

再現性の問題で考え方がどう変わったか

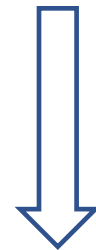
村山航

University of Tübingen

高知工科大学

Timeline

- 2006: Ph. D.
- (Start doing research abroad)
- 2011: “False-positive psychology” published.
- 2013: First teaching (permanent) job
- 2022: Now



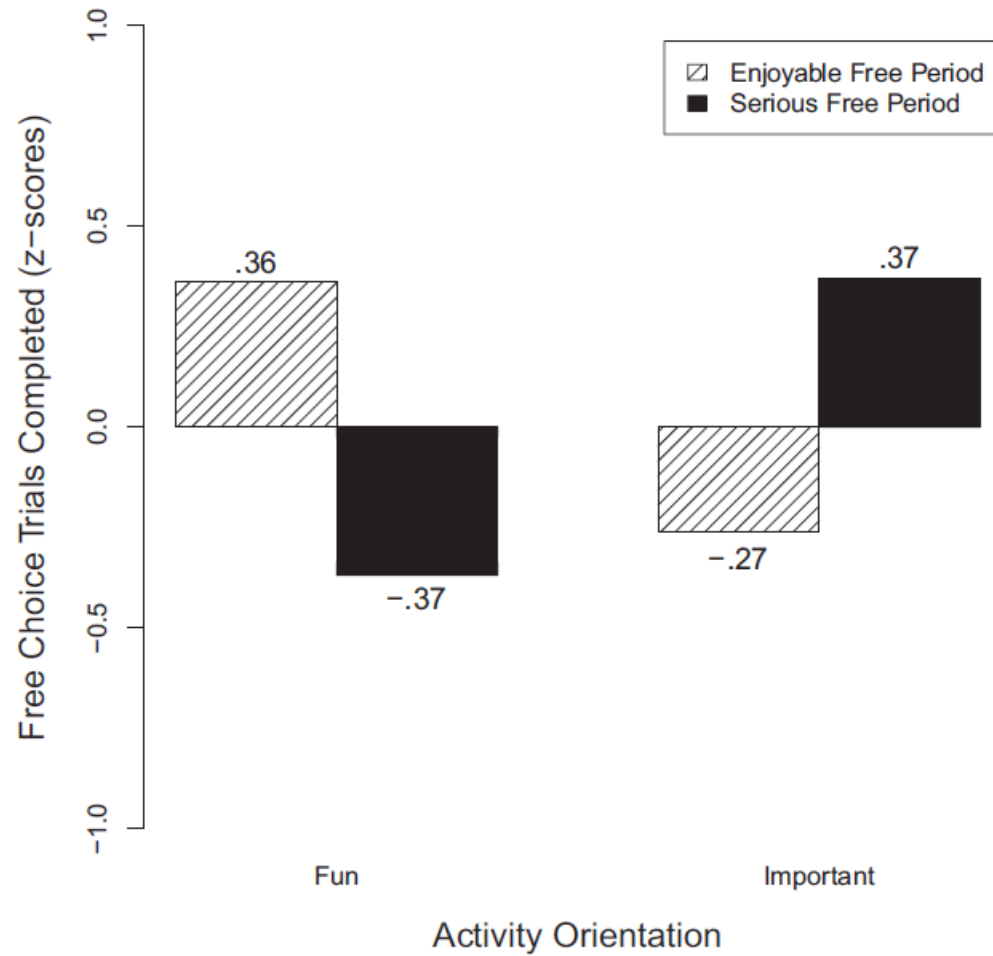
どのようにマインドセット
が変わったか？

Replication crisis が意味するところ

- システム・方法論の転換に目が行きがち
- 根本的な人間観の転換

