批判的合理主義研究

Studies in Critical Rationalism

2024

Vol. 15, No. 2

日本ポパー哲学研究会事務局機関誌編集部

(2024 年 12 月号)

CONTENTS

＜第 34 回年次研究大会自由論題発表要旨＞

学習におけるエラーの働き：記憶の実験心理学からの示唆　岩木　信喜　2

<第 34 回年次研究大会自由論題発表報告>

生成 AI と世界3　蔭山　泰之　10

生成 AI を理解できるか？　永島　秀文　24

法的シンギュラリティ　宇佐美　誠　28

＜追悼論文＞

恩師ヨセフ・アガシ（1927-2023 年）　立花　希一　38

＜査読付き投稿論文＞

ポパーの相互作用主義における存在論：心の主観性と実体性を超えて　池田　健人　51

＜会員活動報告＞　62

<第 34 回年次研究大会自由論題発表報告>

学習におけるエラーの働き：記憶の実験心理学からの示唆

岩木信喜（聖学院大学）

1. 問題の所在

ヒューム（2004、2010）は、帰納が論理的には成り立たないことを論証したが、同時に、事実として人間は帰納的に思考していると言える理由も説明した。ポパー（1974）は、前者をヒュームの論理的問題、後者を心理学的問題として区別している。本報告では、後者の心理学的問題に関するヒュームの仮説に焦点を当て、記憶の心理学実験で得られた事実に基づいて吟味することにする。

1.1. 習慣が可能にする信念形成と帰納的思考

太陽が東から昇るとか、火は熱いというような経験的知識はヒュームによれば「事実の問題」と呼ばれる。これらの知識は、過去に経験した太陽や火の生き生きとした印象に起源をもつ観念で構成される。また、例えば、火と熱の間には因果関係があるはずであるが、その具体的な理由は感官には現れないため、火と熱を経験する人間にはその理由は不明である（ヒューム、2004、p.24）。人が直接的に経験できることは、火と熱の「恒常的連接」だけであり、これはあくまでも過去経験であることから、未来（あるいは未経験の事例）に言及するために、つまり帰納的に思考するために、「習慣」の形成が必要とされる（厳密には全てのケースに該当するわけではない）。すなわち、この「習慣」というアイデアは、「なぜわれわれは、千の事例から、ある推理を引き出すのであり、そしてなぜ、そうした推理は、いかなる点においてもそれらの事例と異なっていない一つの事例からは引き出すことができないのかという、この難問を説明する唯一の仮説である」と、ヒューム（2004、p.40）は主張する。習慣が形成されることにより、火の印象から「熱」観念への移行（習慣的移行）というある種の「期待」が可能になると考えられている。このように、人が帰納的に類似事象の生起を期待するとなぜ言えるのかという問いに対し、ヒューム（2004、p.212）は「習慣のみが、あらゆる事例において、未来が過去に合致すると想定するようにと、心を決定するのである」と述べる。つまり、ヒュームは恒常的連接に基づく習慣形成を通じて、同時に、自然の斉一性原理をも学習して前提することにより、知識あるいは観念連合を材料とする帰納的思考が可能になると考えているのである（ヒューム、2004、pp.211-212）。ただし、厳密には、一度きりの経験であっても、それが十分に納得のいく実験であるような場合、十分な習慣によってすでに確立されている「自然の斉一性原理」が反省の作用を通じて単一事象に適用されることによって、信念が形成されうるとも考えている（ヒューム、2010、pp.67-68）。したがって、単一経験からその習慣化を経ずに信念を構成することは可能である。

信念は、「現在の印象と関係を持つ、すなわち連合する生き生きとした観念」（ヒューム、2010、p.61）であり、習慣による信念の形成には二通りの仕方がある。一つは、二つの対象の恒常的連接の経験によるケースであり、もう一つは、一つの観念が頻繁に心に現れる

ケースである（ヒューム、2010、pp.73-74）。前者の場合、信念は類似印象と観念との習慣的連接に起源をもつ（ヒューム、2004、p.42）。「いつも印象にひき続いて起こるものを経験するうちに、やがて印象が信念のもとをなすものになる」（ヒューム、2010、p.65）のである。後者のケース、ある一つの観念が頻繁に心に現れる場合というのは学習者に反復経験をさせる教育において典型的に認められる。いずれのケースも、観念が「しだいになだらかに、勢いをもって現れるようになり、そして、心にしっかり保たれる点とたやすく入り込む点の両方で、なにか目新しく、見慣れぬ観念と区別されるようになるに違いない。まさしくこの点で二種類の習慣は一致する」（ヒューム、2010、p.74）。

なお、反復経験による実在の信念形成について、いくつかの事例が挙げられている。手足を後天的に失った患者における手足の所有感覚、家族を失った人における故人の存在気配、面識のない著名人の話を繰り返し聞いたことによる面識感覚は、単純反復活動による信念形成の例とされ、「教育が真実には因果的推理と同じ根底である習慣及び反復に殆ど基づいて築かれている」と、ヒューム（1948、pp.188-189）は述べる。この議論に基づくなら、日常生活で反復経験する交通ルールなどの社会規範やよく目にする漢字熟語の読み方（後述する実験で用いた課題材料）などもまた、多くの場合、反復経験を通じて信念となった観念の例とみなせるであろう。

1.2. 信念の強さ

信念の強さの程度とその原因について述べる。火の印象を知覚すれば熱観念への習慣的移行がスムーズにおこり、これには例外がない（火は常に熱い）。このような場合、習慣は知識の「確実性」の認識を導くとされる。すなわち、「一つの特定の種類の出来事が、あらゆる事例において、他の種類の出来事とつねに連接しているとき、…われわれは、一方の対象を『原因』、他方の対象を『結果』、と呼ぶ。一方の対象のうちには、それが間違いなく他方を生み出し、最大の確実性と最強の必然性をもって作用するための何らかの力があると、そのように想定する」（傍点原文のまま。ヒューム、2004、p.67）と考えられている。したがって、ヒュームは、例外のない類似事象の繰り返しによる習慣形成においては、最大の蓋然性（確率 1）を意味する確実性の認識が導かれると仮定している。

しかし、確実性に到達する信念ばかりがあるわけではない。信念の強さには程度に違いがある場合もある。「偶然の蓋然性」が影響する場合がそうである。細工のないサイコロを投げて 1 の目が出るケースと 2 以上の目が出るケースでは、出現頻度に違いがあり、したがって、信念の程度にも違いが生じる（ヒューム、2004、p.51）。同様のことが「原因の蓋然性」にも当てはまることが指摘されている。下剤になる大黄や睡眠薬としての阿片が常に効いたわけではないという例が挙げられている。これらは、因果関係が不規則で不確実であると認識されてきた事例であり、外観上は厳密に類似する原因から異なる結果が生じることが知られているケースである。ヒュームは普通という意味での平均的事象の考えを用いて以下のように説明する。「われわれの過去の経験は必ずしも斉一的ではないということである。あるときは、一つの結果がある原因に続いて生じ、別の時には、他の結果が続いて生じる。このような場合、われわれは、最も普通のものが存在するだろうと、つねに信じる」（ヒューム、2004、p.215）。さらに、事象頻度を考慮に入れて、「われわれは最も平常的と知られたものを優先させ、これらの結果が生じるであろうと信じるにもかかわらず、他の諸結果も看過すべきではなく、それらの各々に、それぞれが見出された頻出の多少に比例して、個別的な重みと権威とを付与せずにはいない」（ヒューム、2004、p.53）と述べる。また、可能性が 3 つ以上の

ケースにも言及し、「このとき非常にたくさんの観察が一つの出来事に一致するのであるから、それらはこの出来事を強化し想像力に対して確証し、われわれが信念と呼ぶ感情を産み出し、かくしてその対象に、それと反対の出来事を上回る優先権を付与する」（傍点原文のまま。ヒューム、2004、p.53）とも述べる。つまり、ある原因からの移行先として複数の結果（観念）がありうる場合、もっともありそうな平均的な結果への推移が起こることで信念が形成されるが、その強さは過去経験の相対頻度に基づいて調整されるということであろう。

1.3. 信念の変更容易性

信念は、類似印象からある観念への推移の繰り返しによって強化され、堅固になった観念であるということが繰り返し述べられている。

- ある対象から移行しうる観念が複数ありうる場合でも、「このとき非常にたくさんの観察が一つの出来事に一致するのであるから、それらはこの出来事を強化」する（ヒューム、2004、p.53）。
- 「信念とは、ある対象に関して、想像が単独で獲得できるものに比べてより生き生きとした、活発な、迫力のある、堅固な、揺るぎない想念に他ならない」（ヒューム、2004、p.44）。
- 「現在の対象からの推移（筆者注：他の対象への習慣的移行）は、あらゆる事例において、それと関係づけられた観念に強さと堅固さとを与える」（ヒューム、2004、p.49）。
- 「それ（筆者注：信念）が虚構や単なる想念よりもはるかに強制的な結果を心に対して及ぼす」（ヒューム、2004、p.214）。
- 教育に典型的にみられるような単純反復学習を通じた観念の獲得に関し、「いろいろのものごとについて幼時から習慣づけられてきた意見や考えは、いずれも深く根ざしていて、理性や経験のどんな力をもってしても根だやしするのは不可能である。この習慣の影響は、原因と結果の恒常的な、分離しがたい結びつきから生じる習慣に近いというだけにとどまらず、これをしのぐ場合も多いほどである」（ヒューム、2010、p.74）。

以上のように、ヒュームにおいては、信念は習慣形成を通じて強化され、堅固さを増すと仮定されている。ここから推理して、信念は習慣形成を通じて変更されにくくなると想定できるであろう。

1.4. ヒューム仮説

ヒュームの解釈は次のようにまとめられる（図 1 参照）。

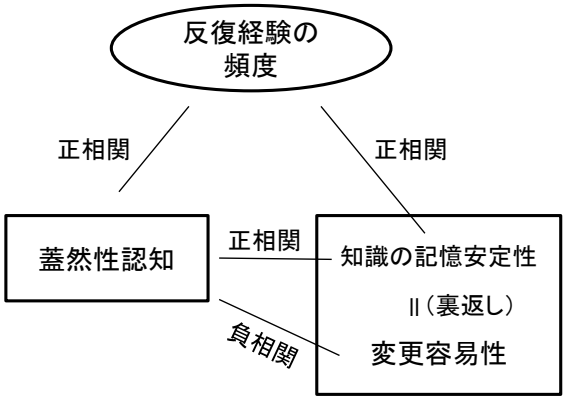

図1. 経験的知識に関するヒューム仮説

(1) 習慣形成によって、ある対象の知覚から未経験の対象の知覚を予期すること、すなわち、なめらかな推移が実現されるという仕方で、信念形成と帰納的推論が可能となる。(2) 信念の強さにはバリエーションがある。2 つの対象の恒常的連接に例外がない場合、形成された信念は確実性（確率 1）を有するに至る。ある対象からの移行先である第二の対象が二つ以上ありうる場合は、過去経験の相対頻度に応じて蓋然性が定まり、移行先としてはもっとも普通の（蓋然性が最も高い）対象が選ばれる。(3) 信念は、二つの対象の連接の繰り返しによって強化され、堅固になる。したがって、第二の対象が複数ありうる場合でも、推移先として選ばれた対象の相対頻度

が多いほど、それに応じて信念は強化され、変更しにくくなる（安定性が高まる）と考えられる。

以上のように考えると、蓋然性認知と知識の安定性はともに反復経験の頻度と正相関するのであるから、それら二つもまた正相関すると予想される。また、知識の安定性と裏返しの関係にあるのが変更容易性であるから、これは蓋然性認知とは負相関の関係にあると予想される。

2. 実験的検討

2.1. 目的

ヒュームに従えば、反復経験は知識の蓋然性認知と安定性の両者を強めると考えられる。しかし、筆者の知る限り、反復経験がそれら両者をつねに強めるのかどうかについて、心理学領域で包括的に検討した研究は見当たらない。そこで、反復経験による知識をもう少し広くとらえて、経験に基づく知識とし、その蓋然性に関する認知と安定性（記憶固定化による変更の困難さ）に関する以下の仮説を検討することにした。その際、蓋然性認知は自らの反応に関する確信度によって測定し、記憶の安定性はエラー反応を矯正するための正答フィードバックに基づくエラー修正試行の割合によって測定した。

以下の三つの仮説を検討した。一つ目は、確信度が高い知識ほど堅固である（記憶として固定化されている）ので、それが誤りであったときに正答フィードバックによる矯正が難しいという仮説である。二つ目は、誤った知識が忘却された場合よりも憶えている場合の方が記憶はより固定化されているので、正答フィードバックによる矯正が難しいという仮説である。三つめは、誤った知識の忘却を促すと、固定化の程度が低減するので、正答フィードバックによる矯正が容易になるという仮説である。

2.2. 仮説検証 1：確信度が高い知識ほど堅固なので、それが誤りであったときに正答フィードバックによる矯正が難しいか？

心理学では、手がかり情報に対して別の情報を関連づける学習のことを対連合学習という。また、経験を通じた知識の修正は、先行学習の記憶が後続学習で得た知識で置換されることであり、次のように整理できる（図 2 参照）。手がかり A に対して C を連合させるべきところ、それに先行させて A と B を連合させたとする。このとき、A に対する B の連合が先行学習であり、B を C に置換する作業が後続学習である。先行学習の後続学習に対する阻害作用があるとき、これを順向干渉と呼ぶ。ヒュームの主張は次のように換言できる。手がかりとしての問題（A）に連合しているある回答（B）の主観的正しさの程度、すなわち、確信度が高いほど、別の情報（C）への置換が阻害される（A-C ペアの学習成績が低減する）はずである。

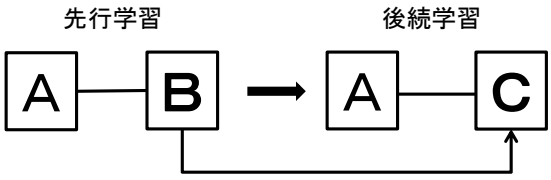

図2. 対連合学習の順向干渉パラダイム. 先行するA-Bペアの学習が後続A-Cペアの学習を阻害する.

