Why and how to replicate research? On a confidence crisis in life and social sciences

Dr. Mateusz Hohol

Copernicus Center for Interdisciplinary Studies
Jagiellonian University

Section of Cognitive Science, Institute of Philosophy and Sociology, Polish Academy of Sciences, Warsaw

IFiS PAN

JAGIELLONIAN UNIVERSITY IN KRAKÓW

Copernicus Center

Journal of Computational Neuroscience (2018) 45:163–172
https://doi.org/10.1007/s10827-018-0702-z

Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail

Marcin Miłkowski[1] · Witold M. Hensel[2] · Mateusz Hohol[3]

Received: 28 April 2018 / Revised: 5 October 2018 / Accepted: 17 October 2018 / Published online: 31 October 2018

Abstract

Replicability and reproducibility of computational models has been somewhat understudied by "the replication movement." In this paper, we draw on methodological studies into the replicability of psychological experiments and on the mechanistic account of explanation to analyze the functions of model replications and model reproductions in computational neuroscience. We contend that model replicability, or independent researchers' ability to obtain the same output using original code and data, and model reproducibility, or independent researchers' ability to recreate a model without original code, serve different functions and fail for different reasons. This means that measures designed to improve model replicability may not enhance (and, in some cases, may actually damage) model reproducibility. We claim that although both are undesirable, low model reproducibility poses more of a threat to long-term scientific progress than low model replicability. In our opinion, low model reproducibility stems mostly from authors' omitting to provide crucial information in scientific papers and we stress that sharing all computer code and data is not a solution. Reports of computational studies should remain selective and include all and only relevant bits of code.

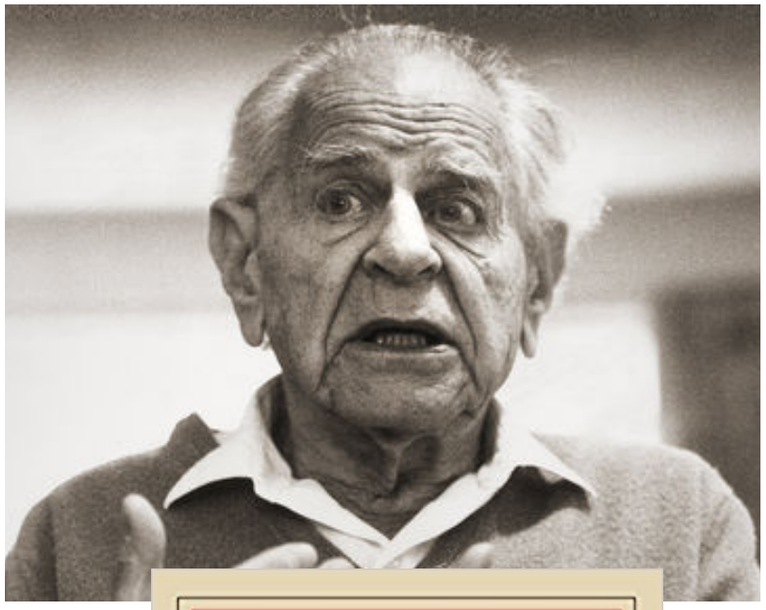
Keywords Replication studies · Computational modeling · Methodology of computational neuroscience · Direct and conceptual replication · Replication and reproduction

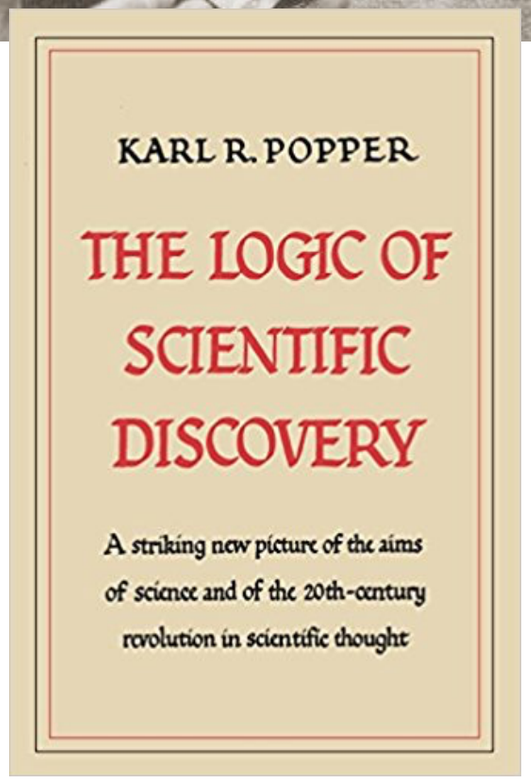
Dr. Witold Hensel (UwB)

Prof. Marcin Miłkowski (IFiS PAN)

Why is it so important?

