

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—

森 口 佑 介

上越教育大学

Developmental science: Replicability and robustness in developmental studies

Yusuke MORIGUCHI

Joetsu University of Education

The replicability and robustness of psychological research has been questioned in recent years. However, few researchers have addressed the issue in developmental science. In this article, I discuss the replicability and robustness of developmental science. Two examples are reviewed. First, I introduce experiments on infants and young children. Several factors that may affect the results of the experiments, therefore revised provided difficulty in replication. The ManyBabies project was launched to resolve the issue. Second, I introduce longitudinal research the that examines is long-term developmental trajectory. Longitudinal research may cost time and money, and therefore robustness rather than replicability may be assessed. Finally, I argue the situation in domestic developmental science, and propose that more efforts should be done to improve reliability in developmental research.

Key words: developmental science, replicability, robustness, infants and young children, longitudinal studies

キーワード：発達科学，再現性，頑健性，乳幼児研究，縦断研究

1. はじめに 「発達で突き崩されないものがあるか」

数年前に開催された発達科学に関する研究合宿で、招待講演者から発せられた言葉である。世界的に著名な知覚心理学者で、発達科学にも詳しい講演者は、知覚研究の例を出しながら、研究の頑健性を説いた。たとえば、成人を対象にした知覚研究では、ランダムドットステレオグラムの頑健性は確かなもので、繰り返し再現されており、心理学において確かな知見の一つである。翻って、発達心理学や発達科学において、特定の図形に対して同じような錯視を生じることと同じようなレベルの、確かな現象や行動があるだろうか。とりわけ、研究へのアクセスが他の年齢層よりも難しい乳幼児の研究で、それだけの再現性は見られるのか。

合宿の参加者はこの問いかけに対して、いくつ

かの例を挙げた。筆者個人としては、乳児の知覚研究の例をいくつか思いついたが、専門の研究者に後で聞いたところによると、それだけの頑健性があるとは言い難いとのことであった。他の合宿参加者は、後で詳しく検討するが、発達研究で最も有名な例の一つである新生児模倣を挙げていた。だが、この例も招待講演者を納得させるほどのものではなかった。

無論、成人の研究と乳幼児の研究を同列に論じるのは無理があるだろう。成人の研究であっても、たとえば一部の物理現象と比べると、その確かさは比べものにならない。異なった研究領域における科学の在り方を同列に論じることには限界がある。しかしながら、それであっても、乳幼児研究において突き崩されないものがあるのか、という点については深く考えさせられた。

本論文では、発達科学における再現性問題の現状について考えてみたい。近年の心理学再現性の

全般的な問題や統計的な問題等についてはこの特集号の他の論文で詳しく論じられているので（池田・平石，2016；大久保，2016），それらについては繰り返さない。本論文では，まず，本研究領域における事情を紹介し，乳児研究と縦断研究という2つの例を出しながら再現性の問題について考えていく。その後，筆者の個人的な実践や国内の現状について考えてみたい。

2. 発達科学

本研究領域の事情について簡単に紹介しよう。従来，発達現象を扱っていたのは，発達心理学であった。我が国にも会員数4000人を超える日本発達心理学会があるし，アメリカにはアメリカ心理学会発達部門が，イギリスにはイギリス心理学会発達セクションがある。しかしながら，近年，従来の発達心理学は，一部において「発達科学」へ移行しようとしている。たとえば，田島・南（2013）は，発達心理学が，心理学にとどまらず，総合科学としての発達科学へ再編されつつあることを指摘している。発達心理学と発達科学の違いはいくつかあるが，主な違いは学際性である。従来の心理学的手法や理論だけではなく，生物学や神経科学，医学，工学などの多様な研究領域が，多様な角度から発達現象に迫るようになってきた。事実，脳機能計測や神経心理学的手法，計算機シミュレーション，遺伝子多型の検討など，多様な手法を用いた乳幼児の発達に関する基礎的研究が国際乳児学会や国際学術誌の *Developmental Science* 誌で報告されている。また，これらの多様な手法が用いられることによって，発達現象が単層的ではなく，複層的に扱われるようになった点にも違いがあるだろう（板倉，2014）。発達科学の新規性について疑念を呈する向きもあるが（中尾，2014），従来の発達心理学では見られなかったような研究が増えていることに疑いはないだろう。

発達科学が「科学」を標榜するには，再現性を含めた科学の基礎的要件の遵守が必須であると言える。これまでの発達研究において再現性はどの程度検討されてきたのだろうか。発達科学は，広範な対象（胎児，乳幼児，児童，青年，成人，高齢者）を含み，それらのすべてについて言及す

