

新進研究者 Research Note

科学哲学からみた心理学における再現性のタイプについて

On the types of replicability in psychology from the point of view of
philosophy of science

山銅 康弘

Abstract

I discuss the types of replicability in psychology from the viewpoint of philosophy of science. Specifically, the discussion between the papers of the philosophers of science Radder and the psychologist Schmidt are outlined. After that, I show that the types of the replicability by Schmidt is difficult from both the viewpoint of methods and purposes, and based on the suggestion from the philosophy of experiment by Radder, I clarify that the view of science established on the basis of the hypothetico-deductive method which Schmidt and others implicitly consider is inappropriate in psychology.

(1) 研究テーマ

科学哲学からみた心理学における再現性のタイプについて議論を行う。具体的には科学哲学者 Radder と心理学者 Schmidt の議論を概観する。その後 Schmidt による再現性の枠組みが方法の観点からも目的の観点からも困難が生じることを示す。また Radder の実験の哲学からの示唆を踏まえることで、Schmidt などが暗黙裡に考えている仮説演繹法的な科学観が心理学においては不適切であることを明らかにする。

(2) 研究の背景・先行研究

実験科学の分野において再現性が問題になっている。心理学の有力ジャーナルで掲載された 100 本の論文を追試したところ、元の結果を再現できたと判断できたのが半数にも満たなかったと報告した Open Science Collaboration(2015)は科学における再現性の危機 (replication crisis) を分野の垣根を超えて提起することとなり、さらにはその内容が一般紙にまで取り上げられ学際的、社会的な広がりを見せることとなった。

問題は多層的である。例えば上記において「判断できた」と遠回しな表現を用いたことには理由がある。Open Science Collaboration(2015)は再現で

きたことの指標として統計的有意性や効果量，追試者の主観的評価などを用いており，ある実験結果を再現するとは実際いかなることなのかを示す絶対的な証拠が存在せず，この問題の難しさがうかがえる。

また科学者が用いる「再現性」や「再現可能性」といった用語の不一致もこの問題をより一層難しくしている。Open Science Collaboration(2015)で用いられている“Reproducibility”という言葉はここ最近で定着しつつある用法とは異なっている。この研究では新しいデータを用いる直接的追試（詳細については後述）を行っているが，現在では“Reproducibility”は同じデータを用いて，記述された同じ分析方法で結果が再現できることを指し，訳語としては「再現可能性」をあてることが多い。対してデータや方法など何らかの異なる操作を行い，オリジナルの研究と同様の結果が得られることは“Replicability”と呼び，「再現性」と訳語を分けているようである（e.g. The Turing Way Community et al., 2019; 高橋, 2018）。本稿ではこれを踏襲し，以後取り上げる文献で使用されている言葉は適宜筆者が判断し現在での用法で言い換えていることに注意されたいⁱ。

「再現性の危機」は上記の再現可能性と再現性の両方が問題となっている。しかし再現可能性の問題は哲学的というよりも，いかに科学者が実験手順の十分な記述やデータの透明性を担保できるような研究環境を構築できるかという制度設計に関する話が主で，社会科学的側面が強いため今回は扱わない。本稿において中心的な話題となるのは再現可能性ではなく再現性である。

再現性の問題は科学哲学において中心的な話題として取り上げられることは少ない。それは実験の哲学が科学哲学の歴史において過小に評価されてきたことと軌を一にしている。しかし数少ない例外としてオランダの哲学者 Hans Radder が精力的に再現性についての議論を展開してきた。本稿では再現性のタイプにはいくつあるのかという問題を扱い，心理学を対象として取り上げることで昨今の再現性の危機と呼ばれるものについて科学哲学の観点から分析を行う。

具体的には再現性概念の議論の変遷について，主に物理学や化学を対象としたうえで三つの概念を提示した Radder(1992)，およびそれを参考に二つの概念に縮約した心理学者の Schmidt(2009)，またこれに対して批判的なコメントを加えた Radder(2015)を軸に議論を行う。その理由として Schmidt の提示した枠組みが現在の心理学では共通了解なものとなりつつあり（e.g. Makel et al., 2012; Earp & Trafimow, 2015），その影響関係に Radder がいることで，再現性という概念が科学においても哲学においても興味深いものとなっているからである。本稿では Schmidt の議論を主に Radder の提示し

た観点から批判を行うが、最終的には以上に挙げた文献すべてについて批判するという立場をとる。

Radder(1992)は再現性を「何を再現しているのか (of what)」という観点から3つに分類している。一つ目は「物的実現 (material realization) の再現性」、二つ目は「実験プロセス全体の再現性」、三つ目は「実験結果の再現性」である。

