

人を対象とした 行動学研究における 再現性問題

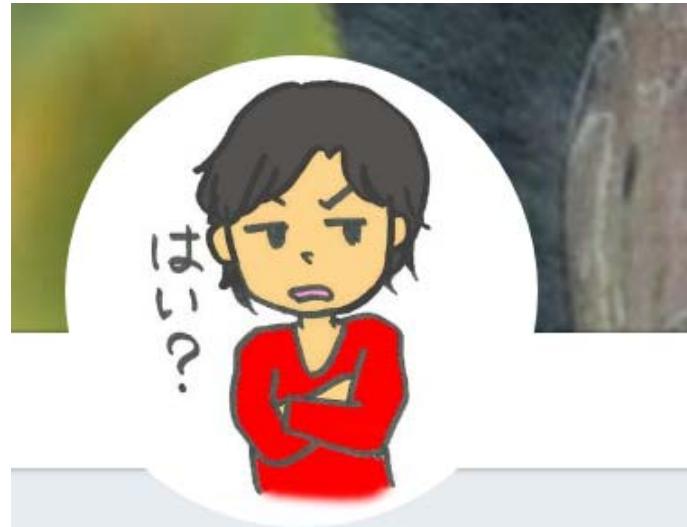
EC2018 招待講演

三浦麻子（関西学院大学・大阪大学）



講演者について

- 専門：社会心理学
- 所属：関西学院大学文学部総合心理科学科・大阪大学大学院基礎工学研究科
- コミュニケーションを通じて新しい「何か」が生まれるメカニズムに关心があり、インターネット上の人間行動に関する研究を多くしている



miura asako

@asarin

三浦麻子 a.k.a. asarin/社会心理学者/KG文心理+阪大Σ数理教員/関西生まれのインターネット育ち/ボーントラキチ/各所全力/当意即妙なボケツッコミ愛好/毒舌キャラ/論争嫌い/本質を見抜いてばすっと言いたい/マキャベリアン/人生直球勝負/ダメな狼が好き/hub/No data, No Life

◎ 京阪神一円

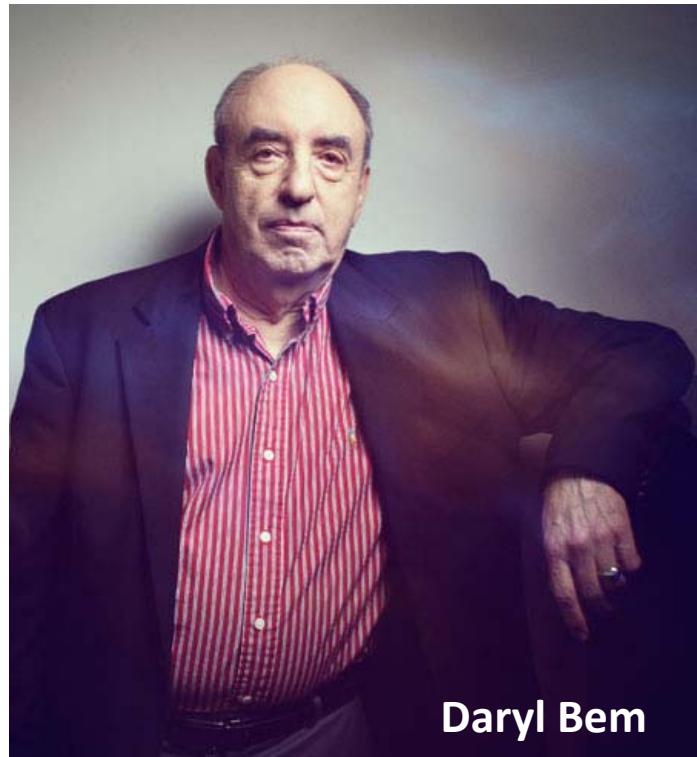
🔗 team1mile.com/asarinlab

📅 2007年4月に登録

講演概要

- 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. 本講演では，心理学のような**人を対象とした行動学研究において，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としてどんなものがあるか**を述べ，それらを**解消するために現に行われている取り組み**について紹介する.

心理学における再現性問題 きっかけ その①



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

心理学のトップジャーナル(JPSP)に,
かつて大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。

きっかけ その①

The screenshot shows a PLOS ONE article page. At the top, the PLOS logo and "A Peer-Reviewed, Open Access Journal" are displayed. Below the logo, there are links for "View this Article", "Submit to PLOS", "Get E-Mail Alerts", and "Contact Us". The article title is "Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect". The authors listed are Stuart J. Ritchie, Richard Wiseman, and Christopher C. French. The editor is Sam Gilbert. The article was published online on March 14, 2012, with the DOI [10.1371/journal.pone.0033423](https://doi.org/10.1371/journal.pone.0033423). The PMID is 22432019 and the PMCID is PMC3303812. A note at the bottom of the page states: "This article has been cited by other articles in PMC." The abstract section discusses nine recently reported parapsychological experiments that appear to support or describe three pre-registered independent attempts to exactly replicate or 'retroactive facilitation of recall', which examines whether performance influenced by a post-test exercise. All three replication attempts failed to (combined $n=150$; combined $p=.83$, one-tailed) and thus do not support

追試による再現性検証論文
(再現に失敗) をJPSPが「追
試は掲載しない」という理由
でリジェクトし大騒ぎに
↓
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

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

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

心理学における再現性問題 きっかけ その②



SCIENCE NEWS NOVEMBER 2, 2011 / 8:41 PM / 7 YEARS AGO

Dutch psychologist admits he made up research data

Kate Kelland

3 MIN READ



LONDON (Reuters) - A Dutch psychologist has admitted making up data and faking research over many years in studies which were then published in peer-reviewed scientific journals.

Diederik Stapel, a psychologist working at Tilburg University in the Netherlands, said he had "failed as a scientist" and was ashamed of what he had done, but had been driven to falsifying research by constant pressure to perform.

現役バリバリで活躍していた著名な社会心理学者が、
10年にわたりデータのねつ造・改ざんをしていたことがわかり、
論文は取り下げとなり、大学は解雇された。

きっかけ その②

RETRACTION

Retraction

D. A. Stapel, S. Lindenberg¹

* See all authors and affiliations

Science 02 Dec 2011;
Vol. 334, Issue 6060, pp. 1202
DOI: 10.1126/science.334.6060.1202-a

Article

Info & Metrics

eLetters

PDF

Our Report "Coping with chaos: How disordered contexts promote stereotyping and discrimination" (1) reported the effects of the physical environment on human stereotyping and discriminatory behavior.

On 31 October 2011, Tilburg University held a press conference to announce findings of its investigation into possible data fraud on the part of author Stapel. These findings of the university's interim report (2) included fabrication of data in this *Science* paper. Therefore, we are retracting the paper, with apologies from author Stapel. Coauthor Lindenberg was in no way involved in the generation of the data and agrees to the retraction of the paper.



Stapelは後に自身が繰り返した不正を詳細に記述し、
その時の心境を明かした書籍を出版している

ダメ押し

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

You are currently viewing the abstract.

