

新しく古い心理学のかたち

高橋 康 介
中京大学

Old and new direction in psychological science

Kohske TAKAHASHI
Chukyo University

In this commentary paper, I discuss the significance of building an evidence evaluation system to evaluate the strength of evidence and the consistency between phenomena and theories in psychological science. I also discuss the usefulness of introducing the concept of “evidence level,” which is popular in evidence-based medicine. We should review the evidence-level further by referring to various external factors involving the theoretical predictability. In addition, transparency in the review process and sharing the evaluation criteria would be important.

Key words: reproducibility crisis, evidence level, GRADE, theory

キーワード：再現可能性の危機, エビデンスレベル, GRADE, 理論

はじめに

著者は現在こそ心理学部に所属して研究教育に勤しんでいるものの、もともとは文学部（宗教学専攻）から情報学という経歴で、体系的な心理学の教育を受けたことはない。専門は何かと問われれば「認知心理学」と答えるところではあるが、最近では霊長類学者や人類学者と一緒にフィールドワークを行うなど（高橋, 2018a）、とても心理学のメインストリームに位置するとは自分自身でも思っていない。そんな私が「心理学研究の新しいかたち」についてコメントを発するというのは、非常におこがましい気持ちで一杯である。

一方で、同僚の池田功毅氏と一緒に「TODARADIO」¹⁾なるコンテンツを発信して再現

可能性問題について語っていたり、R/RStudioを使ったデータ解析の再現可能性についての書籍（高橋, 2018b）を執筆したりと、はからずも心理学の再現可能性問題に触れる機会が少なからずあったことも事実である。

特にフィールドでの実験研究（Takahashi, Oishi, & Shimada, 2017）の再現性問題は悩ましい²⁾。フィールドワークはいわば一回性の学問であり、目下のところ話題となっている再現性と親和性は低い（小島, 2016）。私自身、一方でフィールドワーク、そしてもう一方で再現性問題に触れる中で、人類学者や霊長類学者とも議論を重ねながら、再現性問題を中心に心理学のエビデンスをどのように捉えればいいのかということにずっと頭を悩ませてきた。エビデンスというものの意味について考えざるを得ない状況にあった。そのような立場にいる私だからこそ書けることがあるかもしれない。本稿はそのような立場からのコメント論文である。

1) TODARADIOという命名は、著者と池田氏が所属する中京大学にも縁の深い心理学者、戸田正直氏（1924-2006）に由来する。著者自身、戸田先生が遺された言説、理論に大いに惹かれ、著作すべてを古本屋などで収集しているほどである。再現可能性問題を直接扱った議論は見当たらないが、その思想は再現可能性問題に立ち向かう若い心理学者にも大いに参考になるはずである。詳細については以下のウェブサイトにてラジオを聞いてほしい。特に#002の冒頭では命名の由来などが語られている。 <https://kohske.github.io/research/TodaRadio/index.html>

2) この論文の研究結果が再現できないということではない。しかしフィールド実験の直接的追試は環境面、コスト面において現実的には厳しい。

予知能力論文はなぜ掲載され、 そして受け入れられなかったのか

ご存知のとおり、2011年に「予知能力は存在する」という論文（Bem, 2011）が権威ある社会心理学ジャーナルに掲載され、その後大きな物議を醸した。当時、掲載されたという事実は私自身も認識していたが、掲載されたからといって予知能力が存在することが事実だとは思いませんでしたし、出版システムのどこかに不備があるか（実際、ソーシャル事件的なパフォーマンスだろうと思っていた）、話題作りに過ぎないと考えて特に気にもかけていなかった。おそらく多くの心理学者は同じような感想を抱いたのではないだろうか。同様に今後「テレパシーは可能」とか「透視できる」とか、そういった話が権威ある学術雑誌に掲載されたとしても、鵜呑みにして信じる人は少ないだろう（これらのトピックが実験超心理学分野で真剣に取り組まれていることは認識している）。

なぜ予知能力論文は掲載され、しかし受け入れられなかったのか。掲載された理由は単純で、雑誌の定める（当時の）掲載基準を満たしていたからに過ぎない。ここでの基準とは「ストーリーのインパクト」「統計的有意性」といったものであろう。ではなぜ受け入れられなかったのか。研究者であれば、論文掲載が決してその内容の真実性を保証するわけではないことを知っている、はずである。鮫島の言葉を借りれば『科学というのは、なにか唯一無二の「真実」を「発見」することではなく、人が自然を理解するために世界の見方やモデルを仮説として構築・提案し、その証拠を様々なアプローチで検証・更新するというプロセス』（鮫島, 2016）であるから、権威ある学術雑誌に掲載された論文であっても真偽の判断を保留する権利は誰にでもある。その中で「予知能力」の話は、あまりにも「あり得ない」ので、受け入れられなかった。