るのは現実的ではない。そこで，本論文では，社会的なインパクトや学術的なインパクトは強いが，再現性の検証が難しい2点に絞って考えてみたい。1つは，新生児や乳児を対象にした実験的な研究である。「こんなに小さい赤ちゃんに，こんな能力が備わっているのか！」という意味において，インパクトは極めて大きいと考えられるが，乳児の気質や実験室の状況などによってばらつきが大きく，結果の再現は容易ではない。もう1つは，子どもの発達に関わる縦断的な研究である。こちらも「子どもの頃のある能力が，成人になったときの年収を予測する！」という結果は社会的にも学術的にも極めて重要であるが，予算や人的資源などの面から追試が難しい。

3. 乳児の実験的研究に関する問題

まず，乳児の実験的な研究に関する再現性の問題を考えてみよう。その例として，ここでは新生児模倣の議論を紹介する。新生児模倣とは，生後まもない新生児が，他者の口の動きを模倣する現象のことである（Meltzoff & Moore, 1977）。乳幼児を対象にした研究は19世紀末から本格的に始まり，ボールドウィンやピアジェなどによって飛躍的な発展を遂げた。しかしながら，その時代では，乳幼児は無力で無能な存在だと考えられていた。1960年代以降，研究手法や理論の進展により，乳幼児に様々な能力があることが明らかになり，「有能な赤ちゃん」という考え方が定着しはじめた（森口，2014a）。その中でも，新生児模倣が与えるインパクトは大きく，発達心理学の教科書にも度々登場する。

現在では，新生児模倣を根拠として，様々な認知発達理論が提唱されている。たとえば，Meltzoffは，新生児模倣を含む乳児の模倣能力に関する知見を基に，「ライクミー」理論を提唱している（Meltzoff, 2007）。この理論は，新生児模倣が心の理論などの他者理解の発達の基盤にあるという考えである。Meltzoff以外にも，新生児模倣を基盤として発達理論を構築する研究者は少なくない。

では，この新生児模倣は，どれだけ確かな現象なのだろうか。Meltzoffらが新生児模倣を報告した *Science* 誌の論文には2つの実験が報告されて

いるが、参加者数は十分ではなかった (Meltzoff & Moore, 1977)。実験1では6名、実験2では12名である。現代の視点からすれば検定力が十分ではないとして即座に不採択になるだろうが、その点は置いておこう。手続きとしては、新生児は、実験者の「舌出し」や「口をあける」などの4つのジェスチャーを提示され、提示後の行動が記録された。そして、数名の大学生が、それぞれのジェスチャーの後の乳児の映像を、4つのジェスチャーに近いと思う順序でランク付けした。その結果として、舌出し行動の映像後は、舌出しのジェスチャーが選択されやすく、口をあける行動の映像後には、口をあけるジェスチャーが選択されやすいことが明らかになった。実験1では、新生児が実験者のジェスチャーを見ていないと思われる場合に実験者が繰り返しジェスチャーを産出した点が結果に影響を与えた可能性があるため、実験2では、この点を統制した。その結果、実験2でも新生児模倣が見られた。

この論文が出た後に、複数の研究者によってこの結果は追試されている。Meltzoff and Decety (2003) によると、13の独立した研究機関によって新生児模倣は再現されているという。ある程度頑健な結果であるようである。その一方で、再現できないという研究も多数報告されている。Meltzoff and Moore (1977) の論文が出た2年後に、発達研究では世界最大の学会である Society for Research in Child Development で、Meltzoff and Moore (1977) の実験1が再現できないという研究が発表されている (Hamm, Russell, & Koepke, 1979)。また、Hayes らは、この領域では権威のある *Developmental Psychology* 誌に、新生児模倣の追試研究を報告した (Hayes & Watson, 1981)。彼らは、予備的研究において全くうまくいかなかったため、Meltzoff らに実験のコツなどを直接聞き、本実験に進んだと述べている。実際のところ、乳幼児研究には、論文では表現できないコツがあるので、この点を考慮したのは重要である。Meltzoff らの実験2と同じ方法で、11名を対象に検討したが、全く結果を再現できなかった。さらに、彼らは新生児模倣が手続き上のミスで生じているアーチファクトである可能性を指摘している。