高い主観的確率（＝確信度）を割り当てられる知識は記憶に固定化されていて根強いため、たとえ誤りであっても矯正されにくいという仮説に直接的に関連する研究がある。「hypercorrection effect of high confidence errors（高確信度エラーのハイパーコレクション効果）」と呼ばれる現象を見出した Butterfield & Metcalfe（2001）の研究である。彼らの実験では参加者は一般的知識に関する問題に回答し、同時に、回答の確信度も推定することが要求され、そのあとで正答が呈示された。ヒュームの仮説によれば、図 3 に示したように、高確信度エラーは根強い

（安定した記憶である）ので、低確信度エラーよりも修正されにくいはずである。つまり、エラー反応の確信度とその修正率は負相関を示すと予想された。しかし、実験結果は逆であり、高確信度エラーは低確信度エラーよりも修正されやすく、正相関が確認された。この結果はヒューム仮説に不利である。また、このハイパーコレクション効果は極めて頑健であることがわかっている（Metcalfe, 2017 のレヴューを参照されたい）。

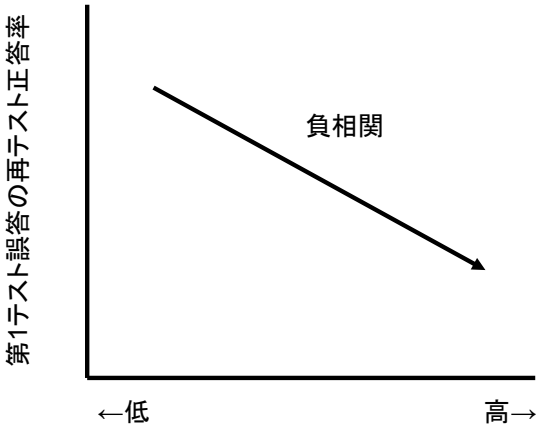

図3. 順向干渉に基づく予測

2.3.　仮説検証 2：誤った知識が忘却された場合よりも憶えている場合の方が記憶は固定化されているので、正答フィードバックによる矯正が難しいか？

この点を検討した Iwaki, Takahashi, & Kaneko (2024) の漢字読み課題を用いた実験を紹介する。実験参加者は日本人大学生であり、実験 1 では 30 名、実験 2 では 34 名が参加した。日本語の漢字熟語発音課題を使用し、熟語を呈示して読み方をキーボード入力によって回答させた。同時に、回答に関する確信度を 0（絶対に間違い）～100（絶対に正しい）の範囲で入力させた。誤答試行では、正答フィードバック直前に与えるエラーの視覚的呈示の有無を操作した（参加者内要因）。5 分間の保持期間後に再テストを行い、初期テストのエラー試行だけを実施した。そのあと、初期テストのエラーを想起するセッションも行った。分析では、エラーの確信度（低、中、高）に応じて試行を分類した。

参加者は漢字熟語（A）の読み方（B）を回答したあとで正答 (C) をフィードバックされたので、エラー試行では A-B ペアを A-C ペアに置換しなければならない。実験の結果は予想の逆であった。誤りを想起できた試行の方ができなかった試行よりもエラー修正率が高かく、十分に大きな効果量が認められた（実験 1 で dD = 1.46、実験 2 で dD = 1.50）。この結果もヒュームの仮説に否定的である。

この実験では重要な知見が得られた。これまでは知識の安定性とその変更容易性は表裏の関係にあると想定していた。しかしながら、この実験では、ハイパーコレクションが発生したのはエラーを正しく想起できる場合に限られていた。つまり、エラーを想起できないときは、エラーの反応確信度と変更容易性との間に相関関係が認められなかったのである。そこで、図 4 に示したように、知識の安定性と変更容易性を独立させることにした。

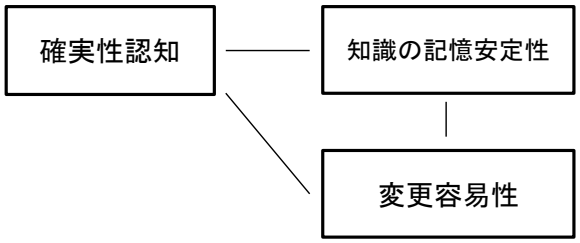

図4.「知識の記憶安定性」と「変更容易性」を独立させた仮説

2.4.　仮説検証 3：誤った知識の忘却を促すと、固定化の程度が低減するので、正答フィードバックによる矯正が容易になるか？

ヒュームの仮説から、誤った知識の記憶固定化レベルを低減すれば、別の知識に置換されやすくなるはずである。つまり、正答 (C) が経験される前に置換されるべきエラー (B) の忘却を促しておけば、順向干渉の低減によ

って C の学習成績が向上すると考えられる。この予想について、Iwaki, Nara, & Tanaka（2017）は漢字熟語の読み課題を用いて検討した。

日本語の漢字熟語の読み課題を用い、72 人の日本人大学生が参加した。参加者は、回答直後に即時フィードバックを受けるグループと回答の 1 週間後に遅延フィードバックを受けるグループに分けられ（参加者間要因）、さらにフィードバックの前に誤りが視覚的に呈示される条件とされない条件を設けた（参加者内要因）。即時フィードバック群は初期テストの 5 分後に再テストを受け、遅延フィードバック群はテストの 1 週間後に提示された遅延フィードバックの 5 分後に再テストを受けた。再テストでは初期テスト誤答試行だけを用い、漢字熟語の正しい読み方と、初期テストにおける誤答を想起させた。

主要な結果は以下の通りであった。まず、遅延フィードバック条件では実際に即時フィードバック条件と比較して誤答想起成績が悪かった、つまり、遅延フィードバック条件の方が誤答の忘却傾向が有意に認められ、その効果量は十分に大きかった（Cohen’s d = 1.09）。しかし予想に反して、再テストの成績に条件差はなく、遅延フィードバックの優位性は認められなかった（Cohen’s d = 0.12）。また、正答フィードバック時に初期テスト誤答を呈示して強制的に誤答を想起させれば順向干渉が復活するはずであり、誤答呈示条件の再テスト誤答想起率は実際に誤答非呈示条件よりも有意に高く、その効果量は十分に大きかった（Cohen’s d = 1.10）。しかしこれについても予想に反し、誤答想起率が高まったにもかかわらず、再テスト正答率は低下しなかった（Cohen’s d = 0.34）。

ヒューム仮説に従えば、遅延条件ではエラーの忘却傾向が生じるので、それに応じて順向干渉が低減される結果、最終テストのエラー修正率が改善されるはずであった。しかし、結果はこの予想に対してネガティブであった。

2.5. ヒューム仮説の修正

実験結果を受けて、図 1 の経験的知識に関するヒュームの仮説を修正したものが図 5 である（図 1 の一部を図 5 に再掲した）。要点は次のとおりである。まず、(1) 認知された

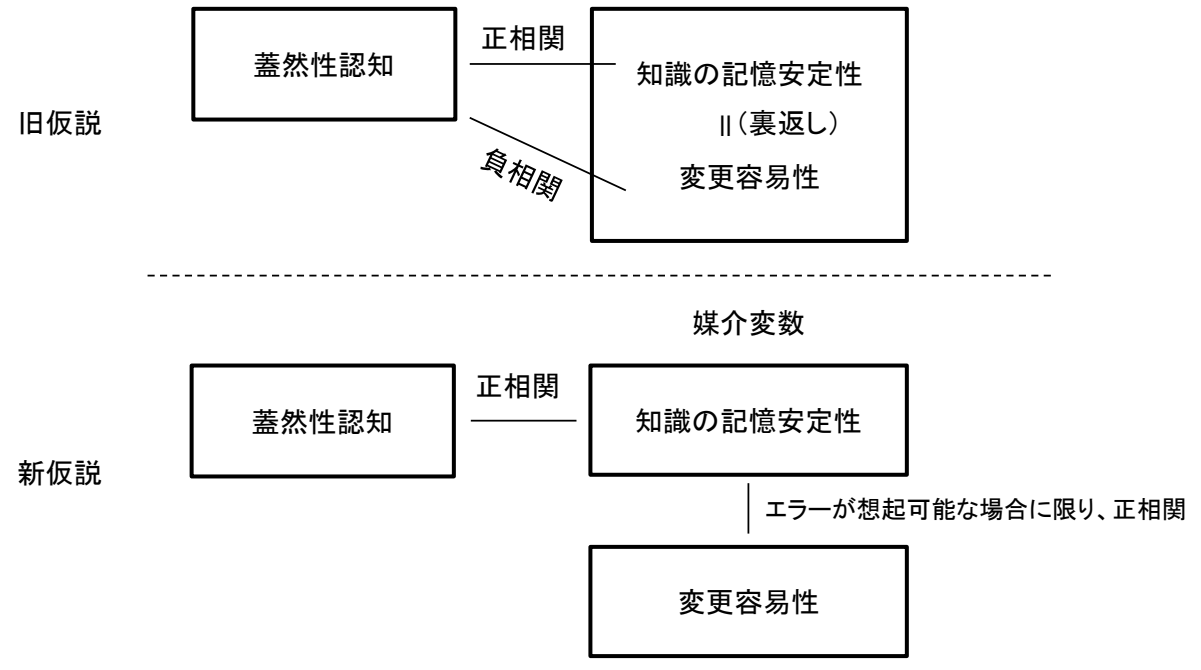

図5. 経験的知識に関するヒューム仮説の修正案

蓋然性の高さと知識の変更容易性は正相関の関係にある。当初は、蓋然性が高く認知される知識は堅固で変更しにくいと考えていたが、実際にはそのような知識ほど変更しやすかったのである。また、(2) 知識の記憶が安定的である（固定化されていて忘却されにくい）ことと変更容易性は独立の関係にある。当初はそれらが裏返しの関係であって、安定した記憶は固定化されているので変更しにくいとみなしていたが、そうではなかった。蓋然性認知と変更容易性の正相関（ハイパーコレクション効果）がエラーの記憶が想起できる場合に限って現れ、エラーを想起できない場合には無相関だったからである。その意味で、(3) 知識の記憶安定性について、蓋然性認知と知識の変更容易性をつなぐ媒介変数として想定し直した。媒介変数として機能しうるメカニズムは今後の重要な検討課題である。

3. まとめ

心理学実験の主要な結果は次のとおりであった。(1) 誤った知識は確信度が高いほど根強いと仮定したが、正答フィードバックによる矯正が低確信度の誤りよりも容易である。(2) 誤った知識が忘却された場合よりも憶えている場合の方がその記憶はより堅固と仮定したが、正答フィードバックによる矯正が実際には相対的に容易である。(3) 誤った知識の忘却を促すと順向干渉レベルが低減すると仮定したが、正答フィードバックによる矯正が実際には容易にならない。これらの事実はすべて順向干渉に基づく予想に反するものであった。そこで、ヒュームの仮説を図 5 に示した案ように修正した。

しかしながら、人の帰納的思考の存在を説明するヒューム哲学の核心部分にはまだ反証は及んではいないと考えられる。帰納的思考というものは「ある対象の知覚から生じる別対象の出現期待」が習慣を通じて形成されることにあるとヒュームは考えているのであり、この点は否定されてはいないからである。そうは言っても、ヒュームの考えが心理学実験データと部分的にでも矛盾を示していることは事実である。本報告から明らかなように、ヒューム哲学は心理学と親和性があり、ヒューム哲学が扱う問題の検討に心理学実験を利用する試みは価値があると思われる。

【文献】

1. Butterfield, B., & Metcalfe, J. (2001) Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1491-1494.
2. ヒューム, D. 斎藤繁雄・一ノ瀬正樹（訳）(2004). 人間知性研究　付・人間本性論摘要　法政大学出版局.
3. ヒューム, D. 大槻春彦（訳）(1948). 人性論（一）　岩波書店.
4. ヒューム, D. 土岐邦夫・小西嘉四郎（訳）(2010). 人性論　中央公論新社.
5. Iwaki, N., Nara, T., & Tanaka, S. (2017) Does delayed corrective feedback enhance acquisition of correct information? *Acta Psychologica*, 181, 75-81.
6. Iwaki, N., Takahashi, I., & Kaneko, S. (2024) How does error correction occur during lexical learning? *Psychological Research*, 88, 1272-1287.
7. Metcalfe, J. (2017) Learning from errors. *Annual Review of Psychology, 68*:6.1-6.25.
8. ポパー, K. R. 森博（訳）(1974). 客観的知識―進化論的アプローチ―　木鐸社.

