

動物心理学における再現可能性の問題

澤 幸 祐, 栗 原 彬
専修大学

Reproducibility in animal psychology

Kosuke SAWA and Akira KURIHARA
Senshu University

The reproducibility and reliability of research are fundamental tenets in science, and in animal psychology. In the field of animal psychology, researchers have used a number of different species in various tasks and settings, such that considerations of the reproducibility are necessary compared with human research. Furthermore, using the appropriate statistical analysis and improving experimental design, a concrete theoretical background underlying each research question seems only to be important for improving the reproducibility between experiments in which the same species were used, but also in the situation where different species have been used. Because it is sometimes difficult to standardize the tasks and settings among investigations in animal psychology, theoretical consideration should help improve the reproducibility of research, as well as the validity of the interpretation of results obtained. Such efforts would also contribute to reduce the unnecessary use of animals from the perspective of animal welfare.

Key words: reproducibility, animal psychology, comparative psychology, comparative cognition, inter-species replication, intra-species replication

キーワード：再現可能性, 動物心理学, 比較心理学, 比較認知科学, 種間追試, 種内追試

1. はじめに

心理学において、研究結果の再現可能性が大きな話題になっている（総論については池田・平石, 2016を参照）。査読を経て学術誌に掲載された論文が報告している結果が、追試によっても再現されないということは、元々の研究結果の信憑性に重大な疑義が生じるだけでなく、その結果を踏まえて行われる後続の研究の方向性にも大きな影響を与えうる。多くの研究者は、すでに報告されている研究とまったく同じ手続きによる直接的追試を行いたがらないため、論文として刊行されたものについては、一定の留保をしながらも「正しいもの」と考えて先に進むことを好む。このとき、「正しいもの」と考えていたことがまったくもってでたらめであればどうだろうか。後続する研究を行うためのコストは、多くの場合無駄になってしまうし、仮に後続する研究において一定の結果が得られたとしても、その結果の解釈には

大きなゆがみが生じてしまう。研究結果の再現性が担保できないということは、「巨人の肩の上に立つ」という表現を借りれば、上に乗るための巨人が実はハリボテであるということであり、極めて危うい状況となってしまう。

動物心理学においても、再現可能性の問題は対岸の火事ではない。Marc Hauser による論文捏造の一件は記憶に新しい (Couzin-Frankel, 2014)。研究結果が捏造されていれば、当然のことながら再現は難しい。現在刊行されている論文のなかに、完全な捏造とは言わないまでも、不適切な実験手法や統計解析によって得られたものが混じっている可能性も決して低くないだろう。また、ヒト以外の動物を対象とするという動物心理学の特性ゆえに生じうる問題も存在する。そこで本論文では、動物心理学において再現可能性問題がどのような意味を持つのか、再現可能性の高い研究のためには何が必要なのか、そしてそもそも実験結果が再現されるとはどういうことなのかに

ついて論じ、動物心理学において再現可能性の高い研究を行うために必要なことはなんなのかを考察する。

2. 現代動物心理学が目指すもの

動物心理学は、その名が示すとおり「動物についての心理学」である。すなわち、動物に関するものであれば学習であろうが発達であろうが社会であろうが、動物心理学の研究対象となりうる。したがって、動物心理学といってもその研究領域は極めて多岐にわたり、実験や観察など手法も多様である。日本動物心理学会での発表を見ても、行動観察や行動実験、神経科学的研究や遺伝子解析と、よって立つものが異なる研究が混在している。もちろん現代心理学において多様な研究手法が用いられるのは動物心理学に限ったことではないが、動物心理学においては共通項が「ヒト以外の動物を研究対象とする」という点のみであるように思われる。では、「ヒト以外の動物を研究対象とする」ことによって、動物心理学は何を目指しているのだろうか。大きくわけて、三つの目的が考えられる。

第一の目的は、研究対象として関心を寄せている動物種の心理学的側面を明らかにするというものである。イヌを対象とした研究においてイヌの心理学的特徴を明らかにする、あるいはニホンザルを対象としてニホンザルの心理学的特徴を検討するといったものは、この目的に沿ったものであり、「この動物のことが知りたいからこの動物を研究する」というスタイルになる。この目的に基づいて研究を行ううえで、再現可能性が重要な意味を持つとすれば、「ある研究によって示された結果が、同一種を用いた研究によっても再現される」ということであろう。

第二の目的は、研究対象としている動物種のみを問題とするのではなく、系統発生の観点から複数種の比較を行うものである。比較心理学、あるいは比較認知科学と呼ばれる領域の研究では、複数種間の比較が重要である。異なる環境に適応した複数種の動物において、どのような認知機能の相違や相似があるのかを明らかにし、進化や適応といった観点から関心のある心理学的属性の起源を検討することになる。複数種の比較という意味

では、ヒト以外の動物種間の比較のみならず、ヒトとそれ以外の動物種の比較もまた、この目的に従って行われているといえよう。この目的に従った研究においては、再現可能性は複数の意味を持つ。まず、AとBという二つの種を用いた先行研究があったとき、先行研究におけるAあるいはB種のそれぞれの結果が、後続する追試によっても再現されるかという問題がある。これは、前述の第一の目的に沿って行われた研究における再現可能性とはほぼ同義である。次に、先行研究によって確認されたAとBの関係性が、後続する研究によっても再現されるかという問題である。比較研究によって明らかになる重要な知見は、Aという種そのもの、あるいはBという種そのものについてではなく、むしろA種とB種の相違や相似という関係性である（澤，2008）。この関係性が進化や適応といった観点から解釈される。複数種間の関係性を検討するうえで問題となる再現可能性の根源は、単一種に関する複数の研究間での再現可能性とは別に議論する必要がある。

