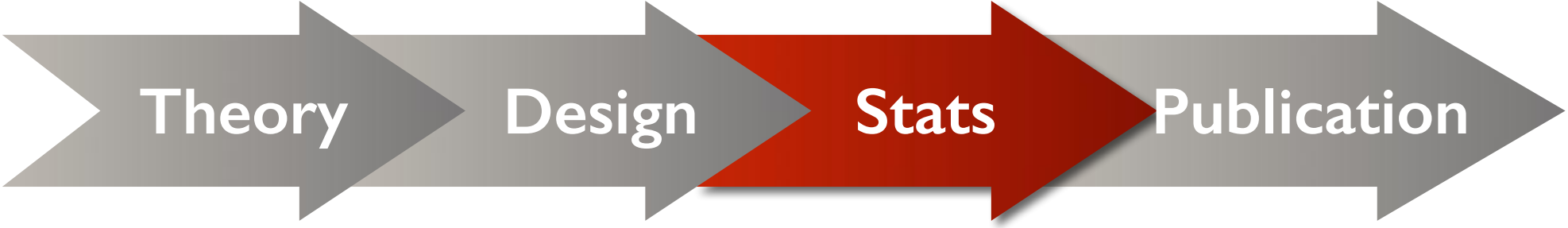
* … but does not suppress creative, off-mainstream, high risk/high gain research (that is clearly labelled as such)

A *weak* **scientific field …**	A ***strong*** **scientific field …**
Does not care about power analyses (or does „post-hoc-a-priori power analyses“ that match the acquired sample size)	Does realistic a priori power analyses, so that „fails“ are also meaningful
Uses unreliable measures and weak designs	Uses reliable measures, within-person designs, and sequential designs to achieve high power
Has a few outstanding „superheros“	Collaborates to achieve this goal
Does no preregistration of confirmatory hypotheses	In confirmatory mode: Preregisters its hypotheses and analysis pipelines to avoid *p*-hacking and HARKing



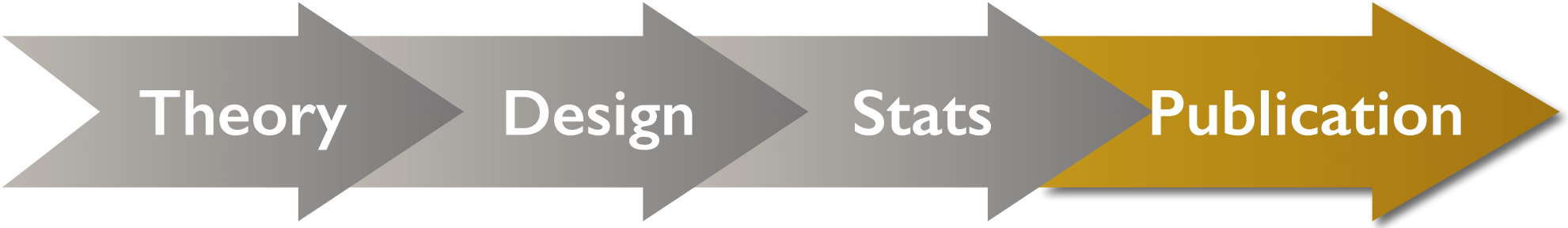
In *exploratory mode*:

A *weak* **scientific field …**	A ***strong*** **scientific field …**
(Mis-)uses statistics built for purely confirmatory research (i.e., *p*-values) to go for fishing expeditions	Uses appropriate statistical methods for exploration (no *p*-values!), such as: • Machine learning • Efficient cross-validation methods for controlling false-positives
By doing so (and failing to correct for multiple testing), it often declares random noise as „a discovery“	By doing so, rarely declares random noise as „a discovery“
Presents exploratory findings as confirmatory	Acknowledges the tentativeness of all exploratory findings



In *confirmatory mode*:

A *weak* **scientific field …**	A ***strong*** **scientific field …**
Uses weak standards of evidence (e.g., $p < .05$) for the declaration of a „discovery“	Uses strong standards of evidence (e.g., BF > 30 or $p < .005$ for the declaration of a „discovery“; e.g., Benjamin et al., 2017)
Cannot quantify evidence for H0; cannot kill zombie theories (Ferguson & Heene, 2012)	Uses tools that can also quantify evidence for „no effect“ by explicitly modeling H1, such as Bayes factors, proper Neyman-Pearson testing, or equivalence testing; e.g., Lakens, 2017.
Uses weak evidence as basis for bold claims	Uses strong evidence as basis for strong inference

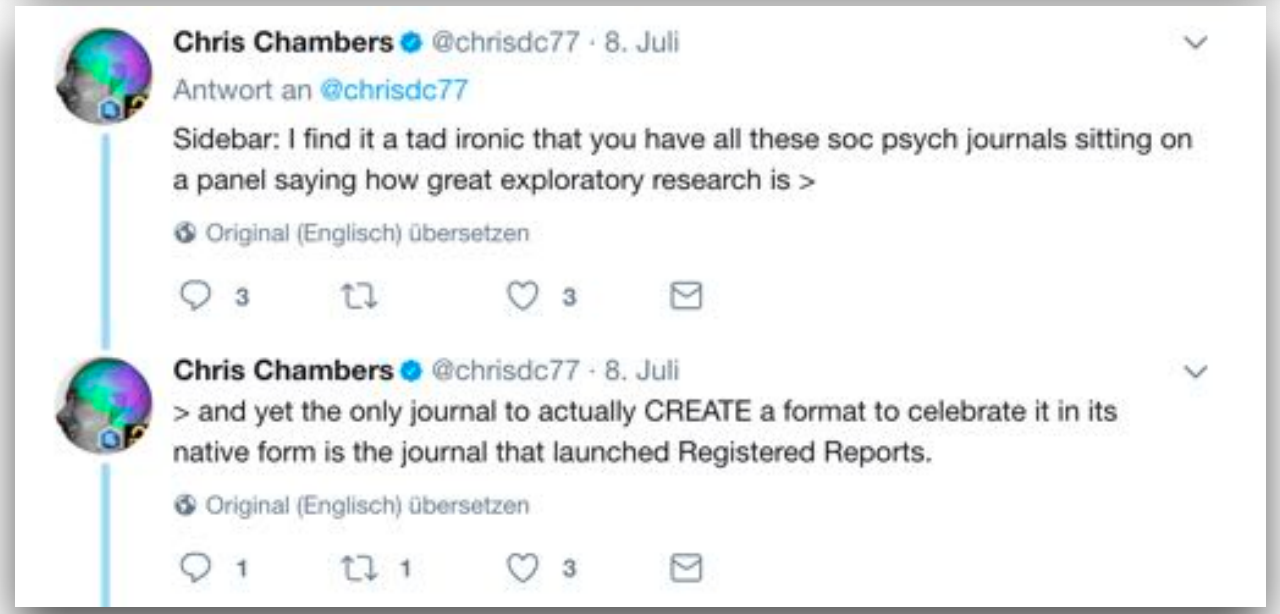


A *weak* **scientific field …**	A ***strong*** **scientific field …**
Presents anything as confirmatory; has no systematic way of distinguishing confirmatory and exploratory research	Has a clear distinction between exploratory and confirmatory results; uses only registered reports for truly confirmatory research
Publishes only „positive“ results	Counter publication bias by publishing registered reports with any outcome
Regards replications as boring	Regards replications as vital
Chooses three or four persons that secretly decide upon what is published	Provides open peer review, embraces post-publication peer review

