

人間の観察研究における再現可能性の問題

小 島 康 生

中京大学

Reproducibility of observational research in human subjects

Yasuo KOJIMA

Chukyo University

The reproducibility of data has most frequently been discussed with respect to experimental research in social and cognitive psychology. This article addresses how the reliability and validity of an observational study can be confirmed and whether an observational study can be replicated in a natural setting, as well as in an experimental (or semi-experimental) setting in developmental psychology. In a natural setting, the subject's cultural and historical background, as well as immediate factors, including the physical environment and social context, may need to be considered to understand individual behaviors and social interactions among two or more people. Therefore, it can be quite difficult to perform a meaningful replication study. Researchers in the various domains of psychology should employ various measures, including the use of alternative procedures other than a replication study, in order to ensure the reliability and validity of their findings.

Key words: observational study, human, interpersonal interactions, natural settings, reliability, validity

キーワード：観察法，人間，対人的関わり，自然場面，信頼性，妥当性

1. 原稿依頼を受けて

正直に白状してしまうと、実のところ筆者は、今回、本特集号の原稿依頼を受けるまで、心理データの再現可能性やそれをめぐる諸々の問題が、国内外においてこれほど白熱した議論を巻き起こしていることを知らずにいた。その後、本号の掲載論文を拝読し、加えて少しばかり関連文献にも目を通すなかで、この再現性の問題は、少なくとも筆者の専門領域である発達心理学においてはまだそれほどホットな話題にはなっていないこと、ほかにもこのテーマにさほど大きな関心を寄せていない分野があることを強く実感した。件の再現性問題は、領域ごとにかんがりの温度差をもって受け止められているのではないかというのが正直な感想である。

2. 実験研究に偏った再現性の議論

再現性と聞いてまず頭に浮かんだのは、大学の

学部で以前に担当していた2年生対象の実験演習のことである。心理学の一般的な実験を1年にわたっていくつもこなす、心理学系の大学ならどこでも開講している授業で、筆者は短期記憶の実験を担当していた。パソコンのディスプレイに順次提示される刺激を記憶し、その場で筆記再生するというオーソドックスな実験である。刺激内容が何種類かに分かれているほか（平仮名、漢字熟語など）、提示される単語と単語の時間間隔もいくつかの条件に分かれており、それらの変数が再生単語数にもたらす影響を調べようとするものである。

この授業を半期の間に各回20～30名ずつ、5～6グループで実施し、実験結果をレポートにまとめさせるのだが、筆者が担当していた数年間、おそらく50前後のグループで同じ実験を行ったが、結果は常にほぼ同じであった。本号でも各執筆者が述べているとおり、同じ手続きで追試研究を実施し、同様の結果が再現されることをもってデータの頑健さが担保されるのだとするなら、こうし

た実験はまさしくそのお手本のようなものだといえよう。これまで心理学の多くの研究がそうした追試をきちんと行わず、またデータの収集方法や分析のあり方、発表の仕方に至るまで、疑問符のつく研究実践（Questionable Research Practices）を放置してきたのだとすると、それは間違いなく大問題である。

だが心理学は、これら実験系の研究だけで占められているわけではない。奇しくも今回、本号の執筆論文を読んで最も強く感じたのは、ほとんどが実験室実験を前提に議論が展開されていることへの不満ないし違和感であった。周知のとおり、実験研究は、仮説を定立し、周到に準備された手続きにしたがってデータを集め、その仮説を検証する形で進められるものが多い。しかし、分野やテーマによっては、それらと対極をなす、いわゆる仮説生成型の研究も存在する。観察法を用いる研究では、そうしたタイプのもの、すなわちほとんど白紙に近い状態から出発し、観察された事象をていねいに記述・記録することを通して新しいものの見方を説き起こすようなタイプのものが比較的多いように思える。以下では、そうした観察研究を念頭において、再現性の問題にどう向き合えばよいのか、考えてみたい。

3. 観察研究における再現性の問題 —相互作用の観察を例として—

もちろん、統制された実験場面で行われる観察研究はある。森口（2016）で取り上げられていた新生児模倣の研究（Meltzoff & Moore, 1977）はまさにそうである。だが、このような実験研究では、観察すべき行動は通常一つ、ないし少数に限られており、条件に応じてそれらの行動の生起の有無や、出現した行動の分類を行うなど、分析の方向性もある程度決まっていることが多い。一方それに対し、多種類の、しかも判断に慎重さを要する行動を同時に扱い、分析の方向性も必ずしもはじめから定まっていなかった研究もある。発達心理学でよく取り上げられる母子関係や夫婦関係、仲間関係、友達関係など対人的な相互作用の観察はその典型例といえる。