動物心理学の第三の目的は、第二の目的の延長線上に位置付けることが可能なものであり、動物研究によって得られた知見をヒトに適用するというものである。ヒトを対象とした研究を行ううえでは、遺伝的背景や過去の経験、成育歴や強化履歴を統制することが困難である。そのために、ラットやマウスといった統制が容易な動物を対象として、ヒトへの適用を目指した基礎研究が行われることがある。あるいは、ヒトには実施することが困難な実験手続きを行うための代替手段として動物を用いるような研究も存在し、疾患モデル動物や薬物が生体に与える影響を検討する研究などが、この目的に従って行われている。こうした研究においては、ラットならラット、マウスならマウスという同一種間の実験結果の再現性に加え、動物で得られた結果がヒトで再現されるかという問題も存在する。異種間の再現可能性という意味では第二の目的に沿った研究についても問題となるが、ヒトへの適用を主な目的とした研究においては臨床的な関心が強いこともあり、そこには倫理的な話題が介在しうる。動物に対してどこまでの実験的介入が倫理的に許されるのかは重要な問題ではあるが、本論文の目的を越える議論が必要になるため、ここでは同一種間での再現可能

性と異種間での再現可能性の問題に焦点を当てて論じる。

3. 同一種間での再現可能性

筆者はかつて、大学院生だったころにラットを対象に味覚嫌悪学習の実験に取り組んだときに、味覚嫌悪学習の再現に失敗したことがある。味覚刺激と内臓不快感の対提示を行うことによって味覚刺激に対して嫌悪が学習されるという味覚嫌悪学習は、Garciaによって報告されて以来、繰り返し報告されてきた頑健な現象である (Garcia, Kimeldorf, & Koelling, 1955)。追試できなかった原因は、筆者の実験手技上の不手際が原因であることは言うまでもない。同様に、筆者はこれまでにさまざまな実験事態を経験してきたが、過去の研究において報告されている基本的な現象の追試がすべて最初から可能であったわけではない。いくたびかの予備実験を行い、その結果に基づいて実験に用いるパラメータを調整し、最終的に自らが関心を寄せている新しい手続きを導入することが可能となる土台を固めるという経験は、多くの研究者が行ったことのあるプロセスであろう。百発百中といっているほど再現ができるほどに頑健な現象もあるが、多くの場合は実際に実験が実施される実験環境ごとに微妙な相違が存在する。先行研究と全く同一の手続きを用いても結果が再現されず、微妙なパラメータ設定が必要となることは、再現可能性が担保されていないと解釈することもできる。しかし現実的には、多くの研究者は研究室ごとにパラメータの違いがあること自体は許容していると思われるし、ある現象を報告した論文を追試しようとして失敗したときに、予備実験を通じて研究者が行っていることは、当該現象を確認するために自らの実験環境において必要な条件は何かを明らかにするという作業である。

ある現象を確認するために必要な条件の探索という点から、同一種間での再現可能性が動物心理学、特に動物を用いた学習研究のなかで議論となった例として、回顧的再評価 (retrospective revaluation) に関する論争がある。古典的条件づけ (classical conditioning; Pavlov, 1927) において、回顧的再評価とは、ある刺激 (標的刺激) に対して訓練を行った結果として獲得された条件反応

(conditioned response: CR) が、それ以外の刺激への訓練が後続することによって変化するという現象の総称である。回顧的再評価には、後続する別刺激への操作と同方向に標的刺激への CR が変化する正の回顧的再評価と、逆方向に CR が変化する負の回顧的再評価がある。前者の代表として媒介条件づけ (mediated conditioning) があり、後者に含まれるものとしては隠蔽からの回復 (recovery from overshadowing) や逆行阻止 (backward blocking) が含まれる。

媒介条件づけは、Holland によって活発に報告されていた現象である (e.g., Holland, 1981, 1983, 1990)。例えば Holland and Forbes (1982) は、ラットを対象として二種類の条件刺激 (conditioned stimulus: CS) の複合を無条件刺激 (unconditioned stimulus: US) と対提示したのち、一方の CS を単独提示するという手続きを行った。二種類の CS を A, B と表記し、US との対提示を +, 単独提示を - と表記すると、この手続きは AB+, B- と表記される。実験の結果、単独提示された CS B に対しては消去手続きを行ったことになるため CR が減弱するが、単独提示されていない CS A についても、統制群と比較して CR の減弱が見られた。この結果は、CS A に対しても、CS B に対する消去の効果が媒介することで消去が起こったという意味で、媒介消去 (mediated extinction) と呼ばれる。媒介消去は、AB+ 試行によって CS A と CS B の間に要素間連合 (within-compound association; Rescorla & Durlach, 1981) が形成され、B- 試行中に要素間連合によって活性化された CS A の表象が CS B と同様に消去経験の影響を受けると解釈された。要素間連合という理論的発想自体はすでに存在しており、いわば媒介条件づけは現象としては新しかったものの、その解釈に必要な理論的な土台はすでに存在していたわけである。