人間行動はとても複雑である

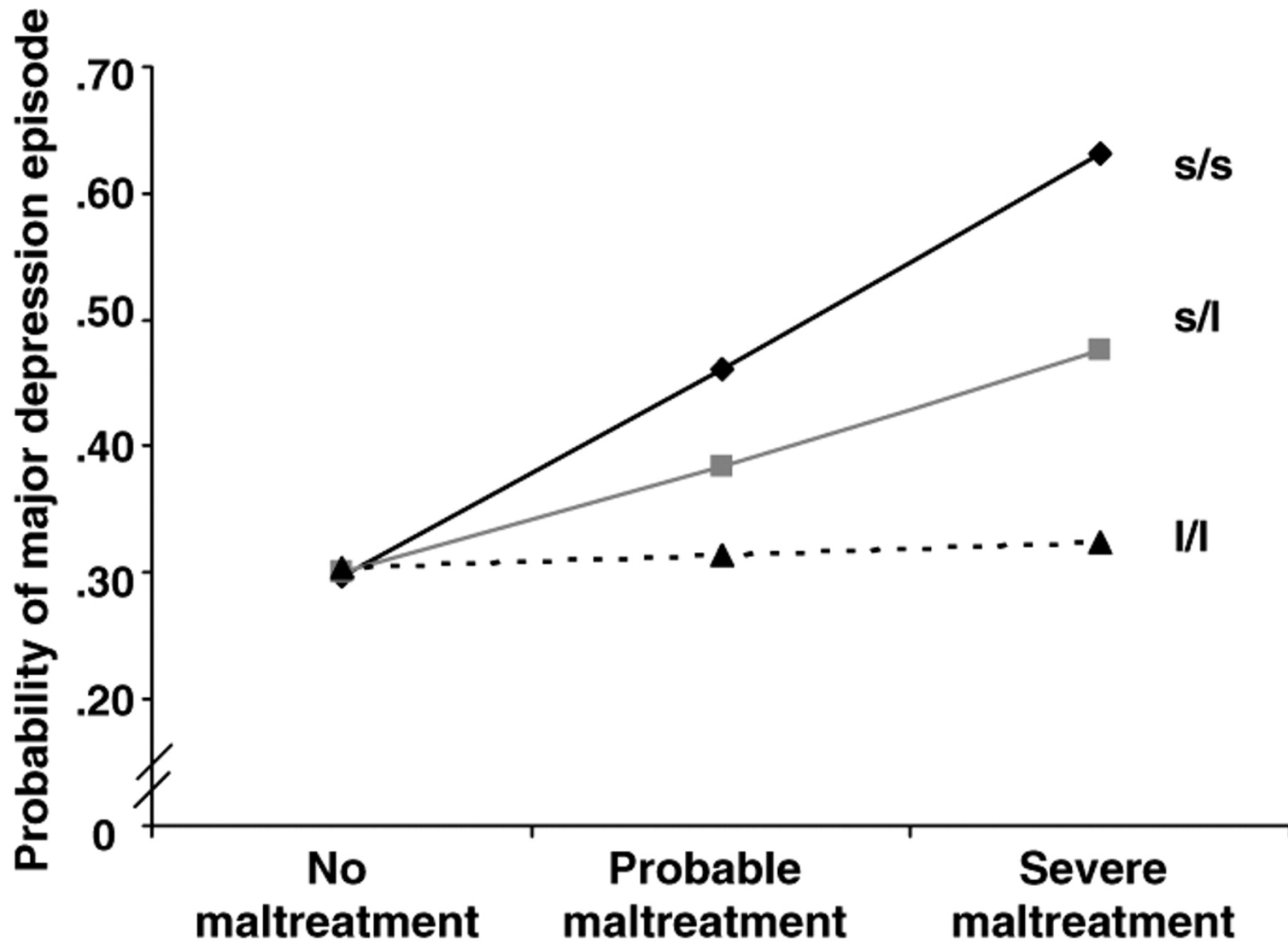
- (単純な反応は別として) 現実の人間行動は複雑であるので, , ,
 - 単純な操作でそう簡単に現実の行動は変わらない。
 - 妥当な「行動指標」なんて簡単に得られない。
 - 複雑な行動パターン (e.g., 相互作用) なんて簡単に予測できない。
 - 人為的な刺激の結果が日常生活の行動に一般化するなんて簡単には言えない。
- なのでそういった類の研究には厳格な目が必要である。
- そしてそうした複雑な行動をもっともダイレクトに扱っているのが社会心理学である。



Higgins et al., (2010, JPSP, Study 4a)

N = 52 (reduced from 63), 2x2x2 between (!) design.