【謝辞】

本報告の作成にあたり、高橋功先生（山陽学園大学）、金子紗枝子先生（比治山大学）、諸冨隆先生（北海道大学名誉教授）から様々

なご指摘、ご意見をいただきました。謹んで
感謝申し上げます。

生成 AI と世界３

蔭山泰之(日本 IBM)

はじめに

現在、注目を集めるようになっている生成 AI は、これまでのどのテクノロジーよりも現代社会に大きなインパクトを与えるようになってきている。この生成 AI について、さまざまな意見や見解が出されてきているが、ものすごいスピードで発展しているため、これの将来について的確な見通しを立てることはきわめて難しい。

ポパーは、周知のように人間活動の産物が客観的な世界３を構成すると唱えたが、この生成 AI はまさに典型的な世界３の住人であると言えるだろう。とすると、世界３の観点から生成 AI を見ることによって、なんらかの気づきや知見が得られるのではないか。また逆に、生成 AI を対象とすることで、ポパーの世界３、三世界論が持つだろう現代における意味、意義をあきらかにすることができるのではないか。以下においては、こうした期待のもとで論を進めていく。

１．生成 AI の登場

１－１．AI の変遷

現状では AI の技術は秒進分歩とでも言えるスピードで進歩、変化しているが、いわゆる人工知能と呼ばれる技術はコンピュータ誕生直後から存在した。現在の AI は第 4 世代と言われているが、これまでの世代は次のとおりである[1]。

AI の第一世代は、1950～1960 年代にかけての時期で、ルール/アルゴリズムによる論理演算や推論/探索に基づく AI だった。この簡単なパズルは解けたものの、複雑な問題にあたってはすぐに限界に突き当たった。

次に第二世代の AI は、1980～1990 年代にブームを迎えた知識ベースシステムである。論理演算に専門家の知識を加えた知的処理が可能になったが、適切なルールを与える人手作業が大変で、この作業の限界がブームの終焉をもたらした。

第三世代の AI は、ビッグデータを統計処理する機械学習の一種であるディープラーニング（深層学習）をベースとした AI である。2010 年代以降注目されるようになってきた。

そして、現在の第四世代の AI は、2020 年代以降から本格化してきた膨大なパラメータをもつ大規模言語モデル(LLM. Large Language Model)や拡散モデルにもとづく生成 AI である。これにより、高度な知的能力が見られるようになった[2]。

１－２．ルールベースからディープラーニングへ

このように見てくると、AI の発展の速度が第三世代以降、加速して来ていることがわかる。これは、AI のアーキテクチャが根本的に変わったためである。第三世代以降の AI は、データから特徴量を抽出して知識を獲得できるようになった点が第二世代までの AI と決定的に異なる。AI が急速に発展するようになったのは、あまり人手を介さずに AI が知

[1] 松尾豊、『人工知能は人間を超えるか』、角川書店、2015 年、第 2、3、4 章。今井翔太、『生成 AI で世界はこう変わる』、SB クリエイティブ、2024 年、第 2 章参照。

[2] 現在の生成 AI は第 3 世代と同じくディープラーニングに基づいている技術なので第 3 世代と第 4 世代を分けない見方もあるが、生成 AI を第 3 世代とは異なる第 4 世代と見る見方は、その性能と影響力がこれまでの AI とは比較にならないほど大きいためである。

識を獲得できるようになったディープラーニングによるところが大きい。

１－３．AIにおける学習

これまでのコンピュータシステムでは、データを処理するルール、つまりデータ処理の手順を人がプログラムのかたちでコンピュータに与えていた。このため、ルールという知識を与える人手作業がボトルネックになっていた。

ルールベース以前のAIでは、AIが対処すべきさまざまな場面での「こうなったらどうするか」を人が予想し、想定して、逐一その条件と対応ルールを入力しておかなければならなかった。写真から犬と猫を判別する場合、犬の特徴、猫の特徴をいちいち入力しなければならなかったし、自然言語処理でも単語の意味と構文解析から、さまざまな文章での意味の違いを処理しなければならなかった。このため、人手による入力作業が膨大になってしまい、人間の作業の限界がそのままAIの性能の限界になってしまっていた。

ところがディープラーニングによって、機械への知識付与という人手作業のボトルネックが解消され、AIが自動的に知識を獲得する機械学習ができるようになった。こうしてAI自体が、膨大な画像データから犬の特徴、猫の特徴を導き出して、判別できるようになった。また自然言語処理では、こういう文章が来たら、次のことばはこうなる確率が高いという推論を行えるようになったわけである。

そのような機械学習にはいくつか種類がある。まず、人が正解データを用意してその正解データと推論の差をもとに学習する教師あり学習がある。これは人がAIをテストして成長させることに相当する。次に、データの中から自動的に特徴を発見し、グルーピングなどを行う教師なし学習があるが、これはアルゴリズムの一種である。

また、機械が自律的に環境を探索して得た経験データとタスクの成功である報酬から学習する強化学習がある。これは、将棋や囲碁などでは正解を与えることはできないが、上手い手だったかどうかを結果から評価しながら学習して、名人の域に達していくプロセスである。

そして教師あり学習の進化形として、教師データも機械が自動的に生成して、その正解から学習する自己教師あり学習がある。これは、AIが自分で用意した穴埋め問題などを解くことにより、知識を蓄えていくプロセスである。これにより、人手が介在する必要性が大幅に減ってきているわけである。

現在の生成AIは、以上見てきたような学習をいろいろと組み合わせて、さまざまな知識を獲得できるようになっている。

１－４．ディープラーニングの基本的な仕組み

現在の生成AIは高度な知的振る舞いが可能になってきているが、その基本的な仕組み、動作原理は、簡単に述べると次のようになっている。

ディープラーニングは、元々は人間の脳の構造を模したニューラルネットワークという構造で実施される機械学習である。このニューラルネットワーク上のひとつひとつのノードでは、数値化されてベクトル表現された入力値を重みづけして調整し、その結果にバイアスを加えて積和計算している（図１参照）。そして活性化関数を通して出力値を生成し、次の層に渡している。これを複数の階層で繰り返して出力する。そして、出力層の推論値と実際の値を比較して、その差の自乗を損失と捉えて、勾配降下法という計算法でその損失が最小になるように各ノードで計算する重みを調整して

いる[3]。

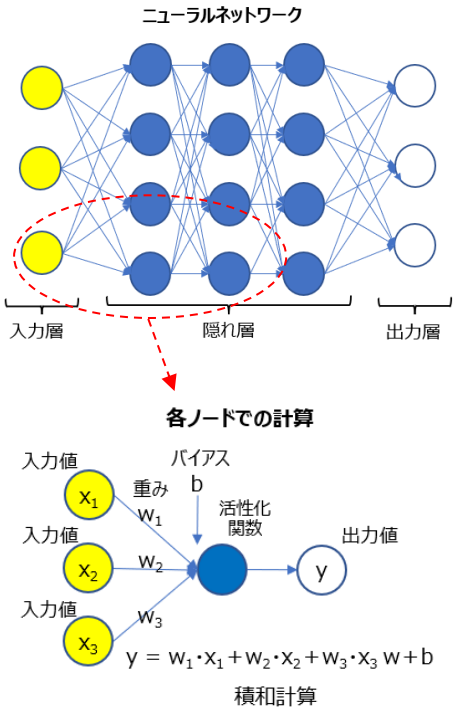

図 1

現在では、ニューラルネットワークの各ノードで行われているこのようなデータ処理がきわめて大規模に行われることによって、驚異的な知的能力が発揮できるようになっているのである。

２．反証主義から見た生成 AI

２－１．生成 AI の特徴

このような生成 AI には、さまざまな特徴がある。まず、生成 AI は文脈が読める。つまり、文脈に応じて振舞いを変えられるのである。これはアテンション（注意機構）によるものだが、この特徴のために、生成 AI は、たんにデータを受動的に受け取るのではなくて、積極的にデータの特徴を捉えることができるようになっている。

次に生成 AI は誤りから学べる。つまり、バックプロパゲーション（誤差逆伝播法）により、自らが出した結論の誤りをニューラルネットに逆方向に伝播させて、誤りから学ぶことができる。生成 AI ではこの誤りのフィードバックを大規模なかたちにして知識を獲得し、精度を向上させている。

そして、生成 AI では能力が創発される。言語モデルのパラメータ数を巨大にすると、それまで見られなかった能力が現れるようになるという現象が起きているのである。

以上見てきたポジティブな特徴以外に、生成 AI にはネガティブな特徴もある。生成 AI からの多数の正解の中に、しばしば虚偽の回答が含まれているが、これはハルシネーション（幻覚）と呼ばれている。生成 AI は真偽を確かめて回答するのではなく、確率的に予測された回答を返しているだけなので、嘘をついているという自覚すらないだろう。学習を重ねていくことで、ハルシネーションは減っていくと考えられるが、根絶できる保証は今のところはない。

生成 AI はさまざまな技術から成り立っていて、これらの特徴を示しているが、そのうち反証主義、批判的合理主義の観点から注目すべきはアテンションとバックプロパゲーションである。

２－２．アテンション：注意機構

アテンション（Attention）とは、機械翻訳を実行する計算の仕組みであるトランスフォーマー（Transformer）にとって、もっともコアとなるテクノロジーであり、これによって言語処理の性能が劇的に向上した[4]。

トランスフォーマーでは、単語の意味を数値でベクトル表現し、その意味に文脈上の他の周りの単語の値を追加（embedding）でき

[3] 岡之原大輔、『ディープラーニングを支える技術』、技術評論社、2022 年、第 6 章。丸岡章、『概説人工知能』、筑摩書房、2024 年、第 4 講。立山秀利、『ディープラーニング AI はどのように学習し、推論しているのか』、日経 BP 社、2024 年、第 4 章。

[4] 吉川和輝、「AI に人間らしさをもたらした大規模言語モデル」、日経サイエンス、Vol.53, No.5, 2023 年、32-39 ページ。

るようになっている。たとえば、以下の図２のように、A、B、C と読める文字の並びと、12、13、14 と読める文字の並びが交差している場合で見てみよう[5]。

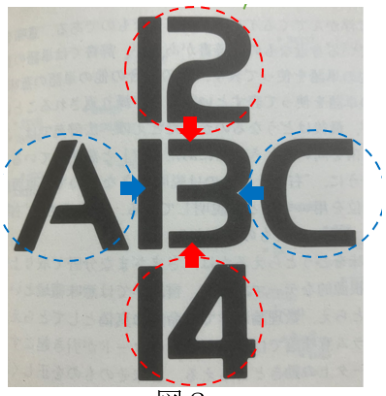

図２

真ん中の文字は、縦の並びにある場合は、上下にある 12 と 14 の情報の一部が注入され、横の並びにある場合は、左右にある A と C の情報の一部が注入される。これによって、文脈に応じて真ん中の文字は、13 と解釈したり、B と解釈したりできるようになる。つまり、アテンションによって、文脈に応じた判断が可能になっている。

次に、以下の二つの例文で考えてみよう。同じ bank という言葉には、文脈に応じて別々の意味が割り当てられている。

I drove across the road to pay the other bank.

(別の銀行に払うために車で道路を渡った。)

I swam across the river to get the other bank.

(対岸に着くために川を泳いで渡った。)

この二つの例文で言えば、単語“bank”に文脈上の他の単語のベクトル値が載せられて、結果として意味の違いが区別できるようになっている。このように、アテンションとはポパーが言うサーチライトである。生成 AI はポパーが言うバケツではない[6]。

ポパーは素朴な帰納主義を論駁するために、観察を繰り返す前に視点が必要だとして、さまざまな図３のような図形の反復を挙げたが[7]、この例で言えば、色、大きさ、形状などのどの特徴に着目すべきかを生成 AI は自ら判断して先に進むことができるようになっているわけである。

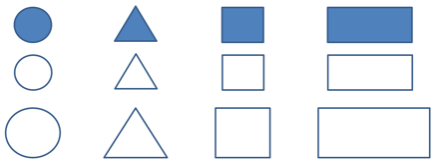

図３

２－３．フレーム問題の解決？

以上のアテンションによって、人工知能の課題だったフレーム問題も、限られた条件下では解決できるようになったと言えるのではないだろうか。

フレーム問題とはダニエル・デネットが提起した人工知能の限界を示した思考実験で、AI を搭載したロボットが、爆弾が仕掛けられた状態のバッテリーの状況を適切に判断して、うまく持ち出せるかどうかという問題である[8]（図４参照）。

図４

[5] 丸岡章、『概説人工知能』、前掲書、158-160 ページ。

[6] カール・R・ポパー、森博訳、『客観的知識』、木鐸社、1974 年、379-402 ページ。

[7] カール・R・ポパー、大内義一、森博訳、『科学的発見の論理』、恒星社厚生閣、1971/72 年、515-517 ページ。

[8] 鈴木貴之、『人工知能の哲学入門』、勁草書房、2024 年、52-59 ページ。

図に説明されているように、1 号機から 3 号機までのケースが考案されているが、どれもうまくバッテリーが持ち出せないという結論になって、AI の限界を示していた。しかし現在の生成 AI は、2 号機のように副次的に発生しうる事項を無限に考えたりはしないし、3 号機のように目的と無関係な事項をすべて洗い出したりもしない。アテンションによって、注目すべきところと、そうでないところを見分けながら問題を解けるようになっている。つまり、文脈に応じて振舞いを変えられるようになっているのである。

また、生成 AI はさまざまな知識を学習できるようになっているので、たんに与えられたデータだけでなく、その背後に隠された意味まで洞察できるようになっている。なので、たとえば生存者バイアスの知識を学習させることにより、エイブラハム・ウォルドが洞察したように、図 5 のような爆撃機の破損状況のデータから、防御を強化すべきは、コックピットと垂直尾翼[9]だと、生成 AI も適切に回答できるかもしれない。

図 5

2－4．バックプロパゲーション：誤差逆伝播法

従来、コンピュータ利用は人から機械への一方向の指示だけだったので、実行結果からのフィードバックを受け取ってそれに対応するのはもっぱら人の役割だった。センサーからのフィードバックを受け取って制御に利用するシステムもあるが、それらは想定された範囲のフィードバックだった。想定外の事態が生じたときに、それに対応してプログラムを変更するのはやはり人だった。

だが、現在の生成 AI では、推論された予想値と実測値のあいだの差異を損失として、バックプロパゲーションにより、ニューラルネットに逆伝播させてノードのパラメータを調整し、その差異を小さくすることができる[10]（図 6 参照）。

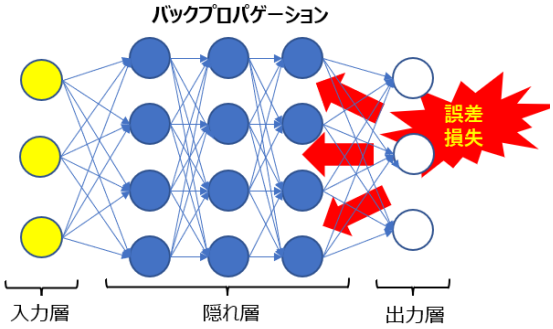

図 6

バックプロパゲーションのそもそもの目的は、各ノードのパラメータを調整する計算量を減らすことだったが、結果として現在の AI は結果からのフィードバックにより、自動的に自己自身を変えられるような仕組みになっている。

このようなバックプロパゲーションは、ポパーが言う偽の逆伝送（*Retransmission of Falsity*）[11]と見なすこともできるだろう。こう見ると、生成 AI は試みと誤りの方法、推測と反駁の方法で学習しているとも言える。生成 AI はこの方法を獲得したことによって、飛躍的な学習能力を備えることになったの

9 コックピットと垂直尾翼をやられた爆撃機はそもそも帰還できないので、データが得られないわけである。

10 丸岡章、『概説人工知能』、前掲書、117-118 ページ。岡之原大輔、『大規模言語モデルは新たな知能か』、岩波書店、2023 年、6 章。

11 カール・R・ポパー、森博訳、『果てしなき探求』、岩波書店、1978 年、下巻、81 ページ。

である。

２－５．生成 AI の能力創発

生成 AI では、「生成される単語や文章に確率を割り当てて予測する言語モデルをベースにしているが、ChatGPT などは、そのモデルの規模が大きいことが特徴である。そのため大規模言語モデルと呼ばれているが、その実現に必要なニューラルネットワークのパラメータ数は、2024 年時点では公開されている限りでも、すでに数千億のレベルにまで達している。