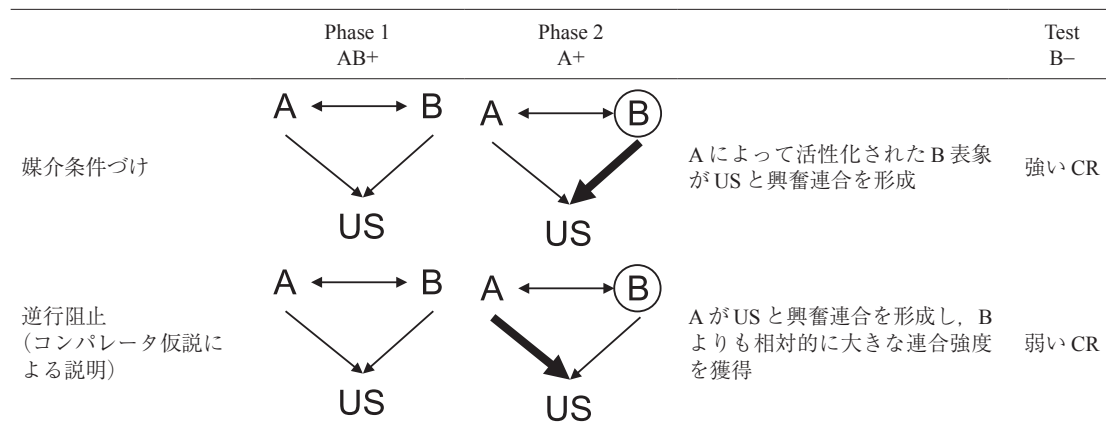
一方で、負の回顧的再評価については、理論的観点のみならず、再現可能性という点からも議論があった。負の回顧的再評価のなかでも特に関心を集めたものに逆行阻止 (backward blocking; Shanks, 1985; Miller & Matute, 1996) がある。逆行阻止とは、Kamin が報告した阻止 (blocking; Kamin, 1968) の手続きにおける Phase 1 と Phase 2 の順序を逆転させたものである。阻止の手続きでは、Phase 1 に CS A が US と対提示され (A+),

Phase 2 においては CS A と CS B の複合刺激が US と対提示される (AB+)。テストにおいては、CS B が単独で提示され、CR の程度が観察される。Phase 2 だけを取り出せば隠蔽 (overshadowing) の手続きであり、CS B を US と単独対提示 (B+) した統制手続きに比べて、B に対する CR は減弱するが、隠蔽手続きに先立って A+ 試行を行う阻止手続きでは、B に対する CR がさらに減弱する。この現象は、Rescorla-Wagner モデル (Rescorla & Wagner, 1972) をはじめとする多くの学習理論のベンチマークとなり、新たに連合学習理論を作る際には阻止現象が説明できるように設計するという流れがあった。逆行阻止手続きは、この阻止手続きの訓練順序を逆転させたものであり、最初に AB+ 試行を行ったのちに A+ 試行が行われる。この手続きによっても、適切な統制群と比較して B に対する CR が減弱することをもって逆行阻止が生じたと解釈される。ラットを用いた実験において逆行阻止が生じるという報告はあったものの、Rescorla-Wagner モデルをはじめとする阻止を説明する理論の多くは、逆行阻止を説明することができなかった。これに対し、コンパレータ仮説 (comparator hypothesis; Miller & Matzel, 1988) と呼ばれる理論は、阻止を説明することができるだけでなく、逆行阻止についても説明することができた。理論的背景があり、現象が確認されているならば、高い説得力を持ってこれらを受け入れる

ことになるはずなのだが、逆行阻止を含む負の回顧的再評価の諸現象は追試できないという報告が現れた。例えば Holland (1999) は、7つの実験を行った結果、隠蔽からの回復も逆行阻止も確認できなかったのみならず、全く逆の結果、すなわち正の回顧的再評価が得られたことを報告している。逆行阻止を例に取れば、AB+, B+ 手続きによって、CS A に対する CR は減弱するのではなく増強したわけである。この結果は、先に述べた媒介条件づけの観点から予測されるものである (表 1)。

ここに至って、正と負の回顧的再評価は、それぞれを支持する実験結果があり、それぞれを説明する理論的背景があるという状況となった。もし負の回顧的再評価が再現できないままであれば、少なくともラットを用いた古典的条件づけ事態においては正の回顧的再評価を支持する理論が採択されて決着するところである。しかし実際には、負の回顧的再評価を再現するためにはどのような条件が必要かを検討した研究が報告されるようになった。例えば Balleine, Espinet, and González (2005) は、訓練に用いる CS に対する知覚学習経験が逆行阻止の検出に有効であることを示し、CS 間の弁別性という要因が重要である可能性を報告した。また、理論的な観点からも、負の回顧的再評価を説明できるように Rescorla-Wagner モデルを拡張する試みが行われるなどの発展があっ

表 1 回顧的再評価に関する異なる結果とその説明理論。A, B は CS のラベルを指し、矢印は興奮連合を示している。丸で囲まれた B は、手続き上提示されていないが、連合によって表象が活性化されていることを意味している。Phase 1 において AB+, Phase 2 において A+ という手続きによって、特に Phase 2 において刺激 B と US が興奮連合を形成するのか、刺激 A と US の興奮連合が刺激 B に対する CR 表出を抑制するのかが理論的な争点のひとつとなった。



た (Van Hamme & Wasserman, 1994)。現在では、古典的条件づけにおける負の回顧的再評価は、確認するために強い条件設定が必要であるものの、現象としては存在するものとみなされているように思われる。