5-HTTLPR



Caspi et al. (2003, Science)

Article | [Published: 16 March 2022](#)

Reproducible brain-wide association studies require thousands of individuals

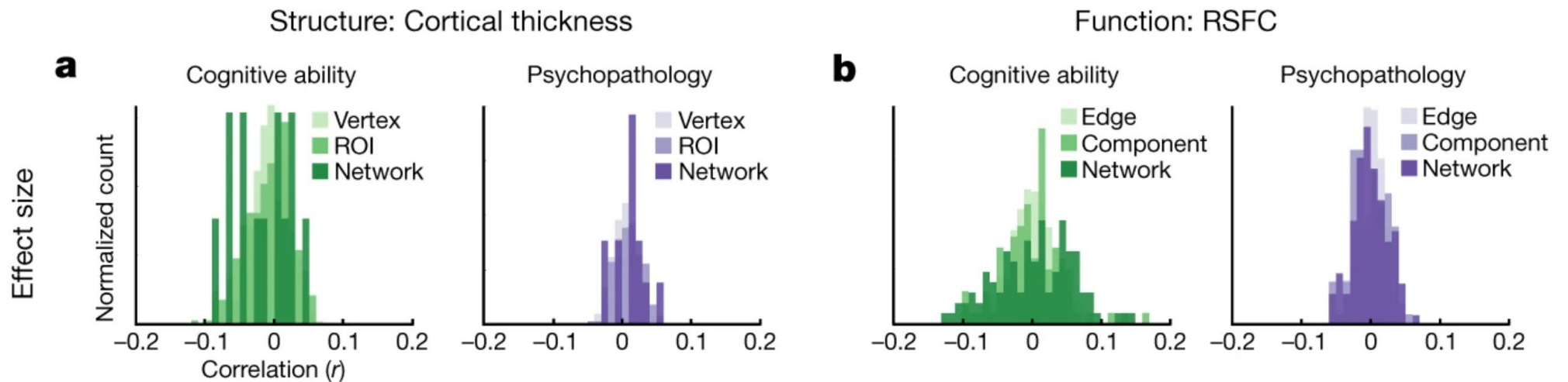
[Scott Marek](#) , [Brenden Tervo-Clemmens](#) , ... [Nico U. F. Dosenbach](#)  [+ Show authors](#)

[Nature](#) (2022) | [Cite this article](#)


7158 Accesses | **761** Altmetric | [Metrics](#)

Abstract

Magnetic resonance brain through well-replicated (e.g., lesion studies) and f



Small Effects: The Indispensable Foundation for a Cumulative Psychological Science

Friedrich M. Götz^{1,2}, Samuel D. Gosling^{3,4},
and Peter J. Rentfrow¹

¹Department of Psychology, University of Cambridge; ²Department of Psychology, University of British Columbia; ³Department of Psychology, University of Texas at Austin; and ⁴Melbourne School of Psychological Sciences, University of Melbourne

Perspectives on Psychological Science
2022, Vol. 17(1) 205–215
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1745691620984483
www.psychologicalscience.org/PPS



Abstract

We draw on genetics research to argue that complex psychological phenomena are most likely determined by a multitude of causes and that any individual cause is likely to have only a small effect. Building on this, we highlight the dangers of a publication culture that continues to demand large effects. First, it rewards inflated effects that are unlikely to be real and encourages practices likely to yield such effects. Second, it overlooks the small effects that are most likely to be real, hindering attempts to identify and understand the actual determinants of complex psychological phenomena. We then explain the theoretical and practical relevance of small effects, which can have substantial consequences, especially when considered at scale and over time. Finally, we suggest ways in which scholars can harness these insights to advance research and practices in psychology (i.e., leveraging the power of big data, machine learning, and crowdsourcing science; promoting rigorous preregistration, including prespecifying the smallest effect size of interest; contextualizing effects; changing cultural norms to reward accurate and meaningful effects rather than exaggerated and unreliable effects). Only once small effects are accepted as the norm, rather than the exception, can a reliable and reproducible cumulative psychological science be built.