ではなぜこの話が、あり得ないのか。その「あり得なさ」はどこに起因するのか。結論を言えば、それは既存の確立された理論や現象的経験からあまりにも逸脱しているからであろう。仮に予知能力を真実として受け入れるのであれば、マクロな時間が一方向に進むという物理学の理論を書

き換える必要がある。現象的経験からも予知能力を支持するようなエビデンスは薄弱である。そういったものが重なり、「予知能力」は、仮に論文に掲載されたとしても、直感的に否定される。幸いだったことは「予知能力」の場合は既存の強力な理論との接点があったということである。「予知能力」という魅力的な仮説とそれをサポートするデータは、既存の理論体系とうまく接合しない。このような理論との接合、不接合による仮説の検証と構築はまさに科学の醍醐味であり、肩を登れる巨人はこのようにして作られなければならない。そして心理学にとってまさに不幸（あるいは不満と言うべきか）であったことは、「予知能力」と接合しないその既存の理論が心理学の理論ではなく、物理学の理論であったということである。

老人プライミング（Bargh, Chen, & Burrows, 1996）を含む社会的プライミングも再現性の危機の渦中にある。老人プライミングはしかし「予知能力」とは異なり発表されるとともに受け入れられないということはなく、魅力的な現象として心理学の一部で強い影響をもたらしてきた。今でこそその効果の疑わしさが顕在化しているが、多くの人に「ありえること」として受け止められ、この現象から様々な議論が展開されてきた。社会的プライミングの真偽はここでは問わないが、この経緯から理解するに老人プライミングは、「予知能力」のケースで起こったような理論的不整合を引き起こさなかったということになる。第一に、老人プライミングに矛盾するような物理学の理論は存在しない（もちろんこの現象を予測する物理学の理論も存在しない）。そして心理学の理論、例えば観念運動の理論と決して矛盾するものではない（再びもちろん、観念運動理論は老人プライミングを予測するものでもない）。心理学的な理論は非常に寛容であり、現象による反証可能性が低く（竹澤, 2018）、言い換えれば新規な現象を排除する力が弱い。新たな一つの魅力的な現象の報告は、理論的検証にさらされることなく、統計的有意性により信頼済みマークの付いたエビデンスとなり、時には理論のようなものが作られ、時にはNull Field（他の理論と接合していないという意味でNull Islandと呼ぶべきか）となっていく（池田・平石, 2016）。

医学におけるエビデンスレベル

ここで少し話題をかえて、医学（特に臨床医学）におけるエビデンスレベルの話を紹介しよう。とはいっても私自身は医師でも医学研究者でもなく、分野外の人間として多少聞きかじったことに過ぎないので、不正確なこともあるかもしれない。

体調を崩して医者に行くと、診察室の書棚に「ほにゃらら診療ガイドライン」というものを目にすることがある。EBM普及推進事業 Minds のウェブサイト³⁾では、診療ガイドラインの定義として以下のように記載されている。

診療上の重要度の高い医療行為について、エビデンスのシステマティックレビューとその総体評価、益と害のバランスなどを考量して、患者と医療者の意思決定を支援するために最適と考えられる推奨を提示する文書。（福井次矢・山口直人監修『Minds 診療ガイドライン作成の手引き 2014』医学書院、2014、3頁）

現代の医学ではEBM (Evidence-Based Medicine) が重視されており、質の高い信頼できるエビデンスにもとづき診療の指針を決定する。当然、各々の臨床医自身が膨大な研究の情報を収集してレビューし、最新の医学研究の質や信頼性を判断することは不可能である。このため専門家によるレビューを経て質や信頼性の評価がなされたエビデンスを集約し、ガイドラインとしてまとめたものが各種団体から発行されている。ガイドラインに対する評価システムも確立されているようである。

一例として、読者にも馴染み深いであろう「痛風」のガイドラインを見てみよう。Minds のウェブサイト上で公開されている「(旧版) 高尿酸血症・痛風の治療ガイドライン 第2版」(日本痛風・核酸代謝学会編集)⁴⁾の「第3章 高尿酸血症・痛風の治療 2. 高尿酸血症の治療 2. 尿酸降下薬の種類と選択」では、冒頭に「ステートメント」として7項目が掲げられており、それぞれ