この一連の流れは、報告された結果が再現できなかった場合に研究者がどのように対処するかについて、健全な作業が行われたものと筆者は考える。理論的背景をもって現象が報告され、対立する理論と現象が報告され、どちらの現象が生じるのかの境界条件を実験的に検討し、あわせて理論に改訂が加えられるという、極めて生産的な流れに沿って研究が進んだ。もちろん実際には、同一種間での再現可能性の問題がこのように解決されるケースばかりではない。うまくいった理由はいくつか考えられるが、ラットという被験体を用いた古典的条件づけ研究においては用いられる刺激のバリエーションが限定的で統制が容易であったこと、そして実験手続きとの接続が容易な節約的理論が存在したことが大きいと思われる。古典的条件づけ理論に含まれているパラメータは、基本的には刺激の物理強度やCS-US対提示回数によって相対的に決定されるものが多く、実験手続きが理論の中身と容易に対応づけができる。実験場面の差異が大きくなると、実験間で結果に不一致が起こったときに解釈が困難になる。また、多くの古典的条件づけ理論は、仮説構成概念の導入が最小限に抑えられており、得られた実験結果が何を意味しているのかの理論的解釈が研究者間でぶれにくい。例えば Rescorla-Wagner モデルにおける仮説構成概念は連合強度のみと言ってよいと思われるが、連合強度は反応強度と対応すると仮定されているため、観察された反応について条件間に差が見られた場合には連合強度に差があったのだという解釈を行わざるを得ず、この点については研究者間で解釈がぶれることはない。実験手続きのなかで具体的に決定されるものが理論の中身や仮説構成概念と直接的にむすびついていればいほど、理論のなかのどの部分に操作を加えたか、結果的にどういう変化が生じると予測されるかについての解釈は一致しやすくなると思われる。手続きと理論の中身が直接的に対応するような理論的背景の存在は、たとえ先行研究で用いられていたパラメータをそのまま用いた直接的追試

に失敗しても、追試研究で用いた条件間での差に帰着させることで、理論そのものの妥当性を議論することを可能にする。ラットを用いた古典的条件づけ研究に限らず、抑制的な理論構築は再現可能性を高めることに寄与するだろう。

4. 異種間での再現可能性

先に述べたように、動物心理学のなかでも種間比較に重点を置く比較心理学や比較認知科学といった分野では、異なる動物種から得られた結果の比較が行われる。異種間比較においては、検討対象となる認知機能の存在を示す行動が一方の種では見られるが他方の種では見られないことや、あるいは両方の種で見られることを示すことが必要となる。異種間比較研究において、再現可能性を考えることで特に大きな問題となるのは、種が異なることに起因する実験環境の不一致や課題の不一致であろう。ラットで行われた先行研究を、ラットを用いて再現する際には、動物の持つ特性が実験間で同一であるため、大きな問題は生じない。しかし、異種間比較研究において関心が寄せられる動物の組み合わせは多様であり、身体サイズや感覚器官の特性が大きく異なることは少なくなく、実験に用いられる刺激への反応性に影響する差異が実験結果にも影響することはありうる。

筆者はかつて、スンクス (*Suncus murinus*) を用いて条件性風味選好 (conditioned flavor preference) の研究を行ったことがある (Sawa & Ishii, 2012)。スンクスは、かつては食虫目とされていたように雑食性ではあるが野生ではミミズや昆虫を好んで食べる哺乳類である (現在はトガリネズミ目)。ラットやマウスと同様に実験動物化されており、生後10日ほどの個体が親の尻尾にかみつ、他の個体がまたその尻尾にかみつくことで連なって移動する「キャラバン行動」が研究対象となったり (Tsuji & Ishikawa, 1984)、嘔吐ができないラットに対してスンクスは嘔吐ができることから、味覚嫌悪学習事態において嘔吐感の研究に用いられている (e.g., Parker, 2006)。条件性風味選好は、バニラ風味やアーモンド風味といった風味刺激をCS、サッカリンのような甘味刺激やエタノールやスターチといったカロリーを含む刺激をUSとす

る古典的条件づけの一種であり、風味刺激に対して選好が獲得されることが知られており、ラットやマウスでは広く知られた現象である（総説として Capaldi, 1996；Sclafani, 1991）。

Sawa and Ishii (2012) では、アーモンド風味とレモン風味の刺激を CS、ショ糖溶液を US として条件性風味選好の獲得と US 事後提示による CR 減弱の検討をunks に対して行った。雑食性のラットやマウスに比べて野生環境での食物選択の幅が狭いと思われたunks において、食物選択に重要な学習性の食物選好がどの程度見られるかに興味をもつてのことであった。unks は、体長が 15 cm ほどで体重は雌雄によって異なるがおおよそ 30 g から 50 g 程度であり、サイズとしてはラットより小さく、マウスより大きい。実験装置自体はラットやマウスのものと大きく変える必要はないと考え、通常飼育されているケージを実験場面として刺激提示を行う方法をとった。結果としてはunks においても条件性風味選好の獲得が観察され、US 事後提示による CR の減弱も生じはしたのだが、ラットを用いた過去の研究と比較して様々な相違があった。