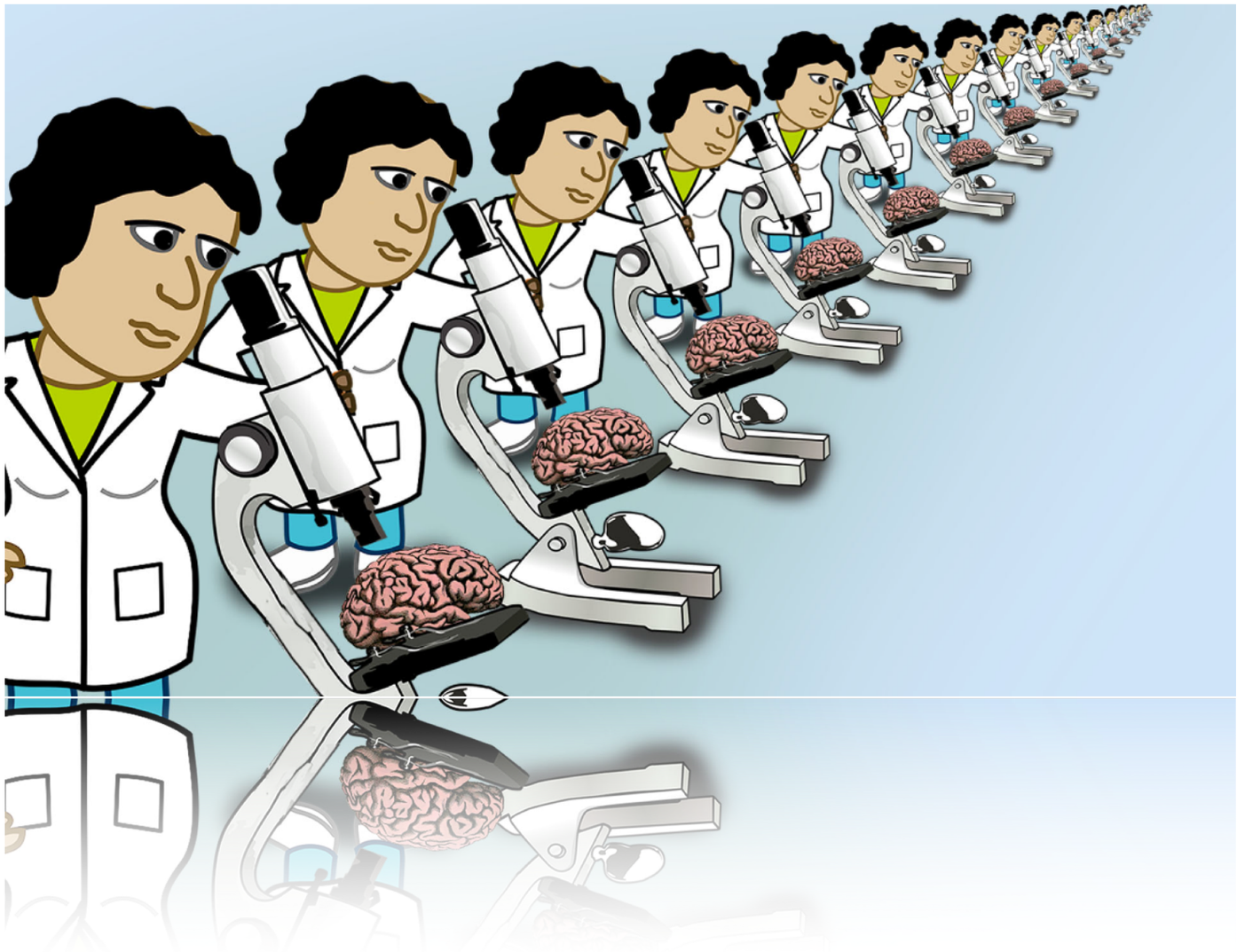
“Any controversy over the question whether events which are in principle unrepeatable and unique ever do occur cannot be decided by science: it would be a metaphysical controversy” (Popper, 1959)





Brain & Cognition Studies

- Experimental psychology
  - social psychology
  - cognitive psychology
- Neuroscience (neurobiology)
  - empirical (patch clamps, neuroimaging)
  - computational



Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

PLOS ONE

Publish | About | Browse

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect

Stuart J. Ritchie, Richard Wiseman, Christopher C. French

Published: March 14, 2012 • https://doi.org/10.1371/journal.pone.0033423

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Perspectives on Psychological Science
2016, Vol. 11(4) 546–573

DOI: 10.1177/1745691616652873
pps.sagepub.com
SAGE

A Multilab Preregistered Replication of the Ego-Depletion Effect

M. S. Hagger,* N. L. D. Chatzisarantis,* H. Alberts, C. O. Anggono, C. Batailler, A. R. Birt, R. Brand, M. J. Brandt, G. Brewer, S. Bruyneel, D. P. Calvillo, W. K. Campbell, P. R. Cannon, M. Carlucci, N. P. Carruth, T. Cheung, A. Crowell, D. T. D. De Ridder, S. Dewitte, M. Elson, J. R. Evans, B. A. Fay, B. M. Fennis, A. Finley, Z. Francis, E. Heise, H. Hoemann, M. Inzlicht, S. L. Koole, L. Koppel, F. Kroese, F. Lange, K. Lau, B. P. Lynch, C. Martijn, H. Merckelbach, N. V. Mills, A. Michirev, A. Miyake, A. E. Mosser, M. Muise, D. Muller, M. Muzi, D. Nalis, R. Nurwanti, H. Otgaar, M. C. Philipp, P. Primoceri, K. Rentzsch, L. Ringos, C. Schlinkert, B. J. Schmeichel, S. F. Schoch, M. Schrama, A. Schütz, A. Stamos, G. Tinghög, J. Ullrich, M. vanDellen, S. Wimbarti, W. Wolff, C. Yusainy, O. Zerhouni, and M. Zwienenberg

Behavioral and Brain Sciences

Making replication mainstream

Rolf A. Zwaan [a1], Alexander Etz [a2], Richard E. Lucas [a3] and M. Brent Donnellan [a4]

https://doi.org/10.1017/S0140525X17001972 Published online: 25 October 2017

Related commentaries (36) Author response

- Context is too variable
- The theoretical value of direct replications is limited
- Direct replications are not feasible in certain domains
- Replications are a distraction
- Replications affect reputations
- There is no standard method to evaluate replication results