「エビデンス」「コンセンサス」「推奨度」として数値や記号で評価が記載されている。例えば「2 尿酸排泄低下型に尿酸排泄促進薬……(略)……を選択することを基本原則とする」はエビデンス3, コンセンサス2, 推奨度Cとなっている。最終的な評価だけでなく、そのような評価に至る理由、関連する詳細な情報が解説されており、末尾に文献リストが添付されている。これを見れば診療指針を考える際にどのようなエビデンスがあり、推奨度がどの程度なのか容易に判断できる。

心理学研究にこれを当てはめれば、さまざまな現象、効果、理論（予知能力、社会的プライミング、……）について、一定の判定基準のもとで専門家集団のシステマティックレビューがなされ、その信頼性についてのグレードが明記されているという状況を想像すればわかりやすいだろう。実学である臨床医学のガイドラインのようなシステムを、人間の心の理を解き明かすことをゴールとする心理学に直接移植することは現実的ではないが、再現可能性の問題は、結局のところエビデンスの信頼性に帰結するものであり、心理学側が学ぶ点も多いかもしれない。もう少し深掘りしてみよう。

ガイドラインの策定においては、エビデンスレベルや推奨度をどのように評価するかが問題となる。この際もさまざまな指針があるようだが最近ではGRADEアプローチという指針が踏襲されることが多いようである。ここでは関連する項目についてのみ言及するが、ものすごく興味を持った読者は「Minds 診療ガイドライン作成マニュアル 2017」⁵⁾が公開されているので読んでみるのもいいだろう。またGRADEアプローチの解説書も市販されている（相原、2018）。エビデンスレベル（表1）とは、主に研究デザインにより分類される、研究がもたらす結論の強さのことである。表1を眺める限り、多少の改変や項目の追加、削除は必要だが、エビデンスレベルの概念は心理学研究にも適用できそうである。

では研究デザインだけでエビデンスの信頼性は決まるだろうか。心理学者であれば、例えば二重盲検でサンプルサイズが大きなランダム割当実験研究（エビデンスレベルは高い）で統計的には

3) https://minds.jcqh.or.jp/s/about_guideline

4) <https://minds.jcqh.or.jp/n/med/4/med0052/G0000210>

5) https://minds.jcqh.or.jp/s/guidance_2017_0_h

表1 エビデンスレベルの1例 (The Oxford 2011 Levels of Evidence を要約)

1 (高い)	ランダム化試験, n-of-1 試験のシステマティックレビュー
2	ランダム化試験, 効果量の大きな観察研究
3	非ランダム化比較コホート研究
4	症例集積, 症例比較, ヒストリカルコントロール研究
5 (低い)	メカニズムに基づく推論

有意な効果が出ていたとしても、効果量が小さいとか複数の実験間の整合性が弱いとか、その他さまざまな要因によって**感覚的に**その研究の信頼性を低く見積もることもあるだろう。医学研究でも心理学と同様、研究デザインだけでは信頼性が決まらない。そして診療ガイドラインの策定の際はこのようなプロセスも明文化し、スコア化している。

GRADEアプローチでは上記の研究デザインに基づくエビデンスレベルに加えて、エビデンスの確実性を評価するための外的要因 (GRADEの8要因) を取り入れている。ここでは確実性を下げる要因 (GRADE-downの5要因) として結果の非一貫性、エビデンスの非直接性、そして**出版バイアス**などが含まれる。確実性を上げる要因 (GRADE-upの3要因) として、効果量が大い、用量反応勾配が認められる、などが含まれる。以上のように、非常に複雑かつ膨大な作業を経て、ガイドラインが作成されていく。重要なことは、研究の質の評価に関する全てのプロセスが明文化され、透明化され、共有されているということである。もちろん医学も一枚岩ではないだろう。しかし、患者という目の前の対象の生死を左右するという点からも、医学における信頼性評価システムの必要性は心理学のそれよりもずっと強いに違いない。

心理学は信頼性評価システムと向き合えるか

すでに述べたように医学研究のシステムを心理学に直接移植することは現実的ではない。それでも、このようなエビデンスの質の評価システムは、再現可能性問題 (= 信頼性ガタガタ問題) に直面する心理学において参考にするべき点も大いにあるように思える。個々の研究、現象、効果についてエビデンスレベルや信頼性を、個人の直感的