条件性風味選好の実験では、味覚刺激や風味刺激が用いられる。ラットやマウスといった動物では、実験に用いられる新奇な食物刺激に対しては新奇性恐怖 (neophobia) と呼ばれる現象が確認される。実験初期においては食べたことのない食物刺激の摂取量は低く抑えられ、有毒でないことが確認されると徐々にその摂取量が増加していくという傾向が一般的であり、その背景のひとつが馴化 (habituation) であると解釈される。しかしながら、Ishii et al. (2000) では、unks が酸味溶液に対する馴化を示さなかったことが報告されており、Sawa and Ishii (2012) でもショ糖溶液に対する摂取量は訓練を通じて大きく増加していくことはなかった。また、ショ糖溶液に対する選好はラットやマウスでは極めて強く、水とショ糖溶液の間での選択を行うと 1:9 程度でショ糖溶液を好む。しかしunks では、8 試行を経たあとでも水とショ糖溶液の選択比率は 2:3 程度であった (Sawa & Ishii, 2012)。このように、ラットやマウスといった動物種とunks では、条件性風味選好という実験事態を導入するにあたって前提となる部分に大きなかい離があり、こうしたかい離は

実験結果やその解釈に直接影響する可能性は否定できない。ショ糖溶液への選好の低さは、もし条件性風味選好が再現できなかった場合には、用いた刺激が不適切だったのかunks において条件性風味選好の獲得はそもそも不可能なのかを鑑別できないことになる。事実、Sawa and Ishii で報告されている CR はラットを用いた研究に比べて相対的に小さく、有意な CR が確認されるまでに必要だった試行数もまた、ラットを用いたものに比べて多かった。また、Sawa and Ishii で検討した US 事後提示による CR 減弱効果は、その理論的解釈として US 表象の馴化が挙げられており (Rescorla, 1973)、仮にunks では味覚刺激に対して馴化が生じないとすれば、US 事後提示効果がunks において再現されなかった場合に、馴化の欠落によるものなのか実験の勝手際などの外的要因なのかを鑑別できなくなる。異なる動物種において実験結果を再現する場合には、このような問題が不可避免的に生じる。

上記の例では、unks とラット等の比較を行ううえで、用いる刺激や実験事態がほとんど同じであったにも関わらず、種の特性によって結果の解釈や再現性に関して議論が起こることを紹介した。比較心理学、比較認知科学においては、そもそも実験事態を共有できないような種間比較を行うことも少なくない。鳥類や魚類、哺乳類と系統発生的に距離が離れている種の比較を行う場合には、共通する実験状況で同じ現象を扱うことは極めて困難である。また、言語指示が可能なヒトで確認された現象を、刺激性制御による訓練に基づいて課題を実行する動物で再現しようとする場合にも、結果の不一致が生じた場合には解釈が難しくなる。ラットやマウスと同じように放射状迷路にヒトを入れることで比較を行った研究 (e.g., Glassman et al., 1994) のように、一方の種において用いられる課題にもう一方の種を合わせるといった試みも行われてはいる。しかし、それぞれの種が持つ進化的背景などを考えると、一方の種にとって不自然な課題設定を行わざるを得ない状況も考えられ、異種間で研究結果を再現すること、比較することの困難さを完全に払拭するには至らないと考えられる。

5. 統計解析と実験計画に関する問題

実験手法や用いる動物種に関わらず、研究結果の再現可能性についてよく注目されるのが統計的分析の適切さに関する問題である。詳細は本特集の統計手法に関する論文（大久保, 2016）に譲るが、有意な結果が得られるまでデータを付けたし、望む結果が得られたところで実験を打ち切る *p* ハッキングと呼ばれるような問題のあるデータ取得方法（具体的な事例については藤島・樋口, 2016 を参照）をはじめとする問題のある研究実践（Questionable Research Practices）の弊害と動物心理学は無縁ではないだろう。こうした問題を防止するために、例えば Psychonomic Society は統計解析に関する指針（statistical guideline）を提示し、検定力分析の重要性を明示している（Psychonomic Society, 2012）。しかしながら、Psychonomic Society が刊行しているなかでも動物研究が比較的良好に掲載されている *Learning & Behavior* 誌について、2015 年に掲載された動物を対象とした論文を検索したところ、論文の中に検定力分析（power analysis）が登場したのは 1 件のみであり、当該論文（Barela, 2015）は方法論に関するものであった。American Psychological Association（APA）が刊行している *Journal of Experimental Psychology: Animal Learning and Cognition* 誌でも、Editorial では検定力に関する記述を奨励しているものの、2015 年に出版された動物を対象とした論文の中に、事前に検定力分析を行っているものは認められず、得られた結果に基づいて検定力を計算して報告したものが 2016 年 1 号に 1 篇（Austen & Sanderson, 2016）掲載されているのに留まる。動物心理学においては、統計改革の流れは、いまだ道半ばといった様相であるように思われる。

再現可能性を高めるために、検定力分析をはじめとする分析手法を使用すべきであることは間違いないが、動物心理学においてその流れが鈍いことにはいくつかの理由が考えられる。動物心理学においては、用いる研究対象は当然動物であり、動物を対象としている以上は動物の飼育設備などによる制限が、ヒトを対象とする研究よりも厳しい。一度に飼育できる動物の総数は、用いる動物種にもよるが決して多くはない。例えば、対応のない *t* 検定によって群間の平均値の差

