UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

iBMLAB
Eye and Brain Mapping Laboratory

The replication crisis:
is it the Ghost of Psychology future?

Junpeng Lao, PhD

Fribourg Day of Cognition
2016/10/05

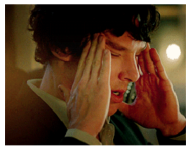
Or: Research in the time of Open Science

Psychology is in crisis!

FEATURE, REPLICATIONS

September 16, 2016

Ten Famous Psychology Findings That It's Been Difficult To Replicate



via giphy

By Christian Jarrett

Every now and again a psychology finding is published that immediately grabs the world's attention and refuses to let go – often it's a result with immediate implications for how we can live more happily and peacefully, or it says something profound about human nature. Said finding then enters the public consciousness, endlessly recycled in pop psychology books and magazine articles.

Replication problem, show headlines
https://digest.bps.org.uk/2016/09/16/ten-famous-psychology-findings-that-its-been-difficult-to-replicate/

Embodiments

- Power posing will make you bolder
- Cleaning your hands will wash away your guilt
- …

Literatures on embodiment that demonstrates surprising links between body and mind raise fast in the past few years (Markman & Brendl, 2005; Proffitt, 2006), unfortunately, many conclusions turn out do not hold.

Harvard psychologist Amy Cuddy and others have published numerous studies that appear to show that our body position can affect our emotional state. One of the reasons this line of research has been so influential is because of Cuddy's TED talk "Your Body Language Shapes Who You Are" which has been viewed many millions of times.

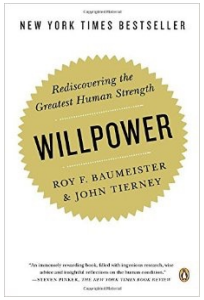
Social priming

- Being reminded of money makes us selfish
- Exposure to words pertaining to ageing will make you walk more slowly
- …

Issues were raised even as early as 2010, which promote Nobel prize-winner Daniel Kahneman to issued a strongly worded open letter to this group of psychologists to restore the credibility of their field by creating a replication ring to check each others' results.
http://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535

Classical results

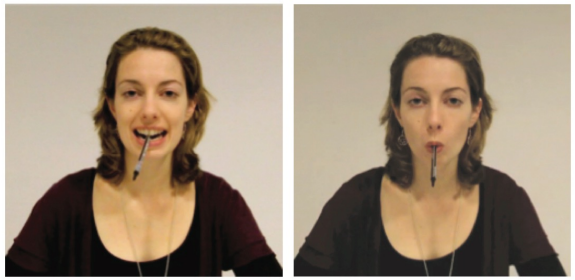
- Self-control is a limited resource



One of the most influential psychological theories of modern times is that willpower is akin to a fuel – the more of it you use in one situation, the less you have left over to deal with other demands.

Classical results

- Smiling will make you feel happier



In 1988, researchers reported that participants found cartoons funnier when they held a pen between their teeth, forcing them to smile, as compared with when they held a pen between their lips, forcing them to pout. The finding appeared to be consistent with the facial-feedback hypothesis – the idea that our facial expression doesn't just reflect our feelings but also affects them – and according to Google Scholar it has been cited nearly 1500 times.

Classical results

- Babies are born with the power to imitate

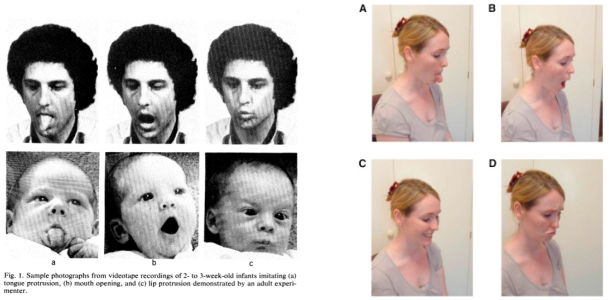
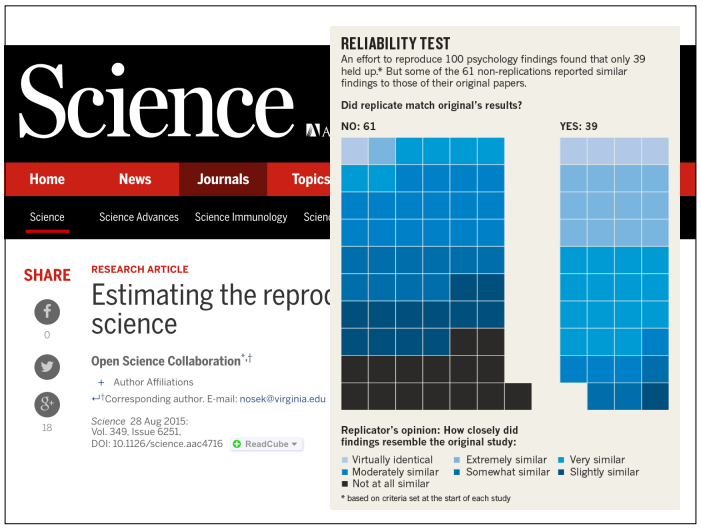


Fig. 1. Sample photographs from videotape recordings of 2- to 3-week-old infants imitating (a) tongue protrusion, (b) mouth opening, and (c) lip protrusion demonstrated by an adult experimenter.

