

心理学研究は信頼できるか？

再現可能性をめぐる

池田功毅, 樋口匡貴, 平石界, 藤島喜嗣, 三浦麻子

2015年11月16日発行 (Ver. 1.0) ●発行元: ちとせプレス

2015年の8月、*Science* 誌に衝撃的な論文が発表されたというニュースが流れました。査読付きの主要な学術誌に発表された心理学と社会科学の研究論文 100 件についてその結果の再現を試みたところ、同じ結果が得られたのは 39% にすぎなかったということです。2010 年以降、心理学の中で研究の信頼性に関する問題がクローズアップされています。何が問題となっているのか、またそれに対してどのような対応が進められているのかを、日本でこうした問題に取り組まれている研究者の方々に話を伺いました。

Section 1

取り組みのきっかけ

——こうした問題に取り組もうと思ったきっかけは何だったのでしょうか。そのときにどう思ったのかも教えてください。

池田功毅 (以下, 池田) :

2010 年に、Daryl Bem の超能力論文⁽¹⁾ が社会心理学のトップジャーナルである *Journal of Personality and Social Psychology* (JPSP) 誌に近々掲載されるというニュースを *New Scientist* 誌⁽²⁾ で読んだことが直接のきっかけでした。

人間に未来予知が可能であるというこの論文の結論は、無論現代科学の視点からはありえない内容です。その掲載を JPSP が認めたということは、いわば心理学が自然科学全体を否定したという意味にもとれました。私はいったい何が起きているのかわからず、非常に困惑したのを覚えています。

Bem 論文によって示された問題は、データ改竄や捏造などとはまったく質の異なる、心理学研究の本質に関わるものでした。Bem 論文は JPSP が制定してい

る正規の査読システムを通過して掲載が決定されました。すなわちそれが意味するところは、これまで「正しい」とされてきた心理学研究の方法は、明確な疑似科学に基づいた虚偽の報告を見抜くことができない、何か科学として根本的な欠陥があるものだという事です。そして何よりも重要なことは、我々心理学者全員が、その間違っただけのシステムの中でこれまで研究を行ってきたという事実です。それは端的に、我々自身のこれまでの研究報告の中にも、多くの嘘が隠されている可能性があることを示しています。

それ以降、私は自身の研究遂行に大きな不安を抱えることになりました。この問題を論理的、現実的に解決しない限り、自信をもって自分の研究成果を報告できないと強く感じたためです。以上のような経緯で、私は再現可能性問題こそ、何よりの優先課題と考えるようになりました。

平石界 (以下, 平石) :

私の場合、こうした問題に注意が向くようになった最初の大きなきっかけは、ハーバード大学教授であった Marc Hauser の「科学的に不適切な行為」(scientific misconduct) を原因とする辞職 (2011 年) であったように思います。Hauser はカプチンモンキーを用いて非常に多くの (新奇な) 比較認知実験を報告していました。しかし彼が辞職する数年前 (2008 ~ 09 年頃) から、「Hauser の結果は、あそこの super monkey でないと出せないと言われている」といった噂話を耳にし、不穏な空気を感じてはいました。しかしさすがにデータの捏造、改竄といった悪質な行為までしているとは思っていませんでした。このニュースは非常に衝撃でした。Hauser 本人への信頼感というよりは、自分の参加している研究領域で、そこまで悪いことが行われているとは、にわかに信じがたかった⁽³⁾。

その後、Bem の超能力論文があり、そこで池田さ

んと Twitter で、Bem は本気で超能力を信じているのか、それとも心理学の進め方の問題点を指摘するために（いわば Alan Sokal⁽⁴⁾ のような形で）、あえて超能力論文を JPSP に投稿、受理させてみせたのかといった話をしたことを覚えています（結論としては、Bem は本気で超能力を信じているらしい、ということになりました⁽⁵⁾）。その後、ニュースを追っているうちに、Bem の問題の「人ごとでなさ」がますます実感されるようになりました。例えば Bem 論文を元に心理学全体を皮肉った “We knew the future all along: Scientific hypothesizing is much more accurate than other forms of precognition—A satire in one part” という論文にはショックを受けました⁽⁶⁾。この論文は、心理学論文のほとんどでは、最初に書かれた予測が支持される結果が出ている。つまり心理学者は、本当に未来を予知している（結果を予知して仮説を立てている）と皮肉ったものでした。Hauser を非難していた自分も、データ改竄ではないものの、科学的には不適切な行為を行ってきたのではないかという恐怖感に襲われました（より正確に言えば、これはまずいのかもかもしれないと感じつつ、でも自分の分析や論文の書き方は、まだまだ誠実にやっている方だと思っていたのが、まったくの勘違いであったと指摘された恐怖に近い）。その後、この危機感・恐怖感は、PPS 誌の特集号読書会によって決定的なものとなりました。

読書会の終了後、三浦さんから、追試研究のための科研申請グループに誘われました。ここで何か動かないわけにはいかないと、参加を決めました。以上が、私がこの問題に関わったきっかけです。

藤島喜嗣（以下、藤島）：

2011 年 9 月、Social Psychology Network の Twitter アカウント⁽⁷⁾からの情報で、Diederik Stapel がデータ捏造を疑われていることをつかみました。Stapel といえば、社会心理学における社会的比較領域の主要な研究者で、社会的認知や自己に関わるテキストで頻繁に引用されていました。同化と対比に関する専門書の編者もしていました。その彼が、データ捏造で大学の取り調べを受けているというのです。

かなりのインパクトでした。私個人は研究で直接的に引用することはありませんでしたが、授業では何件か彼の研究を紹介していました。論文購読を含む専門科目でも学生たちに読ませていました。それを撤回せねばなりません。知り合いの博士課程の大学院生は、Stapel 論文をかなりの数引用していました。彼の研究はどうなるのだろう。2011 年 5 月には社会的比較に関する優れた日本語テキストが刊行されており、私も

熟読していました。Stapel 論文がかなり引用されています。どうなるのだろう。

とはいえ、私は楽観的でした。「それは追試すればいいんだよ」。解決法はわかっていました。研究法の教科書には追試、特に、同じ研究手続きを繰り返す直接的追試よりも新しい手続きを用いて理論を検証する概念的追試が大事だと書いてあります。Stapel の論文には理論的には妥当だと思われるものがかなりありました。彼のデータがダメだったとしても、他の似たような理論から導かれたデータから対処可能だろう。ネットワーク上の 1 つのノードが破壊されても周囲から再生されるように、彼の捏造による欠落も補償されるだろう。そう思っていました。実際、知り合いの大学院生にはそのように伝えて、安心させようとしていました。

しだいにそれが楽観的に過ぎることがわかってきました。実は遡ること 20 年ほど前にも社会的認知で取り下げ騒ぎがありました。その後、その研究が追試されたという話は聞きません。出版バイアスは昔からあって、有意差のないデータは刊行されないし、直接的追試も刊行されません。風の噂で再現可能性の情報を得る、これが 2011 年の社会心理学界での方法でした。だから、その研究を信用してよいかどうかはわからないままです。噂をどうやってつかめばよいのでしょうか。Twitter か？

さらには、Bem 論文の話もありました。2010 年刊行でずいぶん話題になっていました。学生時代に、直観とデータからの示唆とどちらを信用するかといわれたら後者だ、と教わってきました。Bem 論文の掲載をして JPSP のエディターは偉いなあと感じたのを覚えています。怪しい話もデータに関するルールがきちんとしていたら掲載せざるをえない。掲載したうえで追試しながら否定すればいいじゃん、と思っていたのです。しかし、おかしいことになってきました。追試データは論文にならないのです。直感的に怪しい、違う分析をしたら今イチそうだ、追試した人もダメだったといっている、といろいろ証拠が集まっているようでしたが、周囲で騒ぎ立てているだけの様相を呈していました。出版バイアスの見えない大きな壁がそこに立ちはだかってきました。

そうしたら、別のところで、John Bargh の研究に関わる論争が始まりました。Bargh は非意識過程の巨人であり、Cognitive Monster の生みの親です。彼の老人プライミング実験⁽⁸⁾による知覚—行動リンクはその核心の 1 つでした。その老人プライミングが単なる実験者効果だということです⁽⁹⁾。ま、そういうこともあ

るよねと思っていたら、Bargh がけんか腰の反論を始めました⁽¹⁰⁾。「他にもいろんな手続きでデータいっぱいあるし、そもそもオマエが下手なだけだし」。いやいや、両者が同じプラットフォームに立って、一緒に検証すればいいんじゃないの。データ提供して再分析したり、手続きを提供して追試すればいいじゃないの。しかし、それは実現されず、ただ罵倒し合うだけでした。

2010～2012年は社会心理学の歴史で大きく語られるであろう時代となりました。いくつかの事件が、再現可能性問題、出版バイアス、手続きの透明化と共有、というキーワードで合流し、単純だが大きな話になったのです。こういう話は「どげんかせにゃならん」話だし、みんな及び腰だったから旨味もありそうでした。それで読書会もあったし飛びついてみたのですが、みんなが及び腰だった理由がよくわかりました。とにかく大変な話で、理解も得られなければ、お金も得られない、そして制度改革も必要な話だったのです。