判断に負わせるのではなく、分野で共有されたものとして客観的に導き出す透明なシステムをつくることには大きな意義があるだろう。この場合、医学研究のエビデンスレベルをそのまま採用するのではなく、心理学に合わせたものを考案すべきである。

エビデンスレベルについて言えば、研究の価値を統計指標や再現可能性のみに負わせた場合には、観察研究やフィールド実験のような原理的に再現性を保証できない方法論は必然的に評価が低くなる。このことに問題があることは多くの研究者が同意することだろう。確実性とは異なる次元としての指標があつて然るべきである。ただし仮説検証型実験研究と探索型研究の切り分けは明確にするべきだろうから、そのエビデンスが「仮説検証型実験研究」なのか「探索的研究」なのかを明確にし、それぞれの合った評価を行うことが重要である。また確実性評価について言えば、本特集号が扱う「事前審査付き事前登録」や複数の研究室がプロトコルを共有して大規模追試を行う Many Labs プロジェクト (Klein et al., 2014, 2018) などはエビデンスレベルを大きく上げる要因になると考えられる。測定に関する妥当性についての評価も含まれるべきである。

著者自身、TODARADIOの中で「再現可能性トリアージ」というアイデアを発しているが、これは心理学研究の信頼性を評価するようなシステムを作り出すことを念頭においている。おそらく多くの熟練した心理学研究者は、暗にこのような評価基準を自分なりに作り出して他の研究を評価しているのではないだろうか。しかし現状では信頼性評価は各研究者の直感に負うものであり、共通の指針はない。論文化された研究の信頼性についてオープンに議論することは難しい。では心理学では診療ガイドラインのような透明な信頼性評価システムはつくることのできないのだろうか。

結論を言えば、心理学でもエビデンスレベルや

GRADEアプローチを参考にしたエビデンスの信頼性評価システムをつくることはおそらく可能であるし、そうすべきである。心という多様で複雑な対象についての研究の質を拘り定規に決められるものかという批判もあるだろう。自分自身の手で行った研究の信頼性がウェットな議論を介さずにシステムティックに決定されるなど許せないという批判もあるだろう。信頼性の保証などは荒唐無稽な夢物語であり心についての魅力的なストーリーを発することの邪魔をすべきではないという批判もあるだろう。しかしこのような態度が、心理学の再現可能性問題を生み出したのではないか。研究者それぞれの中で確立された、しかし共有されていない信頼性評価システムは、心理学という学問についての各人の価値観を大いに反映したものとなっているはずである。これを明確化し、研究者同士で共有し、摺り合せていくこと自体が、心理学の信頼性とは何かという構成概念を明らかにし、心理学という学問の意義を根拠付け、心理学の進歩に直接的につながるはずである。

ところで医学研究の確実性評価システムについて調べていると興味深いことがわかる。「理論」という言葉がほとんど出てこないのである。この理由は定かではないが、一つは医学が実学であり、事実としての治療効果、診断の正確さの方がより重要で、それを支える理論についてはエビデンスとは切り離して考えるということがあるのかもしれない。もう一つは、エビデンスが既存の理論と整合的であることは当たり前であり、理論と整合しないエビデンスは（少なくとも近代西洋医学では）表に出てこない、ということもあるのかもしれない。

翻って心理学では状況は大いに異なるだろう。臨床現場の心理学は別として、人間の心の理を解き明かすことをゴールとする心理学では、エビデンスは最終的には理論の構築に使われる。逆に理論からはみ出たエビデンスも当然受け入れるべきであり、既存の理論との整合性という視点からエビデンスについての評価がなされ、また逆にエビデンスにより理論が改変されていくことが本来の姿のほうである（鮫島, 2016）。

従って、エビデンスの評価システムを心理学に取り入れる際は、GRADEアプローチの中で「理

論との整合性」も信頼性を左右する問われるべき項目として検討されるべきである。心理学では理論から逸脱する現象や理論から予測できない現象、つまりビックリする現象ほど魅力的であると評価されがちであるが、エビデンスの評価という観点からは、理論からの予測性が低ければ、確実性を下げる要因として取り入れられる。実際に「予知能力」は、直接的に予測する理論は存在しないし、既存の（物理学）理論的整合性の理由から受け入れられなかった。一方で老人プライミングは、直接的に予測する理論は存在しないが、特にそれと整合しない（物理学的・心理学的）理論が存在しないために、魅力的な現象として受け入れられた。信頼性評価において理論との整合性を検証するシステムが機能していれば、「現象を予測する理論的根拠」の項目に引っかかり、この魅力的な Null Field（かどうかはまだわからないが）に進み出そうという衝動は、いささか抑えられていたのかもしれない。