Pick up almost any introductory psychology book and inside you'll read about research conducted in the 1970s that appeared to show that humans are born with the power to imitate.
Earlier this year, however, a methodologically rigorous investigation found no evidence to support the idea that newborn babies can imitate. Janine Oostenbroek and her colleagues tested 106 infants four times between the ages of one week and nine weeks. The researcher performed a range of facial movements, actions or sounds for 60 seconds each including tongue protrusions, mouth opening, happy face, sad face, index finger pointing and mmm and eee sounds. Each baby's behaviour during these 60-second periods was filmed and later coded according to which faces, actions or sounds, if any, he or she performed during the different researcher displays.



Estimating the reproducibility of psychological science
Open Science Collaboration. Vol. 349, Issue 6251, DOI: 10.1126/science.aac4716

- Empirically analyzing empirical evidence
One of the central goals in any scientific endeavor is to understand causality. Experiments that seek to demonstrate a cause/effect relation most often manipulate the postulated causal factor. Aarts et al. describe the replication of 100 experiments reported in papers published in 2008 in three high-ranking psychology journals. Assessing whether the replication and the original experiment yielded the same result according to several criteria, they find that about one-third to one-half of the original findings were also observed in the replication study.

the Ghost of Psychology future

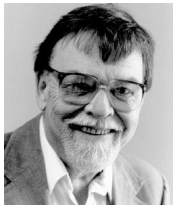
*as seen in A Christmas Carol (2009)

Replication crisis, the back story

- 2010 - Open letter from Daniel Kahneman to Psychologist investigating social priming
- 2011 - Daryl Bem publishes his article supporting Extrasensory perception on JPSP
- 2011 - “p-hacking” introduced by Joseph Simmons, Leif Nelson, and Uri Simonsohn. “the garden of forking paths” introduced by Eric Loken and Andrew Gelman
- 2013 - 36 research groups formed the Many Labs Replication Project to repeat 13 psychological studies
- 2015 - Large replication study by the Open Science Collaboration published on Science

Everybody had tried to replicated some studies and sometimes fail. But how do we get to this point: a full-blown crisis?
http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/

Replication crisis, the problem



Paul E. Meehl
(1920 - 2003)

"…a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing a long series of related experiments which appear to the uncritical reader as a fine example of 'an integrated research program,' without ever once refuting or corroborating so much as a single strand of the network."

1960s-1970s: Paul Meehl argues that the standard paradigm of experimental psychology doesn't work, that "a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing a long series of related experiments which appear to the uncritical reader as a fine example of 'an integrated research program,' without ever once refuting or corroborating so much as a single strand of the network."
http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/

Replication crisis, the problem

"Standard statistical practice is to take your data and work with it until you get a p-value of less than .05. Run a few experiments like that, attach them to a vaguely plausible (or even, in many cases, implausible) theory, and you got yourself a publication. Give it a bit more of a story and you might get yourself on Ted, NPR, Gladwell, and so forth."

- Andrew Gelman

"As most of you are aware…, there is a statistical crisis in science, most notably in social psychology research but also in other fields. For the past several years, top journals such as JPSP, Psych Science, and PPNAS have published lots of papers that have made strong claims based on weak evidence."

“p-hacking” or “researcher degrees of freedom”

“Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant”

http://projects.fivethirtyeight.com/p-hacking/

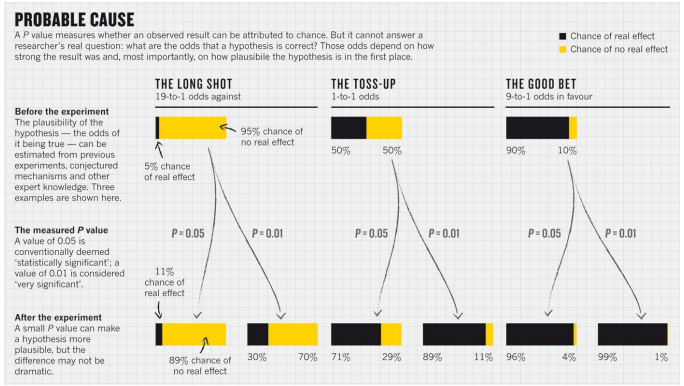
(Simmons, Nelson, and Simonsohn, 2011)

“The garden of forking paths”

“Researcher degrees of freedom can lead to a multiple comparisons problem, even in settings where researchers perform only a single analysis on their data. The problem is there can be a large number of *potential comparisons* when the details of data analysis are highly contingent on data, without the researcher having to perform any conscious procedure of fishing or examining multiple p-values.”