このうち最も説明が容易なのは実験プロセス全体の再現性で、これはオリジナルの研究と同様の理論解釈のもと、同様の手順を用いて実験を行い、結果が再現できることを意味している。次に実験結果の再現性は、同様の理論解釈のもとで行うが、結果に至るプロセスはオリジナルの研究と同一である必要はない。最後に物的実現の再現性であるが、これはオリジナルの研究と同様の手順を用いるものの、追試を行うメンバー内で特定の解釈が正しいという共通の信念を必要とせず、そのため実験手順は日常的な言語で記述されている必要がある。例として物的実現を主題として扱った Radder(2012)ではニュートンの質量とアインシュタインの質量で物体を量り同一の結果が得られることなどを挙げている。

このような着眼点は Radder が注力してきた実験の哲学についての背景を概観しておくことにより理解できる。Radder(2012)によると、ポパーやハンソン、クーンに代表される科学哲学は、科学をあたかも理論的、解釈的活動だけであるように分析され、実験などによる生産的行為、すなわち物的実現は存在しないか、よくて理論の応用としての派生物だとみなしてきた。確かに生産的行為、本稿で扱う範囲に限定すれば実験は理論的な分析をもとに実行され、実行されたプロセスは理論的に記述される。しかし実験によってひとたび作り出されたものには實在論的位置を占めているため、完全に理論に依存的なものではない。

Radder はさらに物的実現と理論解釈の相互依存性についてさらに分析しているが、さしあたりここで留意すべきは、再現性の種類という問題においては理論の側面と生産の側面という二つの次元から見る必要性があるということである。物的実現の再現性とは、理論を共有しなくとも実現されるような、相互依存性をなるべく生じさせないために要請される再現性ということになる。さらなる詳細については続く Schmidt との議論で明確になると思われるため、次に進むとする。

Schmidt(2009)は Radder の議論を踏まえ、まず三つを(1) 物的実現の再現性、(2) 理論的解釈の再現性、(3) 異なる物的手続きによる実験結果の再現性と表現する。そこから、(2)と(3)の違いは小さいと述べており、(1)をオリジ

ナルの研究と同様の手順で行う「直接的追試 (direct replication)」, (2)と(3)を異なる手順で行う「概念的追試 (conceptual replication)」と新たに名付けている。ここで違和感を覚えるのは、おそらく Schmidt の用語法であればむしろ違いが小さいのは同様の手順を行っているという(1)と(2)の方であろう。これは Schmidt が Radder の三つの再現性の例として選んだものが誤っていることⁱⁱに恐らく起因していると思われ、この点については留意されたい。

また Schmidt によるこのような分け方は伝統的に、心理学者である Lykken(1968)などによる追試研究の分類に近く、Lykken は「文字通りの追試 (literal replication)」, 「操作的追試 (operational replication)」, 「構造的追試 (constructive replication)」の三つに分けており、そのうち「文字通りの追試」は再現可能性のことを指すためそれを除いた二者と中身はかなり類似しており、なじみ深いものだと考えているようである。

最後に Radder(2015)であるが、Radder は Schmidt が二番目の再現性である「実験プロセス全体の再現性」ないし「理論的解釈の再現性」を排除していると述べている。ただし Schmidt(2009)は上記のように「違いが小さい」と述べるだけで排除したとは書いていない。上記のように Schmidt が再現性の例示を誤っているため、排除したと解釈するのが妥当だと判断したのだと思われる。その結果、Radder は一番目と二番目の区別が消失してしまっている点を批判している。「物的実現の再現性」と「実験プロセス全体の再現性」は、オリジナルの研究の手順を正確に再現する必要があるという点では同じであるが、理論を共有している必要があるかどうかにより二つの再現性は違いがある。一番目と二番目を区別しなくて良いのは「物的実現の共通言語記述の全ての側面が実験プロセスの理論的記述で考慮されている場合にのみⁱⁱⁱ」可能だと述べている。特に心理学では適切で正確な理論が欠如しやすいということから、理論になるべく依存しない形での再現性が重要になってくるため、物的実現の再現性と実験プロセス全体の再現性は分ける必要があるのだという。次節ではより詳細に彼らがどのような観点から再現性を見ているかを明らかにし、またその妥当性を考える。

(3) 筆者の主張

Schmidt はどのような観点から Radder の「物的実現の再現性」を「直接的追試」と呼びなおし、「実験プロセス全体の再現性」と「実験結果の再現性」を「概念的追試」とまとめなおした、ないしは「実験プロセス全体の再現性」を排除し「実験結果の再現性」を「概念的追試」と呼びなおしたのだろうか。

Schmidt(2009)では続く記述で、再現性の機能について述べている。直接的追試にはサンプリングエラーの制御，内的妥当性，研究不正対策，大きな集団や異なる集団への一般化可能性を満たす機能があると述べており，概念的追試には初期の実験の仮説を検証する機能があるとしている。しかし特段機能から再現性のタイプを考えたという記述はなく，むしろ再現性を二つのタイプに分けた結果割り振られる機能について述べているという意味合いが強い。おそらく Schmidt は実際にどう実験を行うのかという方法の観点から再現性を二つにまとめ上げたのだと思われる。そしてこのような方法の観点からの定義は直接的追試について述べている心理学の文献でよくみられるものである (e.g. 加藤, 2018)。「物的実現の再現性」はオリジナルの研究と同様の手順を踏んで行うのに対し、「実験結果の再現性」は結果が再現できれば良く，プロセスは故意に手順を変えているという点で異なる追試なのである iv.