Journal of Personality and Social Psychology
1988, Vol. 54, No. 5, 768–777

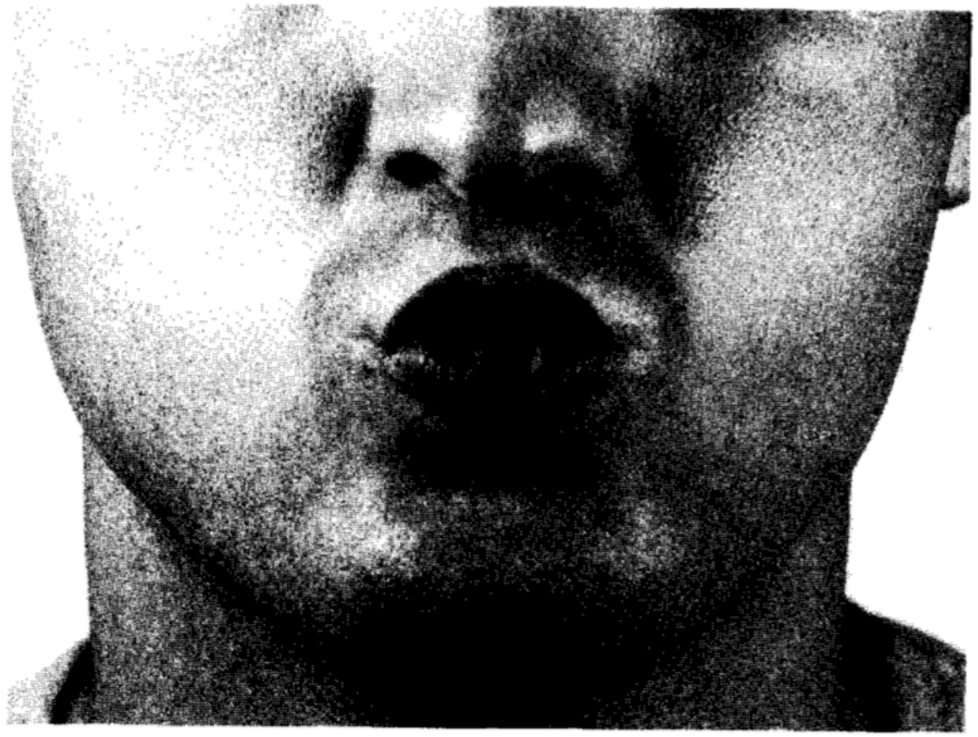
Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis

Fritz Strack
Universität Mannheim
Mannheim, Federal Republic of Germany

Leonard L. Martin
University of Illinois
at Urbana-Champaign

Sabine Stepper
Universität Mannheim
Mannheim, Federal Republic of Germany

We investigated the hypothesis that people's facial activity influences their affective responses. Two studies were designed to both eliminate methodological problems of earlier experiments and clarify theoretical ambiguities. This was achieved by having subjects hold a pen in their mouth in ways that either inhibited or facilitated the muscles typically associated with smiling without requiring subjects to pose in a smiling face. Study 1's results demonstrated the effectiveness of the procedure. Subjects reported more intense humor responses when cartoons were presented under facilitating conditions than under inhibiting conditions that precluded labeling of the facial expression in emotion categories. Study 2 served to further validate the methodology and to answer additional theoretical questions. The results replicated Study 1's findings and also showed that facial feedback operates on the affective but not on the cognitive component of the humor response. Finally, the results suggested that both inhibitory and facilitatory mechanisms may have contributed to the observed affective responses.



aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2016, Vol. 11(6) 917–928

DOI: 10.1177/1745691616674458
pps.sagepub.com
SAGE

Registered Replication Report: Strack, Martin, & Stepper (1988)

E.-J. Wagenmakers*, T. Beek*, L. Dijkhoff*, Q. F. Gronau,*
A. Acosta, R. B. Adams, Jr., D. N. Albohn, E. S. Allard, S. D. Benning,
E.-M. Blouin-Hudon, L. C. Bulnes, T. L. Caldwell, R. J. Calin-Jageman,
C. A. Capaldi, N. S. Carfagno, K. T. Chasten, A. Cleeremans, L. Connell,
J. M. DeCicco, K. Dijksterhuis, A. H. Fischer, F. Foroni, U. Hess, K. J. Holmes,
J. L. H. Jones, O. Klein, C. Koch, S. Korb, P. Lewinski, J. D. Liao, S. Lund,
J. Lupianez, D. Lynott, C. N. Macrae, K. P. Marcum, A. Ozdoy,
A. P. Pacheco-Unguetti, B. Pearson, C. Powis, S. Riding, T.-A. Roberts,
R. I. Rumiati, M. Senden, N. Shnabel, M. Sias, S. Sipes, N. Smith, J. Soto,
T. G. Steiner, J. M. Talarico, Z. M. van Allen, M. Vandekerckhove,
B. Wainwright, J. F. Wayand, R. Zeelenberg, E. E. Zetzer, and R. A. Zwaan
*Proposing authors

If direct replication, maybe there is also undirect?

Protocol vetted by: Ursula Hess

Protocol edited by: Daniel J. Simons

Multilab direct replication of: Study 1 from Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54*, 768-777.

Data and registered protocols: https://osf.io/pkd65/

Citation: Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Jr., . . . Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science, 11*, 917–928.

Why to replicate?
(Schmidt, 2009)

- In psychology (social sciences) replications facilitate:
  - corroboration of hypotheses
  - detection *false positives* (type I),
  - generalization of results
  - justification of existing hypotheses by new experimental setups (procedures)
  - revealing mistakes (or even fraud, cf. Diederik Stapel, Tilburg)

- One type of replication research cannot perform all the above functions
- <u>There are several kinds of replications</u>

The basic distinction
(e.g. Zwaan et al. 2018)

- *Direct replication*: aims to <u>recreate</u> an original study (its samples, measures, procedures, etc.) according to the current understanding of what is needed to produce the phenomenon under investigation.

- *Conceptual replication*): it <u>deliberately modify</u> the critical elements of an original procedure in order to test the robustness of a phenomenon or the generality of a theoretical claim. The main purpose of a conceptual replication is to investigate the target theory or hypothesis in a novel way.

Types of direct replication

- ***<u>Exact replication</u>*** – direct replications performed by the same group of researchers, is to protect the scientific community against false positives, which are likely to occur when the first study is statistically underpowered.
    - They are analogous to repeat measurements in physics, chemistry and medicine.
    - They are strongly recommended when initial findings are either unexpected or loosely based on current theoretical models.
- ***<u>Close replication</u>*** – direct replications performed by an independent team of researchers, also reduce the likelihood of false positives, especially those stemming from experimenter effects and tacit knowledge.
    - They provide information needed to establish the size of an effect, which the original investigators are prone to overestimate.
    - They s enable the research community not only to confirm the existence of an effect but also to disconfirm it.

Is all psychology in crisis?

NO !!!

- social psychology: 25%
- cognitive psychology: 50%

Center for Open Science in Charlottesville, Virginia
Brian Nosek and 269 collaborators

„Hard” psychological/cognitive science research carried out using standardized behavioral procedures

(USUALLY) REPLICATE!