樋口匡貴（以下、樋口）：

取り組みのきっかけは、2013年2月に有志で行った *Perspectives on Psychological Science* (PPS) 誌の再現可能性特集号⁽¹¹⁾ の読書会です。

Bem の超能力論文のような大きな話は承知していましたが、それ以上に自分を動かしたのは、より身の周りの、ひいては自分自身という半径数 10 cm の問題でした。自分自身を含めた身の周りでの研究実践のあり方を考えつつあった当時の私自身にとって、PPS の再現可能性特集号の読書会で得た知識は非常に大きいものだったし、何よりも「問題のある研究実践⁽¹²⁾」(QRPs: Questionable Research Practices) に自分が携わり続けてきたという確信をつかんでしまったことは衝撃的でした。これから自分はどう生きていけばいいのか、どう研究をしていけばいいのか、どうこれまでのことを“贖罪”すればいいのか。こういったあたかも思春期のような感覚が（この時は大学院入学からちょうど 16 年！）、この問題への取り組みを自分自身に対して促すことになりました。

ただ、“研究はこのように進めるべき”“このようなやり方はすべきでない”といったある種の啓蒙的な話題の場合、往々にしてその話題に触れたときから 1 週間もすれば熱が冷めることが多いでしょう。しかし今回の再現可能性の話題については、衝撃を受けたそのときの、“何かをしなれば、変えなければ”という思いを感じているまさにそのときに、同じような感覚を共有できる仲間がいたことは非常に幸運でした。議

論すること自体だけでも非常におもしろくエキサイティングなメンバーがいることは、私自身が再現可能性問題についての取り組みを実際に続けていこうとする大きな要素であったことは間違いありません。

三浦麻子（以下、三浦）：

再現可能性問題について「関心をもった」きっかけは、数年のうちに、Bem の JPSP 論文出版、Bargh と Daniel Kahneman のそれをはじめとする社会的プライミングに関する論争、Stapel らの研究不正問題など、心理学における科学的研究とは何なのか、どうあるべきなのか、を考えさせられる出来事が立て続けに起こったことです。そして、*Perspectives on Psychological Science* (PPS) 誌での特集号の刊行（2012 年秋）があり、こうした出来事が実際に権威ある国際誌の査読システムを変えるに至りそうだと、ということを知ったことです。つまり、単なる対岸の火事でもなんでもなく、この問題が、自らの研究に直接的な影響を与えることになるのだ、ということを実感したわけでした。こうした事態の進行と同業者たちの反応は、Twitter を見ていると手に取るようにわかりました。

ただしそれだけでは、先行研究の追試という形で、また「自ら」この問題に取り組もうとは思わなかった、かもしれません。研究不正や QRPs については襟を正せば、また、査読システムの変化には、ルールに愚直に従ったやり方をすればよいだけです。別に追試に積極的に取り組む必要はないでしょう。

再現可能性問題の他にも、興味をひかれる、つまり「手をつけてみたい」テーマは毎日のように現れますが、そのすべてに手をつけることは（当然ながら）ありません。実際に手をつけるには、「手をつけられる」と思えるテーマであることが必要で、そのためには（少なくとも私にとっては）手を組める人材に見当がつけられることが何より大切です。それが思い浮かばないテーマは「まあ、誰かやるやろ」と手放すことになります。また、私はやれるとなれば「すぐやる課」なので、企画をじっくり温めるようなことはしません。すぐに動きます。

両者がうまく同期した研究テーマの 1 つが再現可能性問題でした。前述の PPS 特集号の刊行を知ってすぐに読書会を企画し、その企画が盛り上がった現場である Twitter と、ちょうど開催された日本社会心理学会の大会@筑波大学（2012 年）などを通してメンバーを募りました⁽¹³⁾。この企画を共同していたか、公表前から参加を表明してくれていたメンバーがほぼそのまま研究プロジェクトのメンバーになっています。池

田さん以外とは企画前から交流がありましたが、共同研究をしたことはなかったし、テーマ的にその後も直接絡むことは難しかろうと思っていたので、良いチャンスだと思いました。池田さんとはそれまでまったく面識がなかった、というかお名前すら存じ上げなかったのですが、平石さんが「このテーマならぜひ彼を」と推薦してくださったのは僥倖でした。

たくらみどおり、この読書会が非常に面白かった（個人的に、多人数でやる読書会が面白いと思った経験はあまりなかったので、これは印象的でした）ので、これは内容もメンバーもいいし、プロジェクトとしてもものにするのができそうだな、というある程度の確信を得たのです。もう少しかっこよく言うと、多士済々な研究者たちと一緒にやるテーマとして、これ以上適切なテーマはない、と思ったのです。何が面白かったか、については、要約すると「組織的な追試というやり方でこの問題に取り組むことには大きな意義がある」と実感した、ということになります⁽¹⁴⁾。

お題目以上の実質的な仕事に実際に着手できた理由は、2回目のチャレンジで科研費をいただくことができたからにほかなりません⁽¹⁵⁾。自分（たち）は「筋が良い」と思っている研究テーマが、それ以外の同業者によって取り組むべき研究テーマだと認められた、ということは大きな意味をもちます。それがなければ、あまり実践を伴わない、啓発的な見かけをした実際は自己満足ないくつかの取り組みに終始し、そのうちに忙しさに取り紛れて盛り下がってしまったかもしれません。

■ 文献・注

- (1) 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.
<http://psycnet.apa.org/journals/psp/100/3/407/>
- (2) Aldhous, P. (2010). Is this evidence that we can see the future? *New Scientist*, 11 November.
<https://www.newscientist.com/article/dn19712-is-this-evidence-that-we-can-see-the-future/>
- (3) Hauser 事件の顛末については、こちらへんが結論になるかと思えます。
Couzin-Frankel, J. (2014). Harvard misconduct investigation of psychologist released. *ScienceInsider*, 30 May.
<http://news.sciencemag.org/people-events/2014/05/harvard-misconduct-investigation-psychologist-released>
- (4) Alan Sokal: 物理学者である Sokal が、物理学や数学などの専門用語をでたらめに用いた論文を作成し投稿したところ、人文社会学系の雑誌に掲載されてしまったという事件がありました。このことによって、これら雑誌の編集者、査読者が、自然科学の概念を濫用していたことが暴露されました。Bem も

また、心理学論文の“作法”をでたらめに用いた論文を JPSP に載せることで、心理学者の研究手法のでたらめさを指摘する意図があったのかと、当時の私は考えました。

ソーカル, A.・ブリクモン, J. (田崎晴明・大野克嗣・堀茂樹訳) (2012). 『「知」の欺瞞——ポストモダン思想における科学の濫用』岩波現代文庫, 岩波書店

- (5) Bem については、拙文を紹介させていただきます。
平石界 (2013). 「裏から読んでも心理学 超能力ってどうなんでしょうか？」『心理学ワールド』 **61**, 32.
http://www.psych.or.jp/publication/world_pdf/61/61-32.pdf
- (6) Bones, A. K. (2012). We knew the future all along: Scientific hypothesizing is much more accurate than other forms of precognition —A satire in one part. *Perspective on Psychological Science*, **7**, 307-309.
<http://pps.sagepub.com/content/7/3/307.long>
- (7) Social Psychology Network
<http://www.socialpsychology.org/>
Twitter: @SocialPsych
<https://twitter.com/socialpsych>
- (8) Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology*, **71**, 230-244.
<http://psycnet.apa.org/journals/psp/71/2/230/>
- (9) Doyen S., Klein O., Pichon C. L., Cleeremans A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PLoS ONE*, **7**(1), e29081.
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029081>
- (10) Bargh, J. A. (2012). Priming effects replicate just fine, Thanks: In response to a ScienceNews article on priming effects in social psychology. *Psychological Today*, May 11.
<https://www.psychologytoday.com/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks>
- (11) 読書会 Web サイト「「実験結果の再現可能性」特集号読書会」
<https://sites.google.com/site/ppsjournalclub/>
- (12) 問題のある研究実践 (QRPs) については、『心理学ワールド』68号の平石・池田論文に詳しい。
平石界・池田功毅 (2015). 「心理学な心理学研究——Questionable Research Practice」『心理学ワールド』 **68**, 5-8.
http://www.psych.or.jp/publication/world_pdf/68/68-5-8.pdf
- (13) (11) を参照。
- (14) 詳しくは『心理学ワールド』68号の三浦論文を参照。
三浦麻子 (2015). 「心理学研究の「常識」が変わる？——心理学界における再現可能性問題への取り組み」『心理学ワールド』 **68**, 9-12.
http://www.psych.or.jp/publication/world_pdf/68/68-9-12.pdf
- (15) 科研費研究 Web サイト「Replicability in Psychological Science (心理学における実験結果の再現可能性)」
<https://sites.google.com/site/kgasasemi/replicability>

Section 2

■ Bem 論文の掲載

——2010 年前後から、心理学研究の信頼性を揺る

がさまざまな出来事が続けて起きていたわけですね。特に2010年のBem論文のインパクトが大きかったようですが、*Journal of Personality and Social Psychology* (JPSP) 誌はなぜこの論文を掲載したのでしょうか。