さらに、乳児研究に特化した雑誌である *Infant Behavior and Development* 誌で、2つの研究が、新

生児模倣が再現できなかったことを報告している (Koepke et al., 1983; McKenzie & Over, 1983)。特に、Koepke et al は、Meltzoff と同じ方法で追試を実施しており、参加者数もほぼ揃えているが (Meltzoff の実験1を6名、実験2を14名で追試)、新生児模倣の証拠は得られなかった。この結果に対して、Meltzoff らはコメントを寄せており、刺激や記録方法、分析方法などの問題点を指摘している。

Anisfeld (1996) は、新生児模倣についてメタ分析を行い、報告されている研究には統計的、方法論的問題点があることを指摘した。新生児模倣の証拠として挙げられている行動にはいくつかあるが、頑健なのは舌出し行動だけだと結論づけている。さらに、再現性の問題からは少し外れるが、この舌出し行動が模倣の証拠として採用できるかについても批判的な検討がなされている。Jones (1996) は、新生児に点滅信号を見せたときに、見せていない場合よりも、舌出し行動が有意に増加することを見出した。これらの結果から、舌出し行動は模倣の証拠というよりは、探索行動として捉えた方がよい可能性を指摘している。

新生児模倣についての議論を見てきたが、現状では、新生児模倣は現象としては確かだが、その行動には別の解釈も可能だというのが妥当な見解だろう。このように、乳幼児研究では、結果の再現性よりも、結果の解釈が問題になることが多いような印象がある (Haith, 1998)。つまり、乳児がある能力を示すような実験結果が得られたとして、その実験結果自体は再現されるものの、もっと単純な解釈 (例えば、刺激の特徴) などでも説明できるという議論である。こういった研究では、当初の結果自体は再現されているので、再現性は担保できていることになる。

だが、実際には、出版バイアスの影響などにより、再現できない結果が報告されていないことも少なくないだろう。ネガティブな結果が出版されやすくなった現代においては、再現性がより問題になってくる可能性が高い。事実、近年の乳児研究で最も注目を集めているものの一つである乳児の道徳性判断の研究でも (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; Kanakogi et al., 2013)、一部の研究成果は再現できないことが報告されている (Salvadori et al., 2015)。さらに、

再現性以前に、乳幼児の研究実践に自省を促す、耳の痛い指摘がなされている。この点を次節で紹介しよう。

4. ManyBabies プロジェクト

実際のところ、自分が関わる研究室以外で、どのように研究が実践されているかはよくわからない部分もある。最近報告された論文で、社会学の研究者（大学院生）が、乳幼児研究を行っている規模の大きい研究室2つと比較的規模の小さな研究室1つに一定期間参加して（明記されていないが、潜入捜査のようなものと思われる）、研究実践の実態把握を行った（Peterson, 2016）。彼は、研究実践がとりわけ困難な領域として乳幼児研究を選び、この領域の研究者が、困難な研究対象（様々な統制が難しい乳幼児）と、要求される研究の規律（統計的有意性の検出など）との間に折り合いをつけるために、様々な方略を使って結果を歪曲している可能性を指摘している。

求められる研究実践とは、事前に標準化された手続きを用いて、平穏で統制された環境下で、実験者は正確に実験手続きを進め、事前に定めた基準でデータ分析に含めるかを判断し、事前に決定したサンプル数を集めるまで実験を実施し、全てのデータが揃ってから解析を始め、統計的に意味がある結果であろうがなかろうが公表するというものだろう。もちろん、実際には、論文に書かれている部分と、論文には記述しきれない、暗黙知とも言える工夫もある。上記の新生児模倣の Meltzoff のところで触れたコツと通じるものだ。しかし、この論文で報告されているのは、工夫は工夫でもかなり灰色もしくは黒色の工夫ばかりである。

例えば、Peterson (2016) が関わった研究室では、実験プロトコルが非常に「柔軟」である。実験者が手続き上エラーを犯した場合も、エラーがなかったように実験を続けたり、研究が進行中であるのに実験方法を手直ししたりする。また、乳児の視線計測実験は、注視行動を2人の評定者が独立してコードするのが通例だが、2人の評定者は難しい例に当たるとその都度話し合いを持っていたそう。これ以外にも、他の心理学研究でも指摘されているように、事前に決めたサンプルになる前から繰り返し統計をかけてみたりするな