What can we do to facilitate replicability?

- Involvement of authors of original research in replication studies
- Preregistration of experimental procedures (or even preregistred articles)
- Posting raw data in publicly available repositories (Open Science Framework, OSF)
- More emphasize of the describing of original experimental procedures in papers

A large-scale survey on finger counting routines, their temporal stability and flexibility in educated adults

Mateusz Hohol[1,2], Kinga Wołoszyn[3], Hans-Christoph Nuerk[4,5,6] and Krzysztof Cipora[4,5]

[1] Copernicus Center for Interdisciplinary Studies, Jagiellonian University, Cracow, Poland
[2] Section of Cognitive Science, Institute of Philosophy and Sociology, Polish Academy of Science, Warsaw, Poland
[3] Psychophysiology Laboratory, Institute of Psychology, Jagiellonian University, Cracow, Poland
[4] Department of Psychology, University of Tuebingen, Tuebingen, Germany
[5] LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany
[6] Leibnitz-Institut für Wissenmedien, Tuebingen, Germany

ABSTRACT

A strong link between bodily activity and number processing has been established in recent years. Although numerous observations indicate that adults use finger counting (FC) in various contexts of everyday life for different purposes, existing knowledge of FC routines and their use is still limited. In particular, it remains unknown how stable the (default) FC habits are over time and how flexible they can be. To investigate these questions, 380 Polish participants completed a questionnaire on their FC routines, the stability of these routines, and the context of FC usage, preceded by the request to count on their fingers from 1 to 10. Next, the test–retest stability of FC habits was examined in 84 participants 2 months following the first session. To the best of our knowledge, such a study design has been adopted for the first time. The results indicate that default FC routines of the majority of participants (75%) are relatively stable over time. At the same time, FC routines can flexibly adapt according to the situation (e.g., when holding an object). As regards prevalence, almost all participants, in line with previous findings on Western individuals, declared starting from the closed palm and extending consecutive fingers. Furthermore, we observed relations between FC preferences and handedness (more left-handers start from the left hand) and that actual finger use is still widespread in healthy adults for a variety of activities (the most prevalent uses of FC are listing elements, presenting arguments and plans, and calendar calculations). In sum, the results show the practical relevance of FC in adulthood, the relative stability of preferences over time along with flexible adaptation to a current situation, as well as an association of FC routines with handedness. Taken together our results suggest that FC is the phenomenon, which is moderated or mediated by multiple embodied factors.

The Reviewer:

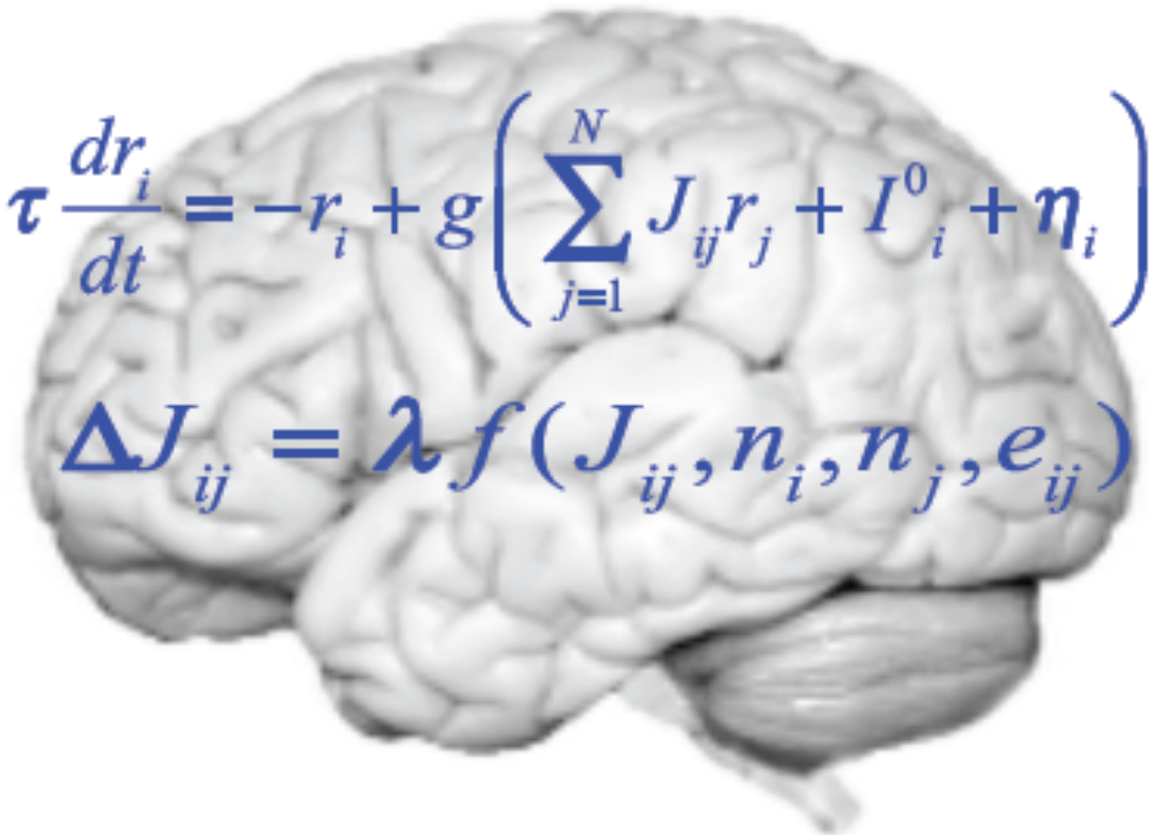
Experimental design

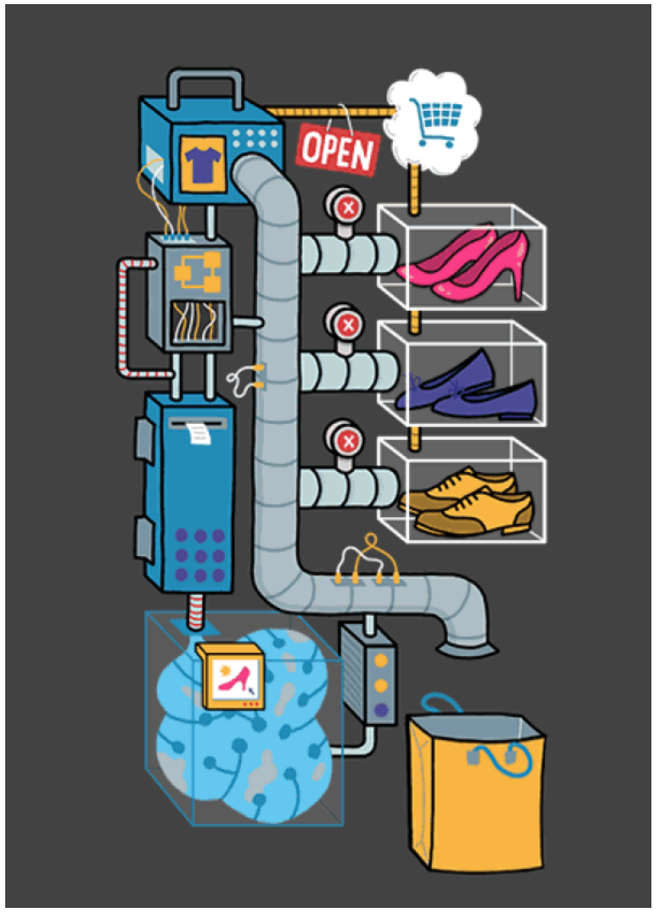
The study is well designed and expertly conducted.

The analyses are sound. The findings are clear cut.

All details are report to replicate the study

Replications & Reproductions in computational neuroscience





Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail

Marcin Miłkowski[1] · Witold M. Hensel[2] · Mateusz Hohol[3]

Journal of Computational Neuroscience (2018) 45:163–172
https://doi.org/10.1007/s10827-018-0702-z

The state of the art

- A number of authors in the computational science community are drawing attention to problems with model replicability and reproducibility (Hutson 2018, Peng 2011, Rougier et al. 2017, Sandve et al. 2013).

- Model reproduction is rarely performed (Legéndi et al. 2013) because successful reproductions do not seem to deliver novel scientific results and causes of failed reproduction may be difficult to discern.

- Instead of reproducing a model with new data, researchers tend to compare new models with previous work.

- No major journal accepts publications related to computational replications or reproductions.

- The journal *ReScience* (https://rescience.github.io), which aims to fill this gap, was established in 2015 but has as few as 22 papers as of August 2018

A few terminology (again:)

- **Repeatability of the model:** The researcher can repeat his own computations by herself
- **Replicability of the mode:** independent researchers, original software (and input data)
- **Reproducability of the model:** independent researchers, their own software (computational implementation)

Association for Computing Machinery; por. Drummond (2009), Crook et al. (2013), Plesser (2018)

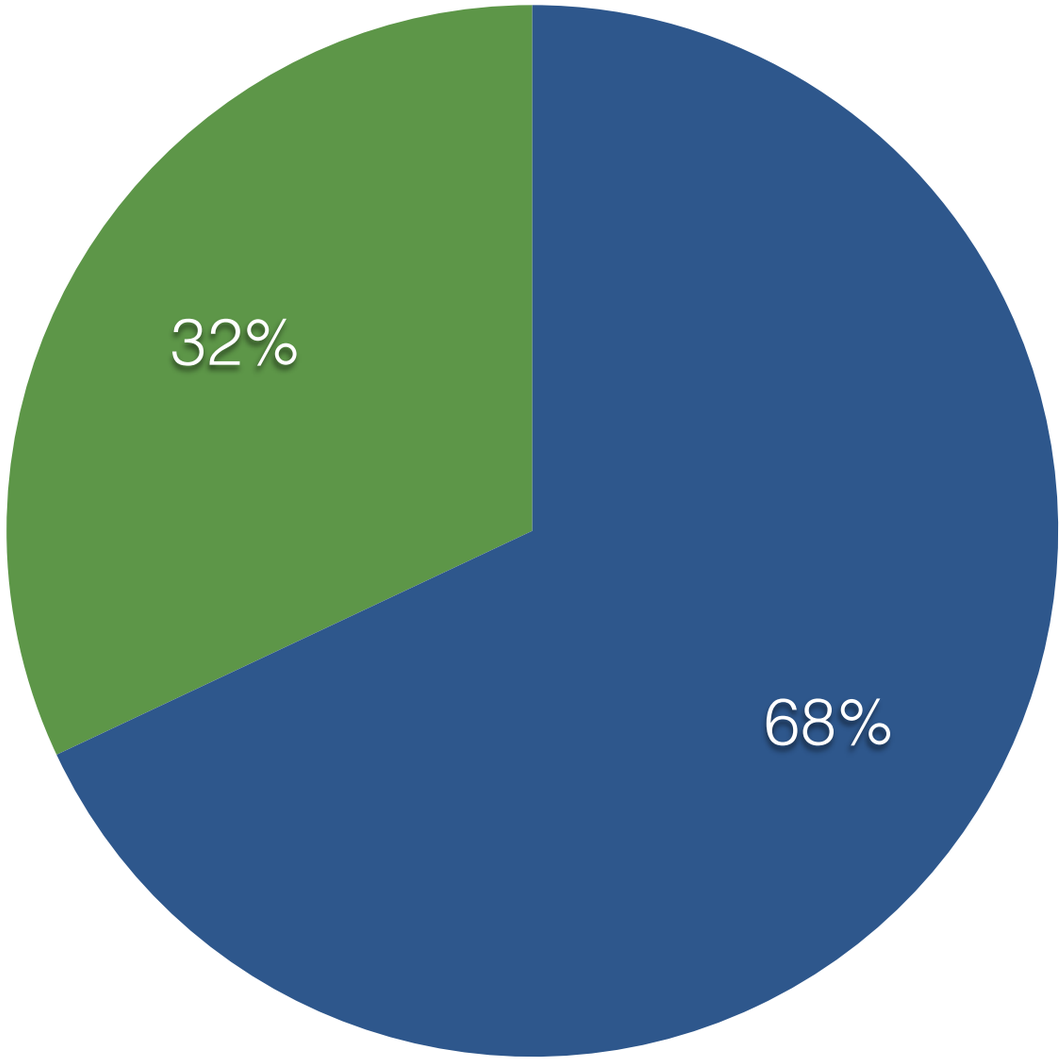
Replicability

Missing data hinder replication of artificial intelligence studies

By **Matthew Hutson** | Feb. 15, 2018 , 12:30 PM

- All code should be shared (Sandve et al. 201).
- This call is largely ignored as only 6% of the 400 algorithms presented at two top AI conferences in the past few years contained the code and only a third had pseudocode, or simplified summaries of the code (Hutson 2018)