かつて筆者は、乳幼児期のきょうだいと母親の相互作用を観察し、子どもに対する母親の行動と

きょうだいの相互作用との関連を分析したことがある（Kojima, 2000）。各家庭に出向いて筆者が持ち込んだおもちゃ（ブロック）で遊び場を構成し、その様子を観察するという、やや実験的な趣向の強い（準実験的と呼んでもよい）観察であったが、この研究は、あらかじめ何かしら仮説があつてスタートしたわけではなかった。家庭訪問を繰り返すなかでそのつど気づいたことや、撮影された映像を後で見ながら考えたことをたよりに、試行錯誤的に分析を進めていったというのが実状で、具体的には、個々の行動をどう切り分けて評定するのが適切か、どのようにすれば筆者のアイデアを数値化できるか、かなりの時間をかけて分析を行った記憶がある。仮説検証型の実験研究を行っている研究者からすれば、なんと行き当たりばったりでいい加減なやり方かとあきれられそうだが、対人的な相互作用の観察研究を行っている者は、一度や二度はこれとよく似たプロセスでデータをまとめた経験があるのではないか。

以上のように、特に対人間の相互作用を扱う観察研究では、データを切り取る行動カテゴリーの策定にかなりの労力を割き、またそこが研究者としての腕の見せどころでもある。テーマによっては先行研究の行動カテゴリーをそのまま援用することもあるが、自分の集めたデータに即して多少それをアレンジすることのほうが多いのではないか。

4. 観察データの信頼性

このように述べてくると、観察研究を行っている者は随分と好き勝手に、いい加減な分析をしていると思われそうだが、決してそのようなことはない。行動観察のデータは通常、複数の評定者が独立に評定し、カッパ係数（Cohen's Kappa coefficient）などの数値が一定以上であることをもって信頼性を担保する（中澤・大野木・南, 1997）。なお、この評定では、当然ながら分析に用いる行動カテゴリーに関して評定者間でコンセンサスをはかっておくことが不可欠である。そのためには、すべての行動カテゴリーに対し操作的な定義を行い、どのような行動ややりとりがどの行動カテゴリーに該当するのか、判断の基準を明確化しておくことが重要であり、先の一致率の評定もそ

れに基づいて実施される。

ところで、行動評定を行う際には、判断が比較的容易なものとそうでないもの（つまり、判断が分かれやすいもの）がある。「微笑む」、「触る」、「叩く」などの行動は、目で見て確認しやすいため、評定者間で判断が分かれることは比較的起こりにくい。それに対し、「指示する」、「教える」などの行動カテゴリーは言語的な働きかけが主で、そのときの口調や声のトーン、前後の文脈など、一見すると冗長で曖昧だが、じつは重要な意味を含む情報を参照する必要がある。したがって、操作的定義や判断の基準を定める際により慎重な対応が求められるのは、これらの行動に関して、である。

5. 対人的相互作用の観察研究は 追試が可能か

追試を行い、分析結果の再現性を確認することは、対人的な相互作用など複雑かつ曖昧な情報を扱う観察研究においても間違いなく重要であろう。一つひとつの行動に対する操作的定義を厳密に整え、判断の基準を明確にするのは、評定者間一致の確認のほか、それにしなくては追試が行えることのためでもある。しかし、たいていのジャーナルでは、字数制限などの制約から、各行動の事細かな説明や具体例、評定に際しての基準にいたるまで、正確な追試が直ちに可能なほどの詳細な記述はなされていないことのほうが多い。先ほど述べたとおり、相互作用の観察データには、時系列的な流れや前後の文脈、話し手の言い回しや声のトーンなど、多数の情報がふんだんに含まれるため、追試を推奨するには、それが実現できるよう相当量の情報を提供しなくてはならない。したがって、論文情報とは別に、そうした情報へのアクセスが可能となるようしくみ作りが必要であろう。具体的には、読者が自由にアクセスできるサイトを開設するなどして、より詳細な観察手続きや行動カテゴリーの決定にいたる紆余曲折、そして何より、実際に観察された行動と行動カテゴリーとの対応関係を示しておくことが有効であろう。