池田：

JPSPの査読が特に他のジャーナルと比べて厳しいわけではないですが、近年特にこのジャーナルで顕著だったのは、複数の(時には10近くの)研究を並べ、そのすべてで同じ方向性の結果が出ていることを示して、知見の頑健性を喧伝する、という流れでした。そしてBem論文⁽¹⁾はまさにこの流れに則って持論を展開しました。9つにも及ぶ実験を報告し、そのうち8つが予測通りの(すなわち未来予知が存在するという)証拠を提出することで、JPSPの編集者・査読者が「方法論上で」反論できない形式の論文としたのです。

当然のことながら、JPSPの編集者たちも、この結論を素直に受け入れたわけではありません。彼らもまた、この論文の知見が通常の因果性に対する一般的信念に反しており、受け入れがたいものであることは十分に承知の上で、それでもなお、他の投稿論文と同様、正規の査読システムによって審査するべきだと考えて、出版に至ったと告白しています⁽²⁾。

編集者たち自身からの明確な反論はありませんでしたが、Bem論文が掲載された号に、心理統計学者であるE.-J. Wagenmakersらの反対論文を同時に掲載した⁽³⁾ことから、彼らの意図が、Bem論文を通じて、統計を含む心理学研究方法と査読システムの改革を促すことにあったのは明白でしょう。

正直に言えば、私はこのニュースを聞いたとき、編集者たちの正気を疑ったのですが、今振り返れば、彼ら、Charles M. JuddとBertram Gawronskiの判断がなければ、それに続く心理学の方法論革命も起きえなかったものであり、その意味で彼らの貢献は(彼らの当初の意図がどうだったにせよ)大きいと言えるでしょう。

——「未来予知」現象を示す結果が繰り返し得られたため方法論上は反論できず、掲載しないわけにはいかなかったわけですね。だとすると、「未来予知」現象が実際にある、ということになるのでしょうか？

上で述べた、同一方向の知見を複数並べるという手法は、一見すると確かに証拠の頑健さを示すのに最適な方法であるように思えるのですが、実際はまったく逆の事実、すなわちその証拠の信用できなさを示して

いる場合が多いのです。というのも、心理学の多くの研究は、低い統計的検定力しか持ち合わせていないからです(例えば次の論文⁽⁴⁾を参照)。統計的検定力とは、仮に予測された効果が実際に存在する場合に、その効果を検出できる確率のことです。通常80%ほどに高めることが望ましいとされますが、心理学の場合、20~40%程度しかない場合が多いとされます。

すなわち検定力が低ければ、実際に当該の効果が存在している場合でも、それを検出することがなかなかできないはずだから、報告の数が増せば増すほど、その中に有意ではない報告が含まれる可能性が高くなってしかるべきです。にもかかわらず、次々に予測通りの報告が積み重ねられていくのであれば、それはありえないほどに美しい——“too good to be true”——結果だと言われても仕方がない(詳しくは次の論文⁽⁵⁾を参照)。言い換えれば、そうした論文では、何らかの疑わしい方法によって統計的有意差を導き出している可能性、すなわちp hackingが行われている可能性が高くなるのです。

Bem論文の追試

——予想通りの報告しか出てこないのは、むしろ「ありえない」と。では、Bem論文はその後、追試されたのでしょうか？

平石：

Bem論文については、*Journal of Personality and Social Psychology* (JPSP) 誌に掲載される前から「どうやらJPSPに載るらしい」という話が広まっており、騒ぎが始まっていました。JPSPという雑誌は、社会心理学、パーソナリティ心理学におけるトップ・ジャーナルですから、そこが超能力論文を載せることは大変なことなわけですが(このあたりは、すでに書きました)。

そのBem論文の追試ですが、特に有名になったものとしては*PLoS ONE* 誌に掲載されたStuart J. Ritchieらの論文を挙げることができるでしょう。なぜこの論文が有名になったかという点、Bem論文のシステムティックなdirect replication(直接的追試)を行ったこと、その結果が(多くの心理学者にとっては予想通り、超能力支持者にとっては予想に反して)ことごとく「超能力が存在する証拠は得られなかった」という否定的(ネガティブ)な結果であったこと、そして投稿した追試論文がJPSPによって門前払いされたこと、の3点によると考えられます。

まずシステマティックな *direct replication* について説明しましょう。池田さんがすでに Bem 論文の奇妙さ（予想通りの結果ばかりが並ぶ論文）について指摘してくれていますが、Bem 論文が出た当時は、これが妙なことであるという共通理解は、少なくとも私の周りではありませんでした。私もまったくそのことに気づいていませんでした。Bem 論文を契機として、今や「常識」となりつつある、従来の心理学研究の手法の危うさが次々と明確にされていったのです。その意味では Bem 論文の追試は、私にとって「追試」の意味について目を開かせてくれたものでした。

この追試研究の特徴は、まず第1に Bem の研究をできる限り正確に、そのまま追試したことです。実験に用いたプログラムも、Bem 本人から提供されたものを利用してあります。それまで、特に社会心理学では、追試というと *conceptual replication*（概念的追試）というものの説得力がむしろ強調されていました。これはある実験で示された結果を、概念的には同じことをやっていると考えられる別の実験で追認するという形です。よい喩えかわかりませんが、例えば「マクドナルドのロゴを見せるとダイエットが気になる」という実験結果について、「ロツテリアのロゴを見せるとお菓子を食べなくなる」という実験で追試するといった塩梅です。しかし概念的追試の場合には、結果が再現されなかったときに「それはマックじゃなくてロツテリアだったからだよ」とか「ダイエットは気になってたけど、お菓子には反応しなかったのだよ。だって使ったお菓子があまりにおいしそうで、そんなの我慢できるはずがない」とか、いくらでも言い訳ができてしまうという問題があります。そこで追試研究の著者らは、まったく同じことをやって、結果が再現されるかを示したのです。

この直接的追試とセットで大きな意味をもっていたのが *pre-registration*（事前登録）です。すなわち著者らは「我々は追試をやるよ！」ということを事前に明らかにしておいて、結果がポジティブであろうとネガティブであろうと発表すると宣言したのです。このようにしておかないと、直接的追試をしても、超能力支持者はポジティブなときだけ報告するし、反対者はネガティブなときだけ報告するということをしてしまいます。自分に縛りをかけることでそれを避けるのが、この方法なわけです。加えて著者らは3つの研究室で独立に追試を行っています。このことで「失敗／成功といっても1箇所だけでしょ」という批判にも対抗できるように準備をしたわけです。

これらの手続きそれぞれが、言われてみればまった

くその通りとうなずけるものだったのですが、自分ではそこまで考えが至っていなかったものであり、「なるほどなあ」と感銘を受けたことを覚えています。

2点目です。実験結果がネガティブであったということです。池田さんが書かれているように Bem は9つの実験を JPSP 論文で報告しています。その中で最も効果量が大きく、Bem 本人も再現可能性が高いとお墨つきを与えていたのが実験9です。この実験では、まず参加者に48個の単語からなるリストを一度ざっと見せます。それから、出てきた単語を思い出して書き出してもらいます（心理学用語で言うと、記憶の再生実験です）。それからコンピュータが48のうち24個の単語を選び出し、これらを何度もタイプしてもらうといった訓練を受けます。すると、後で訓練することになる単語ほど、（訓練の前に）思い出すことができた、よく覚えていたという結果なのです。つまり未来に勉強することになることは、勉強する前からすでによく記憶できるという結果です。書いているとバカバカしくなるような内容ではあるのですが、Bem 論文では統計的に有意に差が出ています。しかし再現論文では、独立に実験した3つの研究室とも、ネガティブな結果でした。

3つ目です。こうしてさまざまな準備をして追試実験を行い結果を得た著者らは、それを JPSP に投稿します。実際、Bem 本人が JPSP 論文の中で追試の重要性を強調していたのですから、最初の投稿先として JPSP を選んだのは合理的な選択であったといえるでしょう（なお Bem は、実際に追試のためにプログラムの提供も行うなど、追試について積極的にサポートをしています）。しかし JPSP はこの投稿を *reject*（掲載拒否）します。しかもそれは、論文の内容について査読者が判断した結果というのではなかったのです。論文を受け取った編集者が、JPSP は追試研究は掲載しない方針であるということで、査読者の判断を仰ぐ前の段階で、*reject* したのです。

論文はその後、複数の雑誌に送られますが（*Science Brevia* 誌、*Psychological Science* 誌、*British Journal of Psychology* [BJP] 誌）、いずれでも掲載拒否の憂き目を見ます。最後の BJP については、査読に回ったものの2人の査読者のうち1名が難色を示し、結果は *reject*。しかしこの難色を示した査読者というのが Bem 本人であったそうです。これらの顛末については、著者の1人である Chris French のブログ記事⁽⁶⁾ に詳しいです。また、こうした掲載までの右往左往は、*New Scientist* 誌でも記事⁽⁷⁾ となりました。

最終的には、この論文は *PLoS ONE* 誌というオー

プン・アクセス誌に掲載されることになりました⁽⁸⁾。実は同時期に *PLoS ONE* 誌には、再現可能性問題を巡るもう 1 人の重要人物、John Bargh の老人プライミング追試実験も掲載されていました。インターネットの力を使い、巨大出版社など“伝統的”な権威から科学を解放しようとするオープン・アクセス誌の、面目躍如たる雰囲気があったことを覚えています。