を検出しようとした場合、中程度の効果（Cohen's $d=0.5$ ）を仮定して、有意水準を 0.05、検定力を 0.8 とすると、両側検定の場合には 1 群 64 個体のデータが必要となり、合計で 128 個体の使用が必要となる。医学や生理学の研究室と比較して小規模な研究室が多い心理学の分野では、たとえばラットやマウスが対象であっても、そう簡単に維持が可能な数ではない。まして、チンパンジーやゾウなどを対象とした研究において、十分な検定力を得るだけのサンプルサイズを個体数の増加によって達成することは、ほぼ不可能であると思われる。実際、2015 年に刊行された *Journal of Experimental Psychology: Animal Learning and Cognition* 誌に掲載された論文の中で、動物を対象とした実験を検索すると、用いられている被験体数は Freestone et al. (2015) においてラットが 60 個体用いられているのが最多であり、げっ歯類とハト以外の種では 10 個体以下の研究が多い。また、動物実験をめぐる倫理的側面は無視できない。動物実験以外の代替手段を講じる（Replacement）、使用する動物の個体数を削減する（Reduction）、動物の飼育や使用環境を改善する（Refinement）という Russel and Burch (1959) による「動物実験の 3R」の発想は広く受け入れられており、動物可能な限り実験に用いられる個体数は削減しなければならない。一方で、検定力不足に伴う偽陽性の増加によって疑わしい研究結果が報告されることは、追試の失敗などによって結果的により多くの実験個体を無駄に使用することにつながってしまう。したがって、できるかぎり少ない個体数でできるかぎり確実な実験結果を報告することが求められる（神経科学分野における議論は鮫島 (2016) を参照）。

そのための方策として、第一に考えられるのがなるべく効果量の大きい現象を扱う、あるいは実験手続きの工夫で大きな効果を得られるようにすることである。用いる動物種によって具体的な方法は異なるが、たとえば味覚嫌悪学習の実験であれば、刺激の摂取量が減少することによって CR が測定されるため、ベースラインとしての刺激摂取量が多いものを CS として選択することは CR の効果量を増加させると思われる。第二の方法としてはデータの変動性をなるべく抑えるような処置が考えられる。実験個体の持つ遺伝的背景や成

育歴などに起因する個体差はデータのばらつきを大きくするが、ラットやマウスについてはある程度の統制が可能である一方、動物種によってはそうした統制が困難なケースもある。飼育環境の統制や実験環境の標準化も、個体差を小さくすることに寄与する。ただし、飼育環境や実験環境の統制は、むしろ実験結果の再現可能性を低下させる可能性があることが指摘されている。Richter, Garner, and Wübel (2009) は、飼育環境や実験環境の標準化によって研究室内でのばらつきは小さくなるものの研究室間でのばらつきが相対的に大きくなり、結果的に研究室ごとの“local truth”が報告されることによって偽陽性が増加し、結果的に再現可能性が低下すると主張している。Richter et al. では、この問題に対する対策として、実験環境を標準化するのではなく系統的に不均一化することを挙げている。第三の方法として考えられるのが、被験体内実験計画と同一個体による反復測定の利用である。同一個体による反復測定については、行動分析学におけるシングルケースデザインの技法が代表的であり、詳細は成書に譲るが(e.g., Barlow & Hersen, 1984; 岩本・川俣, 1990; Sidman, 1960), 実験個体数の確保が様々な理由から難しい場合には検討する価値のある方法であろう。ただし、同一個体から反復測定によって得たデータに対して何らかの統計処置を行う場合には、データの独立性について十分な注意が必要である。

このように、動物心理学において統計学的に望ましいデータの取得のためには様々な努力が求められるが、研究の信頼性や再現可能性を担保するためには必要な作業であるといえよう。直接的追試を重ねることによって信頼できる実験結果の選別を行い、確実な知見を積み重ねていくという方向も確かにあろう。しかし、アメリカにおける実験動物の使用に関するガイドライン(US Department of Agriculture, Animal and Plant Health Inspection Service, 1990)において動物を用いた不必要な追試は行うべきではないと規定しているように、質の低い研究の山から追試によって結果を選別するという方法は倫理的にも経済的にも決して好ましいものではない。また、パソコン上に刺激を提示して反応を取得するといったヒト対象の研究では同じ実験プログラムを利用することで直接

的な追試が可能なケースがあるが、そもそも直接的追試が困難な状況が動物心理学には数多く存在する。げっ歯類やハトのような実験動物として典型的な種であればある程度標準化された実験機材が販売されているが、それ以外の種については研究室ごとに異なる機材が用いられることも多い。まして、関心のある現象を先行研究で用いられているものとは異なる動物種において再現しようとした場合には、直接的追試はそもそも不可能である。直接的追試を行うことが原理的に困難であるという前提を踏まえたうえで、信頼性の高い研究結果を得るための方策を個々の研究者が実行することが求められる。

6. 動物心理学において 再現されるべきものはなにか

一般に、科学において実験結果の再現といったときには、先行研究において用いられた方法をそのまま追試して同じ結果を得ることを指す。ここまでに見てきたように、同一の動物種を用いた研究の場合には、理屈としては先行研究において用いられた方法をそのまま採用することが可能であるが、異なる動物種を用いる場合には、直接的追試は極めて困難であるか、場合によっては不可能である。この意味で、ヒトを対象とした実験心理学における再現可能性の問題とその対策は、そのまま動物心理学に適用すれば問題が解決するというものではない。

直接的追試が原理的に困難な状況が存在する動物心理学において、再現されるべきものとは何だろうか。もちろん、第一義的には実験事実が再現されることが望ましい。しかし、動物心理学において、ひいては心理学において、研究者が関心を寄せている対象は、数値として測定された実験事実ではなく、むしろその実験事実が指し示す仮説構成概念であることが多い。ラットを用いた恐怖条件づけ事態において研究者が関心を寄せるのは、実験事実としての「凍結反応時間の長さ」ではなく、それが指し示す「恐怖の強さ」であろう。イヌのエピソード記憶を研究する場合に重要なのは、正答率そのものではなく、注意深く計画された実験デザインと正答率という実験事実が対応づけられたうえで初めて立ち上がってくる