ど、研究倫理が疑われるような例ばかりが並んでいる。さらに、完全に研究倫理違反の事例も記載されている。乳児実験では、多くの場合、乳児は養育者の膝の上に抱えられる。その際に、養育者の視線が乳児に影響を与える可能性があるため、養育者は目を閉じるように教示される。この論文に掲載された例では、実験者は、論文に記載するためにそのように教示するが、実際には目を開けても構わないと述べたそう。乳幼児研究そのものに対する不信を招きかねない内容である。

Peterson (2016) は、乳幼児研究を含めた心理学研究に、このような、論文に報告されていない研究者の自由裁量部分がかなりあると指摘している。もちろん、このような研究実践を行っている研究室ばかりではないだろうし、研究室の中にはかなり厳しい内部規約を持っているところもある。疑わしい研究室ばかり標的にしたのかもしれない。だが、Peterson の指摘は、研究領域全体の問題として真摯に受け止めるべきだろう。このような「外圧」は、ある意味において研究が進展する好機ともいえる。

この論文を受けて、乳幼児研究の研究実践を見直そうという動きが始まっている。ある気鋭の若手研究者らは、発達科学研究に関わるメーリングリスト上で、乳幼児研究の「最善の研究実践」を集約しており、そこでは実験に参加した全ての参加者の映像をウェブ上で共有してはどうかなどの提案がなされている。個人情報保護との折り合いをどうつけるかが問題になってくる可能性があるが、今後はより透明性のある研究実践に移行することになるだろう。

仮に Peterson (2016) が観察したような研究実践がこれまでの乳幼児研究で一般的なものであったとしたら、これまでの乳幼児研究の知見も信用できなくなってくる。そのため、乳幼児研究においても、過去の重要な知見を再現するためのプロジェクトが始まっている。このプロジェクトは、言語獲得や認知発達の研究者である Frank (2015) らが主導している。彼は、実験室環境（例えば、温度や明るさ）や、実験補助者の容貌（例えば、髭をたくわえているか）などの細かい変数が結果に影響を与えること、だが、その点についての系統だった研究がないことを指摘している。

さらに、彼らが行っている乳児の言語獲得に関

わる実験研究のメタ分析を総括して、平均して、各研究の1つの条件におけるサンプル数が20程度、2条件間比較をした際の条件差の効果量が $d=.5$ 程度であることを明らかにしている。新たな研究で効果量が $d=.5$ が予想される研究では、検定力を80%だと見積もったら、サンプル数が各条件で50以上必要である。Frankは、これらを通して、これまでの乳幼児研究のやり方では偽陽性や偽陰性が出ている可能性は高く、乳児研究の既存の知見には信頼性がないと結論付けている。乳幼児研究の内外から、既存の知見に対する疑義が生じているわけである。

Frankは、乳幼児の研究、特に、実験的な研究では、信頼のあるデータを得るために、3点を考慮すべきだと指摘する。まず、大きなサンプルサイズである。乳児研究では多数のサンプルを得るのに非常にコストがかかるが、偽陽性の結果が出ることに比べれば仕方がなく、幼児研究はそれほどコストがかからないのだから、言い訳の余地はないと述べている。2点目は、研究室内での追試である。乳幼児研究では、研究1で主な結果が報告され、研究2以降はその統制条件になることが多い。そのため、研究2以降も実験条件を含めるべきだという主張である。3点目として、発達の比較の重要性が挙げられている。ある年齢と別の年齢を比較することは、ある年齢の結果が、別の年齢で再現できるかを検討しているともいえる。特に、たとえば3, 4, 5歳を比較する際に、3歳だけチャンスレベル以上の成績が出て、4, 5歳では出ない場合、素晴らしい発見の可能性もあるが、実は偽陽性の可能性もある。

さらに、Frankは、より大きな枠組みとして、ManyLabプロジェクトに着想を得た、ManyBabiesプロジェクトを提唱している。このプロジェクトでは、乳児研究の重要だと考えられる知見を選定し、事前登録をして、多くの研究室で共同して追試を行うことを目的としている。プロジェクトの概要はManyLabsプロジェクトと大きな違いがあるわけではないが、やはり成人の研究と乳幼児研究では考慮すべき点が異なるため、乳幼児研究独自のプロジェクトが必要であろう。また、前述の、研究室で異なる細かい変数と関連して、できる限り研究室環境や実験状況は揃えるものの、完全に一致させるのは難しいと考えられるため、考

