

心理学における 再現可能性問題 － 概説 －

三浦麻子

関西学院大学文学部・大阪大学大学院基礎工学研究科

本資料には[三浦のWebサイト](#)からリンクしています

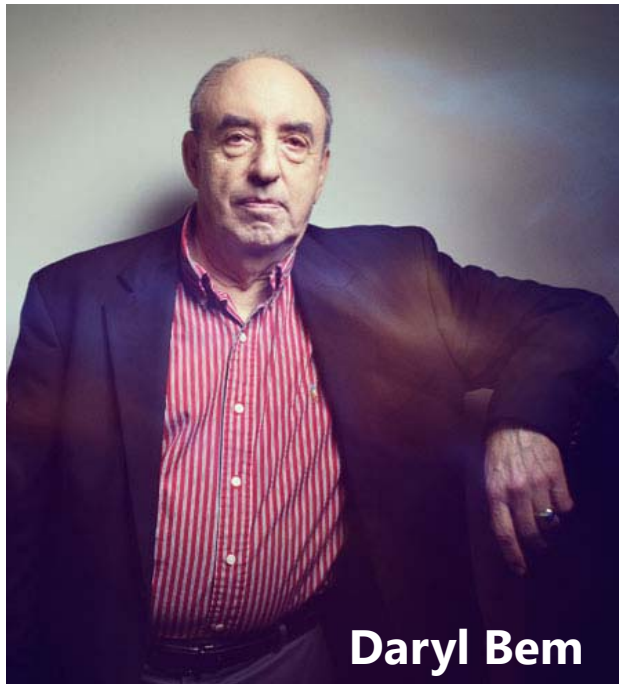
要旨

- ありとあらゆる科学(を標榜する学問)にとって、ある知見が信頼に足るものかどうかを確認するもっとも有力な手段は、一度得られた実験結果についてそれが再現できるかどうかを検証することである。こうした科学の再現性問題は、2014年初頭以来世間を騒がせていた「STAP細胞問題」によって、決して望ましい形だったとは言えないが、一般にも広く知られることとなった。心理学においても、これと同じ時期に、研究の信頼性の著しい低下につながるような、あるいはそれを疑わせるような出来事が相次いで起きたが、その原因の少なからぬ部分は、研究者たちが再現性検証の試みを軽視してきたことにあるとされた。本発表では、心理学において、**研究プロセスそのものに内在するものも含めた再現性を低からしめる問題を概説し、それらを解消するために現に行われている取り組みを紹介**する。

心理学における 再現性問題 きっかけ

それはひとりの大物の「トンデモ」論文から勃発（再燃）した

心理学における再現性問題 きっかけ



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

© 2011 American Psychological Association
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

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of *psi* are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in *psi* performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with *psi* performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about *psi*, issues of replication, and theories of *psi* are also discussed.

Keywords: *psi*, parapsychology, ESP, precognition, retrocausation

JPSPに、かつて「自己知覚理論」を打ち立てた著名な社会心理学者による「人間には予知能力がある」という論文が掲載された。

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407-425.

心理学における再現性問題 きっかけ



[View this Article](#) [Submit to PLOS](#) [Get E-Mail Alerts](#) [Contact Us](#)

PLoS One. 2012; 7(3): e33423.

Published online 2012 Mar 14. doi: [10.1371/journal.pone.0033423](https://doi.org/10.1371/journal.pone.0033423)

PMCID: PMC3303812

PMID: [22432019](https://pubmed.ncbi.nlm.nih.gov/22432019/)

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

[Stuart J. Ritchie](#), ¹ * [Richard Wiseman](#), ² and [Christopher C. French](#) ³

Sam Gilbert, Editor

[Author information](#) [Article notes](#) [Copyright and License information](#) [Disclaimer](#)

This article has been [cited by](#) other articles in PMC.

Abstract

Go to:

Nine recently reported parapsychological experiments appear to support the existence of precognition. We describe three pre-registered independent attempts to exactly replicate one of these experiments, 'retroactive facilitation of recall', which examines whether performance on a memory test can be influenced by a post-test exercise. All three replication attempts failed to replicate the original finding (combined $n=150$; combined $p=.83$, one-tailed) and thus do not support the existence of precognition.

Journal of Personality and Social Psychology
2012, Vol. 103, No. 6, 933–948

© 2012 American Psychological Association
0022-3514/12/\$12.00 DOI: [10.1037/a0029709](https://doi.org/10.1037/a0029709)

Correcting the Past: Failures to Replicate Psi

Jeff Galak
Carnegie Mellon University

Robyn A. LeBoeuf
University of Florida

Leif D. Nelson
University of California, Berkeley

Joseph P. Simmons
University of Pennsylvania

※この問題には、データの捏造・改ざん（Diederik StapelやMarc Hauserなど）など研究不正の発覚も大きな影響を与えているが、話がややこしくなるので、ここでは故意による悪事についてはとりあげない。

Across 7 experiments ($N = 3,289$), we replicate the procedure of Experiments 8 and 9 from Bem (2011), which had originally demonstrated retroactive facilitation of recall. We failed to replicate that finding. We further conduct a meta-analysis of all replication attempts of these experiments and find that the average effect size ($d = 0.04$) is no different from 0. We discuss some reasons for differences between the results in this article and those presented in Bem (2011).

Keywords: psi, precognition, ESP, researcher degrees of freedom, meta-analysis

追試による再現性検証論文（再現失敗）をJPSPが「**追試は掲載しない**」という理由でリジェクトして物議を醸す

↓
結果的に別の研究者らによる再現性検証論文（再現に失敗）がJPSPに掲載された

再現性問題に関する熱い議論が勃発



Perspectives on Psychological Science

aps | ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Journal Indexing & Metrics
View »

Journal Home Browse Journal ▾ Submit Paper ▫ About ▾ Subscribe ▫

Table of Contents

◀ Previous Issue Volume 7 Issue 6, November 2012 Next Issue ▶

Special Section on Replicability in Psychological Science: A Crisis of Confidence?