以上のように、心理学の（特に仮説検証型研究の）信頼性評価システムを構築するならば、再現可能性や統計的指標からなる事実としてのエビデンスの強さと、それを取り巻く理論との整合性、この両者のバランスによって評価するしかないだろう。そして今まで個々の研究者が個人の中で暗に進めてきたこの作業を個人のものとしてせず、エビデンスの強さと理論との整合性を両輪とする信頼性評価システム⁶⁾として構築していくというまさにそのことが、「新しく古い心理学のかたち」なのかもしれない。

現代心理学のブルバキ

この「エビデンスと理論による信頼性評価システム」はしかし、現在のところ理想論に過ぎないだろう。Many Labs プロジェクト（Klein et al., 2014, 2018）によって不確実な現象の再現性に白黒がつけられている。事前審査付き事前登録により再現性の有無がより確実にされてきている（平石ら, 2019；佐々木・米満・山田, 2019；柚取・国里, 2019）。このようにエビデンスの強さにつ

6) Prediction market (Dreber et al., 2015；山田, 2016) のような手法も信頼性評価のフレームワークとして有用だろう。

いては、徐々に検証がなされている。しかし、現在提案されている理論群を用いてエビデンスの評価を行うことは現実的には難しいと思われる。心理学にも、もちろんたくさんの理論がある。たまたま手元に「現代心理学[理論]事典」(中島義明(編)・朝倉書店)というものがあるのでパラパラとめくってみると、感覚・知覚から認知、発達、臨床まで、実に様々な理論が紹介されていて、非常に勉強になる。しかしそこで気付かされることとして、たくさんの理論が紹介されているものの、理論同士の関係性がきわめて薄いのである。「心理学の理論は弱い」ということはよく言われる(竹澤, 2018)。この「弱さ」は多義的である。定量的な予測力が低いということ、命題が曖昧で反証可能性が低いということ、そしてここで最も強調したいのが、理論同士の相互依存性が低いということもある。「現代心理学[理論]事典」はまさに文字通り「事典」であり、それぞれの理論の間に密な結合を見出すのは難しい(この書籍を批判しているわけではない。この書籍は本当に素晴らしいものである)。理論同士の結合が密であればあるほど、その結合を通してひとつのエビデンスが負うべき理論群が巨大になるし、理論からの予測力も高まる。理論的整合性や予測性という観点からエビデンスの評価を行うことを考えると、現代の心理学理論は弱すぎる、特に結合が極端に疎である。だからといって、既存のエビデンスの集合から今すぐに強い理論としての密結合理論群を構築することも難しい。

では今からなにをするのか。新しくて古い心理学のかたちを実現するには何が必要なのだろうか。本特集号が扱う事前登録付き事前審査を含む再現可能性への対応や Many Labs プロジェクトなどでエビデンスの強さを上げていくことは間違いなく必要である。一方で、このまま疎結合の孤島理論を乱立し、Null Islandの世界を築くことを続けるのではなく、理論同士のつながりを吟味し、あるいは既存の理論をつなぐために必要なピースをあぶり出すことを行っていく必要がある。信頼性と理論による心理学のエビデンス評価システムを構築して明文化するというゴールは、このような運動の大きな牽引力になるかもしれない。最高に大げさに言えば、ここで行われる密結合理論群の構築は、現代心理学のブルバキである⁷⁾。

おわりに

本稿では再現可能性を考慮したエビデンスの強さ、そのエビデンスと理論との整合性、この両者の上に、共有可能で透明なエビデンスの信頼性評価システムとして構築していくことが、新しくて古い心理学のかたちであるという提案を行った。本稿での主張のほとんどすべては新しいものではなく、戸田正直氏により今から50年も前に著された名文「心理学の将来」(戸田, 1971)⁸⁾の中でも、本稿よりもずっと深い洞察が残されている。本稿の価値が少しでもあるとすれば、それはエビデンスの評価システムとの関係からこのような心理学理論の話を見渡したという一点にある。最後に少し長くなるが戸田先生の言葉を引用して、本稿の結語としたい。

……さしあたって必要とされる理論は、旧来の意味での一般理論ではなく、極めて多数の、他の理論と論理的にかみ合わせることでできる、いくなれば部品の理論群であるといえよう。部品のということとはつまらない理論ということではない。この新科学の対象全体から考えれば、個人の特徴や行動だけを記述する心理学の理論はすべて部品的である。そうではなくて、ここで部品的というのは、それらの理論の1つ1つが、他の理論と組み合わせて使うことができ、またすぐ相互にかみ合わない場合には改訂を許すだけの柔軟性をそなえていることを意味する。