従来の AI 研究は、規模の大きさに頼らずに、いかにして効率よく性能を向上させるかを目指してきた。それは、パラメータ数を無暗に増やすと、学習データに過剰に適合して、新しいデータをうまく処理できない過学習に陥ると考えられてきたからである。

ところが LLM の性能については、モデルサイズ、学習データ量、学習の計算量で決まるスケーリング則が成り立っている。とくにパラメータ数を増やしていくと、図 7 で示したように突然、モデルの性能が劇的に向上し、それまで出せなかった正答を出せるようになる。このように小規模なモデルにはなかったような能力創発（Emergence）という現象が見られるようになっている[12]。

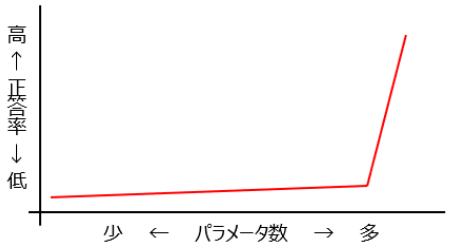

図 7

しかし、なぜ規模を大きくすると能力創発が見られるのか、現時点ではまだその原理は解明されていない。そもそも、学習データで学習した結果、未知のデータに対しても正解が出せるという汎化能力についても、まだその原理はよくわかっていないのである[13]。

ニューラルネットワークの個々のノードは、単なる数値と演算のかたまりにすぎないので、能力創発は、個々のノードには還元できない。ニューラルネットワークは全体として性能を発揮しているわけである。このように、現在の生成 AI は人間の産物であるにもかかわらず、まだまだ知られていない未知の部分を秘めているのである。

２－６．機械学習は帰納か？

ここで、反証主義の立場から AI における帰納の問題に触れておくこう。

人手で開発したプログラムによる演繹的なデータ処理に対して、現在の AI は膨大なデータからルールや特徴量を学習して導き出すので、しばしば帰納的だと言われることがある[14]。では、機械学習は帰納なのだろうか。これは、いわゆる帰納主義者が言う意味での帰納ではない[15]。

まず、機械学習によって膨大なデータからルールや特徴量を導き出せたからといって、それによって真理に近づいたわけでも、真であることが確からしくなったわけでもない。あくまで、学習データに対して相対的に精度の高い推論、予測ができるようになったということだけである。このため、学習データが変われば AI の出力も変わってくる。むしろ生成 AI の機構そのものよりも、学習データの方がきわめて重要になってきている[16]。

---

12 今井翔太、『生成 AI で世界はこう変わる』、前掲書、56~57 ページ。

13 今泉允聡、『深層学習の原理に迫る』、岩波書店、2021 年、第 4 章。出山政彬、「大規模言語モデルとは何か」、日経サイエンス、Vol.53, No.10, 2023 年、46-53 ページ。岡之原大輔、『大規模言語モデルは新たな知能か』、前掲書、6 章。

14 小高知宏、『機械学習と深層学習』、オーム社、2015 年、68-70 ページ。

15 もちろん、帰納ということばを使うべきではないというわけではないが。

16 このため、ビジネスの領域では、質の良い学習

次に、学習プロセスは結果からのフィードバックによってコントロールされている。教師あり学習は正解との差を表す損失によってコントロールされ、強化学習は報酬によってコントロールされているので、この点でも帰納ではない。

さらに、教師なし学習におけるクラスタリング、グルーピング、特徴量抽出、次元削減は、一定のアルゴリズムに基づいたデータ操作であり、この意味では、演繹操作の一種である。

３．世界３と生成 AI

３－１．カール・ポパーの三世界論

ここで、ポパーの三世界論に移ろう。世界３､あるいは三世界論は､1960 年代後半以降、ポパーが展開してきた知識論、世界観、形而上学説である[17]。世界３/三世界論はポパーのそれ以前の議論、理論に比べると、批判も共感もあまりなかったと言えるかもしれない。

反証可能性のような精密な議論や開かれた社会のような合理的な議論に比べると、三世界論はメタファが多用されている[18]。このため、さまざまな解釈が可能で、イメージや説得力に訴えているだけだとの批判を招くこともあった。このように、三世界論はポパー研究者の中でも、しばしば扱いに困るポパーの議論であった。しかし、生成 AI の登場をこの世界３の観点から見ると、いろいろと興味深い示唆が得られてくるのである。

ポパーによれば、次の三つの世界を区別することができる。世界１は物理的なものの世界（唯物論者が唯一認める世界）である。世界２は心的なこころの世界（唯心論者が唯一認める世界）である。そして、世界３は客観的な知的産物・構築物の世界（ポパーオリジナルの世界）である（図８参照)。

図８

３－２．世界３の特徴

世界３には人間が生み出した客観的な知的構築物が住人として存在するが、このような世界３には以下のような特徴がある[19]。

まず、世界３は人間/生命の産物の客観的世界である。世界３は人間が生み出した知識の客観的世界であり、その住人は書籍の内容、理論、問題、議論、社会制度、芸術作品などである。

世界３は世界２によって生み出される。世界３は客観的という意味では世界２からは独立しているが、世界３の住人は世界２の活動によって生み出されるのである。

それでも、世界３はほかの世界からは独立している。世界１と世界２が互いに独立していることは言うまでもないが、世界３もほかの世界からは独立していると考えられる。つまり、世界３の知識は、世界２にある主観的知識（心の状態）とはまったく異なるものである。

そして、世界３は自律している。世界３はとくに世界２から独立しており、さらに自律している。さらに世界３は、ほかの世界と相

データの獲得競争が発生している。

[17] カール・R・ポパー、『客観的知識』、前掲書、第 3 章、4 章で詳細に三世界論が論じられている。

[18] ポパー自身が、世界３はメタファだと述べたことがある。Karl R. Popper, *Knowledge and the Body-Mind Problem*, London, Routledge, 1994, p.25.

[19] カール・R・ポパー、『客観的知識』、前掲書、第 3 章。カール・R・ポパー、小河原誠、蔭山泰之訳、『よりよき世界を求めて』、未來社、1995 年、第１章参照。

互作用するのである。世界３の住人は世界１や世界２の対象と相互作用する。ただし、世界１と世界３は直接相互作用せず、世界２を介して相互作用する。

またさらに、世界３は成長する。世界３は他の世界からは独立して、独自に自律しながら、進歩発展する。この点で、世界３はプラトン的なイデアの世界とは異なるわけである。また、世界３は世界２にフィードバックして世界２を成長させるのである。人間は、世界３からのフィードバックにより成長するのである。

そして最後に、世界３は未知の領域を持つのである。ポパーが「人は自ら生み出したものを知らない[20]」というように、世界３のもっとも重要な論点のひとつである。

３－３．世界３の住人としての生成 AI

生成 AI はインターネットとそこから得られるビッグデータ、いわゆるサイバー空間なしにはありえなかったが、これらはどれも、ポパーの没後に、急速に発達してきたものである。ポパーが現在も存命であれば、必ず生成 AI について論じていただろうし、世界３と結びつけていただろうと思われる。

サイバー空間の構成要素は、どれも世界３の住人としての資格を備えている典型的な世界３の住人である。いわゆるサイバー空間そのものが世界３と言ってよいだろう。サイバー空間なしには、生成 AI もありえなかった。そして生成 AI は、これまでの人間の知的産物のどれよりも、世界３の住人としての特徴をすべて兼ね備えている。

とくに、これまでの世界３の住人よりも生成 AI に特徴的なのが、人に働きかけるその能力である。ポパーが想定していた書物や理論、これまでのサイバー空間上の構築物は、たしかに自律していると言えるが、どちらかというと世界２からの働きかけに答え、反応するかたちで発展して来た。しかし、生成 AI は人に対して働きかける能力がこれまでのどの世界３の住人よりも格段に向上している。このためにインパクトが大きい。世界３が自律している[21]というのは、ポパーの根本的な主張だが、この点で生成 AI はまさに自律しているように振舞うわけである。

３－４．世界３の自律と Unknown

世界３を論じるにあたっては、ポパーはしばしば次のような文明の破壊についての思考実験を持ち出してきた[22]。

① すべての機械と道具が破壊され、それらの使用法などの知識も失われた。だが、書物と人がそこから学ぶ能力は残されている。
② すべての機械と道具が破壊され、それらの使用法などの知識も失われた。しかも書物もすべて失われている。

この思考実験の①の場合、文明はすぐに復興できるが、②の場合、文明の復興は難しいという。

ここで、以下のように書物を生成 AI で置き換えると、この思考実験はより説得力が増してくる。

① すべての機械と道具が破壊され、それらの使用法などの知識も失われた。だが、生成 AI と人がそこから学ぶ能力は残されている。
② すべての機械と道具が破壊され、それらの使用法などの知識も失われた。しかも生成 AI もすべて失われている。

むしろ、生成 AI は書物よりも積極的に人に働きかけて、素早く文明を復興させることだろう。

だが、生成 AI について強調すべきは、むしろそれが抱えている未知、Unknown の領

[20] W・W・バートリー、小河原誠訳、『ポパー哲学の挑戦』、未來社、1986 年、156-202 ページ。

[21] カール・R・ポパー、『客観的知識』、前掲書、133-138 ページ。「学習する」や「理解する」などは、本来、自律的存在である人間や生物に使われる述語だが、AI の分野では、もはや違和感なく機械に対してもふつうに使われるようになってきている。

[22] カール・R・ポパー、小河原誠訳、『開かれた社会とその敵』、岩波書店、2024 年、第 2 巻上、234-235 ページ。カール・R・ポパー、『客観的知識』、前掲書、125 ページ。

域である。

とくに生成 AI の能力創発は、この Unknown にあたると考えられる。これは、ポパーが世界３について論じていた際に、「人は自分たちが生み出したものを知らない」と述べて強調していた部分である。ポパーはこの Unknown を世界３の最大の特徴と捉えていたと思われる。なぜなら、人はこの Unknown から学ぶことで成長もするし、逆にこれによって危険にも晒されるからである。

３－５．Unknown 領域としての世界３

三世界論では、世界２を通して世界１と世界３が相互作用するので、図９の左図のようなイメージが描かれることがある[23]。このイメージを Known/Unknown の観点から捉えなおすと、図９の右図のようにも描くことができる。

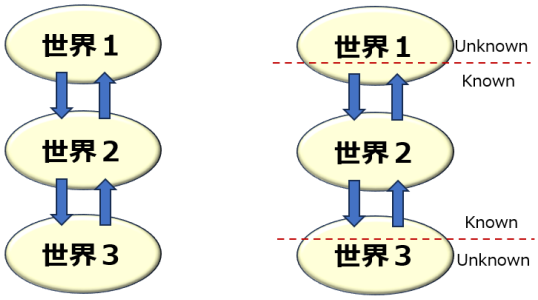

図９

世界１にある Unknown と世界２がやり取りしてそこから世界２が学ぶ以外に、世界３にも Unknown があり、世界２はやはり Unknown 領域から学ぶのである。図９の右図を図 10 のように描き変えると、世界２は、世界１や世界２の Unknown から学ぶということがよりイメージしやすくなるだろう。

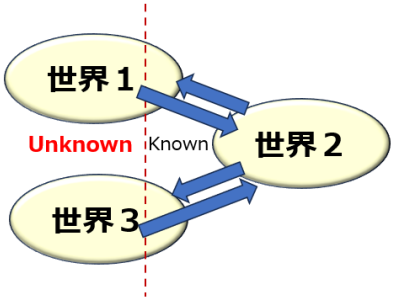

図 10

このように人は世界３の Unknown からフィードバックを受けながら成長するのである。

３－６．生成 AI から見た世界３

能力創発のように、生成 AI は人間が生み出したものであるにもかかわらず、その原理、仕組みが解明されていない側面がある[24]。つまり、生成 AI には未知の部分/領域があるわけだが、これは、ポパーが「人間は自分が生み出したものを知らない」と述べた事態に対応していると言える。

もちろん、世界３だけでなく、世界１にも当然まだ多くの未知の領域が数多く残っている。いずれにせよ、世界１や世界３に未知/Unknown の領域があるからこそ、そこからフィードバックが得られて相互作用することができる。人が学ぶこととは、未知、想定外のことからの反作用によって成長することだと言うことができる[25]。

ここで世界１と世界３の関係に目を向けてみよう。世界３はメタファと言われることもあったが、AI やアバターの進歩をみると、VR や AR などでは、もうすでにサイバー空間上に独自の世界が構築されている。VR や AR などの存在感は今後ますます増していくだろう。もうすでに実在の世界との境界線も曖昧になっている（図 11 参照）。

---

[23] 小河原誠、『ポパー』、第２版、筑摩書房、2024 年、341 ページ。

[24] 能力創発を説明する仮説がいくつか提案されているが、どれもまだ決定的なものではない。今泉允聡、『深層学習の原理に迫る』、前掲書、第５章。

[25] プログラマの能力は遭遇したバグの数に比例すると言われるが、ソフトウェアエンジニアもまさに想定外の事態を通して成長するわけである。

図 11

Real と Virtual の違いが曖昧になっている以上、AI の出力が真かどうかについては、真ということばの意味が変わってきているのかもしれない。

４．フィードバックの哲学

４－１．フィードバックの重要性

世界３から見た生成 AI では、その仕組みにおいても、人間に対する影響という面でも、フィードバックということばがキーワードとなってくる。

生成 AI は、フィードバックの機構を自己に組み込むことによって飛躍的な発展を遂げた。世界３の Unknown からのフィードバックを受けての相互作用によって、われわれは成長するとポパーは述べた[26]。生成 AI は、われわれ人間が生み出したものであるが、この生成 AI がもたらす予想外、想定外の結果を受けて、人間も大いなる発展を遂げられる可能性がある。

では、なぜフィードバックが重要なのだろうか。それは、発展のためと、生存のためと、主体性のためである[27]。

発展のためというのは、真理とか、よりよい状態に近づくためということであり、科学にとって重要である。生存のためというのは、われわれが暮らす変化する環境に適応するためということであり、生物にとって重要なことは言うまでもない。そして主体性のためというのは、狭い主観の世界から脱出するためということであり、われわれが成長するためには、主観の世界から抜け出すことが必要不可欠なのである。