具体的にどのように変わったのか

- 例：研究を評価するとき（査読のときなど）
 - まずは心に一回尋ねる「これは本当にありえる結果なのか」
 - 方法論より前に、まず自分の常識に尋ねる：自分自身が人間行動の複雑さを一番知っているはず！
 - そのうえで、方法論などから結果の堅さを判定する

[nature](#) > [nature human behaviour](#) > [letters](#) > [article](#)

Letter | [Published: 27 August 2018](#)

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

[Colin F. Camerer](#), [Anna Dreber](#), [Felix Holzmeister](#), [Teck-Hua Ho](#), [Jürgen Huber](#), [Magnus Johannesson](#), [Michael Kirchler](#), [Gideon Nave](#), [Brian A. Nosek](#) , [Thomas Pfeiffer](#), [Adam Altmejd](#), [Nick Buttrick](#), [Taizan Chan](#), [Yiling Chen](#), [Eskil Forsell](#), [Anup Gampa](#), [Emma Heikensten](#), [Lily Hummer](#), [Taisuke Imai](#), [Siri Isaksson](#), [Dylan Manfredi](#), [Julia Rose](#), [Eric-Jan Wagenmakers](#) & [Hang Wu](#)

[Nature Human Behaviour](#) **2**, 637–644 (2018) | [Cite this article](#)

58k Accesses | **458** Citations | **2343** Altmetric | [Metrics](#)

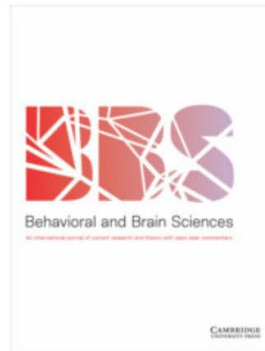
「堅い結果」の基準（例）

- サンプルサイズ（検定力分析や **sensitivity analysis**）
 - **Pre-registration** の有無（仮説・予測が明示されていることが重要）
 - 単純な主効果か，複雑な交互作用か
 - 効果量が小さすぎないか：小さいこと自体は問題ではない。統計的・方法論的アーティファクトによって有意になった小効果量に注意。
 - 刺激依存性が強すぎないか（一般化可能性）。
 - 信頼性の低い実験行動指標に依拠していないか
 - 認知モデリングの場合パラメータの信頼性は？
 - 「心理学の概念 (**construct**)」で遊びすぎていないか
-
- 統計的手法やモデリングの複雑さとは独立！
 - メインの結果に対する意識であって，探索的な検討は別。

Warning signal

- 被験者間要因で1群20人程度の実験
- 複雑な交互作用を示した研究
- 「○○の効果は××が高い人のみ存在する」といった類の結果
- メインの指標が **arbitrary** な **decision** によって決まっている（「評定値が3以上のときを高群とする」「2SD以上の値は除去」みたいな）
- バイオマーカー的な発想を持った研究（神経科学的研究のとき）
- 単一の刺激や **manipulation** に依拠している実験研究

- Preregistration も玉石混交なので注意。



[Behavioral and Brain Sciences](#)

Article contents

[Abstract](#)[References](#)

The generalizability crisis

Published online by Cambridge University Press: **21 December 2020**Tal Yarkoni [Show author details ▾](#)**Article**[Related commentaries](#)[Metrics](#)[Get access](#)[Share](#)[Cite](#)[Rights & Permissions](#)

Abstract

Most theories and hypotheses in psychology are verbal in nature, yet their evaluation overwhelmingly relies on inferential statistical procedures. The validity of the move from qualitative to quantitative analysis depends on the verbal and statistical expressions of a hypothesis being closely aligned – that is, that the two must refer to roughly the same set of hypothetical observations. Here, I argue that many applications of statistical inference in psychology fail to meet this basic condition. Focusing on the most widely used class of model