 **Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?**

Harold Pashler, Eric-Jan Wagenmakers

First Published November 7, 2012; pp. 528–530

All Issues
Current Issue
Online First

<https://journals.sagepub.com/toc/pps/7/6>

「実験結果の再現可能性」特集号読書会

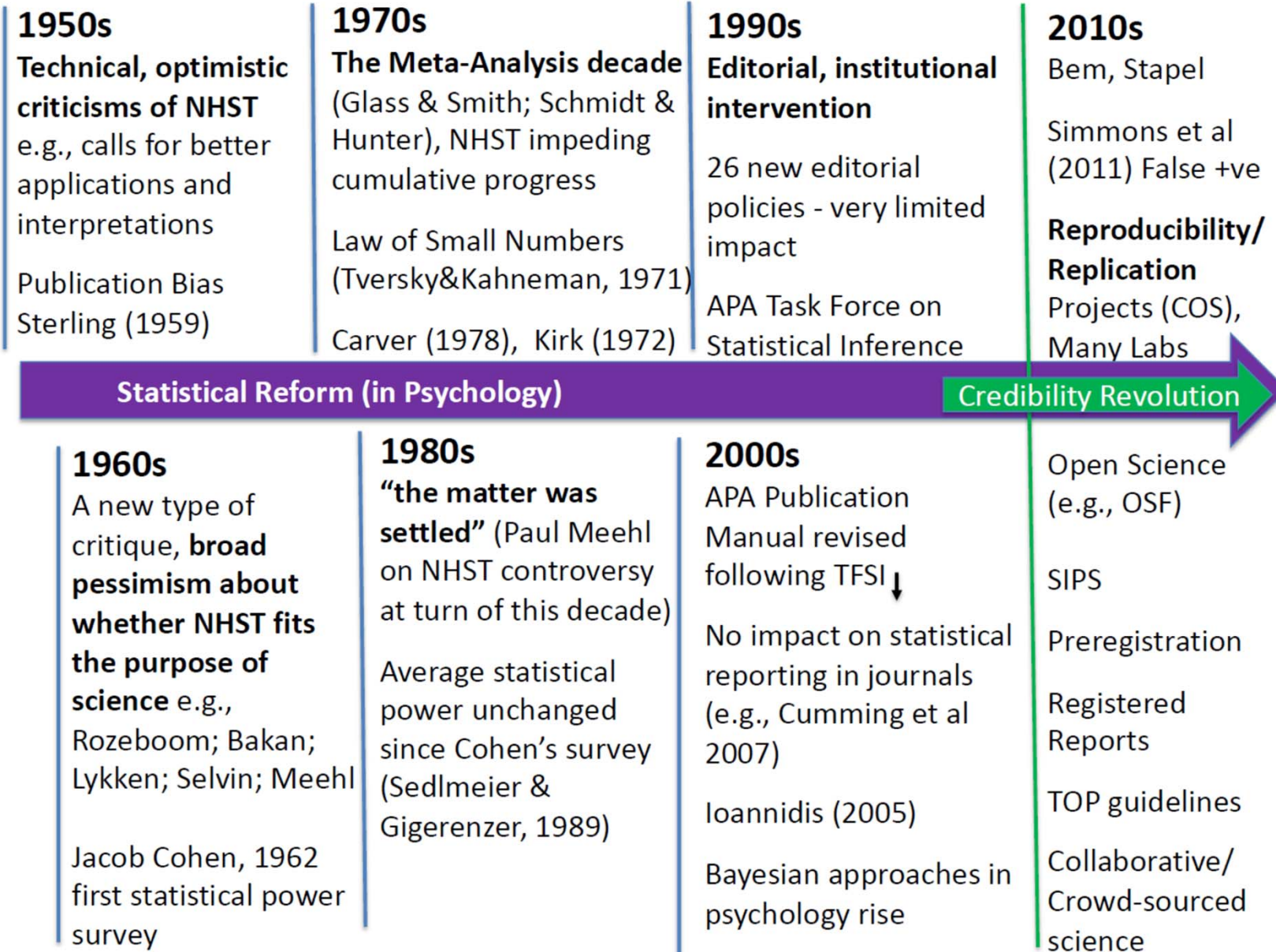
最近、様々な学問領域において、また個別の論文に対して「実験結果の再現可能性」に対する疑義が指摘されています。先日発行されたPerspectives on Psychological Science(PPS)誌11月号では、この再現可能性の問題に対する特集が組まれました。心理学者も再現可能性の問題に関心ではられません。

[Perspectives on Psychological Science, 7\(6\), 2012 Special Section on Replicability in Psychological Science: A Crisis of Confidence?](#)

(この号に限り、どなたでも自由に論文PDFを閲覧・ダウンロードすることができます)

そこで以下の形態にて、本特集号を1日で一気に読む研究会を企画・実施しました。ここでは、その記録をまとめています。

<https://sites.google.com/site/ppsjournalclub/>



Fidler, F. (2019.3.7) [Making science reproducible: Will this time be different? \(Why now, and not then?\)](#)

1950s	1970s	1990s	2010s
Technical, optimistic criticisms of NHST e.g., calls for better applications and	The Meta-Analysis decade (Glass & Smith; Schmidt & Hunter), NHST impeding cumulative progress	Editorial, institutional intervention 26 new editorial	Bem, Stapel Simmons et al (2011) False +ve
<p>主に統計手法の観点から、これまでも既に繰り返 し「ゆゆしき問題」とされ、「是正」が叫ばれてき たトピックだが、浮かんでは消え、を繰り返してき た。しかし今回は本質的な転換（革命）をもたらす ...かもしれない。</p> <p>なぜなら、個々の研究者の不正という規模を超えて、 学問全体への信頼を揺るがす状況だということを自 覚せざるを得ない知見が次々に示されたから。</p>			
Lykken; Selvin; Meehl Jacob Cohen, 1962 first statistical power survey	(Sedlmeier & Gigerenzer, 1989)	Ioannidis (2005) Bayesian approaches in psychology rise	Open Science (e.g., OSF) TOP guidelines Collaborative/ Crowd-sourced science