ポパーが世界３、三世界論で言いたかったこととは、自律している世界３から来るまだ知られていないことについてのフィードバックを真剣に受け取れということであろう。そしてそれを糧にして、自らを成長させよということだった。反証可能性も、経験から得られる Unknown からのフィードバックによって理論を改良して、成長させるためのものであり、科学的知識の成長にとって必要不可欠である。逆に、フィードバックを無視して「すべてわかっている」と思い込むと、そこで成長が止まってしまうのである[28]。

４－２．ポパーの思想を貫くもの

ポパーの思想は科学論から社会論、形而上学にまでさまざまな領域に及んでいて、その全体像を統一的に理解することは難しいが、ここにおいてもフィードバックが理解のカギとなる。要するに、ポパーの思想は、全体として、いわばボトムアップのフィードバック哲学と言えるのである。

まず、ポパーの初期思想の中心にくる反証可能性について言えば、この理論の核心は Openness to the Unknown である。反証可能性は、そのことば上の否定的なニュアンスにもかかわらず、反証された理論を鍛えて、前に進むためにこそ必要である[29]。フィード

---

[26] カール・R・ポパー、『果てしなき探求』、前掲書、下巻、181-182 ページ。

[27] 蔭山泰之、「反証主義の精神」、批判的合理主義研究、Vol.2、No.2、2010 年、2-14 ページ。

[28] ポパーがコペンハーゲン解釈でもっとも反発したのは、End of Road テーゼだった。カール・R・ポパー、小河原誠、蔭山泰之、篠崎研二訳、『量子論と物理学の分裂』、岩波書店、2003 年、10-13 ページ。

[29] Y. Kageyama, “Openness to the Unknown: The Role of Falsifiability in Search of Better Knowledge”, *Philosophy of the Social Sciences*, Vol.33, No.1., 2003, pp.100-121. 蔭山泰之、『批判的合理主義の思想』、未來社、2000 年、第 4 章。ポパーが反証の決定性を主張していたという誤解を、私は第二のポパー伝説と呼んだが、いかなる反証も決定的ではない。それゆえ、反証はただちに理論の放棄につながるわけではない。むしろ、反証というフィードバックを糧にして、理論を前に進めることができるのである。

バックを受け取るためにこそ、未知に対して開かれているべきなのである。反証のようなフィードバックこそ前進するためのトリガーになるからである。

また、ポパーがずっと擁護してきた実在論についていえば、「蹴ったら蹴り返す」というアルフレッド・ランデのことば[30]が、ポパーの実在論をもっとも的確に表現している。この、蹴り返しはまさにフィードバックだと言ってよいだろう。

そしてポパーの中期の思想で、社会論で現れてきたピースミール社会工学についても、フィードバックの重視が読み取れる[31]。そもそもユートピア工学では、プランの実行結果のフィードバックから軌道修正することが最初から想定されていない。プランの実行結果からその誤りを修正するには、フィードバックが有効に活用できるようにピースミールに前進する必要がある。そしてポパーの民主主義論では、民主主義を流血なしに権力者を変えられる制度としているが、これこそ平和的なフィードバックシステムであると言える。

そしてポパーの後期思想では、三世界論が前面に出てくるようになる。その議論において世界３の根本は、それ自体が自律していて未知の領域を持ち、そこから人間、世界２に対してフィードバックを与えてくるというところにある。

４－３. ネガティブフィードバックの重要性

フィードバックの中でもポパーがとくに重視するのは、ネガティブなフィードバックである。ネガティブフィードバックには、誤差、期待外れとから想定外の事態までさまざまなレベルがあるがあるが、どのレベルにおいてもそれぞれ有効に活用することができる。

生成AIにおけるディープラーニングでは、予測値と実測値の差異を損失としてニューラルネットワーク上で逆伝播させ、各ノードのパラメータの修正・調整に活用することで、出力の精度を向上させ、高度な知的能力を得ることに成功したわけである。バックプロパゲーションは、損失の原因となるある特定のノードを探すのではなく、ニューラルネット全体を対象にしている[32]。つまり、反証からその原因を特定することよりも、むしろ反証によって理論体系が全体として改良されていくことこそが重要なのである。

そもそも、「どうしてそうなるのかわからない」ということ自体が実はネガティブフィードバックなのである。人が成長するのも世界３からのネガティブなフィードバックによってであり、これとの相互作用により成長する。

一方、逆のポジティブなフィードバックとは、予想どおりの結果ということだが、これは知識を拡張しないし、成長もさせない。つまり学ぶところがほとんどない。知っていることを返されても、あまり学びはない。逆に、知らないことが返されるからこそ、学びが生じるのである。つまり、真に知識を拡張し、成長させるのは、予想に反する経験なのである。

４－４. フィードバック哲学として見たポパーの思想

---

[30] カール・R・ポパー、『量子論と物理学の分裂』、前掲書、61 ページ。

[31] カール・R・ポパー、久野収、市井三郎訳、『歴史主義の貧困』、中央公論社、1980 年、102-112 ページ。

[32] これは「理論体系のうち反証の原因を特定することはできない」というデュエム・クワインテーゼからの批判に応えてポパーが「理論体系全体が反証される」と述べたことに対応する。K. R. Popper, “Replies to My Critics”, in P. A. Shilp ed., *The Philosophy of Karl Popper*, Vol.II, p.982. Cf., K. R. Popper, *Die beiden Grundprobleme der Erkenntnistheorie*, J. C. B. Mohr, 1979, p.262.

ポパーの思想は、トップダウンかボトムアップかで、従来の哲学思想とはまったく異なっている。

伝統的な哲学は、どれもトップダウンの宣託的である。アルキメデスの点(不動の真理)からトップダウンにすべてを説明・解釈する（図 12 参照)。だからこそ、伝統的に基礎づけということが中心課題になってきた。

図 12

このようなトップダウンの哲学では、現実からの不都合なフィードバックがあっても、無害化して自らの体系に組み込んでしまうことになりがちである。だから、現実に起こっていることはすべて自身にとってのポジティブフィードバックになってしまう。このような哲学においては、ポパーが言うように、自説を支持する事例、つまり実証（Verification）を得るのはたやすいのである[33]。

このように伝統的な哲学はトップダウンで演繹的である。ここでは成長が止まり、現実から乖離していく。現実から乖離しても、主観主義という安全地帯に逃げ込めばいいということになる。

以上のような哲学を批判して、ポパーはボトムアップのフィードバック、とくにネガティブフィードバックに注目せよという（図 13 参照)。

図 13

伝統的哲学の中で、ポパーの思想ほどボトムアップのフィードバックに着目した哲学は、ほかにないだろう[34]。よく帰納主義に対して、ポパーの方法論は演繹主義的だと言われるが[35]、このように見ると必ずしもそうとは言えない。むしろ、ポパーの哲学は経験からのフィードバックを重視するという意味ではボトムアップ主義である。経験がもたらす効果を十分に汲みつくすという意味では、ポパーの思想こそ真の意味で経験主義である。つまり、ボトムアップの経験主義なのである。

５．生成 AI の光と影

５－１．フィードバックの宝庫としての生成 AI

生成 AI が飛躍的に進化したのは、アテンションとバックプロパゲーションによって効率的なフィードバック機構を備えたからである。これまでのコンピュータ利用では、実行結果のフィードバックは主として人が受け取り、それを評価してしていたが、生成 AI 自身がフィードバックを受け取って、自己を変えられるようになった。そうして、人類がこれまで得ることができなかった知的なフィードバックが得られるようになった。

その生成 AI からのフィードバックによっ

---

33 ポパーは、たんに支持するだけの事例は、原則として安っぽくて持つに値しないと主張する。カール・R・ポパー、小河原誠、蔭山泰之、篠崎研二訳、『実在論と科学の目的』、岩波書店、2002 年、184-185 ページ。

34 ポパー自身が若いころ、マルクス主義の流血事件というネガティブフィードバックから強烈に学んでいる。カール・R・ポパー、『果てしなき探求』、前掲書、上巻、55-57 ページ。

35 ミラーは、反証主義を仮説演繹主義というより仮説-破壊主義(hypothetico-destructivism)だとさえ言っている。D. Miller, *Critical Rationalism: Restatement and Defense*, Open Court, 1994, p.111.

て人類も飛躍的に進歩できる可能性がある。ポパーが三世界論で論じたように、人類は自ら生み出したものの予期せぬ帰結によって、進歩を遂げてきた。近代以降の科学技術にはそうした側面がある。現代の自然科学は、言うまでもなく、コンピュータなどの技術なしでは成り立ちえない。科学は人間がつくり出した技術の助けによって飛躍的に進歩してきた。つまり人類が生み出した科学技術との相互作用により、人類の知的レベルは飛躍的に進歩してきたのである。アインシュタインは「私の鉛筆は私よりも賢い[36]」と述べたが、生成 AI は鉛筆をはるかに凌駕する、人類が初めて手にした人間以外の知的存在である。いわばフィードバックの宝庫であり、生成 AI を得て人類は新しい局面に入った。

５－２．ハルシネーションの問題

一方で、生成 AI の宿痾として残り続けるかもしれないのがハルシネーションである。生成 AI が学習を積み重ねて行けば、ハルシネーションは徐々に減っていくだろう。しかし、完全に撲滅できるだろうか。ハルシネーションが完全になくなったことはどのように保証できるのだろうか[37]。

おそらく、それだとは判別できないようなハルシネーションが残ってしまうのではないだろうか。たとえば、「第二次大戦の終結年は 1947 年です」くらいのハルシネーションなら誰でもわかる。けれども、「甲氏は乙財団の資金を不正に流用した」というようなハルシネーションは判別しにくいだろう[38]。すでに誤りを見つける役割の AI もあるが、AI が AI の誤りを見破ったことを人は判別し続けられるだろうか。そもそも判別できないハルシネーションは、ハルシネーションと言えるのだろうか。

ハルシネーションが問題なのは、これが機能的な不具合ではなく、通常の動作の一種だということである。機能的な不具合であればその具体的な原因を特定できるが、生成 AI でハルシネーションを避けるためには、後で触れるように学習データを変えて再学習させるしかない。

今後、生成 AI の性能、能力はますます向上し、人はますますこれに依存するようになるだろう。だが、人には判別できない誤りが人間にとって重大な決断にかかわるような場合、生成 AI をミッションクリティカルなシステムの根幹部分に活用しても大丈夫なのだろうか。今後、致命的な事態が生じてしまった場合の責任の所在が、ますます特定しにくくなっていくと思われる。

５－３．ブラックボックスとしての生成 AI

これまでのコンピュータは、原則としてホワイトボックスであった。コンピュータの実行結果結果を人が不具合と判断すれば、その原因を調査して特定し、誤りを直接修正できた。もちろん、不具合がないことは保証できないが、従来のプログラムでは、不具合の原因を特定して修正することが原理的に可能である（図 14 参照）。

コンピュータに限らず、現在の技術トレンドとしては、不具合の原因を特定できるように複雑さを分割統治の手法で管理しやすくするようになっている。このため、独立した機能モジュールを疎結合したアーキテクチャであるモジュール型設計が主流となって来ている[39]。

---

[36] K. R. Popper, *Knowledge and Body-Mind Problem*, op. cit., p.31.

[37] これは、真理に到達したかどうかは決してわからないというのと同じと思われる。

[38] 実際にこうしたハルシネーションのために訴訟が起きたことがある。

[39] たとえば、長谷川洋三、『自動車設計革命』、中央公論新社、2012 年。ソフトウェア開発では、モジュールの凝集度を高めて、モジュール間の結合度を低くするように設計すべきことは、ずっと以前から言われている。グレンフォード・マイヤーズ、『ソフトウェアの複合/構造化設計』, 近代科学社, 1979

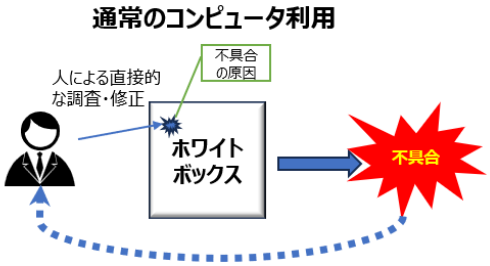

図 14

一方、生成 AI は実態としてブラックボックスで、その中身を直接触ることはできず、再学習により間接的に修正できるだけである（図 15 参照）。

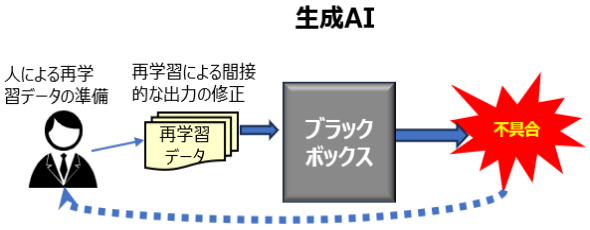

図 15

このように、ニューラルネットワークは、先に触れたモジュール型設計の技術トレンドに反して、複雑になっていく一方である。

こうした生成 AI はウィーナーが次のように言って危惧した状態に近づいているようにも思われる。「もしフィードバックが機械にそっくり仕組まれていて、最後の目標が達成されるまでに検出できないようになっていたら、破局に到達する可能性は著しく大きくなってしまう[40]。」つまり、処理途中の個々のフィードバックが AI に横取りされていて、人には最終フィードバックしか見えないのである。ユートピア工学はブラックボックスなので、前進しているのか、後退しているのか、わからないが、現在の生成 AI もこの意味でユートピア工学に似た側面がある。

５－４．生成 AI からのネガティブフィードバック

以上見てきたように、世界３の代表的な住人である生成 AI には光と影がある。生成 AI の活用が多方面で進む一方で、その危険性に警鐘を鳴らす動きもある。

しかしポパーが、人口問題や食糧問題など現代のさまざまな問題について、科学の進歩で解決されるだろうと楽観的に述べた[41]のと同じく、生成 AI の進歩、発展については楽観視することができて、正の側面を伸ばせると考えられる。つまり、生成 AI からのフィードバックによって人類が飛躍的に進歩できる可能性を伸ばすことで、負の側面も克服できるだろうと考えられる。あるいは、そう期待したい。