それではなぜ JPSP を始めとする“伝統的”ジャーナルは、追試論文の掲載を断ったのでしょうか。そこには人目を引く「目新しさ」「面白さ」に偏った研究の価値判断があったのではないかと考えられます。これは自分でも経験があることなのですが、自他の研究を評価するときに、「何が新しいの？」という基準で判断することが、非常に多くあります。誰かがやったことを繰り返した“だけ”では評価されず、そこに、その研究者なりの新しさ、オリジナリティが加えられていることを強く求めるのです。少なくとも心理学では、その傾向はかなり強固にあるという印象があります。

卑近な例を挙げますと、実は私の卒業論文は、Cosmides & Tooby (1992)⁽⁹⁾ の 4 枚カード問題研究という、進化心理学における古典ともいう論文の追試でした。今にして思えば、これはかなりちゃんとした直接的追試であったと思います。それではなぜ追試をしたのかというと、当時から再現可能性問題に気づいている意識の高い学生であったから、ではまったくなく、指導教員であった長谷川壽一先生からのアドバイスでした。そもそも進化心理学をやっけようかと決心したきっかけがこの論文でしたし、私は素直な学生でしたから、指導に従って追試を行いました。しかし心の中には、自分のオリジナルのアイデアに基づいて研究をしている同級生や先輩への引け目が強くありました。周囲の方々は追試をすることをけっこう高く評価してくださっていたのですが、追試しただけでは論文にはならないという空気も感じていて、卒論がそのまま業績にならないことに対して焦りを感じたことも覚えています。うぶな学部生であっても、目新しく面白い内容でないと雑誌に掲載されないという出版バイアス (publication bias) に気がつくほどに、心理学界隈における新しさへの要求は強い、少なくとも強かったといえるでしょう。

ついでに少し逸話的な話をさせてもらいますと、私が入る少し前から、長谷川研ではクジャクのオスの尾羽の研究をしていました。これも追試研究です。クジャクの尾羽根がなぜあんなにも派手で長いのかという問題は、かのダーウィンをも悩ませた問題でしたが、

Marion Petrie らが、尾羽根についている目玉模様の数が多いオスほどメスにモテる (交尾に至る確率が高い) という鮮やかな研究を発表していたのです⁽¹⁰⁾。ところが長谷川研でこの研究を追試しても、まったくそのような傾向が出ませんでした。

この追試研究もまた、なかなか論文として掲載されませんでした。その原因の 1 つには、目新しさの問題だけでなく、ネガティブ・データは論文になりにくい、という問題もありました。ネガティブ・データというのはつまり「〇〇の関係はなかった」と主張するものです。しかし、何かが存在しないことを証明することは「悪魔の証明」とも呼ばれ、不可能なのです。例えば「白いカラスはいない」と証明するためには、古今東西過去現在未来のすべてのカラスの色を調べないといけなわけです。「目玉模様と交尾成功に関係がある」という Petrie らの論文は 2 年間のクジャクの調査データから書かれたものでしたが、「そうした関係は“ない”」という追試研究が論文として認められるまでには、結局、10 年以上の調査データの積み上げが必要となりました⁽¹¹⁾。

話がずいぶん脇にそれましたが、Bem の追試研究に戻りましょう。一度は追試研究の掲載を拒否した JPSP ですが、実は、後になって別の研究グループによる Bem 論文の追試論文を掲載しています⁽¹²⁾。こちらは Bem 論文の実験 8 と実験 9 の追試を、あわせて 7 つ行い、そのほとんどでネガティブな結果を報告しています (実験 8 も基本的には実験 9 と同じ手続きのものです)。さらに彼らは Bem 論文の実験 8 および実験 9 についての、Bem 本人のもの (2 つ)、彼らによるもの (7 つ)、他の研究者による出版済 / 未出版のもの (10)、あわせて 19 実験の結果をあわせた全体としての傾向を見る分析 (メタアナリシス) を行い、そこでも結果はネガティブであったことを報告しています。

なぜ JPSP が Ritchie らの掲載を拒否し、Galak らの掲載は許可したのかは、わかりません。ただ、この点については時系列を見ると面白いことになっています。

- *New Scientist* 誌の JPSP 拒否記事……2011 年 5 月 5 日
- Ritchie らの *PLoS ONE* への投稿……2011 年 12 月 16 日
- Galak らの JPSP 投稿……2012 年 2 月 8 日
- Ritchie らの *PLoS ONE* 論文：掲載決定……2012 年 2 月 13 日
- Ritchie らの *PLoS ONE* 論文：出版……2012 年 3 月 14 日
- French のぼやきブログ……2012 年 3 月 15 日

- Galak らの JPSP 修正投稿……2012 年 6 月 19 日
- Galak らの JPSP 電子出版……2012 年 8 月 27 日

Galak らが JPSP に投稿した時期と、Ritchie らが *PLoS ONE* 誌から掲載決定を得た時期がきわめて近いのが面白いところです。*PLoS ONE* 誌に追試研究が載るといふ話を聞きつけた JPSP の編集者が、Galak らの論文については査読に回すことにしたのかも？とか、穿った見方をしてしまいますね（根拠は何もありません。単なるお遊びです）。ただ、それぞれの論文のタイトルまで見ると、そうした穿った見方がさらに面白くなるのも、愉快なところ。曰く、

- “Feeling the future”（未来を感じる） Bem のオリジナル → 超能力研究の未来を感じるよ！
- “Failing the future”（先を読み違える） Ritchie ら → 追試やったらトップ・ジャーナルに載ると思ってたけど、先を読み違えたよ……。
- “Correcting the past”（過去を修正する） Galak ら → Bem 論文を載せちまった過去を修正するよ……。

■ 文献・注

- (1) Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, **100**(3), 407-425.
<http://psycnet.apa.org/journals/psp/100/3/407/>
- (2) Judd, C. M., & Gawronski, B. (2011). Editorial comment. *Journal of Personality and Social Psychology*, **100**(3), 406.
<http://psycnet.apa.org/journals/psp/100/3/406/>
- (3) Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, **100**(3), 426-432.
<http://psycnet.apa.org/journals/psp/100/3/426/>
- (4) Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, **105**(2), 309-316.
<http://psycnet.apa.org/books/10109/032>
Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, **14**(5), 365-376.
<http://www.nature.com/nrn/journal/v14/n5/full/nrn3475.html>
- (5) Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, **17**(4), 551-566.
<http://psycnet.apa.org/psycinfo/2012-23130-001/>
- (6) French, C. (2012). Precognition studies and the curse of the failed replications. *The Guardian*, 15 March.
<http://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications?newsfeed=true>
- (7) Aldhous, P. (2011). Journal rejects studies contradicting precognition. *New Scientist*, 5 May.

<https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/>

- (8) Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'Retroactive Facilitation of Recall' effect. *PLoS ONE*, **7**(3), e33423.
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0033423>
- (9) Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press, pp. 163-228.
- (10) Petrie, M., Tim, H., & Carolyn, S. (1991). Peahens prefer peacocks with elaborate trains. *Animal Behaviour*, **41**(2), 323-331.
<http://www.sciencedirect.com/science/article/pii/S0003347205804841>
- (11) Takahashi, M., Arita, H., Hiraiwa-Hasegawa, M., & Hasegawa, T. (2008). Peahens do not prefer peacocks with more elaborate trains. *Animal Behaviour*, **75**(4), 1209-1219.
<http://www.sciencedirect.com/science/article/pii/S0003347207005301>
- (12) Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, **103**(6), 933-948.
<http://psycnet.apa.org/journals/psp/103/6/933/>

Section 3

心理学研究の問題点

—— Bem 論文自体は否定されたけれども、2010 年頃から続いた一連の出来事により、心理学の研究方法に関する問題点が浮き彫りになってきたわけですね。具体的にはどういった問題が明らかになったのでしょうか。

樋口：

心理学の研究におけるさまざまな問題点についての指摘は、Bem 論文が公刊された時期の前後にさまざまな形でなされていました（例えば J. P. Simmons ら⁽¹⁾は、False-positive = 偽陽性と呼ばれる現象に焦点を当て、本当は意味がないのに意味があると主張する心理学者のさまざまな“努力”について論じています）。問題についての体系的な整理は、やはり *Perspectives on Psychological Science* 誌の 2012 年の再現可能性特集号⁽²⁾によるところが大きい。A crisis of confidence と副題のつけられたこの特集においては、心理学における再現可能性の低さを引き起こすさまざまな“問題のあるやり方”が紹介されています。

問題の背景として大きく指摘されているのは、「論文を公刊すること」への圧力と、「どのような論文が

公刊されるのか」という性質です。研究業績が研究者の雇用や給与、さらには助成金の獲得に大きな影響を与えている現状においては、研究者は研究業績、すなわち論文の数を増やすことに非常に強く動機づけられます。しかし科学の世界では、論文はただ書けば公刊されるのではなく、ピアレビューと呼ばれる査読のシステムを通過しなければなりません。研究者は論文を書き上げたらまず関連するどこかの雑誌に投稿する。その論文は、雑誌の編集者によって指名された査読者によって査読され、さまざまな修正要求とともに研究者のもとに返却される。研究者は査読者のコメントに対する返信と修正した論文を再度雑誌に送る。このやりとりを繰り返し、最終的に査読者や編集者の審査に「合格」した論文だけが、ようやく公刊されることとなります。