[View Full Text](#)

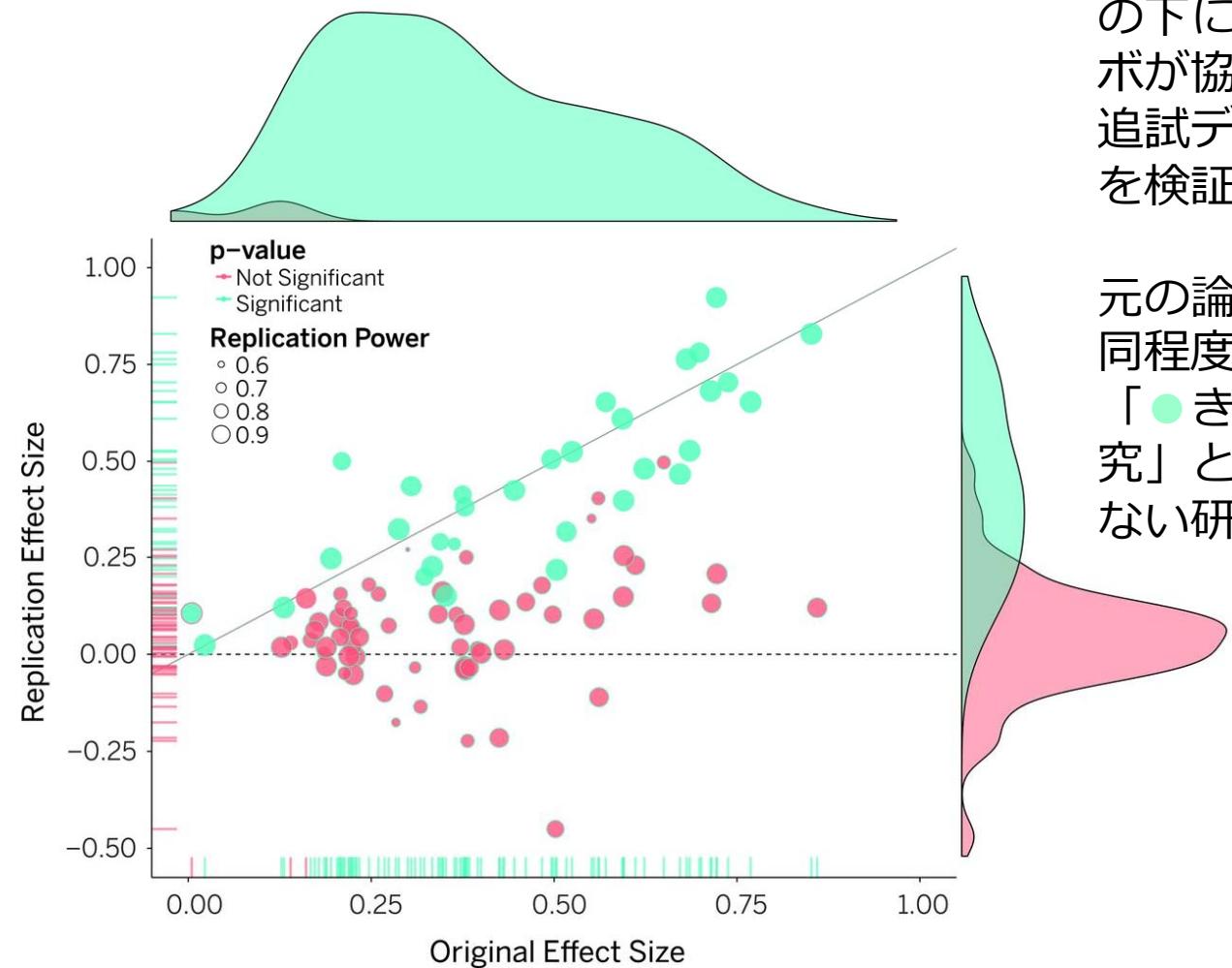


Empirically analyzing empirical evidence

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.

Science, this issue 10.1126/science.aac4716

ダメ押し



Open Science Collaborationの名の下に、世界各国の多くのラボが協力して、同じ手続きで追試データを収集し、再現性を検証

元の論文の効果量に関わらず、同程度の効果量が得られる
「●きちんと再現できる研究」と「●まったく再現できない研究」に二分される印象

そして、百家争鳴



Comment on "Estimating the reproducibility of psychological science"

Daniel T. Gilbert^{1,*†}, Gary King¹, Stephen Pettigrew¹, Timothy D. Wilson²
See all authors and affiliations
Science 04 Mar 2016;
Vol. 351, Issue 6277, pp. 1037
DOI: 10.1126/science.aad7243

Article

Abstract

A paper from the Open Science Collaboration (Research Articles, 28 August 2015, aac4716) attempting to replicate 100 published studies suggests that the reproducibility of psychological science is surprisingly low. We show that this article contains three statistical errors and provides no support for such a conclusion. Indeed, the data are consistent with the opposite conclusion, namely, that the reproducibility of psychological science is quite high.

eLetters

Info & Metrics

Article

Abstract

Gilbert et al. conclude that evidence from the Open Science Collaboration's Reproducibility Project: Psychology indicates high reproducibility, given the study methodology. Their very optimistic assessment is limited by statistical misconceptions and by causal inferences from selectively interpreted, correlational data. Using the Reproducibility Project: Psychology data, both optimistic and pessimistic conclusions about reproducibility are possible, and neither are yet warranted.

eLetters



TECHNICAL COMMENTS

Response to Comment on "Estimating the reproducibility of psychological science"

Christopher J. Anderson^{1,*}, Štěpán Bahnik², Michael Barnett-Cowan³, Frank A. Bosco⁴, Jesse Chandler^{5,6}, Christopher R. C...

*
See all authors and affiliations

Vol. 351, Issue 6277, pp. 1037
DOI: 10.1126/science.aad9163

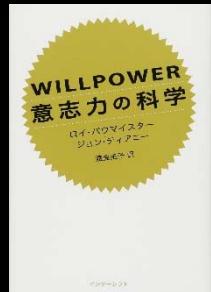
PDF

あの有名な実験も...!?

再現性に疑問符がついている
著名な研究として、

スタンフォード監獄実験
マシュマロ・テスト
自我消耗 ego depletion

を紹介



<https://www.nytimes.com/2018/07/16/health/psychology-studies-stanford-prison.html>

日本でも...

社会心理学研究の再現可能性検証のための日本拠点構築

研究課題/領域番号

15K13122

研究種目

挑戦的萌芽研究

配分区分

基金

研究分野

社会心理学

研究機関

関西学院大学

研究代表者

三浦 麻子 関西学院大学, 文学部, 教授 (30273569)

研究分担者

平石 界 慶應義塾大学, 文学部(三田), 准教授 (50343108)

樋口 匡貴 上智大学, 総合人間科学部, 准教授 (60352093)

藤島 喜嗣 昭和女子大学, 生活機構研究科, 准教授 (80349125)

研究期間 (年度)

2015-04-01 – 2018-03-31



日本でも...