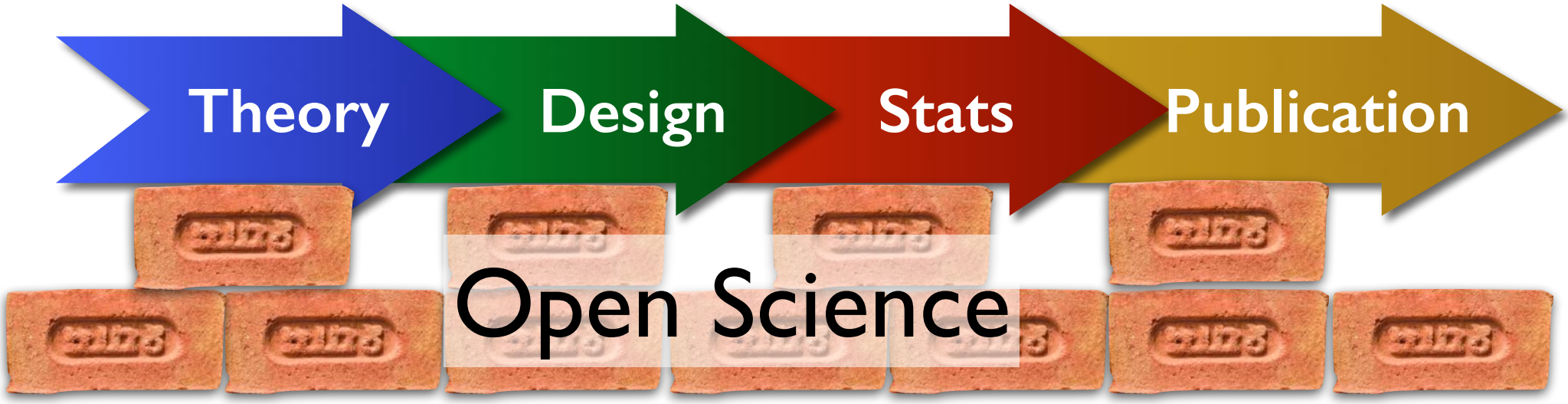
Theory → Design → Stats → **Publication**

- ✓ First journal that publishes only registered reports
- ✓ Review process guarantees adequate power
- ✓ No publication bias