しかしこの方法の観点からの再現性の分類は上手くいかない。なぜなら直接的追試はオリジナルの研究と同様の手順を踏まなければならないわけだが，それはどこまで正確に記述されたオリジナルの実験手順に合わせるか，ということが問題になる。完全に同一の実験を目指そうにも，部分的に同一の実験を目指そうにも失敗することを以下で示す。

まず直接的追試とは完全にオリジナルと同様の実験を行うことだと考えてみる。ここでは心理学で人を対象とした実験について、「誰に対して行うのか」と「誰が行うのか」という二点に注目してみよう。まず「誰に対して行うのか」であるが，完全に同じ性質を持った人々を別の集団から同じ数だけ選択するということが現実的に不可能であることは明らかであろう。またオリジナルの研究で参加した人に対して再度同じ実験を行うことで，完全に同じ性質を持った人々を対象にすることができるように思われるが，既にその実験を行ったという知識はオリジナルの実験の前には持っていなかったはずで，これはもはや同じ性質を持った参加者（被験者）とは言えない。

次に「誰が行うか」であるが，これも上記と同様で，同じ性質を持った別の実験者を選ぶことは難しい，なぜなら実験者は実験を遂行するための何らかの技能を持っているが，その技能は暗黙知であるため明文化できないことが多い。それゆえある実験を再現できなかった場合に，オリジナルの研究の仮説が間違っていたのか，それとも単に実験者の能力が不足していたために再現できなかったのかは一意的に決まらないということを Collins は多くのケーススタディから明らかにしてきた (e.g. Collins, 1985/1992)。

それでは完全に同一な実験はあきらめ，直接的追試は部分的に同一な実験

を目指すものだとすると、今度は概念的追試との区別が判然としなくなる。ある実験が行われ、その実験手順を元に直接的追試を行う際、どの手順は踏襲する必要があり、どの手順は変更、ないし無視しても影響がないのかを判断しなければいけないが、その基準はいかにして設けられるのか。これはもはや再現性を方法の観点からだけでは分析することはできず、その追試実験が何を目的として行われるのかという目的の観点が必要となってくるのである。

直接的追試と概念的追試について、Schmidt(2009)を参照した Earp & Trafimow(2015)は、直接的追試が特定の要因結果の検証を求めるもので、概念的追試は基礎となる理論や現象を検証しようとするものであると目的の観点から記述している。それでは引き続き方法の観点から直接的追試とは部分的に同一な実験を目指すものだとすると、特定の要因結果の検証が求めるという目的の観点を追加で導入し検討しよう。

これによってオリジナルの研究のどの手順は踏襲する必要があり、どの手順は無視しても大きな影響がないのかを判断できるようになると思われる。しかしこれは理論負荷的な行為であることに注意する必要がある。オリジナルな実験が一体どういった理論や現象を検証しようとしているものなのかを、追試者は解釈した上で実験手順を取捨選択し、直接的追試を行わなければならない。ところがこれは先ほどの概念的追試の目的そのものである。従って以上のような観点からでは直接的追試が不可能であると言えよう。Schmidtの直接的追試と概念的追試の枠組みには困難があることが分かる。

(4) 今後の展望

直接的追試の存在に対する過度な信頼の弊害をここでは展望として述べておく。先までの議論は Schmidt による直接的追試、概念的追試の枠組みに対し、Schmidt 的な用法である方法や目的の観点から検討した。この理論的な観点が含まれていないという Schmidt の用法は、言い換えれば「実験プロセス全体の再現性」を排除したという Radder の矛盾のない Schmidt 解釈を採用したうえで検討を行ったということに近いといえる。したがってまだ、直接的追試を物的実現の再現性ではなく実験プロセス全体の再現性として考えるという可能性が残されている。

ところが Radder による実験の哲学の立場からの示唆として、理論解釈と物的実現という二つの次元の問題を元に分析すれば、それもまた困難が生じる。共有された理論解釈のもと、同一の実験手順によって結果を再現するということが科学において重要だという考えは厳密な仮説演繹法を念頭に構築

されているという点に注目されたい^v。Radder がもともと対象として扱っていた物理学や化学においては、かなりの精度で同様の実験手順を行うことができるため、その結果再現できないことが仮説の棄却を意味し、実際的には論文が撤回される^{vi}。