大規模追試による検証

実際、世界は革命的スピードで動いている

[Open Science Collaboration. \(2015\). Estimating the reproducibility of psychological science. *Science*, 349\(6251\), aac4716.](#)

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration^{*,†}

+ See all authors and affiliations

Science 28 Aug 2015:
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

Article

Figures & Data

Info & Metrics

eLetters

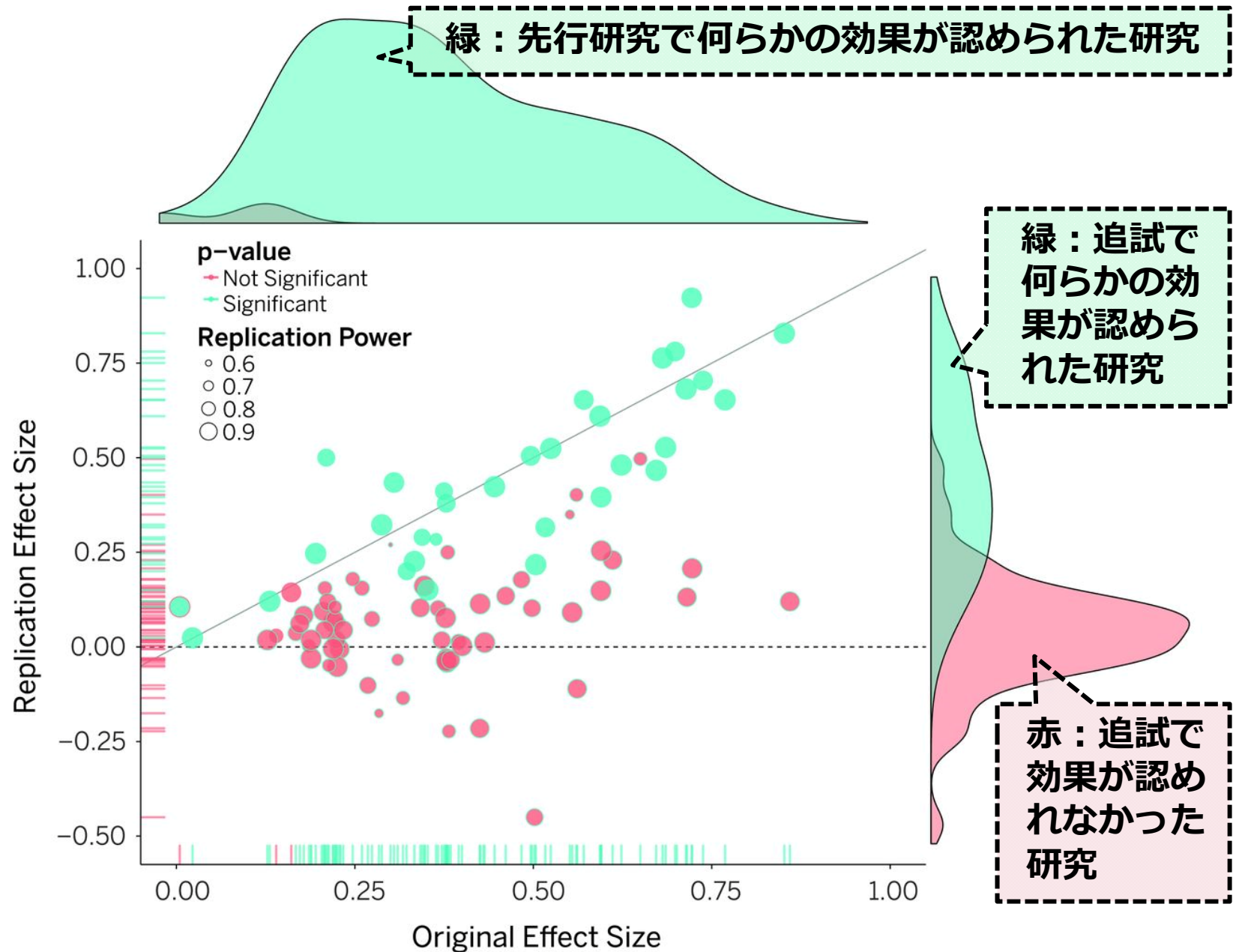
 PDF

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.

2008年に心理学主要誌に掲載された100の実験について直接的追試を実施し、再現性を検証したプロジェクト。2011年11月開始、270名の研究者が参加。

[Open Science Collaboration. \(2015\). Estimating the reproducibility of psychological science. *Science*, 349\(6251\), aac4716.](#)



Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45(3), 142-152.

Investigating variation in replicability: A “many labs” replication project.

 EXPORT  Add To My List    Request Permissions  Database: PsycARTICLES Journal Article