そういう部品の理論とはたとえばどういうものかということ簡単に定義することはむずかしい。しかし、たとえば宇宙船の部品を考えてみよう。たとえビス1本でも、それはビスとしての本来の機能の他に、たとえばアポロという宇宙船全体に課せられた軽量・堅牢・耐熱とか

7) ブルバキ(ニコラ・ブルバキ)は、20世紀中盤に活躍した架空の数学者であり、その正体は若手数学者たちの共通のペンネームであった。彼らは著作「数学原論」を通じて、当時の数学の基礎的な諸部門を集合論の上に厳密に再構築することを目指した(らしい)。内容はさておき、ここで強調したいことは、ブルバキが若き研究者の集団として機能したこと、そして諸部門の接続を試みたことであり、Null Islandで右往左往する個々人という姿とは対極の風景に思える。

8) <https://www.jcss.gr.jp/meetings/archive/toda-hatano-future.pdf>

いう規準を満足しなければならないし、そのピスが使われる船体なら船体の材質に合わなければならない。もう少し具体的にいうと、たとえば幼児の知能発達に関する理論は、抽象的な真空における理論でなく、発達の背景となる社会の文化、両親の性質等々の違いにおいて、それぞれどういう経過をたどるかという答えを出しうものではなくてはならない。初めから答えが正しいことを期待されるわけではない。ただ現実の結果と違っていたら直ちに修正可能な、「開いた」理論であることが要求されるわけである。

謝 辞

本稿を執筆するにあたり、考えるきっかけを与えて頂いた池田功毅氏（中京大学心理学部）、田中悟志氏（浜松医科大学）に深く感謝いたします。またフィールド実験の再現可能性問題について長い時間をかけて多くの議論を行ってきた島田将喜氏（帝京科学大学）、大石高典氏（東京外国語大学）、錢琨氏（九州大学）、田曉潔氏（筑波大学）に深く感謝いたします。本稿の執筆に際しては直接的にはいかなる研究助成も受けていませんが、本稿の内容の一部には、新学術領域「顔身体学」（17H06342 to KT）の研究成果が深く関わっています。

文 献

- 相原守夫 (2018) 診療ガイドラインのための GRADE システム 中外医学社.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- 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.
- Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., ... & Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, 112, 15343–15347.
- 平石 界・斎藤彩乃・西尾真紀・藤井那侑・森 峻人 (2019) 配偶者選好における身体的魅力重視度の男女差は消えたのか 心理学評論, 62, 244–261.
- 池田功毅・平石 界 (2016) 心理学における再現可能性危機：問題の構造と解決策 心理学評論, 59, 3–14.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., ... Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, 45, 142–152.
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1, 443–490.
- 小島康生 (2016) 人間の観察研究における再現可能性の問題 心理学評論, 59, 108–113.
- 高橋康介 (2018a) 認知心理学者のタンザニア滞在記 心理学ワールド, 81, 44–45.
- 高橋康介 (2018b) 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 共立出版.
- Takahashi, K., Oishi, T., & Shimada, M. (2017). Is 😊 Smiling? Cross-Cultural Study on Recognition of Emoticon's Emotion. *Journal of Cross-Cultural Psychology*, 48, 1578–1586. <https://doi.org/10.1177/0022022117734372>
- 竹澤正哲 (2018) 心理学におけるモデリングの必要性 心理学評論, 61, 42–54.
- 戸田正直 (1971) 心理学の将来. In 日本児童研究所編 (Ed.), 児童心理学の進歩 (pp. 336–356) 金子書房.
- 鮫島和行 (2016) システム神経科学における再現可能性 心理学評論, 59, 39–45.
- 佐々木恭志郎・米満文哉・山田祐樹 (2019) 利き手側の良さ—事前登録された Casasanto (2009) の直接的追試— 心理学評論, 62, 262–271.
- 柚取恵太・国里愛彦 (2019) アンヘドニア (anhedonia) と遅延割引：Lempert & Pizzagalli (2010) の追試 心理学評論, 62, 231–243.
- 山田祐樹 (2016) 認知心理学における再現可能性の認知心理学 心理学評論, 59, 15–29.

— 2019. 11. 1 受稿, 2019. 12. 15 受理 —