ただ、その可能性を実現するためには、われわれが生成 AI からくるネガティブなフィードバックをしっかりと受け止めて、これを糧にして前に進んでいく必要がある。さもなければ、つまりネガティブなフィードバックに対して目を閉じてしまっていては、結果的に負の側面が助長されることになってしまうだろう。それゆえ、「どうしてそうなるのかわからない」というネガティブフィードバックを解明していくことこそが肝要だろう。「わからないけど役に立つからそのままでよい」という考え方は、ポパーが危惧していた考え方であり、やはり危ういと言えるのではないだろうか。

年。

40 ノーバード・ウィーナー、鎮目恭夫訳、『科学と神』、みすず書房、1965 年、64 ページ。

41 ポパーは楽観主義者であり、未来に責任を持つために楽観主義者であることが義務であるとさえ述べている。カール・R・ポパー、ポパー哲学研究会訳、『フレームワークの神話』、未來社、1998 年、16-17 ページ。

生成 AI を理解できるか？　大規模言語モデル(LLM)のエンジニア観点での理解

永島秀文（TÜV Rheinland）

chatGPT に代表される生成 AI、あるいは大規模言語モデル（LLM）が社会現象となっている。ここでは LLM の特徴とモデルの位置づけを概括してから、その人工知能への哲学的な批判とその解釈からの理解を試みる。

生成 AI の理解という問いは大きな問いである。生成 AI の中でも LLM に焦点をあて、三段階に分けて考えていきたい。

Q1）人はその動作原理を理解できるか

Q2）人はその出力を理解できるか

Q3）人はその意図を理解できているか

始めの問い Q1 については技術的な始まりを考えるのが適切であろう。

現在の AI ブームは 2022 年に始まったと考えることができる。1990 年代に用意された Deep Learning（深層ニューラルネットワーク）が、自然言語処理に関して大きなブレークスルーを成し遂げた。それが今回のテーマである LLM である。

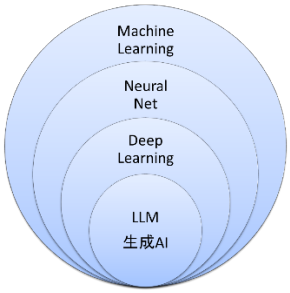

図 1　人工知能の包含関係

ブラックボックス化というのはエンジニアリングの用語であるが、中身が不透明でその作動原理の詳細を扱わないで開発する等で使われる。深層学習からしてすでにブラックボックス化は始まっていた。LLM ではさらに「汎化のミステリー」が加わり、学習していない未経験の事象まで予測できるとされる。即ち、ダグラス・ホフスタッターが 2000 年に主張したように『実際に自ら思考するマシンを、いずれ我々は創造することになるでしょう。しかしながら、そうなったときには、そのマシンを我々自身が設計することもなく、その作動原理についてのごく漠然とした知識すらもつことができなくなるに相違ありません』という事態にあるようだ。

当初、LLM は機械翻訳から出発した。それは、「会話する人工知能」という狙いから開発されたのではなく、日英翻訳のように逐次的な単語の出現確率を予測するという単純な機能が土台になっている。

もちろん、それが日常生活でも十分に使えるレベルになっているのは、驚くべき達成であることは間違いない。

図 2　「持つ」動作に対応する中国語

機械翻訳が「使える」というのは、日本語の「持つ」に対応する中国語を図 2 に示すような文脈による使い分けができてこその満足感なのであろう

LLM はソフトウェアの一種である。ソフトウェア工学（ものづくり）の視点から、その差異を列挙してみる。

設計の特徴　従来のソフトウェアは設計者の実現したい機能とアルゴリズムを明確に指定する。それに対して、LLM は仕様が曖昧であり汎化能力のような思わぬ副産物がある。アルゴリズムもその作動原理の多くは理解が困難である。

内部パラメータ　設計者は内部パラメータとその意味を理解して決定する。その数も把握できるオーダーである。これに対して LLM は数億から数兆に及ぶ内部パラメータ

がある（表１）。それは学習時に自律生成され、設計者には理解不明といっていい状態である。

LLM製品	公開年	パラメータ数
GPT-1	2018	120,000,000
GPT-2	2019	1,500,000,000
GPT-3	2020	175,000,000,000
PaLM	2022	540,000,000,000
GPT-4	2023	（非公開：数千億～数兆）

表１　LLM の内部パラメータ数

以上のように LLM は設計者にとってブラックボックスといっていい。つまり、Q1 に関しては No となる。

次なる問い Q2 についてはどうか？

LLM の出力を理解するとは「人と区別できない」出力をするという判定基準が計算科学の創始者アラン・チューリングによって唱えられていた。チューリングテストである。

チューリングテストとは、

- 人と AI の問いと応答だけを評価するブラックボックステスト
- 対話相手を「人間」と見なせる（AI と人間を区別できない）なら合格となる。
- 「相手が人間的に応答している」ことを「AI を理解した」と見なす。つまり、ヒトの考えを理解して、それにヒトと同様な応答をしているので AI はヒトと同じように「考える」とヒトが判定できる。

しかし、管見によればチューリングテスト自体は現時点の LLM に対して真面目に論じられることはなくなったようである。LLM の出力は人と同格であるとしても異論がないレベルになったと考えられる。

それはローブラ賞という当該テスト合格への活動が誰も取り合わなくなったことにもうかがわれる。Q２への回答は Yes としていいであろう。

見逃せない点はチューリングテストに対して哲学者が批判を行ったことである。その批判は Q3 に関わってくる。

ジョン・サールの「中国語の部屋」がその批判である（図 3）。

図３　中国語の部屋

中国語をまったく解さない人物ジョンが、書類の出し入れ窓口しかない部屋にいる（外部環境から遮断されている）

中国語の文を受け取ったジョンは中国語を適切に変換するマニュアルにより、中国語の文を作成する。ジョンは意味がまったくわからない。

そのマニュアルは英語で書かれた完璧なガイドであり、どのような中国語に対しても適切な中国語を生成できる（調べる手間や所要時間は問わない）

この「仕組み」は果たして中国語を理解していると言えるか？

ジョンが仕組みの主役であれば、答えは No になるだろう。この思考実験に対して、二つの批判があった。

実現問題としてのルベックの批判。マニュアルは膨大過ぎて実現不可能だとする。しかしながら、LLM はそれを実現してしまった。しかも、ジョンは不用になった感がある。LLM は中国語を理解しているかどうかは次の批判が関係するだろう。

記号接地問題として知られるハーナドの批判である。単語の意味を遡ると意味を還元できない単語に突き当たる。青りんごは「青」と「リンゴ」知覚や経験になしに意味理解があるのか？

LLM はあたかも青い、リンゴを体験して知っているかのように回答できる。だがあきらかに LLM は言語処理しかしていないため、現実の青いリンゴを分かっていないはずである。

人は青いリンゴに関して会話する時、相手

がそれを体験していることが前提としているのではなかろうか？

つまり、お互いの意図の背景に体験や情念を前提として人は会話するという観点からすると、LLM はどうやら限定された人間理解のもとでのコミュニケーションを行っているらしい。図 4 の赤枠内を模擬していると考える。泳いだこともないのにおぼれた経験を語り、指導するのが LLM である。

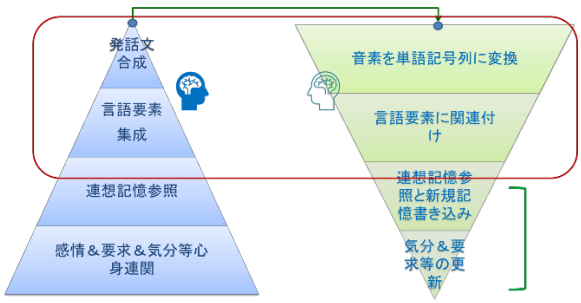

図 4　人間のコミュニケーションの描像

記号接地問題が明らかにしたのは、体験に基づかない記号処理が LLM の作動原理であるなら、「人の理解」をしないまま応答している事態であろう。痛みを知らずして痛みを語るのは「巧妙な模倣」ではないだろうか。

もちろん、通常の社会活動や生活での援用では LLM はなんら支障をきたさないだろう。しかし、相手の心中を理解していない状態は認知科学でいう「心の理論(TOM)」の欠如を示しているかもしれない。概括的には図 4 の右下の括弧部分が相当する。TOM を持つための必要条件は身体を持つことが本質的であるためである。LLM は人の心中で起きていることを推察できないと思われる。

Q3 については No となろう。

本稿では限定的にしか扱えなかったが、こうした迂遠とも思われる哲学的な問いと回答を精密化していくことは LLM の実用化において不可欠である。LLM の尤もらしい虚偽の答えをする傾向がある。いわゆるハルシネーション（妄想）問題が指摘されているが、それ以外の限界や想定外の振る舞いやその影響を適正にモニターする必要がある。

LLM は強力なテクノロジーであり、それを統制することは政府や研究組織、企業の務めである。その急速な普及は社会のあり方を大きく揺るがす可能性が大きい。

無制限、無分別、無思考的な LLM の社会への浸透と拡大はリスクが大きく、利益本位の急速な浸透は社会を揺るがす可能性がある。ポパーのピースミール社会工学的なガバナンスを必要とすると考える。

現在、国際標準化機構(ISO)やアメリカ国立標準技術研究所(NIST)では AI や生成 AI について各種のマネジメントシステムや規定を開示している。

- ISO/IEC42001 Information technology Artificial intelligence Management system
- ISO/IEC TS4213 Information technology Artificial intelligence Assessment of machine learning classification performance
- NIST AI-600 Artificial Intelligence Risk Management Framework

などが知られている。このような国際標準の適用もピースミール社会工学の一助となり、AI 関連製品の開発や運用の現場で十分に考慮すべきと考える。それは一方通行の適用ではなく、事例の蓄積を通じて規格自体の改善や社会への貢献に向けた漸進的イノベーションに寄与することが望まれる。

【参考文献】

1)　丸岡 章，概説 人工知能，筑摩書房（2024）

2)　岡留 剛，概説 生成 AI の基礎，共立出版（2024）

3) Alan D. Thompson , AI-Bubbles https://lifearchitect.ai/

4)　D.ホフスタッター&D･C･デネット編，マインズ・アイ　所収、サール,心・脳・プログラム,TBS ブリタニカ(1984)

5)　今泉允聡,深層学習の原理に迫る,岩波書店（2021）

6)　坪井、海野、鈴木、深層学習による自然言語処理，講談社(2017)

7) S.ウルフラム,ChatGPT の頭の中,早川書房（2023）
8) 小河原 誠,ポパー〔第２版〕,筑摩書房(2024)
9) 永島秀文, ISO 26262:2018 の概説ーシステマチック故障を中心としてー安全工学会誌，Vol.60，No.1（2021）
10) ハーバート・A. サイモン,システムの科学, パーソナルメディア（1999）
11) 岡野原大輔, 大規模言語モデルは新たな知能か,岩波書店(2023)
12) 今井むつみ,ことばと思考，岩波書店（2010）
13) アセモグル＆ジョンソン,技術革新と不平等の 1000 年史，早川書房（2023）
14)トマス・リッド, サイバネティクス全史, 作品社（2017）
15) ISO/IEC TS 4213,Information technology —Artificial intelligence Assessment of machine learning classification performance, ISO(2022)
16) ISO/IEC42001, Information technology — Artificial intelligence Management system, ISO(2023)
17) NIST AI 600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile draft, NIST (2024)
18)サイモン・バロン＝コーエン,自閉症とマインドブラインドネス,青土社(2002)

法的シンギュラリティ

宇佐美 誠（京都大学）

1. 文脈と主題

1.1. AI と法の 3 つの関係

人工知能（AI）が急速に発展し普及するのにともなって、法との関係で多様な諸論点が提起され検討されている。この諸論点は 3 つに大別できる（宇佐美 2023: 181-184）。第 1 は、AI への法適用（law applied *to* AI）と呼ぶべきものに関わる。例えば、自動運転車が交通事故を起こした場合、不法行為責任を誰に帰するべきか。ビッグ・データを収集し利用するに際して、個人のプライヴァシーをどのように保護するか。AI が発明を行ったならば、知的所有権を認めるべきか。ここでは、既存の法律・判例を AI にも適用するべきかという解釈論のみならず、むしろ AI に適用されるべき法律はいかなる内容をもつべきかという立法論がいっそう問われている。

第 2 は、AI での法適用（law applied *with* AI）をめぐる諸論点である。司法において、AI のいかなる利用が許されるか、あるいは望ましいか。その長所は何か、短所は何か。実際、諸外国の司法では AI がすでに様々な仕方で利用されている。日本で紹介・検討されてきたのは、アメリカの予測的司法である。カリフォルニア州・ニューヨーク州などでは、裁判官が保釈判断等を行う際に、再犯の確率を予測する統計的自動判断プログラムが用いられてきた。その代表例は、代替的制裁のための矯正的加害者管理プロファイリング（COMPAS）である。だが、司法での AI 利用は、アメリカに特有な現象ではない。カナダのブリティッシュ・コロンビア州では、5000 ドル未満の少額請求について、オンライン紛争解決（ODR）を行う民事紛争解決法廷（CRT）が運用され、そのなかでアルゴリズムが使用されている。メキシコでは、裁判官が年金請求者の受給資格の有無や適切な年金額を判定する際、EXPERTIUS と名づけられた意思決定支援プログラムが活用されてきた。

ヨーロッパに目を移そう。イギリスでは、ロンドンを除くイングランドとウェールズにおいて、交通違反切符に対する異議申立の当否をオンライン上で判断する交通違反金法廷（TPT）が運用されている。他方、2023 年の「司法官職保有者ガイドライン」では、法的調査や判決作成での AI 利用を推奨しないと明記された[1]。ドイツでは、連邦司法省が 2024 年に「民事訴訟におけるオンライン手続の発展・試用法案」を公表した[2]。指定された地方裁判所において、5000 ユーロ未満の少額訴訟で 10 年間、オンライン手続が試用される予定である。

アジアにおいては、中国での AI の急速な普及が刮目に値する。一部の人民法院（裁判所）においては、意思決定支援プログラムが裁判官に類似の諸事案を参照するよう誘導し、また他の人民法院では、類似の諸事案と大きく異なる判決案を裁判官が入力すると、プログラムが警告を発する。刑事訴訟での被告人尋問に際して、その態度の微妙な特徴をもとに発言の真偽の蓋然性を推計するシステムも、すでに試用されている。訴訟のスマ