[Klein, Richard A.](#) [Ratliff, Kate A.](#) [Vianello, Michelangelo](#) [Adams Jr., Reginald B.](#) [Bahník, Štěpán](#) [Bernstein, Michael J.](#) [Bocian, Konrad](#) [Brandt, Mark J.](#) [Brooks, Beach](#) [Brumbaugh, Claudia Chloe](#) [Cemalcilar, Zeynep](#) [Chandler, Jesse](#) [Cheong, Winnee](#) [Davis, William E.](#) [Devos, Thierry](#) [Eisner, Matthew](#) [Frankowska, Natalia](#) [Furrow, David](#) [Galliani, Elisa Maria](#) [Hasselman, Fred](#) [Hicks, Joshua A.](#) [Hovermale, James F.](#) [Hunt, S. Jane](#) [Huntsinger, Jeffrey R.](#) [Izerman, Hans](#) [John, Melissa-Sue](#) [Joy-Gaba, Jennifer A.](#) [Barry Kappes, Heather](#) [Krueger, Lacy E.](#) [Kurtz, Jaime](#) [Levitan, Carmel A.](#) [Mallett, Robyn K.](#) [Morris, Wendy L.](#) [Nelson, Anthony J.](#) [Nier, Jason A.](#) [Packard, Grant](#) [Pilati, Ronaldo](#) [Rutchick, Abraham M.](#) [Schmidt, Kathleen](#) [Skorinko, Jeanine L.](#) [Smith, Robert](#) [Steiner, Troy G.](#) [Storbeck, Justin](#) [Van Swol, Lyn M.](#) [Thompson, Donna](#) [van 't Veer, A. E.](#) [Vaughn, Leigh Ann](#) [Vranksa, Marek](#) [Wichman, Aaron L.](#) [Woodzicka, Julie A.](#) [Nosek, Brian A.](#)

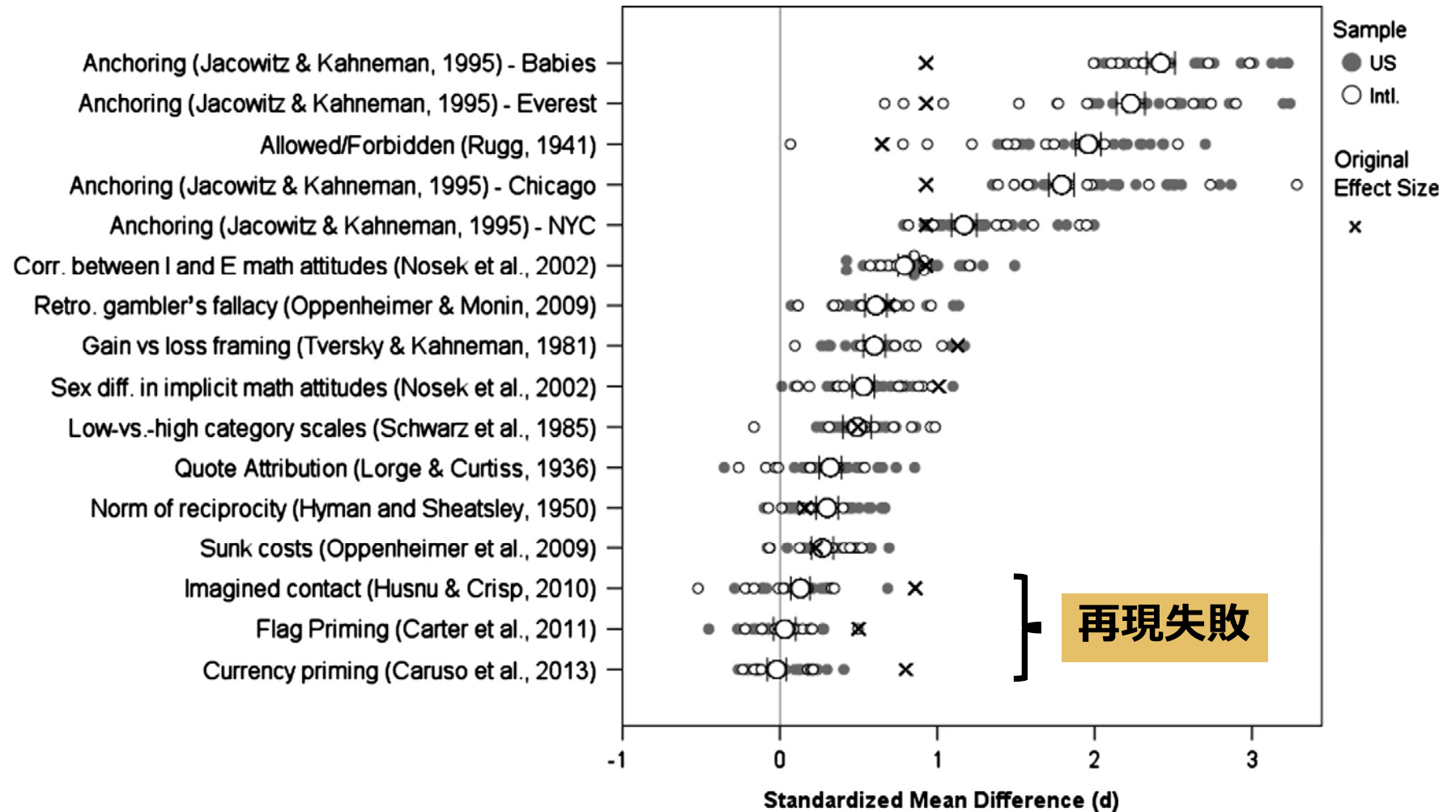
Abstract

Although replication is a central tenet of science, direct replications are rare in psychology. This research tested variation in the replicability of 13 classic and contemporary effects across 36 independent samples totaling 6,344 participants. In the aggregate, 10 effects replicated consistently. One effect—imagined contact reducing prejudice—showed weak support for replicability. And two effects—flag priming influencing conservatism and currency priming influencing system justification—did not replicate. We compared whether the conditions such as lab versus online or US versus international sample predicted effect magnitudes. By and large they did not. The results of this small sample of effects suggest that replicability is more dependent on the effect itself than on the sample and setting used to investigate the effect. (PsycINFO Database Record (c) 2018 APA, all rights reserved)

古典から近年までインパクトの大きな13の実験について直接的追試を実施し、再現性を検証したプロジェクト。2013年2月開始、36名の研究者が参加。

[Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. \(2014\). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45\(3\), 142-152.](#)