心理学評論

Vol. 59 No. 1

2016

特集「心理学の再現可能性:我々はどこから来たのか 我々は何者か 我々はどこへ行くのか」

特集号の概要

巻頭言

目次(原著論文) ▾

目次(コメント論文) ▾



原著論文

- 池田功毅・平石界 心理学における再現可能性危機:問題の構造、現状と解決策
- 山田祐樹 認知心理学における再現可能性の認知心理学
- 森口佑介 発達科学が発達科学であるために:発達研究における再現性と頑健性
- 鮫島和行 システム神経科学における再現可能性
- 澤幸祐・栗原彬 動物心理学における再現可能性の問題
- 大久保街亜帰無仮説検定と再現可能性
- 小塩真司 心理尺度構成における再検査信頼性係数の評価—「心理学研究」に掲載された文献のメタ分析から—
- 藤島喜嗣・樋口匡貴 社会心理学における“p-hacking”の実践例
- 渡邊芳之 心理学のデータと再現可能性

コメント論文

- 小島康生 人間の観察研究における再現可能性の問題
- 松田一希 フィールド研究の再現性とは何か?
- 平井啓 心理学研究におけるリサーチデザインの理想
- 三中信宏 統計学の現場は一枚岩ではない
- 武田美亜 再現可能性の問題から始める心理学研究の「バックヤードツアー」
- 東島仁 研究公正から見た再現可能性問題
- 佐倉統 科学的方法の多元性を擁護する

<http://team1mile.com/sjpr59-1/>

心理学の再現性問題

- なぜ低いのか?
- 本質的な問題
 - 「強い理論」が少ない
 - 一方で（だからこそ）脆弱な「理論」が多く提案される
- 本質的ではない「はずの」問題
 - 不正をしたくなる土壌
 - 不正ができてしまう土壌
 - 抗うことが難しい「認知的バイアス」

不正をしたくなる土壤

私（スター・ペル）はフローニンゲン大学の優雅なオフィスに一人いる。研究データを入力したファイルを開く。そのデータの予想外だった数値である「2」を「4」に変えた。ヒョッとしてと思い、念のため、オフィスのドアを見たが、閉まっていて誰にも見つからなかった。データのたくさんの数値が並んだマトリックスを統計分析ソフトにかけるために、マウスをカチッと鳴らした。私が変更した新しい結果を見た時、世界は論理的状態に戻った。（ページ145）

最初は、論文の中のほんの少しのデータを見栄えよくしようとした、小さな改ざんから始まった。しかし、最後は、1つの論文のほぼ全部のデータをねつ造するまでに至った。まるで、麻薬におぼれるようにねつ造・改ざんの麻薬におぼれていった。研究ネットの動機はいろいろあるとスター・ペルは認めているが、次のような心理状態だった。

「高評価の必要、野心、怠惰、ニヒリズム、権力への欲求、ステータスを失う心配、問題を解決したい欲望、一貫性、出版プレッシャー、傲慢、情緒的孤立、寂しさ、失望、ADD（注意力欠如障害）、解答中毒」（ページ226）。

研究倫理（ネカト）白楽ロックビルのバイオ政治学
https://haklak.com/page_Diederik_Stapel.html

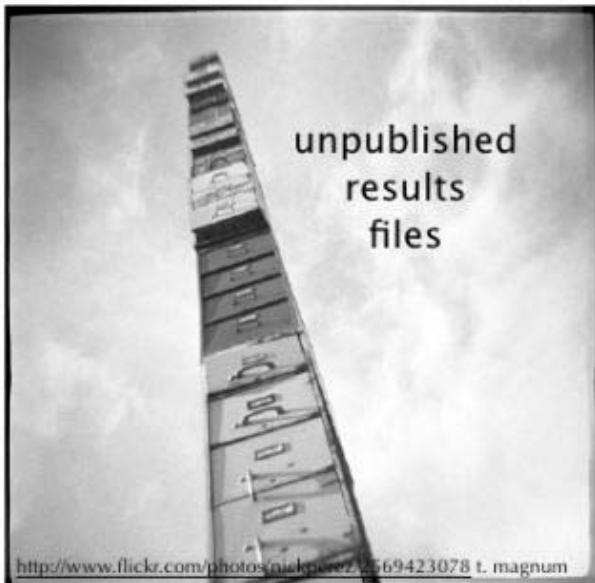


Diederik Stapel on the BrainTrain -- What I Lost And The Importance Of Being Connected

不正をしたくなる土壤

The File-Drawer Problem

「お蔵入り」問題



<http://www.flickr.com/photos/nickperez/2569423078 t. magnum>

- Difficult to publish non-replications and replications
 - Most journals only publish papers that “make a novel contribution”
 - Reviewers/editors tend to hold non-replicating manuscript to higher standard than original.



- Little career incentive to publish a non-replication or a replication

<https://www.slideshare.net/daphthorp/file-drawer-talk>

不正をしたくなる土壤

- 「お蔵入り」問題
 - 仮説が支持された研究でなければ公刊されにくいで論文にしない
→公刊されるのは仮説が支持された研究ばかりになる→仮説が支持されないデータは都合が悪い...という発想に辿り着いてしまう

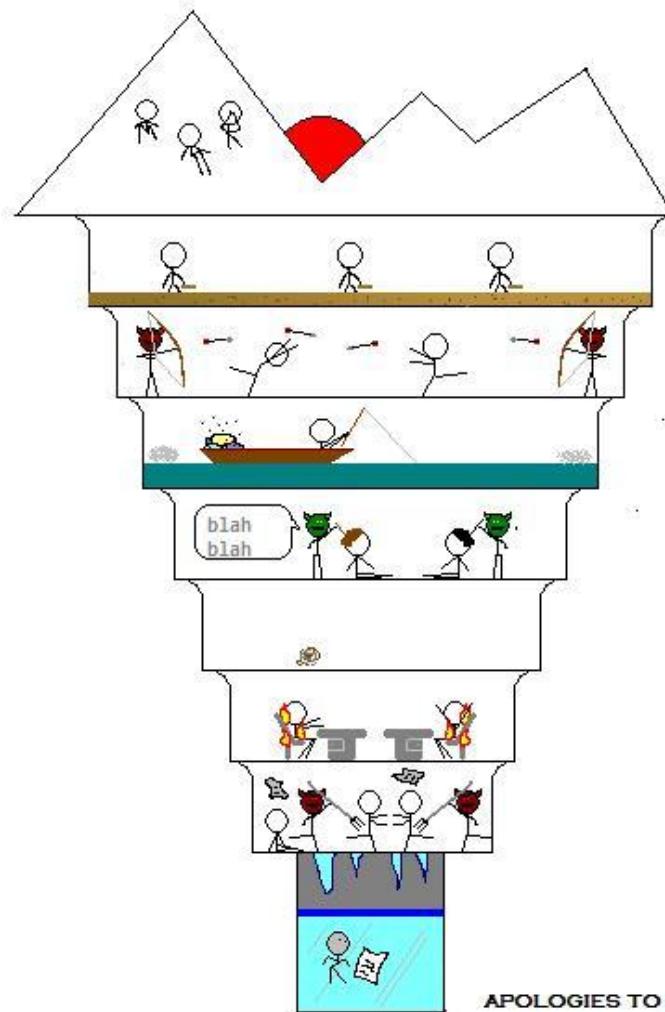
Table 3 Cross-tabulation between statistical results of TESS studies and their publication status (column percentages reported).