えうる限りの潜在的な媒介変数を含めて、事前にプランを立てた解析方針に基づいて解析しようという意向をもっているようである。たとえば、乳児研究では、伝統的に乳児の注視行動をカメラで記録し、人手を使ってコーディングをしていたが、この方法を使う研究室とアイトラッキング装置を使って自動的にコーディングする研究室で違いがあるかを検討するなどの例が挙げられている。このような媒介変数を明らかにすることによって、後進の研究者がどの点に注意してデータ収集をすればよいか明らかにになるだろう。現在、Frankらは、これらのプロジェクトに参加する研究者を広く募集している。興味があれば、ぜひ彼のブログ (<http://babieslearninglanguage.blogspot.ca/>) にアクセスしていただきたい。

乳幼児研究者としては、このようなプロジェクトが進むにつれて、インパクトのある乳幼児研究の論文が減らないことを祈るばかりだ。

5. 再現性と頑健性

さて、次に、実験室実験以外の研究について見ていこう。先述の*Developmental Science* 誌には実験室実験が多いが、伝統的な発達心理学の有力雑誌である*Child Development* 誌と*Developmental Psychology* 誌には、実験室実験以外の研究も多数掲載されている。科学的な研究は実験室実験に限るものではないし、実験室実験だけで子どもの発達を捉えきれないことは明らかなので、本論文ではこのような研究についても見ていこう。

近年、発達心理学や教育心理学のみならず、経済学や教育社会学などの領域を巻き込んで、子どもの頃のいかなる変数がある後の発達を予測するかという議論がなされている。学術的な研究という意味合いだけではなく、教育政策や社会福祉政策、経済政策などの観点からも、これらの研究は非常に注目を集めている。かつてはIQなどの知的能力が主に検討されていたが、近年は非認知能力とか、社会情緒的能力などと言われる、知的能力とは区別される子どものスキルが注目されている (Heckman, 2006)。とりわけ、近年、筆者の専門でもある、幼児期や児童期における自己制御能力が、青年期における学力や友人関係、成人期における年収や社会的地位、健康状態、犯罪の程度

を予測することが示され、子ども期の自己制御能力を高めようという動きが世界的に加速している (Moffitt et al., 2011)。我が国ではこのような自己制御研究に対する動きは極めて鈍いが、いくつかの国々で大型のプロジェクトが動いている。

ここで考えなくてはならないのは、このような長期縦断研究をどれだけ追試できるのかという点である。このようなプロジェクトを立ち上げられない我が国は置いておくとしても、子どもの頃から30歳や40歳、さらには70歳や80歳まで追いつけるのは資金面でも人的資源の面でも、至難の業である。このような場合に、再現性を求めることにどれだけ意味があるのだろうか。

再現性が難しい場合にとりうる手段はいくつかあると考えられるが、本論文では頑健性の評価について紹介する。Duncan et al. (2014) は、発達研究と経済学研究における再現性と頑健性について検討している。この研究では、発達心理学のトップ雑誌である *Child Development* 誌と *Developmental Psychology* 誌、経済学では実証的な研究が多く収録されている *Journal of Human Resources* 誌と *American Economic Journal: Applied Economics* 誌から、2011年8月時点での最新の50の実証的論文と、時代の変化を探るために、1991年8月時点での50の実証的論文 (*American Economic Journal: Applied Economics* 誌は発刊されていないので除外) を選定し、それらの研究の再現性と頑健性を評価した。

再現性については、メタ分析が含まれているか、先行研究を追試するような研究が含まれているかを評価した。まず、メタ分析については、ここで調べられた5つの雑誌には2011年にも1991年にも一つも含まれていなかった。また、追試に関しては、追試が主目的な研究は2011年も1991年もどの雑誌も5本以下であり、部分的に追試したとみなされる研究は、*Child Development* 誌にはほぼ含まれていないが、他の雑誌では5-8本程度の論文が収録されていた。どの雑誌も、20年前であろうとなかろうと、多いとは言えないものの、追試研究がある程度なされていることが示唆される。

次に、頑健性については、3点から検討されている。1点目は、同じ効果を、様々な統計手法(例えば、重回帰分析や操作変数法)を用いて評

価しているかという複数の推定方法である。2点目は、複数のデータセットを用いているかである。通常の発達研究では容易ではないが、同じ論文内で、異なったデータで類似した結果が出ているかが検討される。3点目は、下位集団分析で、集団全体での分析のほかに、性別や人種などの下位集団で分けて分析するというものである。異なった下位集団で同様の傾向が認められれば、その結果にはある程度の一般性があると考えられる。これらの分析の結果、発達研究においては、2011年でも1991年でも、複数の推定方法や複数のデータセットを用いていないことが明らかになった。また、*Developmental Psychology* 誌では、2割程度の研究が下位集団分析を実施していることが明らかになっている。一方、経済学の雑誌では、複数のデータセットを用いた研究は少なかったが、残りの2点に関しては、概ね半数程度の研究がそれらの手法を用いていた。