ただ、人間の観察データに関しては、研究者間の情報共有に際していっそうの倫理的配慮が必要

である。「こういう流れで、こういったやりとりがあった場合にこのカテゴリーに分類する」といった情報を他者に伝えるのには、本来、実際の映像を提示するほうが、誤解が少なく判断の食い違いも生じにくい。だが一方で、ビデオの映像は個人情報であるため、協力者の同意なしに他者に閲覧を認めるわけにはいかない。個人情報の漏洩を防ぎつつ他の研究者と情報共有をはかるためには、研究者としての適切な倫理観が求められよう。

以上、基本的には観察研究の追試は可能であるし、その取り組みは今後いっそう活発に行われるべきとは考えるが、一方で、集めたデータの仔細な検討とていねいな記述・記録の蓄積を通じて、新たな切り口での現象の理解と理論構築を目指す仮説生成的な研究では、追試という営み自体に抵抗を感じる研究者も多いのではないか。以下では、自然（あるいは自然に近い）場面での非参与型の観察研究のほか、人類学、社会学の分野に端を発し、いまや心理学の世界でも目にするのが多くなったエスノグラフィーや参与観察の研究にも触れながら、そうしたアプローチによる研究での追試研究についてどう考えるのか、仮に追試が難しいとなれば、何が代替手段になりうるか、さらには広く心理学全体を見通して、研究成果の信頼性や頑健性をどのように確立していけばよいのかを考えたい。

6. 自然場面での観察研究

観察研究の醍醐味のひとつは、人が生活する日常場面に身を寄せ、自然な動きや関わり合いを観察し、ときにはそれらの人々とディープに関わりあうなかで新たな知見への到達をめざすところにある。このような生態学的妥当性、すなわち本人のありのままの姿を観察することを重要視する研究は、当初、実験室での研究に対するアンチテーゼとしての意味あいが濃かった（Bronfenbrenner, 1979/1996）。だが、自然場面の研究と実験研究とは本来、補完しあい、互いに影響を与え合いながら進められるべきもので、近年は、心理学の諸分野でその論点からの活発な議論が進められつつある（その一例として、伝・諏訪・藤井, 2015など）。

自然場面と一言と言っても、その内容は多種多

様である。たとえば、子どもの社会性の発達をテーマとする研究では保育所や幼稚園での観察がよく行われるが（例えば、加藤ら、2012；野澤、2013など）、「教室」という空間は、限られた資源（同じおもちゃ、同じ絵本など）を背景に、同一のメンバー間で展開するやりとりを観察対象とするという意味では、「実験」とは言わないまでも、設定された場面とみることができよう。一方、後に述べるように、ショッピングモールや街なかなど、より自由度の高い場面での人の行動観察となると、状況はいっそう煩雑で、「自然」の度合いはこちらのほうが強い。

ここではまず、専門分野の異なる読者にも想像しやすいよう、二つの例をあげよう。ひとつは、幼稚園での子どもどうしのトラブルの研究である。この種の研究では、観察者は園児とある程度の距離を保ち、やや離れたところから観察する、いわゆる「非参与的な観察」（中澤ら、1997など）を行うのが一般的である（倉持、1992；高坂、1996など）。仮にビデオカメラでの撮影が可能だったとすると、実験場面での観察と同じく、撮影された映像を複数の評定者で独立に評定し、トラブルが始まった経緯やトラブルの種類、他者の介入、解決に至るプロセスなどをあらかじめ決めた行動カテゴリーに沿って分類していくのが一般的な分析プロセスであろう。このようにしてデータの信頼性を確認するのは、先に述べた実験的（ないし設定場面での）観察とまったく同じである。ただし、自然場面ではビデオカメラが使えないことも多い。その場合は、複数の観察者が観察場面に立ち会い、その場で独立に評定を実施して評定の一致率を確認するのが常套手段である。

もうひとつ例をあげよう。筆者はかつて、地域のショッピングモール内で子ども連れの家族の観察を行ったことがある（小島、2001）。所定のエリアを通行していく家族を対象に、少し離れたところから親子間の関わり（父親、母親のどちらがベビーカーを押していたか、誰と誰が手を繋いでいたかなど）を筆記記録し、子どもの人数や年齢に応じて親の行動や夫婦間の役割分担の特徴を描き出そうとしたものである。この研究は、「ベビーカーを押す」、「手を繋ぐ」など、判断の分かれにくい、比較的シンプルな行動のチェックで済むものだったため、観察も評定もすべて筆者一人