Theory → Design → Stats → **Publication**

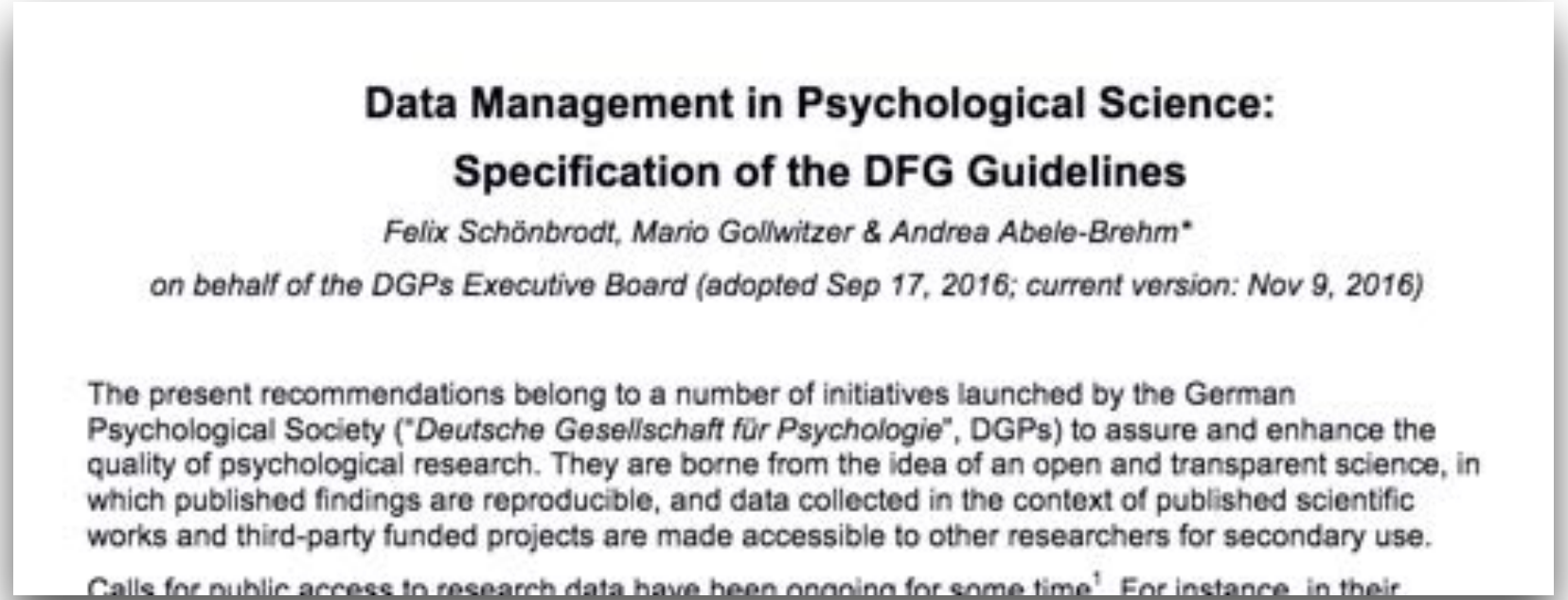


https://twitter.com/chrisdc77/status/883685374140309513



A *weak* **scientific field …**	A ***strong*** **scientific field …**
Hides its data; no verification or reproducibility possible; wastes resources by not reusing existing data	Provides **open primary data** for reproducibility and reuse, without access restrictions (of course respecting privacy and third-party rights)
Precludes direct replications by not providing original material	Provides reusable **open material**; moves away from copyrighted material and develops free instruments
Hides the research results behind paywalls ➙ cf. Ebola crisis	Has **open access** journals with **zero APCs**

German Psychological Society fully embraces open data



Data Management in Psychological Science:

Specification of the DFG Guidelines

*Felix Schönbrodt, Mario Gollwitzer & Andrea Abele-Brehm**

on behalf of the DGPs Executive Board (adopted Sep 17, 2016; current version: Nov 9, 2016)

The present recommendations belong to a number of initiatives launched by the German Psychological Society ("*Deutsche Gesellschaft für Psychologie*", DGPs) to assure and enhance the quality of psychological research. They are borne from the idea of an open and transparent science, in which published findings are reproducible, and data collected in the context of published scientific works and third-party funded projects are made accessible to other researchers for secondary use.

Calls for public access to research data have been ongoing for some time[1]. For instance, in their