Pearson χ^2 test of independence: $\chi^2 (6) = 80.3, P < 0.001$.

Franco, Malhotra, & Simonovits (2014)
<http://science.sciencemag.org/content/345/6203/1502.full>

	Null (%)	Mixed (%)	Strong (%)
Not written	64.6	12.2	4.4
Written but not published	14.6	39.0	34.1
Published (non-top-tier)	10.4	37.8	38.4
Published (top-tier)	10.4	11.0	23.1
Total	100.0	100.0	100.0

The Nine Circles of Scientific Hell



Questionable Research Practices (QRPs)

I Limbo

問題のあるやり方を見て見ぬふり

II Overselling

過大に自分の研究を売り込む

III Post-Hoc Storytelling

後づけで話をつくる

IV P-Value Fishing

こっそりと望むp値を手に入る

V Creative Outliers

外れ値を都合のいいように使う

VI Plagiarism

剽窃・盗作

VII Non-Publication

都合の悪いデータを公開しない

VIII Partial Publication

都合のよいデータだけを選んで公開

IX Falsification

データの捏造

Neuroskeptic(2012)

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

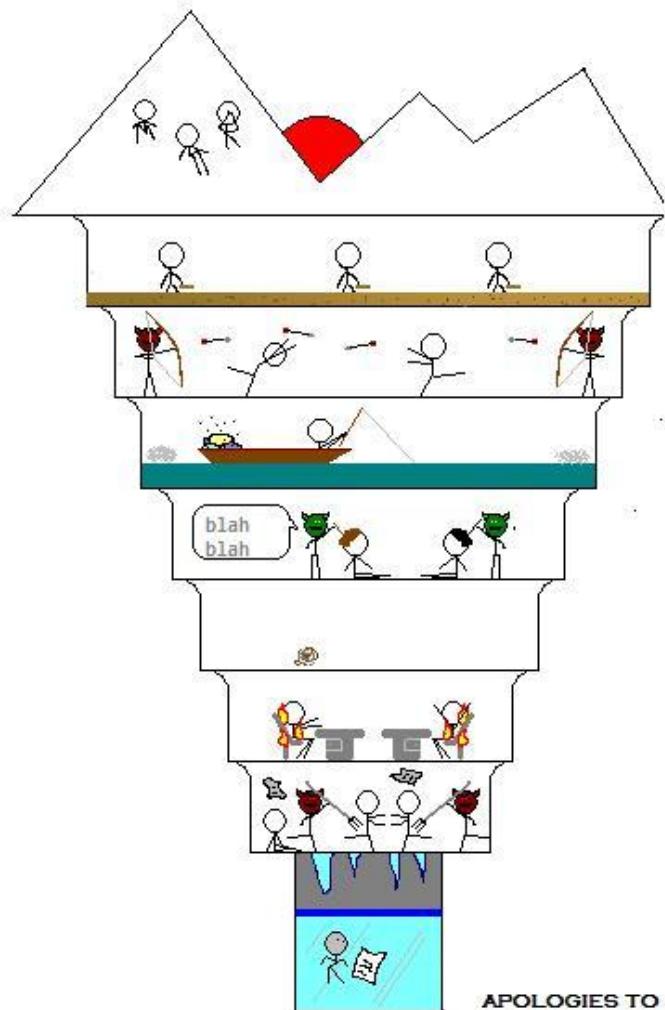
表 2: 各 QRPs に関する自己告白率と弁明可能性評定

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探る	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. <i>p</i> 値を「切り捨て」て報告（例：5.4%を 5%に）	22.5	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

John, et al. (2012)

不正だという自覚の薄い不正



I Limbo

問題のあるやり方を見て見ぬふり

II Overselling

過大に自分の研究を売り込む

III Post-Hoc Storytelling

後づけで話をする

IV P-Value Fishing

こっそりと望むp値を手に入る

V Creative Outliers

外れ値を都合のいいように使う

VI Plagiarism

剽窃・盗作

VII Non-Publication

都合の悪いデータを公開しない

VIII Partial Publication

都合のよいデータだけを選んで公開

IX Falseification

データの捏造

Neuroskeptic(2012)

p-hacking

Japanese Psychological Review
2016, Vol. 59, No. 1, 84–97

社会心理学における“*p*-hacking”の実践例¹⁾

藤島 喜嗣¹, 桶口匡貴²

¹昭和女子大学

²上智大学

Case studies of “*p*-hacking” in social psychology

Yoshitsugu FUJISHIMA¹ and Masataka HIGUCHI²

¹Showa Women's University

²Sophia University

There is currently an ongoing debate about reproducibility in social psychology. One reason for low reproducibility is the excessive use of questionable research practices, called “*p*-hacking”. We present two direct replication studies of social priming and embodied cognition that failed to replicate the original findings under the circumstances of high statistical power. However, a variety of *p*-hacking attempts made it possible to obtain some false-positive findings based on the data from these two studies. We note that selectively reporting the results and deriving the hypothesis after the results are obtained may disguise the presence of *p*-hacking, and argue that pre-registration of studies and fair publishing of negative results could inhibit *p*-hacking.

Key words: reproducibility, *p*-hacking, direct replication, pre-registration, publication bias

キーワード：再現性, *p*-hacking, 直接的追試, 事前登録, 出版バイアス

p-hacking

研究再現性問題 特集

アスタリスク～真実の石を求め（すぎ）て



樋口 匠貴（ひぐち まさたか）
広島大学人間科学研究科博士課程後期修了。
博士（心理学）。現在、上智大学総合人間科学部
教授。現在の主な関心は社会的感情の機能について。
最近の主な著書に「*社会と健康的心理学*」（共
著者、ナカニシヤ出版）がある。



藤島 富嗣（ふじしま ゆしつ）
一橋大学人間社会学研究科博士課程単位取得
退学、修士（社会学）。現在、昭和女子大学人間
社会学部教授。主な研究領域は社会的認知、対人
関係における自己や他者の内的状態の推測に關心
がある。著書に「*現代の認知心理学6 社会と感情*」
（共著者、北入出版房）がある。

樋口・藤島（2018）

QRPs の実際

この QRPs について、我々著者らは「実際に QRPs をやってみた！」という報告を行っている^[4]。ここではその試みを簡単に紹介しよう。

一つ目の事例は、ウイリアムズとバージ^[4]による、空間的距離のプライミングが感情反応の強さに影響するとした研究の追試である。ウイリアムズとバージは、空間的距離が遠く知覚される事象には感情反応が弱まり、近くに感じられる事象には感情反応が強まると考えた。すなわち、遠距離プライミングをすると不健康な食品に対しても健康的な食品に対してもカロリーの推定値は変わらないと予測されるのに対し、近距離プライミングをすると不健康な食品に対してよりカロリー推定値が高くなると予測したのである。元研究では、実験参加者に対して方眼紙上に遠い距離の2つの点（遠距離プライミング）または近い距離の2つの点（近距離プライミング）、あるいは中程度の距離の2つの点（中距離プライミング）のいずれかを記入させた。その後、アイスクリーム、チーズバーガー、リンゴといったいくつかの食品の写真を見せ、そのカロリー量の推定を求めた。元のこの研究自体は、予想通りの結果を得ている（だからこそ刊行されたのだ）。