で行った（詳細は、拙稿をご覧ください）。

さて、最初に紹介した幼稚園での子どもの観察と二つ目の屋外での親子の観察は、いずれも自然場面での非参与型の観察だが、これらは追試が可能だろうか。幼稚園にしてもショッピングモールにしても、同じ（あるいはかなり似た）場所で、同様の手続きで観察を行うことはもちろん可能である。しかし、実験室での観察と比べると、状況を均質化することは格段に難しい。幼稚園での観察に関して言うと、教室の広さやクラス内の子どもの数、置かれている玩具の種類や数は園ごとに異なっていると考えるのが自然であろう。教室内の保育士の人数は統制できないうえ、トラブルが起きたときに保育士がどう対応するかも園ごとに方針が異なるであろう。一方、ショッピングモールでの観察の場合、周囲の構造物、人の混み具合、その日の気温や天候など、ここでも観察結果に影響を与える変数は無数にあり、それらを再現することはやはり不可能と考えるのが妥当であろう。筆者が先の観察を行ったのは今から20年近く前のことだが、その当時に比べていまは、ベビーカーや抱っこ紐など育児具の性能もかなり進んでおり、それが結果に影響することは十分に考えられるし、「イクメン」という言葉が知れ渡った現在のほうが、父親から子どもへの関わりかけが積極的に行われているという結果が得られるかもしれない。以上のように考えると、幼稚園でもショッピングモールでも、自然場面での観察に関しては、直接的な追試を行える可能性はかなり低く、仮に無理をしてオリジナル研究の観察条件に近づけようとする、自然観察がモットーとする生態学的な妥当性（Bronfenbrenner, 1979/1996）が損なわれ、研究自体が成り立たなくなる可能性すらある。

では、そのように再現が難しい（あるいは、追試を前提として行われていない）研究の成果には意味がないのか。少なくとも筆者は、そうは思わない。これに関連して、Finkel, Eastwick, and Reis (2015) は、観察条件や手続きの完全な再現が不可能であることを前提とした「緩やかな追試研究（close replication）」を推奨しており、オリジナル研究の結果との一致や差異を取り上げ、双方の研究の背景の違い（物理的・社会的環境、地域性、時代性など）に着目することで、現象のより深い

理解をめざすことができるのではないかと論じている。たしかに、状況が完全に同じでないことを逆手にとって、それに即した考察を積み重ねることがオリジナル研究、新研究の双方に利益をもたらす可能性は十分にある。

このような追試が実現するためには、一つひとつの研究論文において、観察事態を取り巻く周辺的环境や文脈、観察者と被観察者の関係、観察者の立ち位置、記録の方法、分析の方向性やプロセスなど、じっさいにたどった経路を時系列に即してくまなく記し、それらが結果にどう反映したかを深く考え抜くことが不可欠である。そもそも、人の行動、特に人と人の相互作用は、周囲の文脈、文化、歴史的事情を“地”にして、展開している。自然場面での観察であるからこそ、それらを意識した結果の解釈やまとめ方が必要であり、それが観察研究における追試（直接的追試ではなく、差異に根ざした追試とでもいおうか）の可能性を拓くものと考えられる。

7. 参与観察やエスノグラフィー

最後に、より特殊な観察研究の一例として、エスノグラフィーや参与観察のことを述べておきたい。これらの手法の特徴は、研究者が比較的長期にわたって研究フィールドに身を寄せ、そのフィールドの人々と密接にかかわりながら、観察されたこと、聞き取ったことを（場合によって）研究者の主観も交えて記述する点にある。紙数の関係上、詳細は他にゆだねるが（Cooper et al., 2012; 佐藤, 2006; やまだら, 2013）、鋭い観察力と分析プロセスの仔細な説明、文章の論理性により、読み手を説得するに足る、すぐれた研究もある。

さて、このような研究において追試は可能だろうか。おそらくそれは、先の非参与型の観察研究以上に難しいであろう。そもそもエスノグラフィーや参与観察は、現場の人々に密着し、内部者に近い立場になりつつ、それでいて一歩下がった距離感で主観も交えて見えたことをまとめる営みである。つまり、これらのアプローチでは、他者による追試という発想自体が存在しない。