※ 本稿は、日本ポパー哲学研究会第 34 回年次研究大会のシンポジウム「AI と認識論：技術・知識・社会の共進化」（東京）での報告に敷衍と更新を加えたものである。その初期の版を、JST-RISTEX 研究開発プロジェクト（JPMJRX21J1）の研究会合（オンライン）で発表した。また、本稿の内容の一部を、第 11 回経営学・法学・経済学国際会議（アテネ）での報告において扱った。これらの機会に示唆に富む質問・コメントを下さった多数の方々にお礼を申し上げたい。本稿の研究は、上記プロジェクトの中間成果の一部である。

[1] Courts and Tribunals Judiciary, “Artificial Intelligence (AI): Guidance for Judicial Office Holders,” 2023. https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf

[2] Bundesministeriums der Justiz, „Entwurf eines Gesetzes zur Entwicklung und Erprobung eines Online Verfahrens in der Zivilgerichtsbarkeit“. https://www.bmj.de/SharedDocs/Downloads/DE/Gesetzgebung/RefE/RefE_Erprobungsgesetz_Zivilprozess.pdf?__blob=publicationFile&v=3

ート化をいっそう推進するため、最高人民法院は 2022 年、司法府を支援する有能なシステムを 2025 年までに設立し、いっそう効果的なシステムを 2030 年までに確立するという目標を公表した。

イギリスのガイドラインやドイツの法案は、AI での法適用の典型例である。他方、COMPAS や EXPERTIUS では、限定された場面において、裁判官がアルゴリズムに法的判断をほぼ委ねているように思われる。さらに、中国の 2 種類のプログラムは、範囲を限定することなく裁判官がアルゴリズムにかなりの程度まで依存していると解される。これらは、AI での法適用からさらに先へ進み、私が AI による法適用（law applied *by* AI）と名づける第 3 の範疇に近づいている。AI による法適用は、AI での法適用と連続的である。しかし、法的判断がアルゴリズムに委ねられる前者においては、後者よりも先鋭化した形で、AI 利用の当否が問われることになる。

AI による法適用における鍵的概念は、ロボット裁判官である。ロボット裁判官は通常、人間に代わって訴訟指揮を行い判決を下す AI として理解されているが（e.g., Ulenaers 2020）、この理解はやや狭隘にすぎる。敷衍しよう。AI が人間から職業機会を奪って大失業を招くか否かという周知の論争では、AI が労働者との関係でもちうる 2 つの機能が注目されてきた（参照、宇佐美 2020: 88-94）。AI はおもに労働者の業務を補完する（supplement）と考えるならば、大失業の可能性を否定するいわゆる楽観論に帰着しやすい。他方、AI はしばしば労働者に代替する（substitute）と考えると、大失業の可能性を肯定する悲観論へとつながる。このような大失業の可能性の有無という量的争点でなく、個々の業務遂行の如何という質的論点を考えるときには、補完と代替の中間に、AI が労働者を従属させる（subjugate）という機能を位置づけられる。例えば、ある生命保険会社が、営業員をみな解雇して、オンラインでの保険料の自動見積プログラムを運用し始める場合、営業員は AI に代替されたのである。それに対して、営業員は雇われ続けているが、新規契約の勧誘から、保険料の見積、さらには保険内容の更新の助言にいたるまで、プログラムの提案にほぼ自動的に従っているならば、AI に従属していると言える。人間裁判官が法廷から去るという代替の状態のみならず、AI の提案をほとんどつねに受け入れて判決を下すという従属の状態も、ロボット裁判官の範疇に含めるのが適切だと考えられる。

1.2. 法的シンギュラリティ

AI による法適用をめぐって、英語圏の法学界で関心が近年急速に高まっている概念がある。法的シンギュラリティ（legal singularity）である。その定義は論者によって大きく異なる。ベンジャミン・アラリーら（Aidid and Alarie 2023: 3）は、「ほぼすべてのタイプの法的不確実性に対して、即時に需要に応じて取り組み、解決を与えることができる安定的かつ完成的な法秩序」と定式化する。また、「ほぼいかなる法的問題も瞬時かつ正義にかなった解決を得るだろうという意味で、法が機能的完全性に達するだろうという考え」（8）であるとも言う。ここでは、不確実性の克服や解決の即時性が、法的シンギュラリティの核心とみなされている。別の論者（Goldsworthy 2019: 286）は、正しい判決がつねに下される司法の状態を予言して、「発達した深層学習システムの到来は、あらゆる法的問題への単一の「正」解を見出すことができるだろう」と述べる。

これらの諸定義はやや夢想的だと私には思われる。より現実主義的な定義を考案するための第一歩として、技術的シンギュラリティの定義を確認したい。I・J・グッドからヴァーナー・ヴィンジを経てレイ・カーツワイルにいたるまで、多くの論者が技術的シンギュラリティの到来を予言してきた。これは標準的には、人工超知能（ASI）、すなわち人間

を超えた知能をもつ AI が出現する仮定上の将来の時点だとされる。ここでは、特定の領域で知的業務を行う特化型 AI でなく、人間が担ってきたあらゆる領域の知的業務をこなす汎用型AIが想定されている。その上で、人間の最高の知能水準さえも超える ASI の出現が予言されているわけである。

技術的シンギュラリティの定義を参考にしつつ、現実主義的視点から、法的シンギュラリティを、AIが平均的裁判官を超えた能力をもつ仮定上の将来の時点として定義しよう。この定義には 2 つの特徴がある（cf. Volokh 2019: 1138–1140）。第 1 は、あらゆる領域の知的業務でなく、訴訟指揮や判決作成に視野を限定するという領域限定性である。この定義では、汎用型 AI がたとえ実現しえなくても、特化型 AI による法的シンギュラリティは可能である。第 2 の特徴は、最も優秀な裁判官でなく平均的裁判官を超える水準を想定するという水準穏当性である。このように定義された法的シンギュラリティは、技術的シンギュラリティと比較して、到来の蓋然性がより高く、また到来する場合にはその時点がより近いだろう。

法的シンギュラリティの評価をめぐっては、激しい論争が展開されている。一方には、その到来を称揚するか歓迎する擁護者がいる（Alarie 2016; Goldsworthy 2019; Aidid and Alarie 2023）。他方、おもに法の支配を根拠として、この時点の到来を非難したり警戒したりする批判者は、決して少なくない（e.g., Brownsword 2020; Deakin and Markou 2020; Weber 2020）。しかし、既存の論争には、看過できない間隙が見られる。法的シンギュラリティを的確に評価するためには、人間による判決と AI のそれとの比較が不可欠だと考えられる。人間の判決に仮に重大な欠点が潜み、AIがその欠点に対して効果的に対処できるならば、この事実は、法的シンギュラリティを肯定的に評価する重要な一理由となるはずである。そして、アメリカをはじめ諸外国には、判決の種々の欠点を剔抉する多数の科学的研究や統計的調査が蓄積されている。にもかかわらず、これらの知見を活用した人間と AI の比較は、従来の法的シンギュラリティ論争では皆無に近い。

このような先行研究上の重大な間隙を埋めるため、本稿は、人間の判決がはらむ難点に関する科学的・統計的知見を参照した上で、AI はその難点を解消または緩和できると論じる。そうすることにより、新たな観点から法的シンギュラリティの評価を試みたい。次節以降では、まず予備的考察として、いかなる技術がロボット裁判官を可能にするか、またどのような条件の下でロボット裁判官の判断が信頼に値するかを確認する（2.）。次に、人間の判決に隠れた種々のバイアスとノイズを明るみに出す諸研究を概観する。その上で、AIはバイアスを予防または軽減し、ノイズを消去することを指摘する（3.）。最後に、残された課題に言及するとともに、法的シンギュラリティ後の司法のあり方を展望する（4.）。

2. AI 技術とその利用範囲

2.1. いかなる AI 技術か

法的シンギュラリティが近未来に到来するならば、それはいかなる AI 技術によって可能となるか。この問いに答えるための第一歩として、法的推論の基本構造を確認しておきたい。法的推論の中心をなす法的三段論法では、大前提をなす全称の法規範命題に、小前提となる特称の事実命題を包摂することによって、結論が導出される。例えば、「人を殺した者を、死刑または無期もしくは 5 年以上の懲役に処するべきである」（参照、刑法第 199 条）に、「浅田太郎は石川花子を殺した」を包摂することによって、「浅田太郎を 7 年の懲役に処するべきだ」が導かれる。

裁判官は、事案に適用されるべき法規範命題を発見するとともに、事実認定を通じて具体的な事実命題を確定した上で、前者への後

者の包摂を通じて結論を導出する。中国に見られる通り、事実命題を確定する過程において、AI は特定の場面で補助的に試用されている。また、法規範命題を発見し、その適用指針を把握するため、現下の事案に類似した裁判例を探索する際、AI がすでに広く活用されている。中国での 2 種類の意思決定支援システムが、これに該当する。また、アメリカ・イギリスなどのリーガルテック企業は、法律事務所向けに、AI を用いた裁判例の情報提供サーヴィスを行っている。

次に、AI の発展史を短く回顧しよう。1950 年代後半から 1960 年代まで続いた第 1 次ブーム、1980 年代から 1990 年代にかけての第 2 次ブームを経て、2010 年代以来の第 3 次ブームが現在まで続いている。第 2 次ブームでは、医師・法律家等が行う推論を模したエキスパート・システムが開発された。当時、AI に法的三段論法を行わせる研究が盛んに進められ、今日でも、EXPERTIUS などでエキスパート・システムが用いられている。他方、第 3 次ブームを特徴づけるのは、ディープ・ラーニングによる強化学習（深層強化学習）とビッグ・データである。この組み合わせによって、以前よりも格段に多様かつ精確な情報処理が、いまや可能となっている[3]。

判決は、いかなる AI 技術によって作成されうるか（宇佐美 2023: 187, 189）。主文は、判決予測の技術を用いて作成できるだろう。多数の裁判例を機械に学習させた上で、判決を隠した事案を示して推測させるという判決予測の研究が、長年にわたり重ねられてきた。近年、莫大な数の判決について深層強化学習を行わせた結果、通常は認定事実のみをデータに含めるにもかかわらず、概して 80%以上の正答率が達成されている。ヨーロッパ人権裁判所の他、アメリカ・イギリス・フランス・ドイツ・インド・中国等の裁判所について研究が行われ、その手法はますます精密化し多様化しつつある。裁判例での認定事実とともに法規・判例も AI に学習させ、またアルゴリズムの精度をいっそう向上させるならば、新事案での正答率は、さらに上昇して 100%に近づくだろう。そうなれば、AI に判決の主文作成を委ねることができる。

しかしながら、判決予測の技術のみでは、ロボット裁判官を実現しえない。判決がもつ顕著な一特徴は、詳細な判決理由を提供することにあるから、AI は、適切な判決理由を執筆できなければならないのである。この課題は、2022 年に登場した大規模言語モデル（LLM）の生成 AI によってやがて達成されるだろう。AI に膨大な数の法令や裁判例を学習させた上で、唯一無二の新事案について、適切な法規を適用し、また類似の裁判例と整合した判決理由を執筆させるわけである。ロボット裁判官の判決案が平均的裁判官のそれと比べて遜色がない段階にいたれば、人間裁判官は、その判決案を確認した上で、特に必要な場合にのみ判決理由に手を加えればよい。これは、法的三段論法における法規範命題の同定と結論の導出を AI にほぼ委ねることを意味する。さらに、AI が心証形成を行って事実命題を確定できるようになった暁には、法的三段論法の全段階をほとんど任せられる。ロボット裁判官の出現である。

2.2. ロボット裁判官はいつ信頼できるか

法令・裁判例のビッグ・データの深層強化学習がロボット裁判官の技術的基礎だとすれば、その判決案が信頼に値するためには、データ・サイズが決定的に重要となる。この厳然たる技術的条件から、3 つの含意が導かれる（参照、宇佐美 2023: 188）。第 1 は社会

[3] 生成 AI の登場は第 4 次ブームを画するという見解もある（参照、総務省 2024）。だが、AI の研究開発史を、社会的耳目や経済的投資の波でなく AI 技術の種類の動態的変遷という観点から捉えるならば、ビッグ・データの深層強化学習に基づく生成 AI は、第 3 次ブームの一環として位置づけられるべきだろう。

の規模に関連する。他の条件を一定とすれば、ある法体系内に居住する全人口が大きいほど、また紛争・犯罪の件数が多いほど、判決数が増えるから、AI の判決案はいっそう信頼に値する。第 2 は法体系の成熟度に関わる。他の条件が一定であれば、近代法の歴史が長く、法化が進んでいるほど、法令数・判決数がより多いから、AI の提案をいっそう信頼できる。第 3 は事案の種類に関係している。法哲学においては、ロナルド・ドゥウォーキン（Dworkin 1978: Ch. 4）以来、民事訴訟における容易な事案と困難な事案という区別が、広く用いられてきた。ドゥウォーキンによる定義から離れ、法的三段論法にそくして再定式化するならば、容易な事案とは、適用するべき法規範命題を困難なしに同定できる事案であるのに対して、困難な事案とは、法規範命題をただちに同定できない事案をさす。民事訴訟か刑事訴訟か行政訴訟かを問わず、容易な事案は裁判例の大多数を占めるだろうから、AI の判断を信頼してよい。他方、困難な事案は、成熟した法体系では少数にとどまるはずだから、AI の判断を信頼することができない。

第 3 の含意は、法的シンギュラリティをめぐる論争に対して重大な意味合いをもつ。この仮想的時点の提唱者たちは、あらゆる事案において、ロボット裁判官が瞬時に紛争を解決し、不確実性を消去し、正解を提供すると想定している。あらゆる事案という想定は、批判者たちの間でも共有されている。このような法的シンギュラリティ観を全面説と呼ぶことができる。容易な事案でのみロボット裁判官の判断が信頼に値するという上記の含意は、全面説がじつは根拠を欠くことを意味する。むしろ、大多数を占める容易な事案ではロボット裁判官に委ねてよいが、少数の困難な事案ではそうでないとするいわば大半説こそが、採用されるべきである。