「エピソード記憶の有無」である。この意味で、あえて乱暴に言ってしまうと、心理学において重要なのは表面的な事実の再現ではなく、むしろその事実が指し示している仮説構成概念に関する再現だと言える。

一方で、こうした立場は極めて危うい状況を作り出す原因ともなる。仮説構成概念とは、その名の通りあくまでも仮説的なものであり、直接観察することができない。曖昧に定義された仮説構成概念の濫用は、どういう結果が得られても解釈を可能にしてしまう。この意味では、ヒトを含む生活体の行動を操作可能な環境との対応によって記述しようとする徹底的行動主義の立場は極めて明快であり、解釈がぶれる余地が少ない。先に紹介したシングルケースデザインの洗練とも相まって、徹底的行動主義に基づく行動分析学的研究においては、実験操作の効果の再現可能性も他分野に比べて高いように思われる。曖昧な仮説構成概念を用いるのではなく、環境の操作と「実験事実」として測定された行動の距離が近いことは、第一義的な再現可能性を高めることに寄与する。

しかしこの立場も、動物心理学において満足のいくものではないと考える研究者は多いと思われる。言語指示を用いることのできない動物研究においては、結局のところ研究に用いられる課題は刺激性制御で語られるものであり、その意味では行動分析的立場を無視することはできない。その一方で、異種間比較においては標準化された実験状況を設定することが難しい以上、異なる実験状況において共通して測定しようとしている媒介物を仮定することは自然な流れに思われる。スキナー箱であろうが直線走路であろうが、注意深くデザインされた研究であれば共通して測定可能だと研究者間で合意形成できるものがないければ、現実問題として異種間比較は実行できないだろう。

仮説構成概念の濫用が再現可能性を損なう一方で、異種間比較を行ううえで仮説構成概念の導入がある程度必要であるのならば、再現可能性を損なわないような仮説構成概念の導入はいかにして可能かを検討しなければならない。そのために必要なのは、極めて当然のことであるが、抑制的な操作的定義に支えられた理論である。複雑な認知的機能に関する仮説構成概念は、ともすれば動物を用いた実験において実際に行われる操作を越え

て導入されるおそれがある。ヒトにおいては無理なく導入できるようなものであっても、動物研究において採用される実験操作とは対応づけが困難な仮説構成概念を導入してしまうと、実験状況が変わった場合には、その概念の挙動が当初の目論見とは異なってしまい、結果的に実験結果が再現できないというケースが生じる。確固たる理論的背景があって初めて、実際に行われる実験操作と仮説構成概念が接続される。前節で述べたような適切な統計処理の採用や十分に注意深くデザインされた実験計画の導入を前提としたうえで、実験状況の違いや場合によっては種を越えて、再現すべき対象としての仮説構成概念が議論の対象になりうるだろうと考えられる。

7. まとめにかえて Romanes と逸話法

動物の持つ心理学的属性、特にその知的能力は、古くから多くの関心を集めてきた。Romanes は、動物の知的能力を検討する第一歩として、動物の持つ知的能力の発露を目撃した人々の証言を集めるという、いわゆる「逸話法」を採用した (Romanes, 1882)。

Romanes の逸話法については、動物が示した行動に対する擬人的解釈に対する批判が行われているが、それ以外にも、問題はいくつか挙げることができる。そもそも、報告された逸話が本当にあったことなのかについては、報告者を信頼するかどうかにかかっている。Romanes は無批判に動物に関して報告される逸話を自らの研究に反映させていたわけではなく、誰が報告したのか、信頼のおける人物であるかを重要視したといわれる。例えば、C. Darwin による報告は、信頼のおけるものとして採用していた。いわば、報告内容の査読を自ら行っていたわけである。しかしこれは、報告内容が真正であることを保証しない。まったくの出まかせ、あるいは見間違えであった場合には、当然ながら同様の動物行動を再現的に確認することはできない。また、多くの逸話は、動物が行った行動をある時点で輪切りにしたエピソードであることが多い。「知的な」行動を示した動物が観察されたとして、その動物がそれまでにどういった経験をし、どういった来歴があったのかは、観察者には知ることができない。逸話に登場

する個体が、過去に特別な経験をしていた結果として「知的な」行動を示したとすれば、その行動を再現的に確認することは困難となるだろう。

現代動物心理学は、Romanesの逸話法にまつわる種々の問題を解決するべく実施されていることは間違いない。野生動物の観察を行う場合には可能であれば映像をはじめとする記録を残し、複数の観察者による一致を検討することが望まれる。実験的統制についても、長足の進歩を遂げている。しかしそれでも、Romanesの逸話法への批判には、現代動物心理学における再現可能性問題について示唆的なところがあるように思われる。それは結果の真実性であったり、(Romanes自身によるのだが) 査読の問題であったり、時間的に輪切りにされたデータの扱いの問題であったりする。また、「動物が示した驚くべき知的能力、人間的行動」という審美性の問題もある。しかしそれ以上に重要なのは、そもそもRomanesが何を目指したかという点にある。Romanes (1882)の序文にも述べられているように、彼は逸話収集で目的が達成されると考えていたわけではなく、集めた逸話を元に知性の起源に関する理論を構築しようとしていた。結果的に、動物心理学の歴史のなかでRomanesについて残ったものが逸話法と逸話法への批判が中心であるのは皮肉なことである。Romanesが試み、そして到達できなかった理論的基盤の構築という課題は、動物心理学において研究結果の信頼性を求め、再現可能性を高めることを目指す我々にとっても重要な宿題である。