査読を通過して論文が公刊されるための重要なポイントの1つは、データが「ポジティブ」であることです。目新しくオリジナリティのある仮説が存在し、その仮説に対してポジティブな（≒有意な）結果が得られた研究ほど、査読に合格し、公刊されやすいようです⁽³⁾。

論文を公刊するためには有意な結果が望ましいとなれば、研究者たちは当然有意な結果が得られるように努力することになります。統計的に有意であるかどうかは p 値 (probability) という値で判断しますが、その p をハックするという意味で “ p hacking” と呼ばれることもあります。もちろん多くの努力は適切な範囲でのものです。しかし中には、いろいろと問題があるやり方 (Questionable Research Practices : QRP = 問題のある研究実践) も存在します。

近年の社会心理学における再現可能性プロジェクトの旗振り役の1人でもある Brian Nosek は、*Perspectives on Psychological Science* 誌の再現可能性特集号に寄せた論文において、目新しくポジティブな結果が推奨されている現状が、以下のような研究実践を増やしていると指摘しています⁽⁴⁾。

- 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。つまり、投資でいうところのレバレッジである。
- ポジティブな結果ばかり報告され、ネガティブな結果が報告されない。
- 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ともとと仮説があったかのように報告する (注: 結果がわかってから仮説を作る = Hypothesizing After the Results are Known の意味で HARKing と呼ばれます。HARKing については N. L. Kerr の論文⁽⁵⁾ 参照)
- 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。

こういった QRPs を指摘した Nosek はこうも述べています。「目新しくポジティブな結果は論文の公刊にはきわめて重要だ。しかしそれは真実とは限らない」(p. 617)。

なお、数々の QRPs が実際にどのようにして行われていくのかは『心理学ワールド』第 68 号における特集「その心理学信じていいですか?」の中の「心理学な心理学研究——Questionable Research Practice」と名づけられた論文に詳しい⁽⁶⁾。この論文では心太と名づけられたどこにでもいるような心理学徒が、とてもナチュラルに QRPs を実践していく様子が描かれています。

また心理学・脳科学に関する有名ブロガー兼サイエンスライターである Neuroskeptic⁽⁷⁾ は、ダンテの『神曲』における地獄篇のパロディを用いて、心理学者による 9 つの問題のある研究のやり方を “地獄” として紹介しています (図参照)。詳細は論文を直接見てほしいですが⁽⁸⁾、ここでは簡単に 9 つの地獄を紹介しておきます。

第 1 の地獄：他の問題のあるやり方を見て見ぬふり、

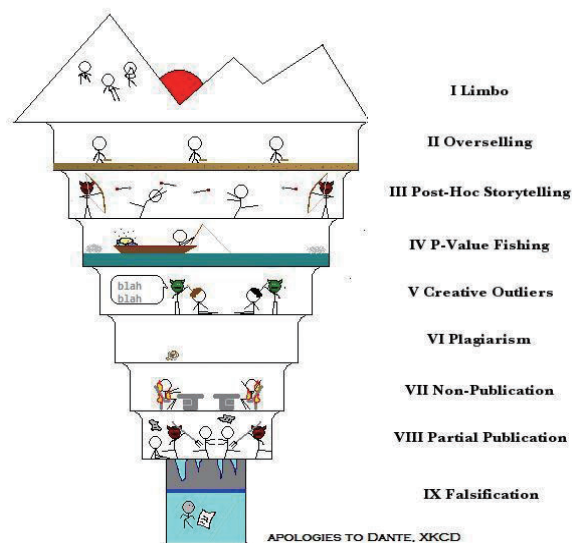


図 9 circles of scientific hell

© Neuroskeptic より許諾を得て掲載。

第2の地獄：過大に自分の研究を売り込む，第3の地獄：後づけで話を作る，第4の地獄：統計的に有意になるように手を変え品を変える，第5の地獄：外れ値を都合のいいように使う，第6の地獄：剽窃・盗作，第7の地獄：都合の悪いデータを公開しない，第8の地獄：都合の良いデータだけを選んで公開する，第9の地獄：データの捏造

心理学者の中で，自分は心太ではない，そしてどの地獄にも落ちたことがないと胸を張れる人ははたしてどれぐらいいるのだろうか。そしてどうすれば心太をはじめとした心理学者は（あるいは心理学は？），地獄に落ちず真実にたどりつくことができるのだろうか。

QRPsを防げるのか

——研究実践上のさまざまな問題点が指摘され，それが多くの研究者にとって他人事ではなかったわけですね。しかも背景には論文公刊をめぐる制度的・構造的な問題もあったと。ことの重大さがわかってきました。では，どうすればQRPsを防ぐことができるのでしょうか。

藤島：

QRPsを防ぎ，再現可能性の検討を促進するにはどうすればいいか，これらの問題について論文が刊行されると同時に，いくつかの学会，雑誌の編集委員会が指針を示しています。例えば，*Psychological Science* 誌に掲載されたG. Cummingの論文⁽⁹⁾は非常に刺激的な論文です。また，最近*Social Psychological and Personality Science* (SPPS) 誌はEditorialに再現可能性に関する方向性も示しています⁽¹⁰⁾。これらを参考にしながら考えてみましょう。

まず，QRPsを用いた p hackingを抑止するにはどうすればいいのでしょうか。根本的な解決としては，Null Hypothesis Significance Test Procedure (NHSTP；帰無仮説検定)を捨て去るという方法があります。大胆に聞えますが， $p < .05$ という結果を得るためにQRPsがなされているのですから，その目標自体をなくすことは理にかなっています。

すでに帰無仮説検定に対する代替手法が存在します。例えば，効果量や信頼区間を記載，少なくとも併記するとか，ベイズ統計学に基づくアプローチに移行するといった方法があります。実際に*Basic and Applied Social Psychology* 誌は帰無仮説検定のみの論文投稿を禁じました⁽¹¹⁾。先述のSPPS誌も帰無仮説検定への偏

重を修正しようとしています。日本の雑誌ではまだ明確な指針は示されていませんが，将来的に修正，移行がなされる可能性はあります。

心理学界全体を考えると，すぐ移行できるかはわかりません。困難もあります。研究者なら当然払うべき労力だと思いますが，新たに統計解析手法をマスターすることは容易ではありません。大学院，学部において心理統計の手法を刷新することに難しさを感じる教員もいるでしょう。その点に関して，日本社会心理学会を始めとしていくつかの学会が新しい方法論についてシンポジウムやセミナーを開催しているのは非常に心強いことです⁽¹²⁾。

そもそも，雑誌で採用する統計手法を変えるというのは大きなルール改定です。雑誌の編集委員会のリーダーシップと実行力が問われます。世界的に見ても，再現可能性の重要性認識には温度差があります。従来の方で問題ないと考える研究者もいますので，合意に時間がかかる可能性があります。また，心理学の研究手法は実験だけではないという点も見逃せません。これらの中には再現可能性それ自体を問うことが難しい研究もありえます。これらをどう扱うかについては複数の意見が対立することも考えられます。

また，帰無仮説検定からの移行は p hackingの駆逐に対して有効なのであって，他の不正に関しては効力をもちません。さらには，新しい統計手法に対して不正な手段が発生しないとも限りません。その意味では限定的な対策ですが，現状と比べると良い状況をもたらすことは間違いないと思います。

再現可能性の検証を促進するためには

次に，再現可能性の検証を促進するにはどうすればよいのでしょうか。まずは，直接的追試やデータの再分析がしやすい環境を整える必要があります。

第1に，実験手続きが詳細に論文内で紹介される必要があります。今から数十年前の研究論文では実験手続きが詳細に記されていました。1論文に実験が1つか2つしか紹介されませんでしたからそれが可能でした。現在では同じぐらいの紙幅で3つ以上の研究が紹介されます。その結果，手続きの紹介が簡略化される傾向にあります。これでは追試は困難になります。雑誌の執筆規程を改めて文字制限を緩和したり（オンライン・ジャーナルなら容易でしょう），付加情報の参照を可能にしたりするような措置が必要です。

第2に，実験刺激や質問項目の公開，共有が必要で

す。Bem 論文の直接的追試が複数ラボで可能になったのは、Bem が実験刺激を提供したからです。このような公開、共有は直接的追試の大前提となります。また、実験データとその分析手順も公開、共有されるとよいと思います。実験データの分析手順は本来実験手続きと密接に関連するもので、追試でも同じ手順で行われるべきものです。この第2、第3の点に関しては、Open Science Framework (OSF) が共有の場を提供しています⁽¹³⁾。これを積極的に利用することで直接的追試が促進されるはずで

さらに、平石さん、樋口さんの発言にもありますが、「ポジティブな研究」が刊行されやすい状況を変えなければなりません。現状では、目新しくオリジナリティのある仮説が美しく支持されてはじめて刊行されます。目新しさとオリジナリティは重要な価値観であり、これを求めることが学問の進展につながることは確かです。ただ、それに偏重するあまり、知見の頑健性の確認や理論の精緻化、特に制限状況の特定がおろそかになった部分があります。Bem 論文の追試に対するいくつかの雑誌の reject (掲載拒否) 対応はその現れです。ここのバランスを修正しなければなりません。