もちろん、Duncan et al. (2014) によって提示されている、研究の再現性や頑健性の評価方法が妥当であるとは言い切れない。雑誌のある号に含まれている追試研究の数を調べるよりは、特定のテーマの研究を選定し、その後追試されているかを調べた方がいいかもしれない。また、頑健性については、複数のデータセットが利用可能な研究は限られている。複数の推定方法が有効な場合もあればそうでない場合もあるだろう。そうではあるものの、追試が難しいような場合に、研究の頑健性を評価することは、今後ますます重要になってくると思われる。データのオープン化が進めば、自分たちの取ったデータと、オープンなデータをそれぞれ解析して、結果が同じ方向性であるかを確認することはできるようになるだろう。また、ここで挙げられている以外にも、今後は評価の基準が増えてくるかもしれない。

発達研究には追試研究が難しい場合が少なくない。その際に、再現性の必要性だけを訴えても無意味である。このような方法を用いて研究の頑健性を評価するのも一案であろう。

6. 個々の研究で何ができるか

ここまで発達研究における再現性と頑健性の問題についてみてきた。ManyLabs や ManyBabies の

ような、広範な再現性プロジェクトは重要だが、それ以外に、個々の研究レベルで何ができるだろうか。個人でできる研究の再現性や頑健性のチェックの方法としては、上述したようにメタ分析と大きなサンプルでのデータ取得だと思われる。特に真新しいものではないが、これらを組み合わせる筆者が最近実施した研究を紹介する (Moriguchi & Todo, 2016)。要点としては、メタ分析をすることによって先行研究での食い違いの要因を突き止め、その点を考慮して大規模なデータを取得することで、研究の頑健性をある程度確保したということである。

筆者は、現在幼児期に見られる空想の友達 (Imaginary Companion) の研究を行っている (森口, 2014b)。空想の友達とは、名前を持ち、数か月間継続して、ある種のリアリティを伴って子どもが相互作用する目に見えない存在のことを言う。幼児期においてしばしばみられる、ふり遊びの一種だと考えられている (Moriguchi & Shinohara, 2012; Taylor, 1999)。この空想の友達はしばしば幼児期に特徴的な現象であり、特に4歳において最も多いという指摘がなされている。だが、研究を俯瞰してみると、必ずしも4歳児において最も特徴的に見られるわけでもなさそうである。ある研究者は、ピークは2歳から3歳の間だと指摘し (Nagera, 1969)、別の研究者は児童期以降がピークであると主張するなど (Pearson, 1998)、知見は一致していない。そこで、4歳時点が本当にピークであるかを検討したいと考えた。

だが、先行研究の結果が一貫していないことを考えると、仮に筆者が何らかのデータを提出したところで、そのデータは混沌に拍車をかけるだけの可能性もあった。そこで、まず、先行研究の結果をメタ分析することにした。詳細は省くが、空想の友達の研究に関する論文を収集し、その中で信頼に足ると考えられる40程度の研究を選定し、空想の友達の割合 (効果量) を算出した。その結果として、効果量が、研究間でかなりばらついていることが示されたため、そのばらつきを説明する要因を探った。年齢の効果を始め、いくつかの要因を調べたが、結果としては、参加者の年齢は効果量のばらつきを説明しなかった。

実はこの結果は当初の予想通りであった。というのも、メタ分析で含めた研究の中で、空想の友

達の基準が様々であったからである。特に、この研究領域では、「現在空想の友達を持っている子ども」と「空想の友達を過去に持っていた子ども」があまり区別されていない。仮に「現在空想の友達を持っている子ども」のピークが4歳だったとしても、「空想の友達を過去に持っていた子ども」も空想の友達を持つ子どもとみなされていたら、年齢差は出ないだろうと予想される。

そこで、「現在空想の友達を持っている子ども」と「空想の友達を過去に持っていた子ども」を区別して、比較的大きなサンプル (2歳から9歳の子ども800名) を対象に、空想の友達を持つ子どもの割合を調べた。そして、対数線形モデルを使って、子どもの年齢や性別などの要因が空想の友達の割合をいかに説明するかをモデル化した。その結果として、800名の子どものデータは、性別と年齢の効果によって最もよく説明されることが明らかになった。特に年齢について解析したところ、2歳から4歳までの間に空想の友達を持つ割合が増え、4歳でピークに達し、5歳以降は線型的に減っていくことが示された。つまり、4歳がピークであるという主張が支持されたのである (Moriguchi & Todo, 2016)。