それに対して藤島と樋口^[4]は、この研究の直接的追試を3研究機関の合同で行い、元研究通りの分析結果をまず報告している。結果は、条件間の差異は見出せず、再現失敗であった。当然のように健康的な食品（リンゴやヨー

グルト）は不健康食品（チーズバーガー やフライドポテト）よりもカロリー推定値は低くなっていた。その一方、空間的距離のプライミングによってカロリー推定値が変化するという結果は得られなかつたし、元研究が示したような健康食品なのか不健康食品なのかによる効果の違いも見られなかつた。

ここからが本番である。この通常の分析に加え、藤島と樋口^[4]は、QRPsをいくつか実践した。まず、研究機関および食品の分割を行つた。研究機関は3箇所、食品は健康食品と不健康食品それぞれ5品目の合計10品目が用いられていたが、まず1研究機関のデータのみを用いることとした。さらに食品別に行った予備分析の結果から、予測と一致しそうな平均値パターンを示す食品のみに限定して分析を行つた。分散分析を行つたところ、「有意に近い」交互作用効果 ($p = .06$) が認められ、その結果から藤島と樋口は「この結果は元の研究の仮説を支持するものであった」(p.88) と結論づけた。さらに3条件存在した空間的距離プライミングの条件を2つに限定して分析を行つた結果、「有意な」交互作用効果が認められ ($p = .02$)、「この結果は元の研究の仮説を支持した」^[4] (p.88) と結論づけた。ついに彼らは探し求めたアスタリスクを手に入れたのである。

この追加分析を行うにあたり藤島と樋口はいちいちもっともらしい理屈をつけていた。たとえば、食品の分割においては「予測通りの結果を示すのは、間食に相当する3品目であり、健

p-hacking

p 値ハッキングが行なわれがちなのは、当初想定した十分な数のデータをとった時点で分析をしてみると「仮説が支持されたと断言はできないが、まったく支持されなかったと言うのもどうかと思うような微妙な結果」、つまり *p* 値が 0.05 をわずかに上回るような結果が得られた場合である。こういう結果を報告すると、こんな「アドバイス」をくれる指導教員や先輩がいるかもしれない。

「じゃあ、もうちょっとデータ足してみたら？」

データを足すこと自体に問題があるとはいえないが、その目的が「仮説が支持された」という確証を得るためにあれば問題である。アドバイスを得た学生が、データを 1 人分足しては検定をし、また足しては検定し、を繰り返したあげく、ある時点で *p* 値が 0.05 を下回ったとしよう。そのとき学生はどうするだろう。ここでデータ収集を打ち切ってしまえば、

「追加で 5 人参加してもらったらちゃんと有意になりました！」

「おおそうか、よかったよかった。やっと論文が書けるな！」

データ（参加者）を1名ずつ足していくって、有意になつたらそこでストップ、という方法で「偶然有意になる」結果を得る確率は「案外」高い

「これは、全体としては有意差が見られなくても、参加者のパーソナリティによっては差があつたりするんじゃないかな？」

という「アドバイス」もありがちだ。これもその影響を検討すること自体に問題があるとはいえないが、その目的が「仮説が支持された」という確証を得るためにあれば問題である。仮説として想定していたわけではないが、参加者のパーソナリティに関するデータを同時にとつておいて、その影響を加味した分析をすると、

「外向性による違いはなかったんですが、誠実性が高いと差がありました！」
「おおそうか、よかったよかった。やっと論文が書けるな！」

となる。「差を出す」と褒められるのだ。

拙著『なるほど!心理学研究法』

第9章「研究倫理:研究者として「やってはいけないこと」」

False-Positive (偽陽性)

本来は帰無仮説が真なのに
偽として棄却してしまう
(本来検証したい仮説を採
択してしまう) 誤り

Unlikely results

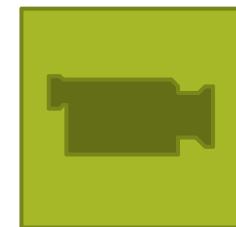
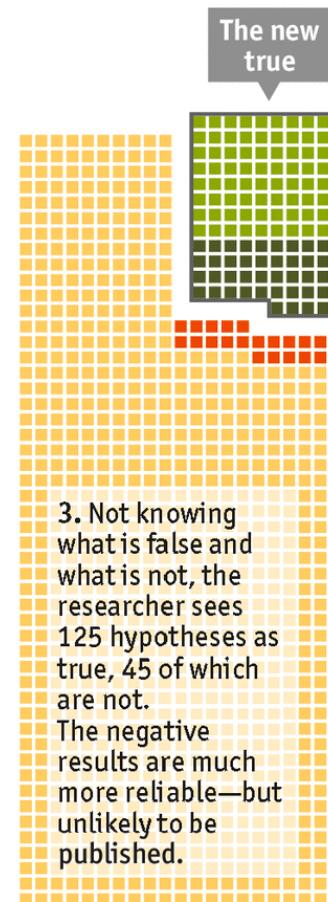
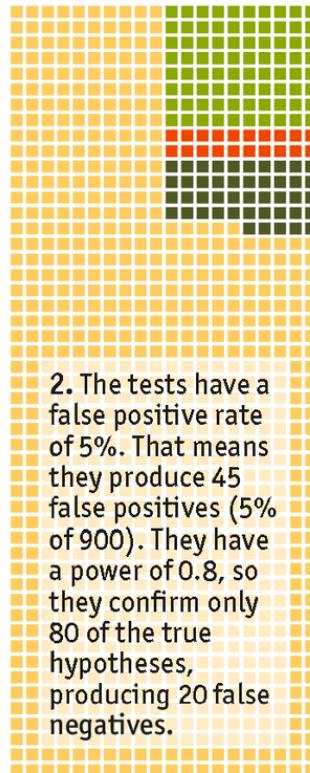
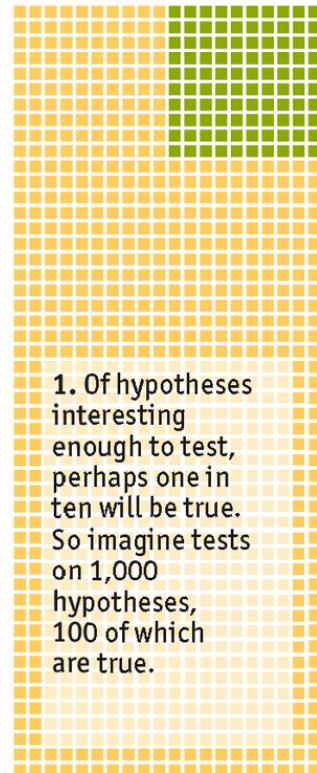
How a small proportion of false positives can prove very misleading