それには、やはり雑誌の編集方針の見直しが必要でないでしょう。その1つとして、直接的追試研究の意義を積極的に認めることが必要です。少なくとも、たくさん引用されるような理論的に重要である知見については早々に直接的追試を行い、再現可能性を検証すべきですし、その結果を出版して共有すべきだと思います。

そのときには、再現できた結果だけではなく、ネガティブ・データもしくは null effect (効果なし) 研究の意義も認めることが必要です。特に、一定の要件を満たしたネガティブ・データは重要な意味合いをもつと思います。具体的には、明確な理論をもち、十分な統計的検定力を備えた実験デザインで実施された研究は価値があるでしょう。それによって、かつてはあると信じられていた現象が存在しない可能性が高まるのであれば論文として掲載する価値はあると思います。例えば、血液型性格診断について、きちんとした手続きで十分な統計的検定力をもって検証した結果、ネガティブ・データが得られた (例えば、縄田健悟の論文⁽¹⁴⁾)、というのはいろいろな意味で意義があると思います。

先ほど、すべてのネガティブ・データに価値を見出すことは難しいと言いましたが、それでもどのようなデータであれ、公開される、もしくは利用可能になるべきだと思います。そうすることで、メタ分析の精度

が上がるかもしれませんし、どれくらいの実験者トレーニングが必要かの指針が得られるかもしれません。現在、null effect を示すデータを公開できる場所として、PsychFileDrawer⁽¹⁵⁾ があります。ここでの公開は研究業績となりませんが、それでも再現可能性に関する重要な研究記録を公開する場として機能しています。

ただ、直接的追試を行うときには、pre-registration (事前登録) が必要だと思います。これは、平石さんの発言にあったように、直接的追試をする前にその実施を宣言するというものです。この所作には、良い結果であれ悪い結果であれ公開するという研究者の姿勢があります。元研究に味方して良い結果だけ公表するのでもなく、元研究に敵対して悪い結果だけを公表するのでもないという公正な態度が保証されるのです。これを突き詰めると pre-review (事前審査) になります。これは、研究結果を事後的に審査するのではなく、研究計画を事前に審査して雑誌への掲載を決定するという制度です。この場合、良い結果であれ悪い結果であれ掲載することになります。出版バイアスに対する雑誌の公正な態度が保証されます。このような pre-review を制度化した雑誌も存在します (例えば *Comprehensive Results in Social Psychology* 誌⁽¹⁶⁾)。

先ほど統計解析のルール変更には困難が伴うと発言したのと同じで、直接的追試研究、特にネガティブ・データの雑誌掲載、pre-review 制度の採用は大幅なルール変更で、合意には困難が伴うと思います。一方で、希望ももてます。2015年6月26日、*Science* 誌にある記事が掲載されました⁽¹⁷⁾。それは、“The Transparency and Openness Promotion Guidelines”⁽¹⁸⁾ に関する記事で、研究の透明性とそれに伴う再現可能性検証の促進を狙ったガイドラインの提言でした。この提言に50の学会と526の雑誌が署名しており、中には心理学系の学会、雑誌も含まれています。すべての学会、雑誌が十全な透明性を確保できてはいるわけではありませんが、再現可能性を含む問題に多くが関心を寄せていることは明るい将来を感じさせます。

■ 文献・注

- (1) Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
<http://pss.sagepub.com/content/22/11/1359.full>
- (2) *Perspectives on Psychological Science* 誌再現可能性特集号
Special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528-654.
<http://pss.sagepub.com/content/7/6.toc>

- (3) Fanelli, D. (2010). Do pressures to publish increase scientists' bias?: An empirical support from US states data. *PLoS ONE*, 5(4), e10271. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010271>
- (4) Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615-631. <http://pps.sagepub.com/content/7/6/615.full>
- (5) Kerr, N. L. (1998). Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217. <http://psr.sagepub.com/content/2/3/196.short>
- (6) 平石界・池田功毅 (2015). 「心理学な心理学研究——Questionable Research Practice」『心理学ワールド』 68, 5-8. http://www.psych.or.jp/publication/world_pdf/68/68-5-8.pdf
- (7) Neuroskeptic のブログ <http://blogs.discovermagazine.com/neuroskeptic/>
Neuroskeptic 氏：再現可能性の問題などについて積極的に発言している匿名ブロガーであるが、その正体はニューロサイエンス領域の研究者であるという噂も。なおニューロサイエンス領域では死んだ鮭の fMRI を測定し、視覚刺激に反応する脳部位を発見したという有名な論文がある。fMRI は脳の活動を測定する機器であるが、もちろん死んだ鮭は視覚刺激に反応するはずがない。ちなみにこの論文ではあえて不適切な統計手法 (= p hacking) が用いられており、オチとして「適切な統計手法を使おうね」という主張がなされている。
鮭論文は以下。
Bennett, C. M., Baird, A. A., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon An argument for multiple comparisons correction. *NeuroImage*, 47(1), S125. <http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf>
- (8) Neuroskeptic, (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, 7(6), 643-644. <http://pps.sagepub.com/content/7/6/643.full>
- (9) Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. <http://ps.sagepub.com/content/25/1/7>
- (10) Vazire, S. (2015). Editorial. *Social Psychological and Personality Science*. <http://spp.sagepub.com/content/early/2015/09/29/1948550615603955.full>
- (11) Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37(1), 1-2. <http://www.tandfonline.com/doi/full/10.1080/01973533.2015.1012991>
- (12) 日本社会心理学会 (2014). 「春の方法論セミナー あなたの実験結果、再現できますか？——false-positive psychology の最前線」 http://www.socialpsychology.jp/sympo/seminar_140317.html
日本社会心理学会 (2015). 「第 2 回日本社会心理学会春の方法論セミナー GLMM が切り開く新たな統計の世界」 http://www.socialpsychology.jp/sympo/seminar_150325.html
*後半で少しベイズ統計について触れられる。
- (13) Open Science Framework <https://osf.io/>
- (14) 縄田健悟 (2014). 「血液型と性格の無関連性——日本と米国の大規模社会調査を用いた実証的論拠」『心理学研究』 85, 148-

156. https://www.jstage.jst.go.jp/article/jjpsy/advpub/0/advpub_85.13016/_article/-char/ja/
- (15) PsychFileDrawer <http://psychfiledrawer.org/>
- (16) *Comprehensive Results in Social Psychology* 誌 <https://osf.io/8mpji/wiki/home/>
pre-registration, pre-review を認める雑誌一覧 (OSF による) <https://osf.io/8mpji/wiki/home/>
- (17) Nosek, B. A., et al. (2015). Promoting open research culture. *Science*, 348(6242), 1422-1425. <http://www.sciencemag.org/content/348/6242/1422>
- (18) Transparency and Openness Promotion (TOP) Guidelines <https://cos.io/top/>

Section 4

Science 誌掲載論文の内容とは

——Science 誌にセンセーショナルな結果が掲載されたようですが、どのような内容だったのでしょうか。

三浦：

この連載が始まるきっかけとなった、第 1 回冒頭で紹介された論文のことですね。これまでの記事をお読みいただければおわりの通り、われわれは数年前からこうした動きを知り、自ら追試に取り組んでいたわけですから、「センセーショナル」とは思いませんでした。しかし、その内容を世界的に著名な 2 大科学雑誌 *Nature* 誌と *Science* 誌が速報し、論文を掲載したこと、そしてそれらを比較的詳しく紹介する日本語記事がオンライン・ニュースメディアで配信されたこと⁽¹⁾で、普段は「学問としての心理学」にそれほど関心をもたない人々も含めた世間の耳目を集めました。自分や他者の心の動きに関心をもつ人々、特に科学の無謬性、つまり「科学的研究には絶対に間違いなどないはずだ」という素朴な信念をもつ人々にとっては、「心理学ってなんていい加減なんだ」と思わせるに十分なニュースだったかもしれません。

ここではその内容のうち、日本語による報道にはない、あるいは正確さに欠ける部分を補足する形でかいつまんでご紹介します。Open Science Collaboration (OSC)⁽²⁾とは、志を同じくする研究者なら誰でも参加できるオープンな共同研究グループで、「集合知」を目指す研究チームだと言えわかりやすいでしょうか。この論文で追試対象となった研究は、*Psychological Science* (PSCI) 誌、*Journal of Personality and Social Psychology* (JPSP) 誌、*Journal of Experimental*