- Journal of Computational Neuroscience (108 articles)
  - The code is shared: 40 of 108 artykułów (**37%**)
- Frontiers in Neuroinformatics (79 articles)
  - The code is shared: 32 of 79 articles (**40%**)
- Biological Cybernetics (55 articles)
  - The code is shared 5 of 55 articles (**9%**)



We investigates
242 articles
(from 2016 to September 2018);
Credits: Piotr Litwin

An empirical analysis of journal policy effectiveness for computational reproducibility

Victoria Stodden, Jennifer Seiler, and Zhaokun Ma

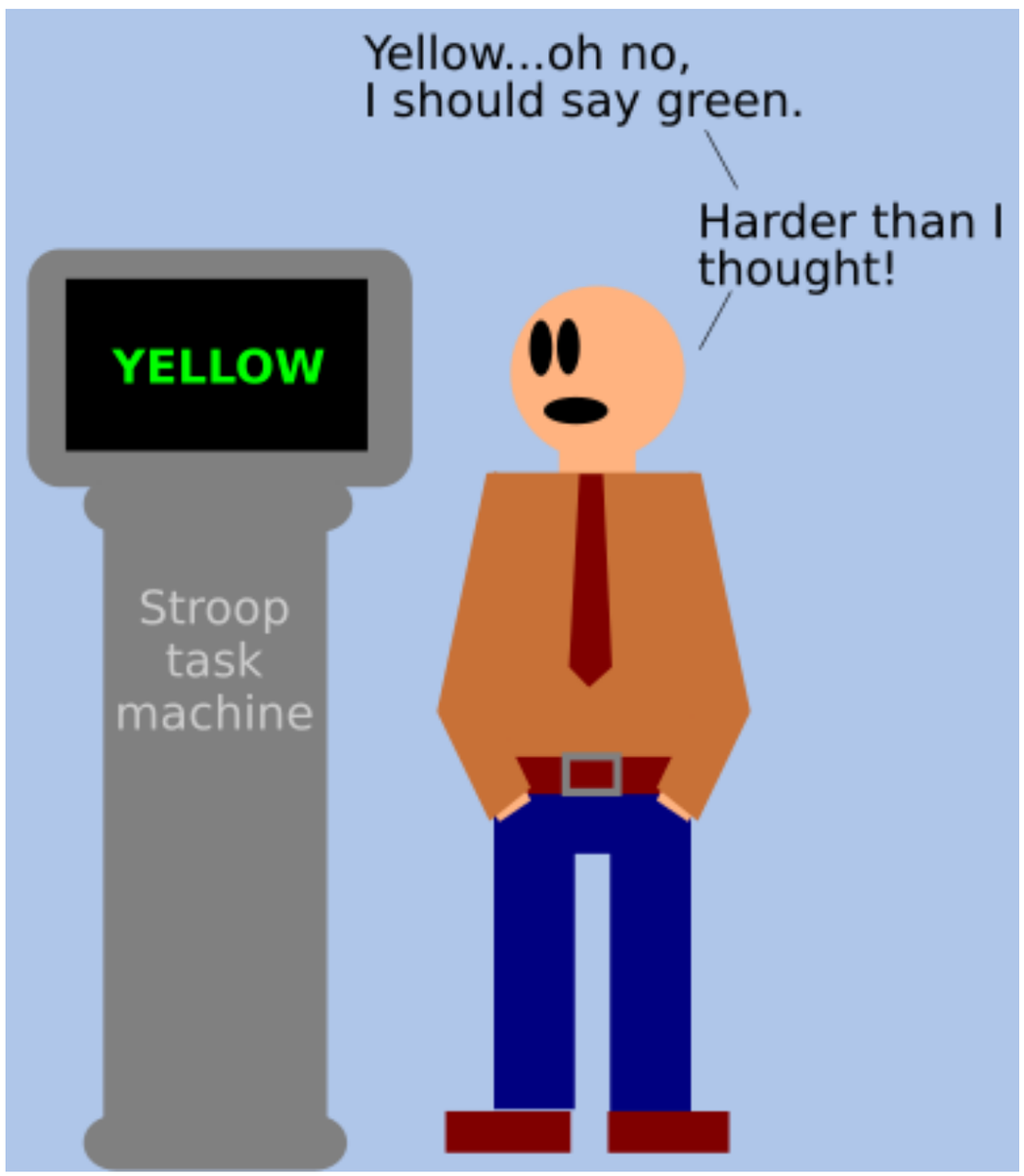
PNAS March 13, 2018 115 (11) 2584-2589; published ahead of print March 12, 2018
https://doi.org/10.1073/pnas.1708290115