- 新しい研究やグラントを書くときも同じ：「堅い研究」を目指す
 - 自分が依拠している理論が旧時代の遺物ではないかをまず疑う
 - アイディアが出たら自分の常識に尋ねる：「本当に結果出るの？」
 - 先行研究の堅さも自分でもう一回尋ねる：「これって本当？」
 - ある程度の自信を持ったら、パイロット
 - 自信を強く持ったら **preregistration**（最初の段階からする必要はない）
 - 生態学的妥当性の高いデータで確かめる方法も考える
 - 「結果が出なくても面白い」研究も考える
- 人事の審査のときも研究の堅さ・**Open Science** へのコミットメントはポジティブに作用する（少なくとも私が関わっているところでは）。

- 「小サンプルの実験×7 = JPSP」という時代は終わりつつある
- 「1つの実験データを入念に検討してストーリーを作り上げる」というタイプの研究も古く感じられる（Neuroscienceではいまだに主流）。
- 研究の数や分析の量ではなく、いかにメインのアイデアを大サンプルで一般化可能性を確保した状態で、シンプルに検討することができるか。
 - 分析や7つの実験を準備する時間をむしろアイデアとデザインの検討に費やすべき。

「堅い研究」のジレンマ

- 「堅い研究」が innovative な研究とは限らない
- むしろ「堅い結果」は boring な結果も多い：ジレンマ
- 「堅くてかつ意義のある」研究を目指すことが重要，けどこれ
がもっとも難しい（健全なこと）。
 - 学生とプロジェクトを考えるときにもっとも苦勞する点。
 - 社会心理の有名な理論自体が実は怪しいことも多い（ここは学生が自分で見出すのではなく，指導教官がしっかりとした目を持つべき）。
 - どうしたら意義のある研究ができるか：これはまたディスカッションの方で。。。

ただ「厳しく」「難しく」なったのか？

- 方法論の厳格化という意味ではイエス。
- 一方、「研究の意義」の多様化
 - 「驚き」＝「面白い」＝「意義のある研究」から脱却
 - 「驚きのある面白い研究」だけが意義のある研究ではない
- 追試研究，有意でない結果の研究への間口は広くなった
 - ただし，すべての追試や **Null Results** が等しく意義があるわけではない（やはり少しは頭を使う必要性）。
- 研究方法のあり方を考えさせる研究（メタサイエンス的研究）

大規模な共同研究の広がり

- 「こんなに人間って簡単に変わるんだよ」というスタイルに変わって、「人間は複雑なんだ」という前提に立って、その問題を克服するために業界全体で努力するような研究のスタイルが評価されるようになった。
 - インパクトのある研究の追試研究 (e.g., Many labs)
 - 結果の一般化可能性（刺激・文化）を幅広く示すようなタイプの研究
 - 大規模サンプルの二次データ・実世界データを使った研究
 - データベースの構築
 - Etc. etc.

Cross-Cultural Consistency and Relativity in the Enjoyment of Thinking Versus Doing

Nicholas Buttrick, Hyewon Choi,
Timothy D. Wilson, Shigehiro Oishi,
and Steven M. Boker
University of Virginia

Daniel T. Gilbert
Harvard University

Sinan Alper
Middle East Technical University

Mark Aveyard
American University of Sharjah

Winnee Cheong
HELP University

Marija V. Čolić
University of Belgrade

Ilker Dalgar and Canay Dođulu
Middle East Technical University

Serdar Karabati
Istanbul Bilgi University

Eunbee Kim
Yonsei University

Goran Knežević
University of Belgrade

Asuka Komiya
Kochi University of Technology

Camila Ordóñez Laclé
Universidad de Iberoamérica

Caio Ambrosio Lage
Pontifical Catholic University of Rio de Janeiro

Ljiljana B. Lazarević and Dušanka Lazarević
University of Belgrade

Samuel Lins
Pontifical Catholic University of Rio de Janeiro

Mauricio Blanco Molina
Universidad de Iberoamérica

Félix Neto
Universidade do Porto

Ana Orlić
University of Belgrade

Boban Petrović
Institute of Criminological and Sociological Research,
Belgrade, Serbia