False

True

False negatives

False positives



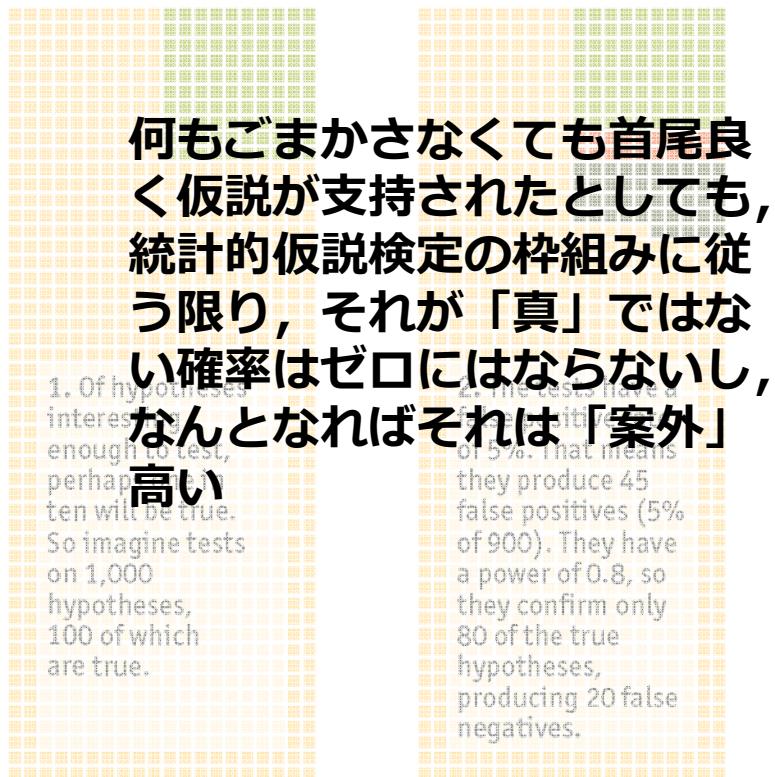
Source: *The Economist*

False-Positive (偽陽性)

Unlikely results

How a small proportion of false positives can prove very misleading

■ False ■ True ■ False negatives ■ False positives



Source: *The Economist*

表1 様々な p 値、検定力、仮説が正しい確率において、有意な結果が偽陽性である確率

統計 検定力 (%)	有意となった結果が偽陽性である 確率 (%)		
	$p=.05$	$p=.01$	$p=.001$
仮説のうち 80% が正しい場合			
20	5.9	1.2	0.1
50	2.4	0.5	0.05
80	1.5	0.3	0.03
仮説のうち 50% が正しい場合			
20	20	4.8	0.5
50	9.1	2	0.2
80	5.9	1.2	0.1
仮説のうち 10% が正しい場合			
20	69.2	31	4.3
50	47.4	15.3	1.8
80	36	10.1	1.1

註：Sterne & Smith (2001), Table 3 の一部を改変して引用。

池田・平石 (2016)

False-Positive (偽陽性)

表 9-2 偽陽性の有意差を得る確率 (池田・平石, 2016)

研究者の自由度	$p < .05$
状況 A : 2 種類の従属変数を分析 ($r = .5$)	9.5%
状況 B : 有意差が出なかった場合、各条件ごとに 10 回の観察を追加していく	7.7%
条件 C : 性別の主効果か性別 × 操作の交互作用を共変量としてコントロール	11.7%
状況 D : 3 条件を測定するも、2 条件のペアの結果しか報告しない	12.6%
A と B を組み合わせた場合	14.4%
A と B と C を組み合わせた場合	30.9%
A と B と C と D を組み合わせた場合	60.7%

Simmons, Nelson, Simonsohn(2011)を改変

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定

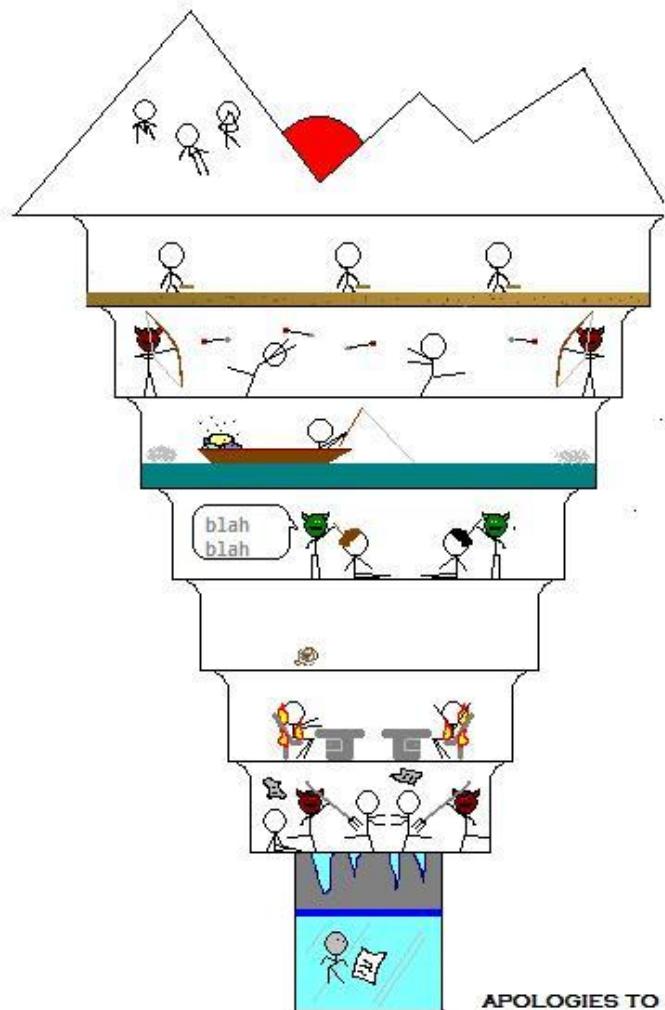
QRPs	自己告白率 (%)	弁明可能性
1. 従属変数を選択的に報告	66.5	1.84
2. 結果が有意かどうかを見てからさらにデータを探る	58.0	1.79
3. 実験条件の選択的報告	27.4	1.77
4. 望む結果が出たので予定より早くデータ取得を終了	22.5	1.76
5. p 値を「切り捨て」て報告（例：5.4%を 5%に）	23.3	1.68
6. うまくいった研究だけを選択的に報告	50.0	1.66
7. データ除外の影響を見てから除外するかどうかを決定	43.4	1.61
8. 予想していなかった結果を、予測していたかのように報告		1.50
9. 実際は不確かなの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

HARKing

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

John, et al. (2012)

不正だという自覚の薄い不正



I Limbo

問題のあるやり方を見て見ぬふり

II Overselling

過大に自分の研究を売り込む

III Post-Hoc Storytelling 後づけで話をつくる

IV P-Value Fishing

こっそりと望むp値を手に入る

V Creative Outliers

外れ値を都合のいいように使う

VI Plagiarism

剽窃・盗作

VII Non-Publication

都合の悪いデータを公開しない

VIII Partial Publication

都合のよいデータだけを選んで公開

IX Fabrication

データの捏造

Neuroskeptic(2012)