※日本語レジュメ（三浦による）



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

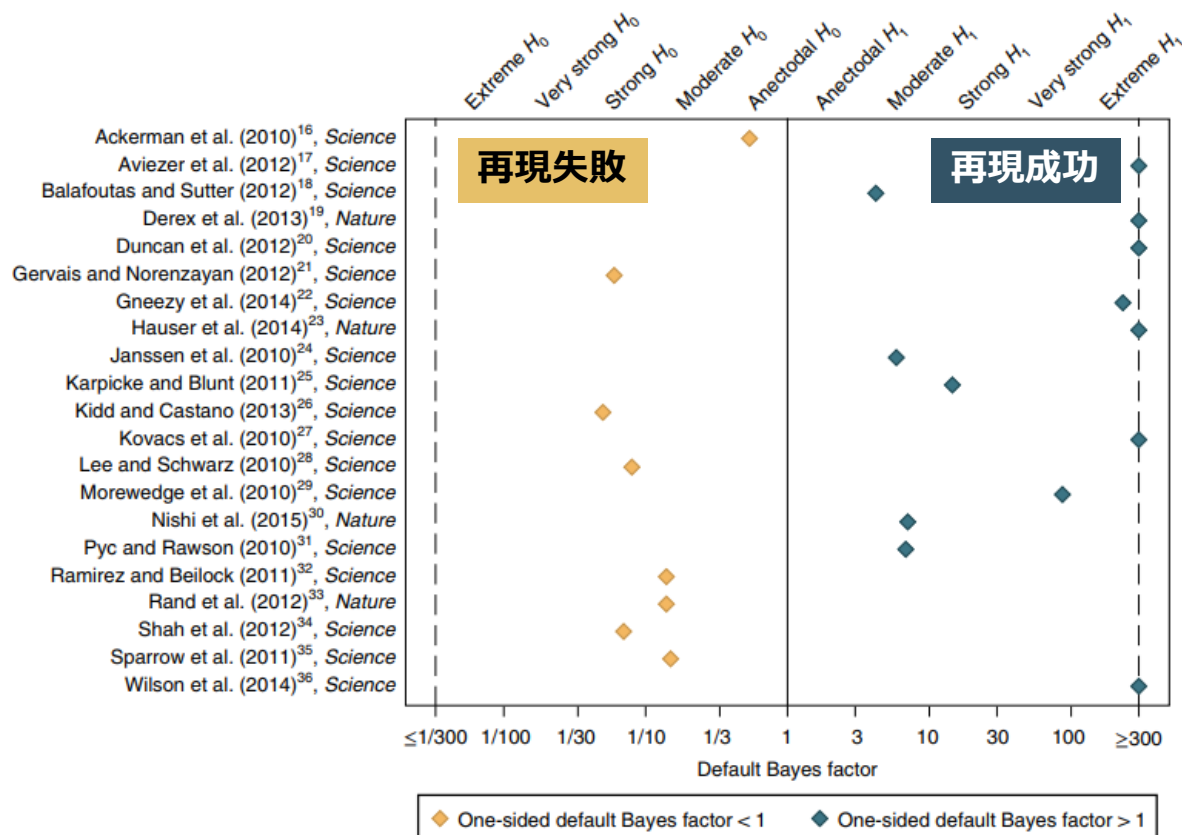
Colin F. Camerer^{1,16}, Anna Dreber^{2,16}, Felix Holzmeister^{3,16}, Teck-Hua Ho^{4,16}, Jürgen Huber^{3,16}, Magnus Johannesson^{5,16}, Michael Kirchner^{3,5,16}, Gideon Nave^{6,16}, Brian A. Nosek^{7,8,16*}, Thomas Pfeiffer^{9,16}, Adam Altmeld², Nick Buttrick^{7,8}, Taizan Chan¹⁰, Yiling Chen¹¹, Eskil Forsell¹², Anup Gampa^{7,8}, Emma Heikensten², Lily Hummer⁸, Taisuke Imai¹³, Siri Isaksson², Dylan Manfredi⁶, Julia Rose³, Eric-Jan Wagenmakers¹⁴ and Hang Wu¹⁵

<https://www.nature.com/articles/s41562-018-0399-z>

Nature誌やScience誌に発表された21件の実験社会科学研究を対象に、当初の研究の約5倍のサンプルサイズを用いて検出力の高い2段階再現実験を試み、ベイズ統計（ベイズファクター）など多様な手法で検証。再現実験の62%で当初の研究と同じ方向の効果が見られた。

For details:

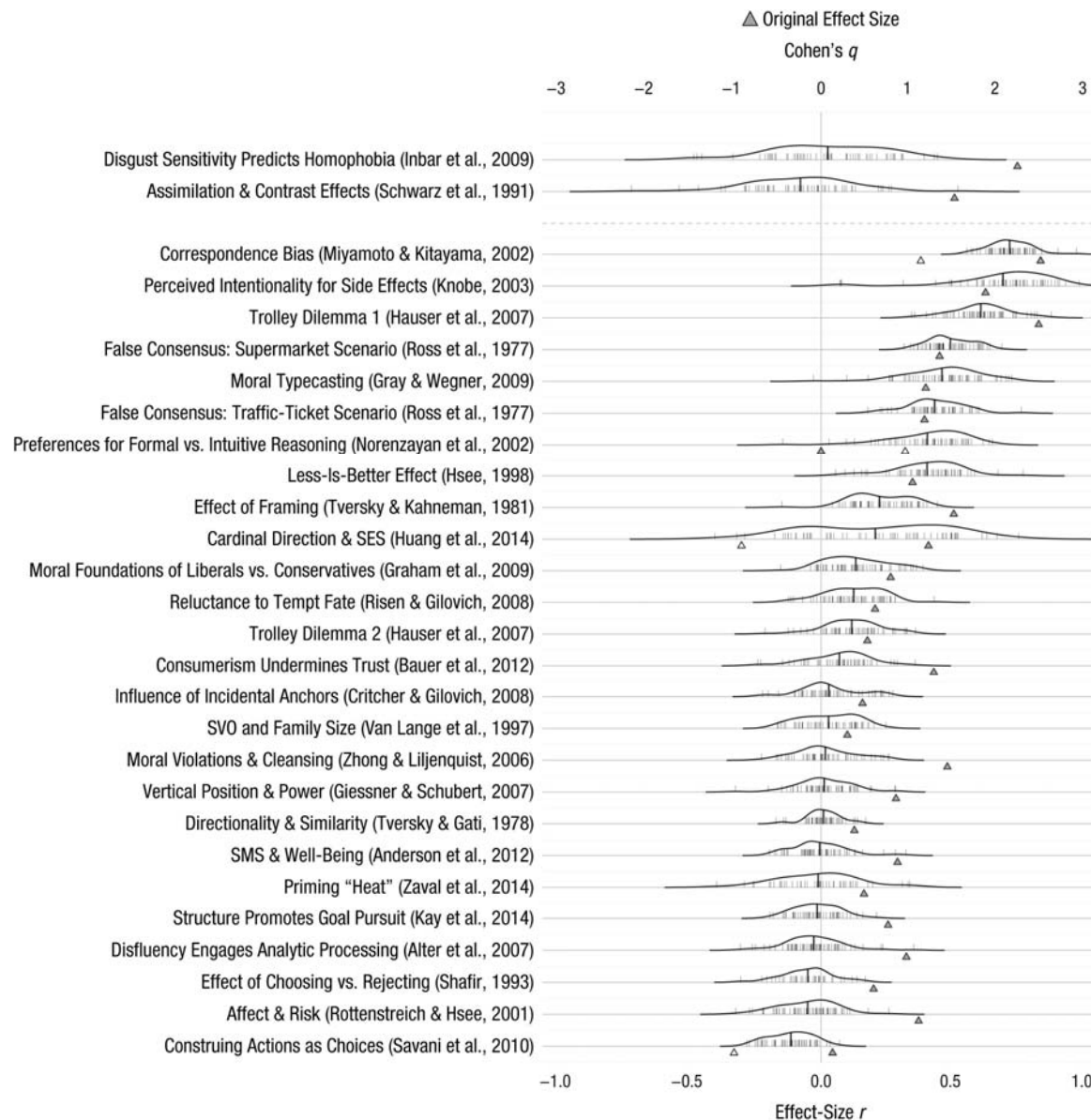
<http://www.socialsciencesreplicationproject.com/>



BFによる心理学的仮説・モデルの評価に関する詳細は岡田 (2018; 心理学評論) 参照

再現失敗の8研究については、著者自身が「なぜ再現不可能だったのか」に関するコメント論文(Correspondence)を寄せている。

Many Labs 2: Investigating Variation in Replicability Across Samples and Settings



▲ オリジナルの効果量(WEIRD))

△ オリジナルの効果量(not WEIRD)

古典から近年まで28件の（社会）心理学研究を対象に、特に文化差を念頭に置いて、36カ国のラボでデータを収集。Web実験。

[Slate1/Slate2](#)で13/15本の実験を一気に実施する方式

For details:

<https://osf.io/8cd4r/wiki/home/>

読書会レジュメ・資料：

<https://sites.google.com/keio.jp/reading-manylabs2/home>

大規模追試プロジェクトの特徴

- ハイインパクトな研究を追試対象として設定
- 参加は公募（クラウドソーシング）
 - 多くのラボで、多くの追試データを一気に集める
- 実験デザインや分析計画を事前登録（後述）
 - 原著者の積極的な協力も求める
- オープンデータ
 - マテリアル、ローデータ、分析プログラム等のすべて

高くない再現性を自覚し、対策をしなければ学問の危機だという意識はひしひしと伝わってきた（ただし、心理学の中でも領域による「温度差」は結構あるようだとわかった）。

なぜ再現されないのか？

本質的な問題と，本質的でない問題と

なぜ再現されないのか?

ここでの「不正」は、
捏造や改ざんレベル
ではないものを意図
している

■ 本質的な問題

- 「強い理論」構築の困難さ
- 一方で（だからこそ）提案される理論は脆弱さを多く抱える

■ 本質的ではない「はずの」問題

- 「不正」をしたくなる土壌
 - 常に「成果」を論文という形で求められる
 - 「仮説がデータによって検証」されることが強く求められる（それを追究することこそが科学だ、という信念が歪んだ形で実践されている）
- 「不正」ができてしまう土壌
 - 研究過程は研究者/ラボ内に閉じている一方、実施後に公刊される論文には、構想段階からの研究過程が含まれている
- 人間の「認知的バイアス」
- **これらが「不正」＝「問題のある研究実践」
(Questionable Research Practices; QRPs)につながる**



QRPs 問題のある研究実践

- 1回の検定力の大きい研究ではなく、何回もの検定力の小さい研究を行う
- ポジティブな結果のみ報告し、ネガティブな結果を報告しない
- 都合の良い結果が得られた段階でデータ収集をやめる
- 都合の良い結果が得られるまでデータ収集を続ける
- 複数の独立変数や従属変数があったとしても、ポジティブな結果が得られた部分のみを報告する
- さまざまな分析が可能であるにもかかわらず、特定の分析結果しか報告しない
- 新たな発見をあたかも元々仮説があったかのように報告する（結果がわかってから仮説を作る）
- 一度都合の良い結果が得られたら、その後それを確かめる追試をしない

[Nosek, et al. \(2012\)](#)

QRPs 問題のある研究実践

表 2: 各 QRPs に関する自己告白率と弁明可能性評価

John, et al. (2012)

QRPs	自己告白率 (%)	弁明可能性
1. 従属変数を選択的に報告	66.5	1.84
2. 結果が有意かどうかを見てからさらにデータを採る	58.0	1.79
3. 実験条件の選択的報告	27.4	1.77
4. 望む結果が出たので予定より早くデータ取得を終了	22.5	1.76
5. p 値を「切り捨て」て報告 (例: 5.4%を 5%に)	23.3	1.68
6. うまくいった研究だけを選択的に報告	50.0	1.66
7. データ除外の影響を見てから除外するかどうかを決定	43.4	1.61
8. 予想していなかった結果を、予測していたかのように報告	35.0	1.50
9. 実際は不確かなのに、性別など人口統計学的変数によって結果は影響されないと報告	4.5	1.32
10. データの改竄	1.7	0.16