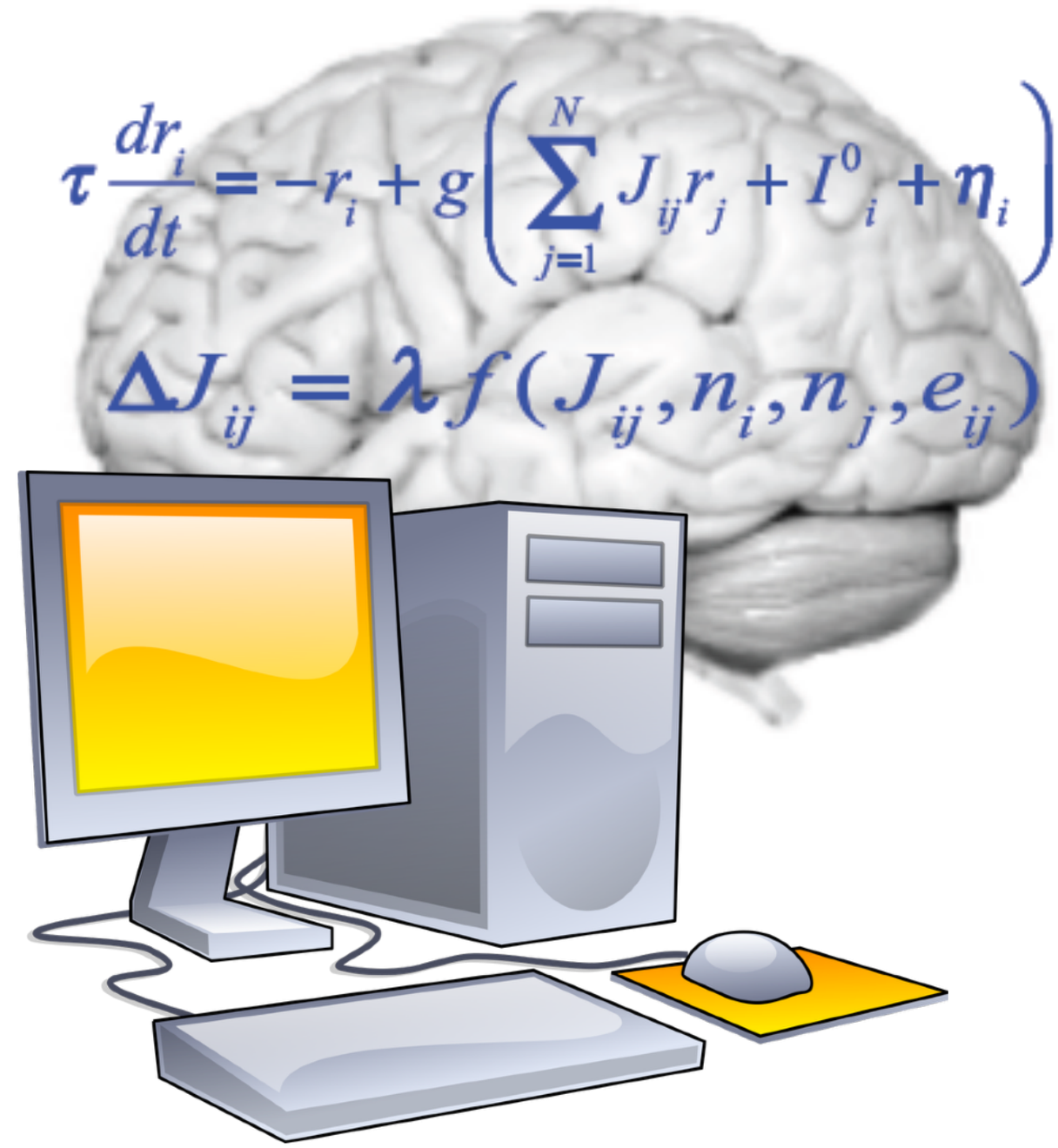
- They investigated the effectiveness of a replicational policy adopted by *Science* in 2011.
- SCIENCE requires authors to make the data and code sufficient to replicate their study available to other researchers upon request.
- The researchers selected 204 computational studies published in *Science*.
- Out of those, 24 papers (about 12%) provided code and data *via* external links or supplementary material.
- The researchers contacted the authors of the remaining 180 studies.
- To start with, 26% of the authors failed to reply altogether while the others often responded evasively—*e.g.*, by asking for reasons, making unfulfilled promises or directing the researchers back to supplementary material.
- In the end, it was possible to obtain data for only 36% of the papers.
- Stodden et al estimated about 25% of the models to be replicable.
- <u>The requirement to share data on demand after publishing is not being followed.</u>

How to ensure model replicability?
(Sandve et al., 2013)

- tracking how results were produced,
- avoiding the manual manipulation of data sets, archiving the exact versions of external programs used,
- using version control to store custom scripts, recording intermediate results (preferably in standardized formats),
- noting random seeds for randomized analyses, storing raw data behind plots,
- connecting textual statements to underlying results and finally providing public access to scripts,
- runs and results.

Neuroscience: experimenting and modelling





ORIGINAL RESEARCH ARTICLE

Front. Neuroinform., 21 February 2017 | https://doi.org/10.3389/fn-inf.2017.00011

Reproducibility and Comparability of Computational Models for Astrocyte Calcium Excitability

Tiina Manninen, Riikka Havela and Marja-Leena Linne*

Computational Neuroscience Group, Faculty of Biomedical Sciences and Engineering and BioMediTech Institute, Tampere University of Technology, Tampere, Finland

- They found that it is impossible to reimplement three of the models, those by Riera et al. (2011, De Pittà et al. (2009) and Dupont et al. (2011) due to insufficient information in published papers

- Relying on the original paper and a subsequent corrigendum, the researchers were able to reproduce the outcomes of only one model of astrocyte activity, *i.e.*, by Lavrentovich and Hemkin (2008).

- They found serious mistakes in the mathematical formalisms presented in two original papers (Riera et al. 2011 and Dupont et al. 2011, which made exact reproduction impossible.

- Prominent computational models of astrocyte excitability—one of the key biological events participating in synaptic transmission—are very hard to recreate and compare.