HARKing

- **Hypothesizing After the Results are Known** 仮説の後付け
 - 出た結果に合わせて仮説を作り替える
 - 心理学研究で対象とする場面には、実験的に操作したり（できたり），そもそも主たる関心としていた独立変数以外にも，従属変数に影響しそうな剩余変数が（無数に）想定可能
 - それらをあれこれとつぶやいて，事後分析で影響をもつことが分かった変数を，あたかも当初から主たる関心のひとつとしていたかのようにストーリーを再構築する

HARKing

HARKingの弊害は様々に論じられているが、おそらく最大の問題点は、それが第一種の過誤を増大させてしまう点にあると思われる。仮に、得られたデータに対して、様々な共変量の投入などの分析を複数行ったとしよう。そして多重比較の修正を無視すれば、その分析のひとつで有意差が得られたとしよう。ここでHARKingを用いず、当初の理論的予測に忠実に従った報告を行うならば、多重比較問題の無視が明らかとなってしまうため、報告の信頼性は損われ、論文の出版につながらないかもしれない。ところがここでHARKingを使えば、当該の有意差が得られた検定が、あたかも当初から予測されていた唯一の分析であるかのように扱うことができるため、多重

比較の問題を「隠蔽」することが可能になる。さらに、心理学の弱い理論のもとでは、事前に厳密な理論的予測を立てることが難しいという事実を考えると、こうして「隠蔽」された多重比較を、論文報告のみから見抜くことは非常に難しいと思われる。このようにHARKingには、言わば「見かけの研究者の自由度」を低め、第一種の過誤の発生確率を高める危険性がある。

池田・平石(2016)

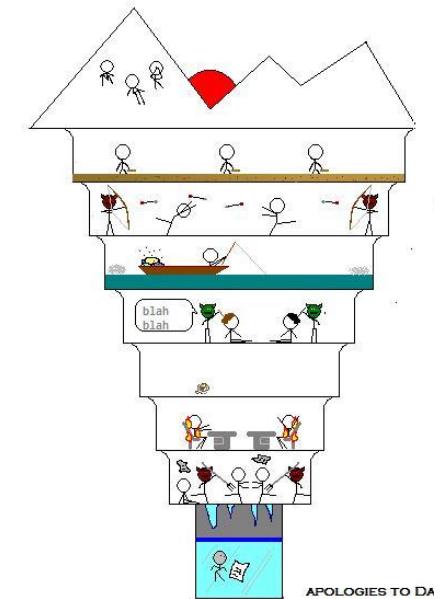
不正をしやすくしている土壤

- 研究というブラックボックス

今まで誰も私の研究過程をチェックしたことがない。彼らは、私を信頼し、私は、すべてのデータを自分自身で作ったのである。あたかも、私の隣にクッキーが入った大きなジャーがあり、母はいない、鍵は掛かっていない、ジャーには蓋もない状況だった。

手を伸ばせばすぐ届く範囲に、甘いお菓子が一杯に詰まったクッキーの大きなジャーがあった、その場所で私は毎日、研究をしていた。クッキーの大きなジャーの近くには誰もいなかつたし、監視もチェックもされなかつた。私がする必要なことのすべては、手を伸ばしてクッキーを取るだけだった。 (ページ164)

研究倫理（ネカト）白楽ロックビルのバイオ政治学
https://haklak.com/page_Diederik_Stapel.html



- それでもさすがに、事後のデータ改竄や捏造はまずい、という程度の自覚は、ほとんどの研究者にある。しかし多くのQRPsは...

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

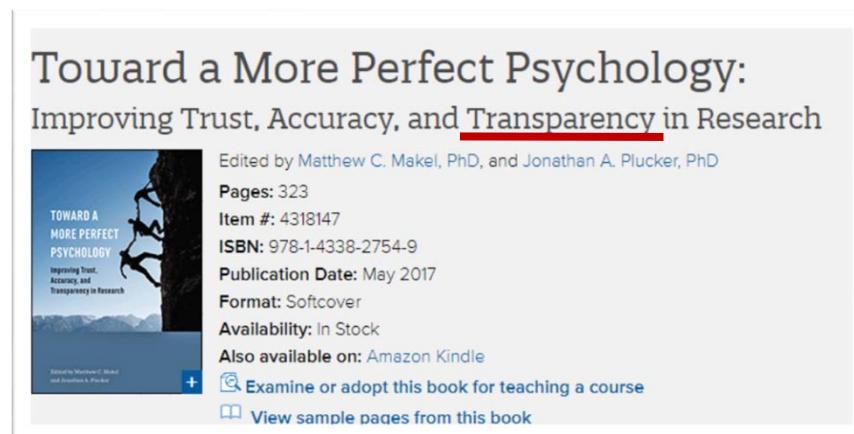
→**False Positive**

- 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

→**HARKing**

本質的ではない（はずの）問題への対応

- 研究者個人にそれを一任するのではなく、コミュニティとして対応→なるべくQRPs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壤を改良できる制度を作ろう！



学会の対応

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ペ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)

学会の対応(APS)



Registered Replication Reports

Multi-lab, high-quality replications of important experiments in psychological science along with comments by the authors of the original studies.

学会の対応(BPS)

We're offering Registered Reports across all eleven of our academic journals

13 July 2018

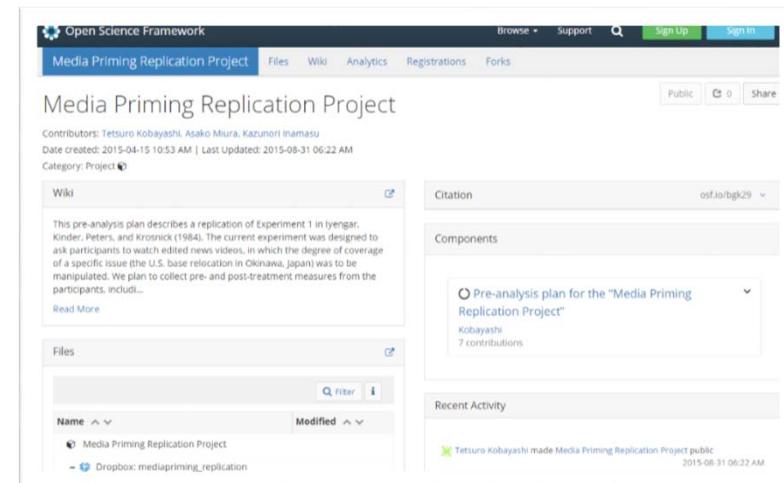
You can now improve the replicability of your work by publishing a Registered Report with the British Psychological Society



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QRPs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、QRPs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する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。

事前審査つきの事前登録制度が画期的原因は、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、その時点以後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。



Open Science Frameworkを利用したプレレジ
(事後にKobayashi, Miura, & Inamasu(2017)として公刊)



Nosek et al.(2018)

Preregistration 事前登録



事前審査つきの事前登録制度が画期的原因は、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、その時点以後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。

<https://cos.io/rr/>



なかなかたいへんらしい...