アメリカの主要大学在籍の5,964名の心理学者を対象にQRPsの使用経験について調査
弁明可能性: 0=できない, 1=おそらくできる, 2=できる

人間の認知的バイアス

- 人間の判断と意思決定は合理的選択理論とは異なった方法で行われている.
- **Confirmation bias** 確証バイアス
 - 仮説や信念を検証する際にそれを支持する情報ばかりを集め、反証する情報を無視または集めようとしない認知的バイアス
- **Hindsight bias** 後知恵バイアス
 - 物事が起きてからそれが予測可能だったと考える認知的バイアス

QRPs 問題のある研究実践

表 2: 各 QRPs に関する自己告白率と弁明可能性評価

[John, et al. \(2012\)](#)

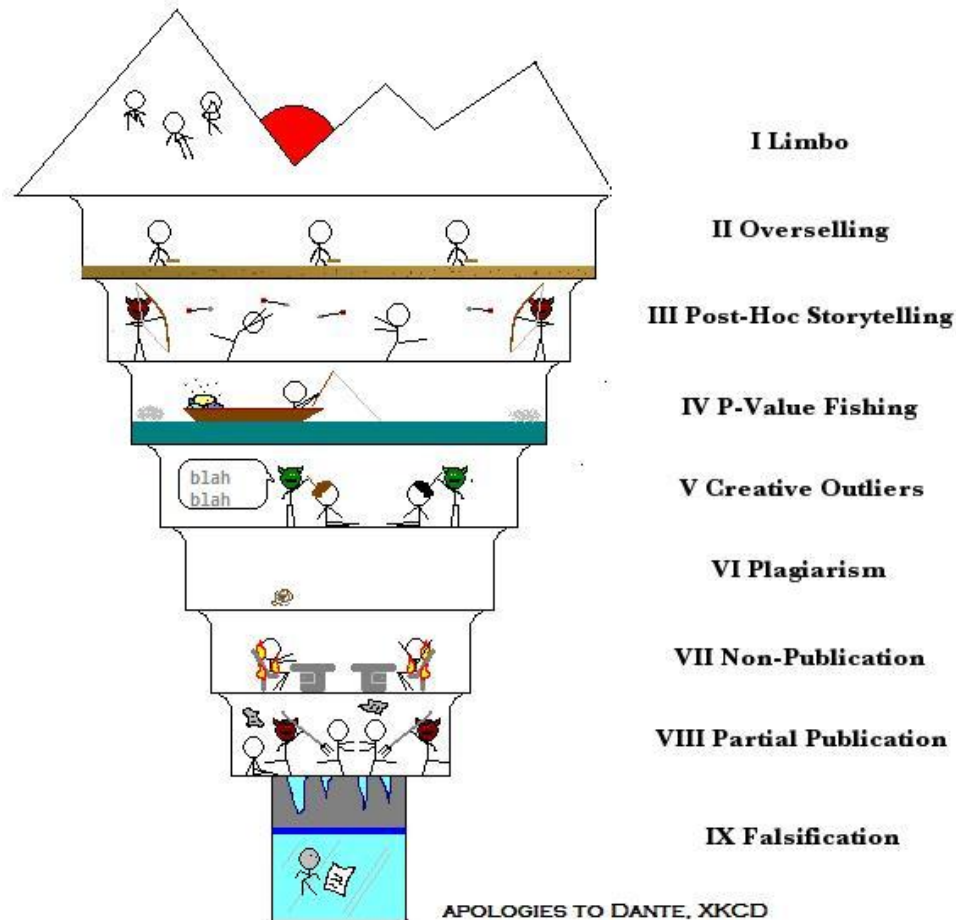
QRPs	自己告白率 (%)	弁明可能性
1. 従属変数を選択的に報告	66.5	1.84
2. 結果の誤った解釈	79	1.79
3. 実験の誤り	77	1.77
4. 望む結果のみに着目	76	1.76
5. p 値の誤った解釈	68	1.68
6. うまくいかなかった実験の報告を避ける	66	1.66
7. データの誤った解釈	61	1.61
8. 予想外の結果の報告を避ける	50	1.50
9. 実際よりも良い結果を報告	32	1.32
10. データの改竄	1.7	0.16

単に研究者の統計/倫理教育を徹底し、自覚を促したところで、QRPsの抑止にはつながりにくいだろう（実際、これまでの「是正」の努力はかけ声だけで終わってきた）。

なぜなら「やったほうがいいことがある」し、「やってもばれない」し、「やりたい気持ちを抑えて合理的に振る舞うことそのものが人間には難しい」のだから。

アメリカの主要大学在籍の5,964名の心理学者を対象にQRPsの使用経験について調査
弁明可能性：0=できない，1=おそらくできる，2=できる

QRPs 問題のある研究実践



Nine Circles of Scientific Hell ([Neuroskeptic, 2012](#))

我が身を顧みても、その罪の軽重によらず、研究を「円滑に」進めるためのヒューリスティックとして、問題のある研究実践が用いられてきたという歴史があることを認めざるを得ない。

→だからこそ「**革命**」，つまり研究者個人の努力目標設定（だけ）ではなく、**研究業界の構造変革**をしないと、本質的な問題への取り組みにつながらない。

研究業界の構造変革

もうあんなことやこんなことは、しない、させない
心理学の再構築を目指して

研究業界の構造変更

- 再現性の高い研究を蓄積できるような基盤を作るために、つまり、研究者個人に一任するのではなく、コミュニティとして対応→なるべくQRPを犯させないような制度を作る。
- 例えば、人倫にもとる研究を抑止するために倫理綱領や倫理審査があるように、不正をしたくなる土壌を改良できる制度を作ろう！

Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research



Edited by [Matthew C. Makel, PhD](#), and [Jonathan A. Plucker, PhD](#)

Pages: 323

Item #: 4318147

ISBN: 978-1-4338-2754-9


Publication Date: May 2017

Format: Softcover

Availability: In Stock

Also available on: [Amazon Kindle](#)

 [Examine or adopt this book for teaching a course](#)

 [View sample pages from this book](#)

研究業界の構造変更

表2 頑健性確保のため各ジャーナル、団体が制定した新ガイドラインの例

	Society for Personality and Social Psychology (Funder et al., 2014)	Psychonomic Society (The Psychonomic Society, 2012)	Psychological Science (Eich, 2014)
統計報告	<ul style="list-style-type: none"> 事前の検定力推定とサンプルサイズ決定方法の報告 効果量と95%信頼区間の報告 	<ul style="list-style-type: none"> 事前の検定力推定とサンプルサイズ決定方法の報告 効果量と95%信頼区間の報告と重視 	<ul style="list-style-type: none"> サンプルサイズ決定方法の報告 効果量と95%信頼区間の報告とメタアナリシスの奨励
QRPsの禁止	<ul style="list-style-type: none"> 修正なしに同一データに対して複数の検定を行うことの禁止 データ分析の結果を見てデータ収集を止めることの禁止 データ分析の結果を見て観察、指標、項目、条件、参加者、行った実験などを除外して報告することの禁止 	<ul style="list-style-type: none"> 修正なしに同一データに対して複数の検定を行うことの禁止 データ分析の結果を見てデータ収集を止めることの禁止 データ分析の結果を見て観察、指標、項目、条件、参加者、行った実験などを除外して報告することの禁止 測定したすべての従属変数の報告 共変量分析は事前に計画し、探索的であると示す 	<ul style="list-style-type: none"> データ収集停止ルールの報告 測定したすべての独立変数、指標の報告 当該論文の研究目的のために分析したすべての従属変数の報告 観察を分析から除外した場合、理由とともにそれを示す
その他	<ul style="list-style-type: none"> 付録での教示等の正確な報告 データ公開 追試の奨励 柔軟で多様な研究評価軸の維持 	<ul style="list-style-type: none"> 帰無仮説が棄却されなかった時は第二種の過誤を疑う ネガティブな結果報告の奨励 ベイズ統計など帰無仮説検定以外の統計手法も受け入れる 	<ul style="list-style-type: none"> 論文の方法と結果の字数制限を撤廃 データ公開の奨励 マテリアル公開の奨励 事前登録の奨励

池田・平石 (2016)

Preregistration 事前登録

透明性と開放性の確保によって、第三者が研究に関する（論文に掲載されている以外の）情報を詳しく知ることができ、研究の信頼性や妥当性を検証できる環境を整えたとしても、そこに QRP^{こと}s が含まれており、またそれが出版バイアスをくぐり抜けるために糊塗されていたとしたら意味がない。第9章で詳しく述べたとおり、QRP^{こと}s は研究に着手した後、データ収集から分析へと進めて行く過程で発生しやすいのだから、そこで不正が行なわれないためには、「正」の研究手続きをあらかじめ定めておき、これを第三者に向けて公開することが効果的だと考えられる。これが事前登録制度（pre-registration system）であり、これを出版する側のシステムに取り入れ、なおかつこの時点で審査を試みが始まっている。これが事前審査つき事前登録制度である。



<https://doi.org/10.1073/pnas.1708274114>



事前審査つき事前登録



事前審査つきの事前登録制度が画期的なのは、これまでの論文は研究が終了してから学術誌に投稿されるものだったのを、データ取得以前に投稿することを求める点である。つまり、研究の目的（論文の序論部分にあたる）、取得予定のサンプルサイズ（実験参加者や調査回答者の数）、研究計画、統計分析手法など（論文の方法部分にあたる）を投稿し、その内容が審査される。そして、審査を通過した研究は、その内容をタイムスタンプ（登録日時）とともに事前に登録することによって、その時点以降の改編が不可能になるよう「凍結」される。その後実際にデータが収集され、得られた結果がどのようなものであろうと、その報告がそのジャーナルに掲載される。

自由と引き換えに真実を手に入れよ!?

1950s

Technical, optimistic criticisms of NHST
e.g., calls for better applications and interpretations

Publication Bias
Sterling (1959)

1970s

The Meta-Analysis decade
(Glass & Smith; Schmidt & Hunter), NHST impeding cumulative progress

Law of Small Numbers
(Tversky&Kahneman, 1971)
Carver (1978), Kirk (1972)

1990s

Editorial, institutional intervention

26 new editorial policies - very limited impact
APA Task Force on Statistical Inference

2010s

Bem, Stapel

Simmons et al (2011) False +ve

Reproducibility/ Replication
Projects (COS),
Many Labs

Statistical Reform (in Psychology)

Credibility Revolution

1960s

A new type of critique, **broad pessimism about whether NHST fits the purpose of science** e.g.,

Rozeboom; Bakan; Lykken; Selvin; Meehl

Jacob Cohen, 1962
first statistical power survey

1980s

"the matter was settled" (Paul Meehl on NHST controversy at turn of t

Average st power und since Cohen (Sedlmeier & Gigerenzer, 1989)

2000s

APA Publication Manual revised following TFSI

Ioannidis (2005)

Bayesian approaches in psychology rise

Open Science (e.g., OSF)

SIPS

TOP guidelines

Collaborative/ Crowd-sourced science

ずっと問題にされ続けてきたのにずっと
追い求められ続けてきたNHST（帰無仮
説検定）が遂に...？

心理学における 再現可能性問題 － 概説 －

**きっかけ
大規模追試による検証
なぜ再現されないのか？
研究業界の構造変革**