この研究では、メタ分析と新しいデータの取得を組み合わせることで、明確な結果が得られたと考えられる。再現性を検討したわけではないが、研究の頑健性は確保できたのではないかとと思われる。

7. 国内の現状

最後に、国内の事情に触れておく。自戒を込めると、我が国の発達研究においては、再現性の問題はほとんど認識されていないのが現状である。ただ、皮肉なことに、我が国においては、追試研究は幅広くなされている。学会で発表されている研究を見てみると、オリジナルな研究はわずかで、欧米の先行研究を追試しているだけか、もしくは、先行研究の変数を1つか2つ操作したものが少なくない。これは、発達研究に限られた問題ではないだろう。日本心理学会などに参加すると、他の領域でも同じような研究が目につく。また、この再現性の問題自体が欧米の後追いである。ある意味において、我が国に必要なのは、再

現性の追求ではなく、オリジナリティの追求であるように思う。

ただ、それであっても、国内の追試研究にはいくつかの問題がある。ここでは2点挙げる。一つは、欧米の研究を追試していても、それを学会か、もしくは、紀要論文でしか発表していないことである。いずれも査読がないか、あったとしても、厳しい査読コメントが並ばないことが多い。さらに、未刊行の卒業論文や修士論文の中には、追試に関するかなりの情報があると推測される。どの程度先行研究が再現されるかは非常に有益な情報なので、査読を経た追試研究が発表される場が国内にあってほしい。筆者としては、心理学評論誌か、心理学研究誌に、追試専門のセクションがあってもいいのではないかと思う（山田、2016も参照）。

2点目として、これは1点目とも関わるが、追試結果の解釈の問題である。日本の研究者がある欧米の研究を追試したとしよう。オリジナルの研究とはほぼ同じ結果が出た場合は良いとして、オリジナルの研究とは異なった結果が出た場合を考えてみよう。この場合、少なくとも3つの解釈が考えられる。1つは、オリジナルの結果が疑わしい場合である。再現性の追求という視点からすると、この解釈に落ち着いた場合は意義深い。2つ目の解釈は、西洋と我が国の文化的事情の違いに影響を受けて、オリジナルの結果が再現できなかった場合である。これは社会心理学などでは顕著だと思われるが、発達研究でもしばしばみられる。3つ目は、研究の方法や精度が悪く、研究に必要な条件を満たしていないため、オリジナルの結果が再現できなかった場合である。

あくまで私見だが、我が国の発達研究を見てみると、あまり1つ目の解釈や3つ目の解釈を選択する研究者は少なく、最も「無難」である2つ目の解釈に落ちついていることが多いように思う。だが、研究発表を聞く限り、研究の方法や精度に問題がある場合も少なくないように思える。つまり、3つ目の解釈をするべきときもある。無論、研究の精度が高い場合は、1つ目や2つ目の解釈を選ぶ必要があるだろう。我が国の追試研究は、結果をしっかりと解釈するべきである。

興味深いことに、研究の方法、結果、解釈に関する重みづけが、文化によって異なるという指摘

がある。ある日本人科学者によれば、日本人は研究の結果重視であり、欧米人は研究の解釈や考察重視だという（Motokawa, 1989）。また、欧米の研究者からも、フランスやスイス、イタリアなどの国々では、解釈が最も重きを置かれ、英語圏では方法、そして、日本では最も結果が重視されるという指摘がある（Reese, 1999）。そのため、北米の研究者が研究に疑義を抱くとき、フランスやスイスの研究に対しては再現性を疑い、日本の研究に対してはその解釈を疑うという。この指摘が科学のどの分野にでもあてはまるのかはわからないが、筆者の印象と一致する。追試研究に限った問題ではないが、我が国の研究は解釈についても少し慎重であってもいいかもしれない。

8. おわりに

本論文では、発達科学の領域における国内外の再現性問題の動向を紹介してきた。全体的に再現性の検証は不十分であり、「突き崩されないもの」があると胸を張って言える状況ではなさそうである。