汚されてしまった私っぽい何か



Yuki Yamada
2018/05/17 03:18

山田祐樹氏（九大）のnote記事
[https://note.mu/momentumyy/n/n73a34fcc528b
?magazine_key=m77f2becf2d56](https://note.mu/momentumyy/n/n73a34fcc528b?magazine_key=m77f2becf2d56)

とはいえ、制度を利用するか
どうかによらず、自戒のため
にも、できる努力はすべき

2. 当方初の審査付き事前登録研究 (registered reports; レジレポ) であったこと。追試は何度もやってきたのですが、レジレポは初めてでした。プレレジ革命だからみんなやろうぜ！と言う分にはいいんですが、それがどんなもんか実際に体験する必要がありました。感想としては、まじキツ。レジレポは第1段階の査読をクリアすると結果がどんなんであれ掲載させるよってやつなんですが、けっこうキツツい査読を受けて第1段階でアクセプトされた際には、クククもらったぜッと喜んだものでした。しかし結果が出揃って論文を完成させた後の第2段階の査読は、単に結果の陽性性を評価しないというだけで普通にガチでした。コメントが数十箇所つきました。まあつまり査読のみで考えると、通常の倍キツイです（普通は論文完成後の1回だけだからね）。キツイだけじゃなく、査読期間も倍なので時間かかります。このあたりはなんとかしないと絶対に流行らんだろうなと確信しました。これを今後の学術出版の主流にしていくには相当の工夫が必要ですね。ちなみに雑誌側も我々の論文が初めてのレジレポだったようです。お互い貴重な経験になりました。

再現性replicability検証を可能にさせる 再生可能性reproducibilityの保証

- 研究の透明性を着手前に確保することで、研究者が実施過程で無用な罪を犯すリスクを低減させると同時に、手続きを詳細に報告する（あるいはデータも公開する）ことで研究の再生可能性を高め、誰でも再現性を検証できるようにする

再現可能性のすゝめ—RStudioによるデータ解析とレポート作成—



高橋 康介 著・石田 基広 監修・市川 太祐・高橋 康介・高柳 慎一・福島 真太朗・松浦 健太郎 編
シリーズ名 Wonderful R 【3】巻
ISBN 978-4-320-11243-8
判型 B5
ページ数 184ページ
発売日 2018年05月11日
本体価格 2,500円



新刊

データ解析とレポート作成の再現性を高めよう。その先には、ボスの笑顔が待っている。
RStudioを使いこなそう。その先には、旨いお鮒が待っている。
Rマークダウンをマスターしよう。その先には、明るい未来が待っている。

本書はRによる再現可能なデータ解析とレポート作成を身につけるための一冊である。Rがデータ解析のデファクトスタンダードとなった現在、Rの使い方やRによるさまざまなデータ解析手法についての情報は数多く入手できる。しかしどれだけ立派なデータ解析であっても、結果が再現できなかったり、間違いだらけだったり、時間がかかりすぎて完成しなかったりしたら、価値は半減である。本書ではRStudioとRマークダウンについての解説を通して、再現性と信頼性を高め効率を上げるような「データ解析とレポート作成というプロセス」について解説している。データ解析に携わるすべての人に、再現可能性の意義を学び、RStudioとRマークダウンによる再現可能なデータ解析とレポート作成を日々の習慣としてほしい。

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^{10,16}, Michael Kirchler^{3,5,16}, Gideon Nave^{6,16}, Brian A. Nosek^{10,7,8,16*}, Thomas Pfeiffer^{10,16}, Adam Altmejd^{10,2}, Nick Buttrick^{7,8}, Taizan Chan¹⁰, Yiling Chen¹¹, Eskil Forsell¹², Anup Gampa^{7,8}, Emma Heikensten², Lily Hummer⁸, Taisuke Imai^{10,13}, Siri Isaksson², Dylan Manfredi⁶, Julia Rose³, Eric-Jan Wagenmakers¹⁴ and Hang Wu¹⁵

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

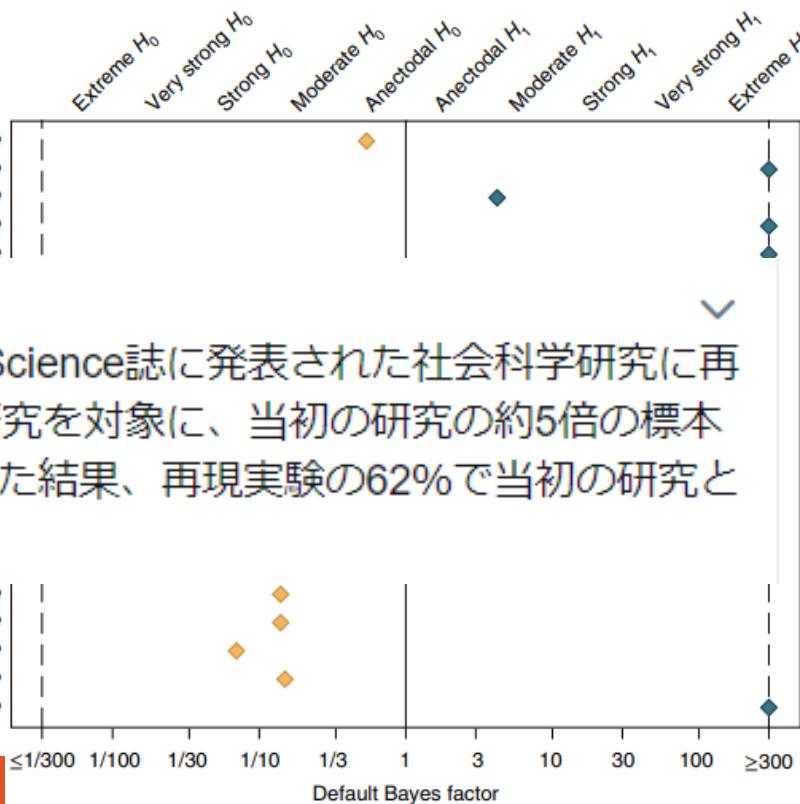
「これならちょっと安心」でただ先に進むというのではなく、再現実験をするのは当たり前、何か問題があれば蓋をせずにみんなで解決、という文化が根付くことが重要



Nature Japan @NatureJapan · 2分

Nature Human Behaviour 「Nature誌やScience誌に発表された社会科学研究に再現性はあるか？」21件の実験社会科学研究を対象に、当初の研究の約5倍の標本数を用いて検出力の高い再現実験を試みた結果、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。

Ramirez and Beilock (2011)³², *Science*
Rand et al. (2012)³³, *Nature*
Shah et al. (2012)³⁴, *Science*
Sparrow et al. (2011)³⁵, *Science*
Wilson et al. (2014)³⁶, *Science*



For details:

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

参考情報

- ・『心理学評論』「心理学の再現可能性」特集号
- ・ヒューマンインターフェース学会誌2018年度1号「特集：研究再現性問題」
- ・サイナビ!「心理学研究は信頼できるか? ——再現可能性をめぐって」
- ・拙著『なるほど! 心理学研究法』北大路書房

心理学評論

Vol. 59 No. 1

特集：心理学の再現可能性
編集：友永雅己・三浦麻子・針生悦子

2016