Massiel Arroyo Sibaja and David Torres Fernández
Universidad de Iberoamérica

Wolf Vanpaemel and Wouter Voorspoels
University of Leuven

Daniela C. Wilks
Universidade do Porto

[nature](#) > [articles](#) > article

Article | [Published: 20 May 2020](#)

Variability in the analysis of a single neuroimaging dataset by many teams

[Rotem Botvinik-Nezer](#), [Felix Holzmeister](#), ... [Tom Schonberg](#) 

[+ Show authors](#)

[Nature](#) **582**, 84–88 (2020) | [Cite this article](#)

42k Accesses | **203** Citations | **2056** Altmetric | [Metrics](#)

The Confidence Database

Dobromir Rahnev^{1*}, Kobe Desender^{2,3}, Alan L. F. Lee⁴, William T. Adler⁵, David Aguilar-Lleyda⁶, Başak Akdoğan⁷, Polina Arbuzova^{8,9,10}, Lauren Y. Atlas^{11,12,13}, Fuat Balci¹⁴, Ji Won Bang¹⁵, Indrit Bègue¹⁶, Damian P. Birney¹⁷, Timothy F. Brady¹⁸, Joshua Calder-Travis¹⁹, Andrey Chetverikov²⁰, Torin K. Clark²¹, Karen Davranche²², Rachel N. Denison²³, Troy C. Dildine^{11,24}, Kit S. Double²⁵, Yalçın A. Duyan¹⁴, Nathan Faivre²⁶, Kaitlyn Fallow²⁷, Elisa Filevich^{8,9,10}, Thibault Gajdos²², Regan M. Gallagher^{28,29,30}, Vincent de Gardelle³¹, Sabina Gherman^{32,33}, Nadia Haddara¹, Marine Hainguerlot³⁴, Tzu-Yu Hsu³⁵, Xiao Hu³⁶, Iñaki Iturrate³⁷, Matt Jaquiere¹⁹, Justin Kantner³⁸, Marcin Koculak³⁹, Mahiko Konishi⁴⁰, Christina Koß^{8,10}, Peter D. Kvam⁴¹, Sze Chai Kwok^{42,43,44}, Maël Lebreton⁴⁵, Karolina M. Lempert⁴⁶, Chien Ming Lo^{35,47}, Liang Luo³⁶, Brian Maniscalco⁴⁸, Antonio Martin³⁵, Sébastien Massoni⁴⁹, Julian Matthews^{30,50}, Audrey Mazancieux²⁶, Daniel M. Merfeld⁵¹, Denis O'Hora⁵², Eleanor R. Palser^{53,54,55}, Borysław Paulewicz⁵⁶, Michael Pereira⁵⁷, Caroline Peters^{8,9,10}, Marios G. Philiastides³², Gerit Pfuhl⁵⁸, Fernanda Prieto⁵⁹, Manuel Rausch⁶⁰, Samuel Recht⁶¹, Gabriel Reyes⁵⁹, Marion Rouault⁶², Jérôme Sackur^{62,63}, Saeedeh Sadeghi⁶⁴, Jason Samaha⁶⁵, Tricia X. F. Seow⁶⁶, Medha Shekhar¹, Maxine T. Sherman^{67,68}, Marta Siedlecka³⁹, Zuzanna Skóra³⁹, Chen Song⁶⁹, David Soto^{70,71}, Sai Sun⁷², Jeroen J. A. van Boxtel^{30,73}, Shuo Wang⁷⁴, Christoph T. Weidemann⁷⁵, Gabriel Weindel²², Michał Wierzchoń³⁹, Xinming Xu⁴², Qun Ye⁴², Jiwon Yeon¹, Futing Zou⁴² and Ariel Zylberberg⁷⁶