ただ、発達科学において、成人の研究と同じ基準を用いることに対する疑念もある。成人の研究のように、簡単にサンプル数を増やすことはできないし、大掛かりな再現性プロジェクトも難しい。心理学全体の基準も必要だとは考えられるが、個別の事情も勘案する必要があるだろう。Duncan et al. (2014) が指摘するように、発達科学は発達科学における再現性や頑健性の評価方法を探らなければならない。

いずれにしても、国内においても国外においても、この問題についてはまだまだ認識が共有されていないので、発達科学が発達科学であるために、関連する学会や学術誌において、これらの点を議論していく必要があるだろう。

謝 辞

本論文は、Michael Frank 氏、実藤和佳子氏、鹿子木康弘氏、篠原郁子氏ほか、多くの発達科学に関わる研究者に情報提供をいただいた。ここに御礼を申し上げます。

文 献

- Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review*, 16, 149–161.
- Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, 50, 2417–2425.
- Frank, M. (2015). *The ManyBabies project*. (retrieved from <http://babieslearninglanguage.blogspot.ca/2015/12/the-manybabies-project.html> on 3 February 2016).
- Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, 21, 167–179.
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26, 30–39.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557–559.
- Hamm, M., Russell, M., & Koepke, J. (1979). *Neonatal imitation?* Paper presented at the Biennial Meeting of the Society for Research in Child Development, San Francisco, March 1979.
- Hayes, L. A., & Watson, J. S. (1981). Neonatal imitation: Fact or artifact? *Developmental Psychology*, 17, 655–660.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900–1902.
- 池田功毅・平石 界 (2016) 心理学における再現可能危機：問題の構造と解決策 心理学評論, 59, 3–14.
- 板倉昭二 (2014) 発達科学の最前線 ミネルヴァ書房.
- Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Development*, 67, 1952–1969.
- Kanakogi, Y., Okumura, Y., Inoue, Y., Kitazaki, M., & Itakura, S. (2013). Rudimentary sympathy in preverbal infants: Preference for others in distress. *PLoS One*, 8, e65292.
- Koepke, J. E., Hamm, M., Legerstee, M., & Russell, M. (1983). Neonatal imitation: Two failures to replicate. *Infant Behavior and Development*, 6, 97–102.
- McKenzie, B., & Over, R. (1983). Young infants fail to imitate facial and manual gestures. *Infant Behavior and Development*, 6, 85–95.
- Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, 10, 126–134.
- Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358, 491–500.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75–78.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108, 2693–2698.
- 森口佑介 (2014a) おさなごころを科学する—進化する乳幼児観 新曜社.
- 森口佑介 (2014b) 空想の友達—生成メカニズムと子どもの特徴— 心理学評論, 57, 529–539.
- Moriguchi, Y., & Shinohara, I. (2012). My neighbor: Children's perception of agency in interaction with an imaginary agent. *PLoS One*, 7, e44463.
- Moriguchi, Y., & Todo, N. (2016). Prevalence of imaginary companion: A meta-analysis and new data. Manuscript Submitted for Publication.
- Motokawa, T. (1989). Sushi science and hamburger science. *Perspectives in Biology and Medicine*, 32, 489–504.
- Nagera, H. (1969). The imaginary companion: Its significance for ego development and conflict solution. *The Psychoanalytic Study of the Child*, 24, 165–196.
- 中尾 央 (2014) 発達科学はいつから発達科学なのか 板倉昭二 (編著) 発達科学の最前線 (pp. 189–208) ミネルヴァ書房.
- 大久保街重 (2016) 帰無仮説検定と再現可能性 心理学評論, 59, 57–67.
- Pearson, D. (1998). *The social acceptability of children hearing voices*. Clinical Psychology, University of Leicester.
- Peterson, D. (2016). The baby factory difficult research objects, disciplinary standards, and the production of statistical significance. *Socius: Sociological Research for a Dynamic World*, 2, 2378023115625071.
- Reese, H. W. (1999). Strategies for replication research exemplified by replications of the Istomina study. *Developmental Review*, 19, 1–30.
- Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PLoS One*, 10, e0140570.
- 田島信元・南 徹弘 (2013). 発達心理学の理論・方法論の変遷と今後の展望—発達科学を目指して 田島信元・南 徹弘 (編) 発達心理学と隣接領域の理論・方法論—発達科学ハンドブック 第1巻 (pp. 1–16) 新曜社.
- Taylor, M. (1999). *Imaginary companions and the children who create them*. New York, NY: Oxford University Press.
- 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.

— 2016. 2. 29 受稿, 2016. 3. 14 受理 —