しかし心理学においては完全に統制された条件を実現させることはできない。閉じた実験系 (Radder 2012) においてのみ、実際の実験手順が理論的記述と一致する。別の表現をすれば実際の実験手順のほぼ全てを記述することができる。心理学は物理学や化学とは異なり参加者の性質などを完全に統制することは不可能なため、閉じた実験系は実現できず、実際の実験手順が理論的記述と一致しない (例えば個々の参加者の性質などは十分には把握できないため全てを記述することはできない)。したがって厳密な仮説演繹法はここでは成立せず、再現ができないことを理由としてオリジナルの研究を撤回する、ということは心理学という科学において要請されることはない。むしろ再現できないことが、オリジナルの研究で扱われている何らかの対象や概念に対する理解を促進させ、新たな研究へとつながっていくという有益な機能さえ持ちうる。

まとめると、厳密な仮説演繹法を持たない科学においては再現できないことがすぐさま科学としての要件を満たさないということの意味しない。現在「再現性の危機」と呼ばれ議論されている問題点は、その全てが真の危機だというわけではなく、心理学のような科学では原動力と呼べるものさえそこには含まれてしまっているのである。

本稿では理論的分析に終始したが、現在では間違っていると判断される実験結果とその理論的解釈が、結果的には後続の研究を後押ししたというケーススタディを用いて、実証的にも明らかにしていきたいと考えている

i 前もって述べておくと、Radder(1992) や Radder(2015)では「実験結果の再現性」を“replicability”と呼び、それ以外を“reproducibility”と呼んでいる。また Schmidt(2005)や Earp & Trafimow(2015)においては直接的追試、概念的追試について“replication”を用いているが、これらは全て本論では「再現性 (replicability)」に統一している。

ii 具体的に理論的解釈の再現性の例として、Schmidt は光の速度には限界があるというアインシュタインの仮説を少なくとも二つの完全に異なる実験系から検証したことが挙げられているが、この例は Radder(1992)では取り上げられていない。

iii Radder(2015), p.9, 拙訳。

iv Radder(1992)では「実験結果の再現性」を異なる実験手順だけでなくオリジナルの実験手順による結果の再現も含んでいると包含的な概念として

述べている。他方 Schmidt は直接的追試と概念的追試を排他的な概念として捉えており、ここからも Schmidt の枠組みは Radder の枠組みの単なる縮約ではないようである。

v 心理学者が心理学に対して仮説演繹法を明示的に使用したものとして Chambers(2017)がある。

vi Collins(1985/1992)は物理学であっても再現できたか否かの判断に社会的な力学が作用することを主張するが、本稿ではこれ以上は深入りしない。

(5) 参考文献

加藤 司, 2018, 「『パーソナリティ研究』の新たな挑戦—追試研究と事前登録研究の掲載について」, 『パーソナリティ研究』, advpub.

高橋 康介, 2018, 『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.

Chambers, Christopher D., 2017, *The Seven Deadly Sins of Psychology A Manifesto for Reforming the Culture of Scientific Practice*, Princeton: Princeton University Press.

Collins, Harry M., 1985/1992, *Changing Order: Replication and Induction in Scientific Practice*, Chicago and London: University of Chicago Press.

Earp, Brian D. and David Trafimow, 2015, “Replication, Falsification, and the Crisis of Confidence in Social Psychology”, *Frontiers in Psychology* 6(621), pp.1-11.

Lykken, David T., 1968, “Statistical Significance in Psychological Research”, *Psychological Bulletin* 70(3, Pt.1), pp.151-159.

Makel, Matthew C., Jonathan A. Plucker, and Boyd Hegarty, 2012, “Replications in Psychology Research: How Often Do They Really Occur?”, *Perspectives on Psychological Science* 7(6), pp.537-542.

Open Science Collaboration (OSC), 2015, “Estimating the Reproducibility of Psychological Science”, *Science* 349(6251), pp.943-951.

Radder, Hans, 1992, “Experimental Reproducibility and the Experimenters' Regress”, *PSA* 1992(1), pp. 63-73.

---, 2012, *The Material Realization of Science: From Habermas to Experimentation and Referential Realism*, Boston: Springer.

---, 2015, “Causality, Theory Ladenness and Reproducibility in Experimental Science”, Conference: 2015 OZSW Conference, At VU University Amsterdam.

Schmidt, Stefan, 2009, “Shall We Really Do It Again? The Powerful

Concept of Replication Is Neglected in the Social Sciences”, *Review of General Psychology* 13(2), pp.90–100.

The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, Anna Krystalli, Alexander Morley, Martin O'Reilly and Kirstie Whitaker, 2019, “The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4)”, *Zenodo*, <http://doi.org/10.5281/zenodo.3233986>

(東京大学)