謝 辞

本論文は、専修大学社会知性開発研究センター心理科学研究センターが行う平成23-27年度文部科学省私立大学戦略的研究基盤形成支援事業「融合的心理科学の創成：心の連続性を探る」(S1101013)の助成を受けた。

文 献

- Austen, J. M., & Sanderson, D. J. (2016). Contexts control negative contrast and restrict the expression of flavor preference conditioning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 95-105.
- Barela, P. B. (2015). The goldfish conditioned withdrawal preparation: Effects of some basic methodological variables. *Learning & Behavior*, 43, 313-321.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon. 高木俊一郎・佐久間徹(監訳)(1997) 一事例の実験デザイン：ケーススタディの基本と応用 二瓶社.
- Balleine, B. W., Espinet, A., & González, F. (2005). Perceptual learning enhances retrospective revaluation of conditioned flavor preferences in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 341-350.
- Capaldi, E. D. (1996). Conditioned Food Preference. In E. D. Capaldi (Ed.), *Why we eat what we eat: The psychology of eating* (pp. 53-80). Washington, DC: American Psychological Association.
- Couzin-Frankel, J. (2014). Harvard misconduct investigation of psychologist released. *ScienceInsider*, 30 May.
- Freestone, D. M., Balci, F., Simen, P., & Church, R. M. (2015). Optimal response rates in humans and rats. *Journal of Experimental Psychology: Animal Learning and Cognition*, 41, 39-51.
- 藤島喜嗣・樋口匡貴 (2016) 社会心理学における“p-hacking”の実践例 心理学評論, 59, 84-97.
- Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, 122, 157-158.
- Glassman, R. B., Garvey, K. J., Elkins, K. M., Kasal, K. L., & Couillard, N. L. (1994). Spatial working memory score of humans in a large radial maze, similar to published score of rats, implies capacity close to the magical number 7±2. *Brain research bulletin*, 34, 151-159.
- Holland, P. C. (1981). Acquisition of representation-mediated conditioned food aversions. *Learning and Motivation*, 12, 1-18.
- Holland, P. C. (1983). Representation-mediated overshadowing and potentiation of conditioned aversions. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 1-13.
- Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, 37, 105-131.
- Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology*, 52B, 307-333.
- Holland, P. C., & Forbes, D. T. (1982). Representation-mediated extinction of conditioned flavor aversions. *Learning and Motivation*, 13, 454-471.
- 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.
- Ishii, K., Matsuo, M., Tsuji, K., & Matsuo, T. (2000). Thirsty drive and habituation process to flavor in the house musk shrews (*Suncus murinus*). *Bulletin Reports for a Grant in-Aid from the Ministry of Education, Sports,*

- Culture, Science and Technology of Japan.*
 岩本隆茂・川俣甲子夫 (1990) シングルケース研究法
 新しい実験計画法とその応用 勁草書房.
- Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium On the Prediction of Behavior: Aversive Stimulation* (pp. 9–32). Coral Gables, FL: University of Miami Press.
- Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, 125, 370–386.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.
- 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.
- Parker, L. A. (2006). The role of nausea in taste avoidance learning in rats and shrews. *Autonomic Neuroscience*, 125, 34–41.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (G. V. Anrep, Trans.). New York: Oxford University Press.
- Psychonomic Society (2012). *Psychonomic Society guidelines on statistical issues*. (retrieved from <http://www.psychonomic.org/statistical-guidelines>)
- Richter, S. H., Garner, J. P., & Würbel, H. (2009). Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nature methods*, 6, 257–261.
- Rescorla, R. A. (1973). Effect of US habituation following conditioning. *Journal of Comparative and Physiological Psychology*, 82, 137–143.
- Rescorla, R. A., & Durlach, P. J. (1981). Within-event learning in Pavlovian conditioning. In N. E. Spear & R. R. Miller (Eds.), *Information Processing in Animals: Memory Mechanism*. (pp. 81–111). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Procasey (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton.
- Romanes, G. J. (1882). *Animal intelligence*. London: Keagan Paul.
- Russel, W. M., & Burch, R. L. (1959). *The principles of humane experimental technique*. London: Methuen.
- 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.
- 澤 幸祐 (2008) 動物心理学における研究パラダイム間の「翻訳」について 動物心理学研究, 58, 73–76.
- Sawa, K., & Ishii, K. (2012). Conditioned flavor preference and the US postexposure effect in the house musk shrew (*Suncus murinus*). *Frontiers in Psychology*, 3, 242.
- Sclafani, A. (1991). Conditioned food preferences. *Bulletin of the Psychonomic Society*, 29, 256–260.
- Shanks, D. R. (1985). Continuous monitoring of human contingency judgment across trials. *Memory and Cognition*, 13, 158–167.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology* (Vol. 5). New York: Basic Books.
- Tsuji, K., & Ishikawa, T. (1984). Some observations of the caravanning behavior in the musk shrew (*Suncus murinus*). *Behaviour*, 90, 167–183.
- US Department of Agriculture, Animal and Plant Health Inspection Service (1990). *Animal welfare act 22*, Riverdale: U.S. Department of Agriculture.
- Van Hamme, L. J. & Wasserman, E. A. (1994). Cue competition in causality judgment: The role or nonpresentation of compound stimulus elements. *Learning and Motivation*, 25, 127–151.

— 2016. 2. 29 受稿, 2016. 4. 7 受理 —