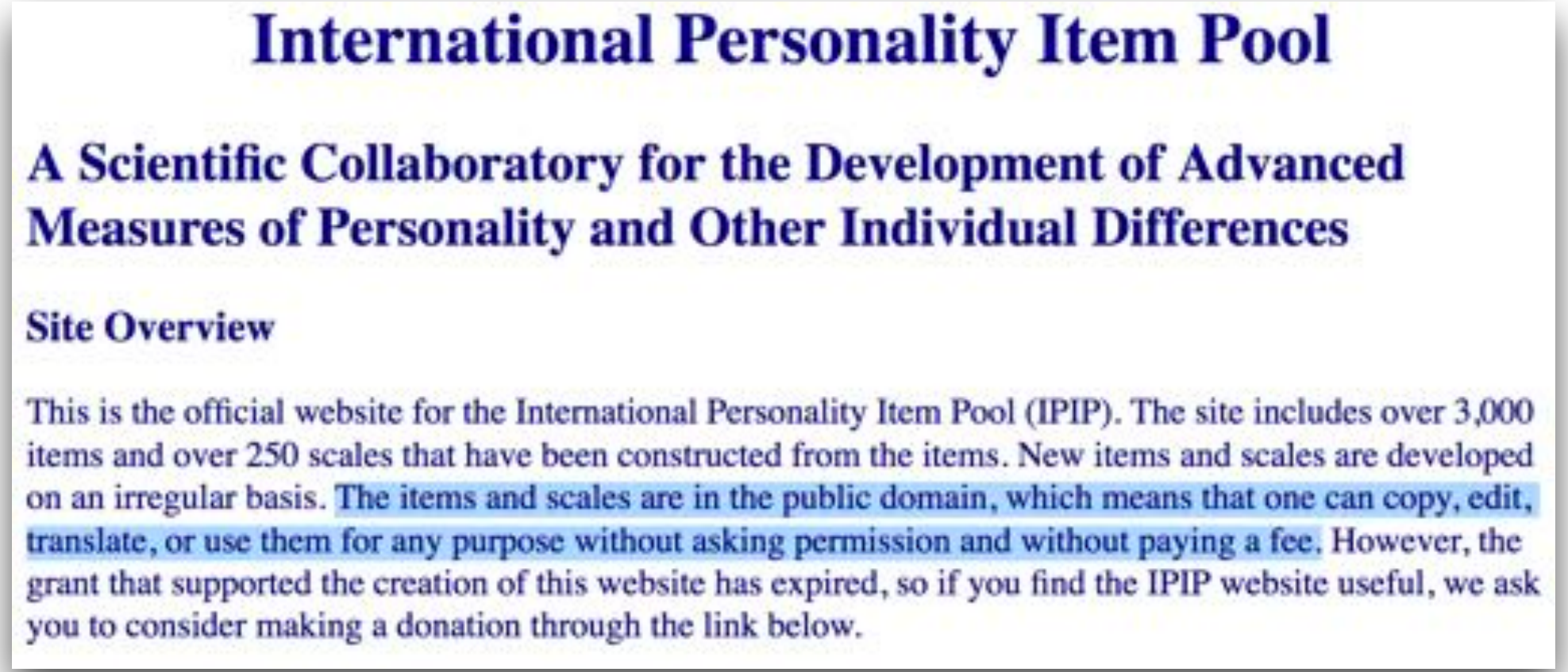
- ✓ Open by default: Primary data are essential part of any empirical publication
- ✓ Exceptions of course possible, but should be justified
- ✓ Recommendations for data reuse
- ✓ Reaching out to APA, APS, SIPS, EAPS, … for discussion and harmonization of guidelines

https://osf.io/preprints/psyarxiv/vhx89

Journals with mandatory open data (or justification why not)

- Advances in Methods and Practices in Psychological Science (http://www.psychologicalscience.org/publications/ampps/ampps-submission-guidelines#DISC)
- Journal of Research in Personality (http://www.sciencedirect.com/science/article/pii/S0092656617300211)
- Judgment and Decision Making (http://journal.sjdm.org/)
- Journal of Cognition (https://www.journalofcognition.org/about/editorialpolicies/)
- Experimental Psychology (http://econtent.hogrefe.com/doi/10.1027/1618-3169/a000355)
- Royal Society Open Science (http://rsos.royalsocietypublishing.org/author-information#Open_data)
- PLOS ONE (http://blogs.plos.org/everyone/2017/05/08/making-progress-toward-open-data/)
- Collabra: Psychology (https://www.collabra.org/about/research-integrity/)

Open, reusable instruments



International Personality Item Pool

A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences

Site Overview

This is the official website for the International Personality Item Pool (IPIP). The site includes over 3,000 items and over 250 scales that have been constructed from the items. New items and scales are developed on an irregular basis. The items and scales are in the public domain, which means that one can copy, edit, translate, or use them for any purpose without asking permission and without paying a fee. However, the grant that supported the creation of this website has expired, so if you find the IPIP website useful, we ask you to consider making a donation through the link below.

http://ipip.ori.org/

- ✓ Gold Open Access
- ✓ Open peer review (optional)
- ✓ mandatory open data & material
- APC: 875 $ (but waivers granted)