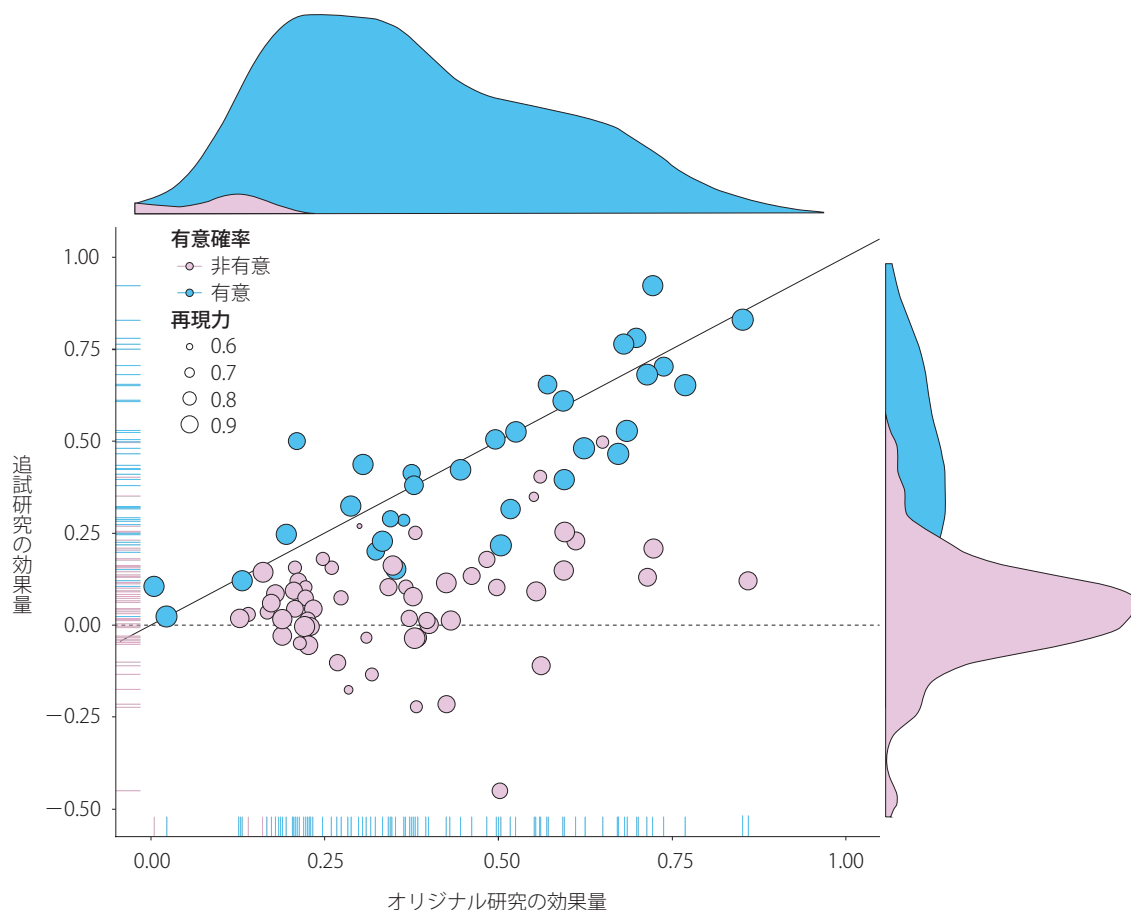


図 オリジナルと追試の効果量の比較⁽⁴⁾

Psychology: Learning, Memory, and Cognition (JEP: LMC) 誌の3誌⁽³⁾で2008年以降に刊行された論文のうち、事前に決められた一定の基準に該当する100本でした。これら3誌はそれぞれ心理学全般、社会心理学、認知心理学のトップ・ジャーナルです。

記事には「再現性が確認できた研究は「100件中の39件」「全体の39%」だった」といった表現が見られますが、これは、追試を実施した研究者たちの主観的評価、つまり「オリジナルの結果が再現されましたか?」という質問に「はい」と答えた比率が39%だった点を取り上げたものと思われます。より客観的なデータを見ると、オリジナルの論文のうち97%は統計的に有意な効果が得られていた（つまり3%は null effect [効果なし] 研究だった）のですが、追試では36%、追試の効果量の95%信頼区間に元論文の効果量が含まれていた（つまり「ほぼ」も含めて再現されたと考えてもよさそうな）研究は47%でした。また、オリジナルと追試をあわせると、もともと示されていた効果の有意性が保たれたのは68%でした。

図は、効果量の散布図です。横軸がオリジナル、縦軸が追試の効果量で、青丸が追試で「有意」、赤丸が同じく「有意ではない」結果だったことを、縦軸でマ

イナスの部分（点線より下）にプロットされている点は、追試ではオリジナルとは逆の効果が得られたことを意味しています。また斜線は縦横が同じ数値のラインなので、これより下の点が多いことは、オリジナルより追試の効果量が小さかった研究が多いことを示しています。また、グラフ上で右へ行くほど赤丸が減り、青丸が増えていることから、オリジナルの効果量が大きいと結果の再現性も高まる傾向が見取れます。さらに、追試を実施した研究者たちには、オリジナルの知見の重要性や意外性、追試実施の困難さや質、オリジナルと追試両方の研究者の技量や経験について評価を求めましたが、こうした要因は再現性にほとんど影響しないことも示されました。

論文は、「科学の究極の目標は真実である。われわれはエビデンスを蓄積することによって、科学コミュニティにおける方法論に自己修正を施し、これの追究のために努力しなければならない」と結ばれています。確かにやや心細い再現性だという感は否めませんが、特段に「センセーショナル」な結果でもないと思います。方法論の脆弱性を認めつつ、そのブラッシュアップを常に心がけることこそが、科学者に求められる科学的行為なのです。

——まさに現在進行形で、今後の展開から目が離せません。かたや日本でも学会などで取り組みが始まっているようですが、今回の科研のプロジェクトではどういったことに取り組まれるのでしょうか。

われわれの科研プロジェクト「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」(挑戦的萌芽研究2015-2017)は、社会心理学領域における再現可能性検証の日本における拠点を構築することを目指しています。

わざわざ国内拠点形成というワンステップを踏まなくとも、世界規模のOSCに個人で参加すればよいのではないかと、思う方もいらっしゃるでしょう。しかし、これまでのところOSCに参加した日本の研究者はわれわれを含めてほぼ皆無です。こうした反応の鈍さを説明する1つの要因は、初期参入コストの高さだと考えられます。コストは種々ありますが、そのうち1つは「言葉の壁」です。標準化された手続きの日本語化、それをを用いた実験の実施には人的・経済的コストがかかります。しかも、そのコストを負担して追試を行っても、それが学界において正当に評価される保証もないとなれば、追試研究の実施は研究者個人のきわめて高い「やる気」だけに依拠せざるをえなくなります。われわれも人間ですから、たった1人で、カネもコネもない環境でやる気を維持するのは困難です。そこで、まずは徒党を組もうということになったわけです。高い意欲をもつ研究者グループが初動部隊となり組織的な検証実験の拠点となる研究者ネットワークを形成すれば、直接的追試への新規参入コストを下げられるのではないかと、と。

このたびの*Science*論文に触れて、危機感を抱かない(社会)心理学者はいないでしょう。少しずつでも積み上げてきたつもりだった研究知見が、その地盤から崩れていくのですから、まるで杭打ちが不完全な施工不良のマンションのようなものです。ここで、わざわざ「(社会)」とカッコ書きを加えたことにも意味があります。心理学の中でも領域によりこの問題への対応には温度差があるからです。知覚・認知心理学あるいは心理物理学など行動指標を使う人たちにとっては、再現性の問題はまだ対岸の火事というか、それほど大きな問題になっていないようです。当該領域を専門とする同僚によれば、おそらくは追試にかかるリソースがそれほど小さくなく、なにか面白い現象が報告されたらまずそれを追試するところからスタートするので、再現性のない研究が一人歩きしにくいという事情があるのかもしれない、とのことでした。

そんな事情もあるせいか、日本の心理学系の学会で再現可能性問題に積極的に取り組んでいるところは、残念ながらまだほとんどありません。現在、私の知る限りにおいて追試研究歓迎を謳っている雑誌は、人間行動進化学会刊行の英文誌 *Letters on Evolutionary Behavioral Science* 誌⁽⁵⁾のみです。また『心理学研究』『社会心理学研究』『教育心理学研究』誌など主要誌に、研究材料や補足資料などをWeb公開する仕組みをもつところもなく、自らの発信する研究成果を再現可能性の検証に供する努力も未だしといった状況です。正直なところOSCとは比較にならないほど小さいですが、われわれの「実力行使」が少しでも制度や仕組みを動かす原動力となればと願っています。

科研費プロジェクトで着手した追試研究の具体的な内容は、日本社会心理学会第56回大会(2015年11月1日)のワークショップでご報告しました。先ほどご紹介したOSCのプロジェクトで追試対象となった研究も含まれています。報告内容や詳しいデータはぜひWebからご覧ください⁽⁶⁾。

最後に、これまでの日本の心理学者たちが再現性の問題を看過してきたわけではないことも強調させていただきます。例えば日本教育心理学会では、1988年の年次総会で開催されたシンポジウムの中でこの問題を取り上げています。『教育心理学研究』『心理学研究』『実験社会心理学研究』誌に掲載された97編の論文について学部学生による追試研究を行い、原論文との一致率を求めたという労作です⁽⁷⁾。残念ながら具体的なデータは記載されていませんが、一致率は全般的にあまり高くなかったとの報告がなされています。

再現性問題は「古くて新しい」問題なのです。心理学研究の再現性の低さが問われるというピンチを、組織的検証の取り組みを端緒につけるというチャンスに変えて、解決に向けて新たな一歩を踏み出します。