Understanding how people rate their confidence is critical for the characterization of a wide range of perceptual, memory, motor and cognitive processes. To enable the continued exploration of these processes, we created a large database of confidence studies spanning a broad set of paradigms, participant populations and fields of study. The data from each study are structured in a common, easy-to-use format that can be easily imported and analysed using multiple software packages. Each dataset is accompanied by an explanation regarding the nature of the collected data. At the time of publication, the Confidence Database (which is available at <https://osf.io/s46pr/>) contained 145 datasets with data from more than 8,700 participants and almost 4 million trials. The database will remain open for new submissions indefinitely and is expected to continue to grow. Here we show the usefulness of this large collection of datasets in four different analyses that provide precise estimations of several foundational confidence-related effects.

むしろこの機会をチャンスだと捉える

- 厳しくはなったが、今までのように「エリートラボによる寡占」状態が崩れつつある。
- 我々のようなアジア人研究者にとっては大チャンス状況！
 - 大規模なコラボレーションの広がり
 - **Open Science movement** と **diversity/inclusion** の流れとの繋がり。
 - 既存の理論・知見に関する懐疑：下剋上のチャンス！

SCORE

Systematizing Confidence in Open Research and Evidence

[Project Overview](#)

[Reproductions](#)

[Replications](#)

[Robustness](#)

[Core Team](#)

Final Call for Collaborators

The SCORE project is assessing the credibility of published social-behavioral science claims and developing methods to better assess credibility. An overview of the project goals and design is available as a [preprint](#). Started in 2019, the project is in its final phase and recruiting collaborators to conduct reproduction, robustness, and replication studies of social-behavioral science findings.



Social media (e.g., Twitter) を通して情報を得る

- (ちなみに私は数年前 **Twitter** に遅れながら参加しました)。
- **Open Science Movement** を支えてきている
- 情報の **Exchange** が劇的に早くなった
 - Preprint
 - 新たな論文, 批判, 新しい統計手法 etc.
 - **Science** の進展を肌で感じ取れる
 - いいアイデアを得るうえで新しい情報はとても重要!
 - 海外の情報も積極的に取り入れるとよい
- 新しいコラボレーションの情報も素早く手に入る。
- 自分の論文も宣伝できる
- ネットワーキングの効果もある





These are not the effects you are looking for: Causality and the within-/between-person distinction in longitudinal data analysis

AUTHORS

Julia M. Rohrer, Kou Murayama

AUTHOR ASSERTIONS

Conflict of Interest: No

Public Data: Not applicable

Preregistration: Not applicable

Page: 1 of 18 Automatic Zoom

1

These are not the effects you are looking for: Causality and the within-/between-person distinction in longitudinal data analysis

Julia M. Rohrer¹ and Kou Murayama^{2,3}

In psychological science, researchers often pay particular attention to the distinction between within- and between-person relationships in longitudinal data analysis. Here, we aim to clarify the relationship between the within- and between-person distinction

Download preprint

Downloads: 823



Leonardo B. Marques, Ph.D. has endorsed this work.



Abstract

In psychological science, researchers often pay particular attention to the distinction between within- and between-person

オンライン実験を積極的に活用する

- Prolific/Amazon m-turk
- サンプルサイズを確保するには最適の方法。
- 意外にまじめにやってくれる
- 意外にいろいろとできる

- ただしお金が意外とかかる

まとめ

- 再現性の問題によって，論文・グラント・研究者の評価の枠組みが変化してきている。
- 1つ1つの変化を追うことも大切だが，より大きな「人間観の変化」を抑えることがもっとも大切。
- この時代の変化を，日本人研究者としてむしろチャンスと捉える
- 時代の変化をソーシャルメディアなどでリアルタイムに追っていく姿勢は重要。
 - 「研究内容」は後追いではダメだが「研究スタイル」は後追いでもついていった方がいい。

- ありがとうございます！
- 宣伝
 - 縦断データと因果推論のワークショップ（3月29 – 4月1日）
<https://www.longitudinaldataanalysis.com/>
 - ポスドク常時募集しています！