- ✓ Gold Open Access
- ✓ No APCs
- ✓ Registered Reports
- ✓ (at least) „strongly encourages“ open data

https://www.rips-irsp.com

Meta-Psychology

Version 1.0

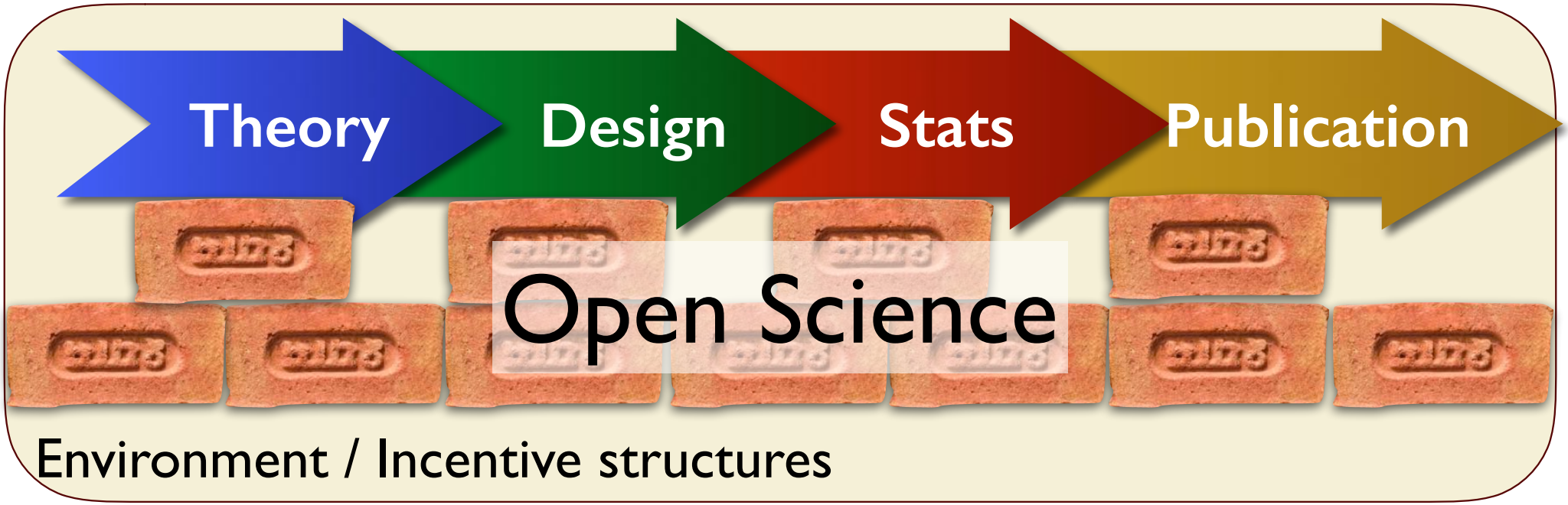
Co-editors-in-Chief
Rickard Carlsson (*Linnaeus University, Sweden*)
Ulrich Schimmack (*University of Toronto, Mississauga, Canada*)

Editorial Board
Åse Innes-Ker (*Lund University, Sweden*)
Marcel van Assen (*Tilburg University, and Utrecht University, The Netherlands*)
Felix Schönbrodt (*Ludwig-Maximilians-University Munich, Germany*)
Moritz Heene (*Ludwig-Maximilians-University Munich, Germany*)
Henrik Danielsson (*Linköping University, Sweden*)
Daniël Lakens (*Eindhoven University of Technology, The Netherlands*)
Yana Weinstein (*University of Massachusetts, Lowell, USA*)