では、追試自体を容認しない、そのような研究は、科学を標榜する心理学の世界では排除せざるを得ないのか。その判断は研究者によって分かれ

るところだろうが、一部には、一般的な仮説検証型の研究とは異なる基準で研究成果の信頼性や妥当性を認めてもよいのではないかとという考えがある。研究者がどのようなプロセスをたどってそのフィールドに参与したか、いかなる経緯で具体的な研究テーマを着想するにいたったか、何を観察し、そこで何を感じ取ったか、どういう分析プロセスでデータをまとめたか、以上のことを読み手があたかも追体験できるように“分厚く”記述し、読み手がその記述に納得すれば、それはある意味でデータの信頼性や妥当性が確認されたといってもよいのではないかとする考えである（麻生, 2009; 佐藤, 2006）。本特集号で多くの執筆者が述べていることからすると、180度発想の違う考え方だが、心理学の世界ではそういった視点もあることをぜひとも頭に留めておいていただきたい。

8. おわりに

人間観察、それも対人的な相互作用の観察に代表されるように、複雑で、瞬間的にポイントを捉えることが難しい行動を扱う研究においても、通常の実験室研究に準じて可能な限り観察条件や手続きを統制した追試を行うべきことはたしかである。だが、これまでに述べたとおり、観察という手法はそもそも、試行錯誤を繰り返しながら仮説を生成しようとする研究との相性がよく、それと関連して、追試を行うことが不可能であったり、追試自体が意味を成さない（追試を想定していない）研究もある。これらの研究では、観察状況や観察者・被観察者の関係性、観察者の立ち位置、データの集め方や分析プロセスに至る一部始終を可能な限りオープンにし、差異を意識した追試を行えるよう配慮することや、その研究論文を読んだ読者が書かれていることを了解し、場合によって追体験できることが、実験室研究での追試に代わる、結果の信頼性、妥当性に繋がるのではないか。

冒頭にも述べた通り、現在の再現可能性の議論は、一部の、しかも実験系の研究テーマでの議論に大きく偏っており、心理学全体を包括するようなものには進展していない。本稿で述べた人間観察、それも自然場面での観察研究での問題を踏ま

えて考えると、あらゆる分野、テーマでの研究の価値を追試と結果の再現可能性で保障していこうという流れには無理があるように感じる。今後、さまざまな専門分野の研究者が意見交換を交わすなかで、自らの分野においてこの科学性をどう保障するか、考える機会が必要なのではないか。

文 献

- 麻生 武 (2009) 「見る」と「書く」との出会い 新曜社.
- Bronfenbrenner, U. (1979). *The Ecology of human development: Experiments by nature and design*. Cambridge MA: Harvard University Press. 磯貝芳郎・福富 護 (訳) (1996) 人間発達の生態学—発達心理学への挑戦 川島書店.
- Cooper, H., Camic, P. M., Long, D. L., Panter, A. T., Rindskopf, D., & Sher, K. J. (Eds.) (2012). *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. Washington, DC: American Psychological Association.
- 伝 康晴・諏訪正樹・藤井晴行 (2015) 特集「フィールドに出た認知科学」編集にあたって 認知科学, 22, 5-8.
- Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology*, 180, 275-297.
- 加藤真由子・大西賢治・金澤忠博・日野林俊彦・南 徹弘 (2012) 2歳児による泣いている幼児への向社会的な反応：対人評価機能との関連性に注目して 発達心理学研究, 23, 12-22.
- Kojima, Y. (2000). Maternal regulation of sibling interactions in the preschool years: Observational study in Japanese families. *Child Development*, 71, 1640-1647.
- 小島康生 (2001) 外出中の家族を対象とした親子の関わりと夫婦間の役割調整—子どもが1人の家族と2人の家族の比較を通して 家族心理学研究, 15, 25-34.
- 倉持清美 (1992) 幼稚園の中のものをめぐる子ども同士のいざこざ：いざこざで使用される方略と子ども同士の関係 発達心理学研究, 3, 1-8.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-78.
- 森口佑介 (2016) 発達科学が発達科学であるために 心理学評論, 59, 30-38.
- 中澤 潤・大野木裕明・南 博文 (1997) 心理学マニュアル 観察法 北大路書房.
- 野澤祥子 (2013) 歩行開始期の仲間同士における主張的やりとりの発達過程：保育所1歳児クラスにおける縦断的観察による検討 発達心理学研究, 24, 139-149.
- 佐藤郁哉 (2006) フィールドワーク増訂版 書を持って街へ出よう 新曜社.
- 高坂 聡 (1996) 幼稚園児のいざこざに関する自然観察的研究：おもちゃを取るための方略の分類 発達心理学研究, 7, 62-72.
- やまだようこ・麻生 武・佐藤達哉・能智正博・秋田喜代美・矢守克也 (2013) 質的心理学ハンドブック 新曜社.

— 2016. 4. 22 受稿, 2016. 4. 27 受理 —