(Gelman & Loken, 2013)

Statistical problems



http://www.nature.com/news/scientific-method-statistical-errors-1.14700
Also, The American Statistical Association released a committee report on the use of p-values
https://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf

---

Statistical problems

- Statistical assumptions (e.g., normal distribution)
- influence of outliners (non-robust statistics)
- Other statistical problem (e.g., effect size, statistical power)

Other problems

- File drawer problem
- Publication bias

Solutions:

Easy  Hard

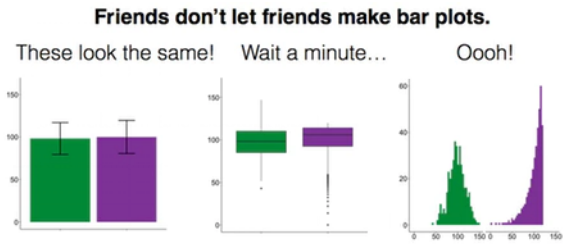
Easy solutions:

- Report more stats (effect sizes and confidence intervals)
- Increase sample size

The minimal you should do.
e.g., Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.

Improve your statistics

- Better data representation
  - no more Barplot #barbarplot



https://en.wikipedia.org/wiki/Replication_crisis#Addressing_the_replication_crisis

Improve your statistics

- Better data representation
  - no more Barplot #barbarplot
  - use histogram
- Non-parametric statistics
  - permutation and bootstrapping
- Robust statistics
  - other robust estimators instead of e.g., mean

https://en.wikipedia.org/wiki/Replication_crisis#Addressing_the_replication_crisis

---

Disclosure-based solution:

Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.

Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.

Authors must list all variables collected in a study.

Authors must report all experimental conditions, including failed manipulations.

If observations are eliminated, authors must also report what the statistical results are if those observations are included.

If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

(Simmons, Nelson, and Simonsohn, 2011)

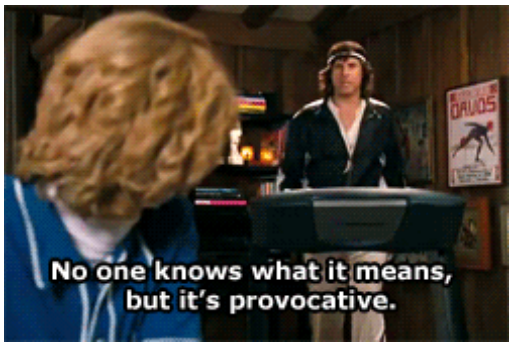
full disclosure

Pre-registration

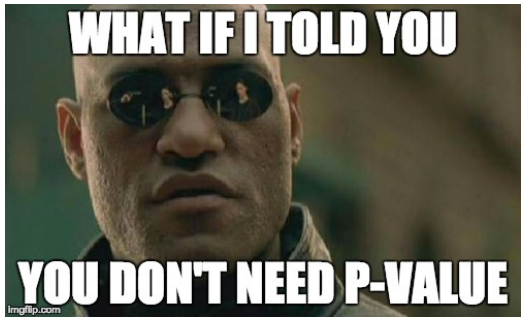
The registered report format requires authors to submit a description of the study methods and analyses prior to data collection. Once the method and analysis plan is vetted through peer-review, publication of the findings is provisionally guaranteed, based on whether the authors follow the proposed protocol.

https://en.wikipedia.org/wiki/Replication_crisis#Addressing_the_replication_crisis
http://www.bayesianphilosophy.com/preregistration/

Advance solutions



No more p-value



http://andrewgelman.com/2016/03/07/29212/

Valid p-values cannot be drawn without knowing, not just what was done with the existing data, but what the choices in data coding, exclusion, and analysis would have been, had the data been different. This ‘what would have been done under other possible datasets’ is central to the definition of p-value.