Advisory board
David Meyer (*University of Michigan, Ann Arbor, USA*)
Eiko Fried (*University of Amsterdam, The Netherlands*)
James Coyne (*University of Pennsylvania, USA*)
Nick Brown (*University of Groningen, The Netherlands*)
Stephen Martin (*Baylor University, USA*)
Paul Bürkner (*University of Münster, Germany*)
Donald Williams (*University of California, Davis, USA*)
Julia Rohrer (*University of Leipzig, Germany*)
Malte Elson (*Ruhr University Bochum, Germany*)
Anne Scheel (*Ludwig-Maximilians-University Munich, Germany*)

- ✓ Full open access, no APCs
- ✓ Non-commercial institutional publisher (Linnaeus U library)
- ✓ Open, citable peer review (with doi)
- ✓ Well-powered null results and direct replications welcomed
- ✓ Registered Reports as option
- ✓ Mandatory open data
- ✓ Open Science badges (including a reproducibility badge)

Accepting submissions in Summer 2017



A *weak* **scientific field …**	A ***strong*** **scientific field …**
Bases hiring decisions on quantity and flawed indicators such as the JIF; is ignorant about open practices	Hiring practices explicitly value quality over sheer quantity; recognizes and incentivizes open practices
Favors „perfect“ papers	Tolerates errors, embraces imperfection, values honesty
Evaluates the outcome, not the process	Evaluates the process, not the outcome

An der **Fakultät für Psychologie und Pädagogik** der Ludwig-Maximilians-Universität München ist zum Wintersemester 2016/2017 eine

Professur (W3) für Sozialpsychologie (Lehrstuhl)

. . .

Das Department Psychologie legt Wert auf transparente und replizierbare Forschung und unterstützt diese Ziele durch Open Data, Open Material und Präregistrierungen. Bewerber/innen werden daher gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie diese Ziele bereits verfolgt haben und in Zukunft verfolgen möchten.

+ 3 additional professorship job descriptions

Am Fachbereich 07 - Psychologie und Sportwissenschaft, Institut für Psychologie - der Westfälischen Wilhelms-Universität Münster ist zum nächstmöglichen Zeitpunkt eine

Professur (W 2 BBesO) für Pädagogische Psychologie

zu besetzen.

. . .

Das Fach Psychologie in Münster legt Wert auf transparente und replizierbare Forschung und unterstützt Open Science Praktiken. Bewerber/-innen werden gebeten, in ihren Bewerbungsunterlagen darzulegen, auf welche Art und Weise sie diese Praktiken bereits verfolgt haben und/oder in Zukunft verfolgen möchten.

The **Department of Psychology at the Faculty of Human Sciences of the University of Cologne (UoC) seeks to appoint a**

Full Professor (W3)

of Social Psychology

to be filled as soon as possible.

. . .

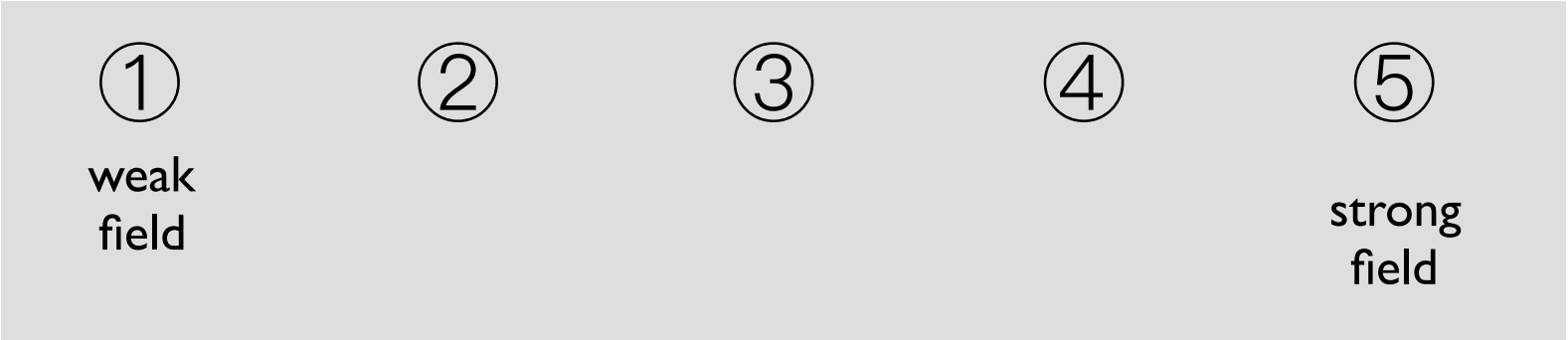
The Department of Psychology aims for transparent and reproducible research (including Open Data, Open Materials, and Preregistrations). Applicants are asked to illustrate how they have pursued these goals in the past and/or how they plan to do so in the future.