——よりよい心理学研究に向けた、意欲的な取り組みですね。最後に、プロジェクトへの意気込みをひと言ずついただけますでしょうか。

池田：

今後、さまざまな心理学のサブ研究領域で、先行研究の再現可能性をチェックし、信頼できる知見だけを選んで教科書を作り直す必要があります。今はまだ余裕がなく全然できていませんが、特に僕の個人的な関心対象である感情研究や進化心理学などについては、今後システムティックかつ継続的に、高い検定力を備えた追試研究を行っていきたくので、できるだけ早く

そのシステム作りを始めたいと思います。その際、すでに述べたように個人の力には限界があるので、国内・国外を問わず、同じ関心をもつ研究者間でのネットワーク作りが重要であると思われるので、今後積極的にその立ち上げを行っていきたいと思っています。

藤島：

再現可能性の検証と直接的追試の促進について日本はずいぶん遅れています。日本のデータは何でも比較文化データだと言われてしまう困難もあります。少しでも追いつく努力が必要ですし、他国の研究者との連携も深めていくべきだと思います。その手がかりとしてこのプロジェクトを生かしたいと思っています。

一連の騒動と今回のプロジェクトは、自分自身の研究を見る目を変えてくれているような気がします。科学的手法をより自覚するようになりましたし、同時に科学的でない手法の意味も少し理解できるようになりました。これらを生かして、新しい中にも安定感のある研究を遂行できればと思っています。

平石：

私個人としては「意気込み」というよりは、研究者という仕事をしている以上、やらねばならないなあ、というくらいの地味な覚悟です。研究という仕事の（私にとっての）良さの1つは、嘘をつかないことが尊ばれる世界であるということです。自分が清廉潔白人間であるなどと言うつもりは毛頭ありませんが、正直であることが尊重される社会の方が、自分は居心地が良い。しかし人間というのは弱いもので、悪気はなくとも、自分に好ましいようにデータを見てしまうことがあるのは、この連載でもすでに述べられている通りです。そういった歪みに気づいてしまったのであれば、少しでも治す努力はしないとね、ということです。ただ人間は弱いものなので、志は高くとも1人でやるのはしんどい。仲間がいるから続けられるという意味で、このプロジェクトにはとても感謝しています。

樋口：

私にとってこのプロジェクトは、まさに自分自身への叱咤になっています。自分および自分の研究室で行っている今の研究のやり方、これは本当に正しい知識を得るために有効なのか、それを常に考え続けることができる装置がこのプロジェクトです。

また、このプロジェクトで目指していることは心理学の世界において非常に重要なことだと確信していますが、それが多くの人に重要だと認識され、心理学界

におけるデフォルトの発想として定着するかどうかについてはあまり楽観的な見方をできていません。再現可能性問題への対策が定着していくためには、やはりその活動が楽しいことである、素晴らしい意味をもつことであると示されるのが重要でしょう。私が今一番大切にしたいのは、再現可能性の問題を考え検討していく活動がとてもエキサイティングで新しい知見を生み出すことができる、真に科学的な作業だということを身を以て示し続けるといふ点です。特に学部生や修士の学生に対して、身を以て示し続けたい。それ自体が、自分自身への叱咤にもなります。

三浦：

「息長く続けられるようガンバリマス！」はデフォルトとして、実験実習や卒論研究などに再現可能性プロジェクトを組み込むことで、基礎的な心理学教育の効果も同時に高めたいと考えています。心理学は教育プログラムが比較的きちんと確立されている学問領域ですが、中でも「自分たちで実験をしてデータを収集・分析してレポートを執筆する」という一連の流れを経験する「実験実習」は非常に重視されており、卒論研究はその延長線上に位置づけられます。再現性検証プロジェクトで対象となった研究を素材として使えば、詳細な情報が手に入るうえに、オリジナルばかりではなく世界各国の追試データも比較対象として手に入ります。こうすることで、前出 OSC 論文の結びにあった「科学的研究とは」という自覚をより強くもたせられるのではないかと考えています。

■ 文献・注

- (1) Baker, M. (2015). First results from psychology's largest reproducibility test. *Nature News*, 30 April.
<http://www.nature.com/news/first-results-from-psychology-s-largest-reproducibility-test-1.17433>
COURRIER (2015). 「有名な心理学研究のうち「再現可能性がある」ものはわずか 39% だった」『COURRIER JAPON』7月7日 (*Nature* 誌の記事紹介)
<http://courrier.jp/news/archives/1061>
Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
<http://www.sciencemag.org/content/349/6251/aac4716>
Sheridan, K. (2015). 「心理学の研究結果、6割以上が再現不可能 検証調査」『AFPBB News』, 8月28日 (*Science* 誌の論文紹介)
<http://www.afpbb.com/articles/-/3058654>
- (2) Open Science Collaboration
<https://osf.io/vmrgu/>
- (3) *Psychological Science* (PSCI) 誌
<http://pss.sagepub.com/>

Journal of Personality and Social Psychology (JPSP) 誌
<http://www.apa.org/pubs/journals/psp/>
Journal of Experimental Psychology: Learning, Memory, and
Cognition (JEP: LMC) 誌

<http://www.apa.org/pubs/journals/xlm/>

(4) (1) の Open Science Collaboration (2015) Fig. 3 より作成。

(5) 人間行動進化学会

<http://beep.c.u-tokyo.ac.jp/~hbesj/>

Letters on Evolutionary Behavioral Science 誌

<http://lebs.hbesj.org/index.php/lebs>

(6) 「実験結果の再現可能性検証に関する諸問題」日本社会心理学
学会第 56 回大会ワークショップ

https://sites.google.com/site/kgasasemi/replicability/jsps_kaken
hi/activity_log/jssp56ws

(7) 寺田晃ら (1988). 「教育心理学研究」の方向性とあり方を考
える——最近の動向を通して『教育心理学年報』27, 20-29.

[https://www.jstage.jst.go.jp/article/arepj1962/27/0/27_20/_
article/-char/ja/](https://www.jstage.jst.go.jp/article/arepj1962/27/0/27_20/_article/-char/ja/)

■ 著 者

池田 功毅 (いけだ・こうき) : 中京

大学心理学研究科・日本学術振興会特
別研究員 PD。主要著作・論文に

Shape and spatial working memory capaci-
ties are mostly independent (*Frontiers in*

Psychology, 6, 581, 2015, 共著), Fearful faces grab attention in
the absence of late affective cortical responses (*Psychophysiology*,

50(1), 60-69, 2013, 共著)。web サイト (<https://sites.google.com/site/kokiikedaresearch/home>),

Twitter: English (@kokiikeda), Twitter: 日本語 (@kokiikedaJP)。



樋口 匡貴 (ひぐち・まさたか) : 上智

大学総合人間科学部准教授。主要著作・
論文に「コンドーム購入行動に及ぼす羞

恥感情およびその発生因の影響」(*社会*
心理学研究) 25(1), 61-69, 2009 年, 共著),

「ビデオフィードバック法によるコンド
ーム購入トレーニングの効果に関する予備的検討」(*日本エ*
イズ学会誌) 12(2), 110-118, 2010 年, 共著)。Twitter (@

HIGUCHI_MA)。



平石 界 (ひらいし・かい) : 慶應義

塾大学文学部准教授。主要著作・論文に
Heritability of decisions and outcomes of

public goods games. (*Frontiers in Psychology*,

6, 373, 2015, 共著), 「進化心理学——理
論と実証研究の紹介」(*認知科学*) 7(4),
341-356, 2000 年)。web サイト ([https://sites.google.com/site/
kaihiraishi/](https://sites.google.com/site/kaihiraishi/)), Twitter (@kaihiraishi)。



藤島 喜嗣 (ふじしま・よしつぐ) : 昭

和女子大学人間社会学部准教授。主要著
作・論文に「自尊感情と自己関連動機に

基づく推論の歪み」(村田光二編『現代
の認知心理学 6 社会と感情』北大路書房,

2010 年), 「社会的影響」(遠藤由美編
『社会心理学——社会で生きる人のいとなみを探る』ミネルヴ

ア書房, 2009 年)。Twitter (@gsd9720)。



三浦 麻子 (みうら・あさこ) : 関西学

院大学文学部教授。主要著作・論文に
「オンライン調査モニタの Satisfice に関

する実験的研究」(*社会心理学研究*)

31(1), 1-12, 2015, 共著), 「東日本大震災
時のネガティブ感情反応表出——大規模

データによる検討」(*心理学研究*) 86(2), 102-111, 2015, 共著)。
web サイト (<http://asarin.team1mile.com>), Twitter (@asarin)。



* サイナビ! (URL 参照) に掲載された記事をもとに
作成しています。

<http://chitosepress.com/category/psychology-navigation/>

* 記載された内容の著作権等の知的財産権は、著者ま
たは著者に権利を許諾した者に帰属します。

* 購入者・利用者は印刷・配布して使用することがで
きます。

* CC BY-ND ライセンスによって許諾されていま
す。ライセンスの内容を知りたい方は [https://
creativecommons.org/licenses/by-nd/4.0/deed.ja](https://creativecommons.org/licenses/by-nd/4.0/deed.ja) でご確認

ください。