Let go of null hypothesis significance testing (NHST)

http://andrewgelman.com/2016/02/04/the-notorious-n-h-s-t-presents-mo-p-values-mo-problems/

“NHST is all about rejecting straw-man hypothesis B and then using this to claim support for the researcher’s desired hypothesis A. The trouble is that both models are false, and typically the desired hypothesis A is not even clearly specified.”

http://andrewgelman.com/2016/09/10/my-talk-at-warwick-england-230pm-thurs-15-sept/

Ultimately the problem is not with p-values but with null-hypothesis significance testing, that parody of falsificationism in which straw-man null hypothesis A is rejected and this is taken as evidence in favor of preferred alternative B (see Gelman, 2014). Whenever this sort of reasoning is being done, the problems discussed above will arise. Confidence intervals, credible intervals, Bayes factors, cross-validation: you name the method, it can and will be twisted, even if inadvertently, to create the appearance of strong evidence where none exists.

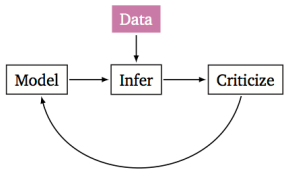
- Instead of p-value:
  - model estimation/coefficient
  - effect size
  - prediction power and cross-validation
- Instead of NHST
  - Multilevel (Hierarchical) Modeling
  - informative Bayesian inference

http://andrewgelman.com/2016/09/10/my-talk-at-warwick-england-230pm-thurs-15-sept/

Be Bayesian

First gather data from some real-world phenomena. Then cycle through Box's loop (Blei, 2014).

1. Build a probabilistic model of the phenomena.
2. Reason about the phenomena given model and data.
3. Criticize the model, revise and repeat.

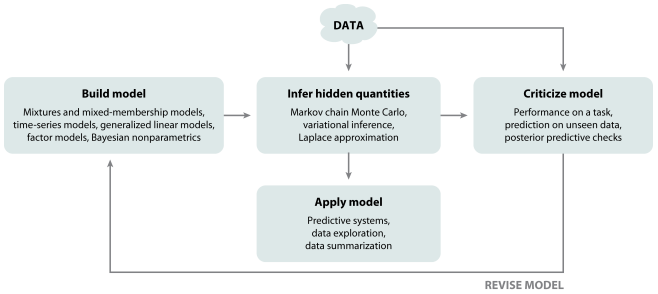


Avoid paradox like the one you see in Frequensitic statistic

Blei, D. M. (2014). Build, compute, critique, repeat: Data analysis with latent variable models. Annual Review of Statistics and Its Application, 1, 203–232.

Box, G. E. (1976). Science and statistics. Journal of the American Statistical Association, 71(356), 791–799.

Better model



Blei DM. 2014.
Annu. Rev. Stat. Appl. 1:203–32

Avoid paradox like the one you see in Frequensitic statistic

Blei, D. M. (2014). Build, compute, critique, repeat: Data analysis with latent variable models. Annual Review of Statistics and Its Application, 1, 203–232.

Box, G. E. (1976). Science and statistics. Journal of the American Statistical Association, 71(356), 791–799.

Pipelines

- Parametrically learn more about the human side of "hyper parameters"
  - what kind of smoothing to use in fMRI
  - mean? trim-mean? median? or other rules to exclude outliners?
  - Normalisation? how is z-scored effecting the result

Why not Deep learning?

apply industrial standard end-to-end inference
Using Deep Learning module such as Tensorflow or Theano to build large (linear) model with multiple layers. Each layer will represent a pre-processing/processing step (such as taking the mean for each subject, etc). The model will then evaluated either using Bayesian statistics or cross-validation.

How to go forward

- Learn to write codes
- Learn Bayesian statistics
- Be open - share your data and code

1 write code, break out your conform zoom of clicking button (e.g., SPSS)
2 once you learn bayesian, many hard thing became easy
3 open source your data and code

open science

Science Code Manifesto

Manifesto Discussion Endorse Resources About

Software is a cornerstone of science. Without software, twenty-first century science would be impossible. Without better software, science cannot progress.

But the culture and institutions of science have not yet adjusted to this reality. We need to reform them to address this challenge, by adopting these five principles:

Code	All source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.
Copyright	The copyright ownership and license of any released source code must be clearly stated.
Citation	Researchers who use or adapt science source code in their research must credit the code's creators in resulting publications.
Credit	Software contributions must be included in systems of scientific assessment, credit, and recognition.
Curation	Source code must remain available, linked to related materials, for the useful lifetime of the publication.

support open source http://sciencecodemanifesto.org/

Conclusion:

- There is a Replication crisis in Psychology.
- Science is about prediction. Being able to replicate is the minimal.
- We need to change the way we do research, the way we perform statistical analysis, and the way we reason/inference our result

Further information

- http://andrewgelman.com/
- http://datacolada.org/
- http://blogs.discovermagazine.com/neuroskeptic/

Be part of the future!

QUESTIONS?

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

iBMLAB
Eye and Brain Mapping Laboratory