Mission Statement
Society for the Improvement of Psychological Science

The Society for the Improvement of Psychological Science (SIPS) brings together scholars working to improve methods and practices in psychological science. Anyone interested in improving psychological research is welcome to join, regardless of experience. SIPS is a service organization aiming to make psychological science higher quality and more cumulative by pursuing activities including:

http://improvingpsych.org/

Summary: My field is a …



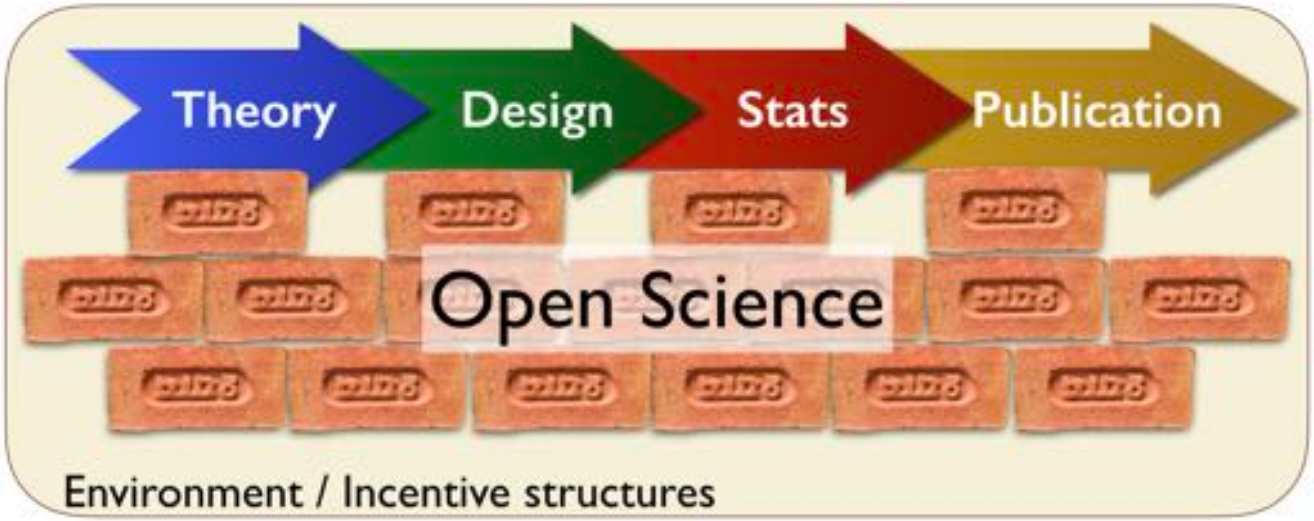
How to build a strong field

- ➡ Accept the necessity to change something
- ➡ Combine bottom-up approaches (e.g., local Open Science initiatives) with top-down changes (e.g., editorial guidelines)
- ➡ Strengthen existing initiatives, get a critical mass of early adopters/ idealists, teach the next generation, be a role model
- ➡ Change the incentive structure

How to build a **strong** scientific field

Lessons from other disciplines and an outline for structural changes

Thanks!



PD Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität
München

CENTER FOR OPEN SCIENCE

RESEARCH TRANSPARENCY

BITSS

www.nicebread.de
www.researchtransparency.org
@nicebread303