ELSEVIER

Neuropsychologia

Volume 111, March 2018, Pages 334-343

Active inference and the anatomy of oculomotion

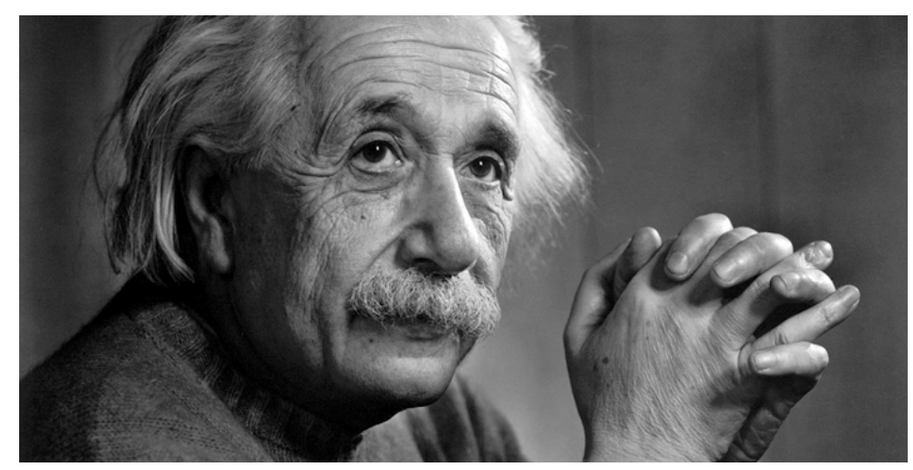
Thomas Parr, Karl J. Friston

Show more

https://doi.org/10.1016/j.neuropsychologia.2018.01.041

Open Access funded by Wellcome Trust

	Wszystkie	Od 2013
Cytowania	207945	88611
h-indeks	213	143



	Wszystkie	Od 2013
Cytowania	121507	37979
h-indeks	112	67

- Two (very long!) sections of their paper on oculomotion to basic equations of the free energy framework instead of specifying exactly how their computer simulation was built, which is described extremely tersely, without any quantitative detail.

- The paper contains both too much (lengthy introductions, complex figures instead of data) and too little information (no experimental data fed into the simulation, no details of the simulation framework).

The standard structure of the scientific article

Experimental psychology and experimental neuroscience

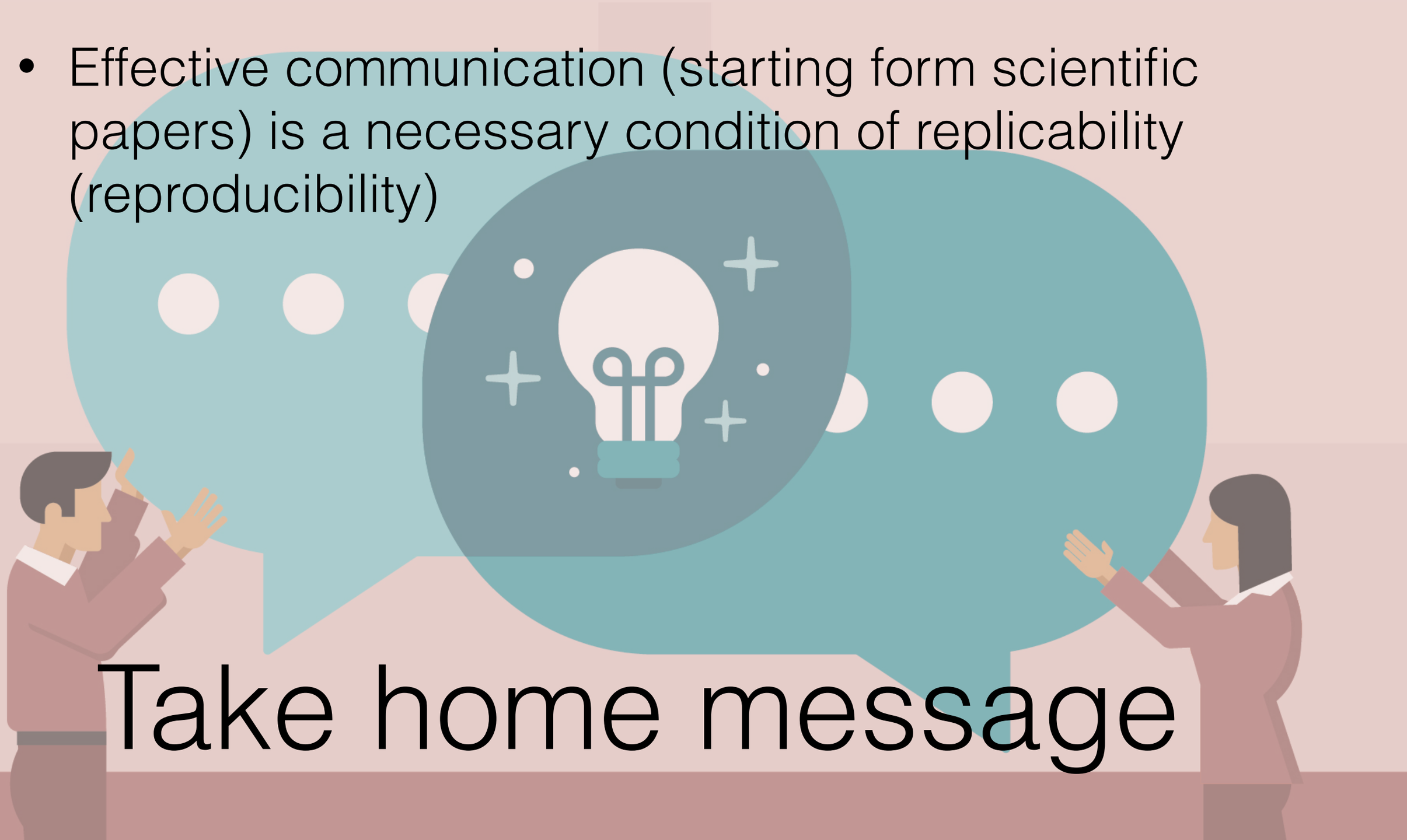
- Introduction
- Methods (participants, procedure, material etc.)
- Results
- Discussion

Computational modelin of the brain / cognitive functions

THERE IS NO THE STANDARD STRUCTURE OF THE ARTICLE !!!

- Replicability (reproducibility) is a necessary condition of reliability, cumulatively and applicability scientific knowledge

- Effective communication (starting form scientific papers) is a necessary condition of replicability (reproducibility)



Take home message

email: mateusz.hohol@uj.edu.pl

web: www.hohol.pl

Thank you for attention!

JAGIELLONIAN UNIVERSITY IN KRAKOW

Copernicus Center

IFiS PAN

NARODOWE CENTRUM NAUKI
2015/19/B/HS1/03310, 2014/14/E/HS1/00803

FNP Fundacja na rzecz Nauki Polskiej