3. バイアスとノイズ

3.1. 判決に潜むバイアス

判決を含む人間の判断がはらむ難点としてただちに想起されるのは、バイアスだろう。バイアスとは、形式的に定義すれば、諸判断が特定方向へと向かうパターン化された望ましくない傾向である。キャス・サンスティーン（Sunstein 2022: 1179）は、行政上の意思決定をアルゴリズムに委ねるべきだと論じるなかで、C バイアス（認知バイアス）と D バイアス（差別バイアス）を区別している。

行動経済学では、人間の認知におけるじつに多様なバイアスが発見され計測されてきた。最初に得た情報に過度に依拠するアンカリング効果、想起しやすい情報から過大な影響を受ける利用可能性バイアス、小規模サンプルが母集団を代表していると誤解する賭博師の誤謬などが知られている。司法上の C バイアスは、例えばアメリカへの亡命の可否を決する移民裁判所で観察されている。裁判官がある事案で亡命申請を認容すると、直後の事案では棄却する確率が増加し、認容が 2 件続くと、棄却の確率はいっそう高まる（Chen et al. 2016: 1193-1205）。ここでは、確率認知に関する賭博師の誤謬に似た心理的メカニズムが、事理判断において生じている。

他方、D バイアスとは、性別、人種・民族、精神的・身体的な障碍、性的な嗜好・同一性などに基づいた特定集団への偏見的な処遇として現れる、諸判断の特定方向への偏りだと定式化できるだろう。アメリカ連邦裁判所の被告人のうち、黒人男性は、類似の状況にある白人男性と比べて 23.4%保釈を認められにくく、ヒスパニック系女性は、白人女性より 29.7％も保釈されにくい（U.S. Sentencing Commission 2023）。また、黒人の被告人に科される刑期は、白人へのそれと比べて平均 19 カ月も長く、ヒスパニックへの刑期は、白人へのそれよりも 5 カ月長い

（Topaz et al. 2023）。

判決に表れる偏向は、C バイアスと D バイアスに限定されない。いわば E バイアス（情緒バイアス）と F バイアス（党派バイアス）にも目を向ける必要がある（宇佐美 2023: 191-192）。E バイアスは、快・不快などの時々刻々と変転する情緒的状態に起因する。著名な研究として、いわゆる空腹な裁判官がある（Danzigera et al. 2011）。この研究は、イスラエルにおいて刑務所収監者の仮釈放を決する、裁判官ら 3 名からなる判定団を対象としている。1 日に 3 回開催される会議で、午前の冒頭に扱われる収監者の仮釈放率は約 65%だったが、その後は次第に低下してゆき、昼食休憩の直前にはほぼ 0%まで下落し、しかし休憩直後には再び 65%まで回復した。午後の最初の会議でも、仮釈放率は同様に下落してゆき、しかし間食の休憩後には再び回復した。

F バイアスは、政治的立場・道徳的信条などの安定的信念に由来する。アメリカ連邦裁判所において、共和党指名の裁判官は民主党の裁判官と比べて、黒人の被告人に対しては類似の非黒人よりも平均 3 カ月長い刑期を科する（Cohen and Yang 2019）。また、女性の被告人には男性よりも平均 2 カ月短い刑期を言い渡す。共和党系裁判官による黒人への特に苛酷な量刑は、人種的偏見が政治的立場によって強化されていることを示している。他方、女性への温情的量刑は、女性が脆弱であり有利な扱いを必要とするという保守的・伝統的なラベリングの帰結だと考えられる。最近の一論文（Vilaça et al. forthcoming）は、別種の F バイアスを浮き彫りにしている。その分析対象は、ブラジルの国営石油企業をめぐる南米史上で最大規模の汚職に対するいわゆる洗車場作戦である。一連の贈収賄事件において、裁判官の政治的立場による量刑の差は、ほとんど見られなかった。しかし、有罪の被告人のうち、政治家は類似の企業関係者と比べて、平均で 73%長い刑期や 154%重い罰金を科されていたのである。論文の著者たちは、これを反政治階級バイアスと呼んでいる。

判決上の 4 種類のバイアスは、いずれも正義に違背する。正義には比較権衡と相関権衡という 2 つの次元があると、私はたびたび論じてきた（例えば、宇佐美他 2019: 14-16）。比較権衡とは、過去の行為や現在の属性について重要な点で類似した諸個人を類似の仕方で処遇することである。他方、相関権衡とは、過去の行為や現在の属性に相応した仕方で各人を処遇することをさす。法の下の平等は、比較権衡の 1 つの具体化であり、また犯罪と刑罰の均衡は、相関権衡を刑事訴訟の文脈で具現している。バイアスのある判決は、比較権衡と相関権衡をともに損なう。例えば、被告人の黒人男性は、類似の状況にある白人男性と同程度の刑期を科されるべきであり（比較権衡）、また犯した罪に見合うよりも重く罰せられるべきではない（相関権衡）。したがって、正義の理念は、判決上のバイアスを消去するか、少なくとも縮小することを求める。

3.2. 判決におけるノイズ

バイアスと同様に深刻な判断の難点として、ノイズがある。ノイズとは、諸判断が収束せず分散する望ましくない傾向をさす。この現象は、ダニエル・カーネマンらの影響力ある著作（Kahneman et al. 2021）によって広く知られるようになった。類似の事案における判決上のノイズは、正義の 1 つの次元である比較権衡に違背し、法の下の平等から逸脱している。

アメリカでは半世紀にわたって、刑事訴訟で裁判官によって量刑が著しく異なるという現象が、学術研究の対象となり、制度改革の焦点ともなってきた。マーヴィン・フランケルは、量刑の恣意性を告発し制度改革を提唱した先駆的著作（Frankel 1973: 21-22）のなかで、偽造小切手の現金化によって 58.4 ド

ルを騙し取った男が 15 年の刑を科された一方で、類似の事案で 35.2 ドルを取った男は 30 日の刑にとどまったという衝撃的な例を挙げている。その後、裁判官の間にある量刑の顕著な格差が、定量研究を通じて明るみに出されていった（e.g., Clancy et al. 1981）。判決ガイドラインが策定された結果、格差はいったん縮小したが（Anderson et al. 1999）、ガイドラインが勧告に格下げされた後、再び拡大している（Yang 2014）。量刑以外でも、様々なノイズが観察されている。例えば、マイアミの連邦移民裁判所では、コロンビアからの難民申請について、ある裁判官の認容率は 5%にとどまったが、別の裁判官の認容率は 88%に上った（Ramji-Nogales et al. 2010）。こうした状況を、論文の著者たちは難民ルーレットと呼ぶ。

ノイズはバイアスといかなる関係にあるか。両者は論理的に別個だが、しばしば同時に発生するので、ノイズの有無とバイアスの有無を掛け合わせた 2 行 2 列の行列を措定できる。難民認定における賭博師の誤謬に似た偏向や、空腹な裁判官の大きく変動する仮釈放率は、同一の裁判官による諸判断に見られるノイズとバイアスをともに例証している（表 1 の A）。また、共和党系裁判官が民主党系裁判官と比べて黒人の被告人にいっそう苛酷だという傾向は、相異なった裁判官の間でのノイズとバイアスを例示する。それに対して、洗車場作戦での反政治階級バイアスは、バイアスのみを示している（B）。難民ルーレットは、適正な認容率が不明であるから、ノイズのみを表す事例だと言える（Γ）。だが、類似の事案では、バイアスもノイズも許されないはずである（Δ）。

	ノイズあり	ノイズなし
バイアスあり	A	B
バイアスなし	Γ	Δ

表 1　バイアスとノイズ

3.3. 処方箋としての AI

人間裁判官が示すバイアスやノイズは、ロボット裁判官によって防止されるか、あるいは縮減される（宇佐美 2023: 193-194, 196-197; cf. Sunstein 2022）。まず、AI は人間がなすのと同種の認知を行っていないから、C バイアスはそもそも発生しえない。例えば、難民申請の認容が何件連続しようとも、それを原因として次の申請を棄却する確率が高まることはない。また、AI が空腹を感じないことからも分かる通り、E バイアスは起こりえない。

D バイアスに関してはどうか。過去の多くの判決が、特定集団に対して差別的であるならば、その判決を学習した AI もまた差別的となる。人間裁判官が、差別的でない人から、極度に差別的な人にまで及ぶとき、その分布に応じて、AI の差別度は中間のどこかに位置するだろう。もっとも、D バイアスを縮小させるプログラムを設計することは、技術的に困難でないと思われる。例えば、被告人が有色人種である事案において、裁判官が判決案を入力すると、人種を問わない類似の事案の中央値・平均値から一定以上の乖離を示す場合には、裁判官に警告を発するというプログラムを考えられる。

F バイアスについても、同様に考えられる。特定の志向を備えた裁判官の判決を選択して学習させたり、一定の傾向を示す管轄地の判決を学習させたりするならば、AI はバイアスをもつだろう。例えば、アメリカの南部諸州で、刑事訴訟での過去の州裁判所判決を学習させる場合には、AI は全米の平均値・中央値よりも長い刑期を言い渡しがちだと予想される。だが、全米のデータを学習させた上で、裁判官の判決案が平均値・中央値から乖離している場合に警告を発するプログラムを、容易に設計できるだろう。

カーネマンら（Kahneman et al. 2021: Ch. 10）は、裁判官を含む専門家の判断について、アルゴリズムによるノイズへの対処を示唆

する。また、サンスティーンは、アルゴリズムによる政策判断がノイズを一掃すると指摘している。彼らが認める通り、ロボット裁判官はあらゆるノイズを除去できる。

バイアスとノイズに関する考察から、ロボット裁判官は、判決による比較権衡・相関権衡への違背を予防するか縮減することを通じて、より正義に適った司法を実現しうると言える。具体的には、民事訴訟か刑事訴訟か行政訴訟かを問わず、法の下の平等を判決のなかで実質化し、また刑事訴訟では、犯罪と刑罰の均衡を促進する。法的シンギュラリティとは、司法が、容易な事案での AI によるバイアスとノイズへの効果的対処を通じて、比較権衡・相関権衡の両面で正義の実現へと近づく将来の時点だと言える[4]。

4 結 論

4.1. 残された課題

前節までの考察は、AI による法適用の諸論点のなかでも法的シンギュラリティの当否に焦点をあわせた上で、人間裁判官とロボット裁判官の比較を通じて、この仮想的時点への評価を試みるものだった。まず、AI への法適用、AI での法適用、AI による法適用を区別した後、法的シンギュラリティを定義づけ、先行研究とは異なって、人間と AI の比較という観点から評価を行うという目的を設定した（1.）。次に、予備的考察として、法的シンギュラリティが到来するならば、それを可能にするのは判決予測の AI 技術と LLM の生成 AI だと予見した。また、法的シンギュラリティ像について、容易な事案でのみロボット裁判官を信頼できるという大半説を唱えた（2.）。これらを踏まえて、人間裁判官の判決に潜む C バイアス・D バイアス・E バイアス・F バイアスに関する科学的・統計的知見を概観するとともに、ノイズの知見を瞥見した。そして、法的シンギュラリティが到来した暁には、容易な事案において、バイアスが予防または軽減され、ノイズが消去されると論じた（3.）。

本稿は、その目的のゆえに、先行研究が着目してきた法の支配との関係では、法的シンギュラリティの当否を検討してこなかった。ロボット裁判官が、特定の観点から理解された法の支配に違背するという近時の一学説に対して、別稿では詳細な検討を加えている（宇佐美 2023: 197-201）。だが、法的シンギュラリティが法の支配に適合するかに関する包括的精査は、今後の課題である。その他、ロボット裁判官に対して示されてきたいくつかの懸念が、法的シンギュラリティにも当てはまりうる。アルゴリズムの設計が民間企業に白紙委任されるならば、司法権の独立が脅かされるのではないか。中国での AI 利用について憂慮されているように、裁判官への政治的統制の手段となるのではないか。これらの懸念への簡潔な応答を他所で行ったが（201-202）、本格的検討が必要である。

4.2. 法的シンギュラリティ後の司法

法的シンギュラリティが到来するならば、その後の司法はどのような姿を現すだろうか。前述の通り、容易な事案でのみロボット裁判官が信頼に値する以上、例えば貸金返還請求・不払賃料請求や交通違反への罰金など、法規範命題の同定も事実命題の確定も比較的容易であるタイプの事案は、ロボット裁判官に委ねることが可能となるだろう。他の多種多様な事案については、人間裁判官が容易な事案と困難な事案を分類しなければならない。その上で、人間裁判官は、容易な事案においては、プログラムが出力する判決案を確認して判決を下せるが、困難な事案では、訴訟指揮や判決作成の過程でプログラムに適宜照会するとしても、基本的には自ら判決を作成しなければならないだろう。

[4] 本稿の元となった研究報告では、訴訟遅延の解決と司法アクセスの促進にも論及したが（参照、宇佐美 2023: 190-191）、字数を勘案して論述を省略する。

こうした司法の描像が大きく誤っていないとすれば、法的シンギュラリティの到来後、人間裁判官は、容易な事案を AI に任せ、困難な事案に精力を傾けることができる。人間裁判官の業務は、以前よりもはるかに創造的となる。また、深刻な公共的問題への立法府・行政府の対処が不十分であるとき、司法府は従来、社会改革の一端を担うよう期待されてきた。機械的業務からかなりの程度まで解放された裁判官は、必要な場合には社会改革的な判決の作成に集中できる。法的シンギュラリティとは、人間裁判官が、人間にしか遂行できない困難だが創造的な役割を十全に果たすようになる時点なのである。

Aidid, Abdi and Benjamin Alarie (2023) *The Legal Singularity: How Artificial Intelligence Can Make Law Radically Better*, Toronto: University of Toronto Press.

Alarie, Benjamin (2016) "The Path of the Law: Towards Legal Singularity," *University of Toronto Law Journal* 66(4): 1–13.

Anderson, James M., Jeffrey R. Kling, and Kate Stith (1999) "Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines," *Journal of Law and Economics* 42(1): 271–308.

Brownsword, Roger (2020) "Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law," in Simon Deakin and Christopher Markou (eds.), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, Oxford: Hart, pp.135-160.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue (2016) "Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *Quarterly Journal of Economics* 131(3): 1181–1242.

Clancy, Kevin, John Bartolomeo, David Richardson, and Charles Wellford (1981) "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity," *Journal of Criminal Law and Criminology* 72(2): 524–554.

Cohen, Alma and Crystal S. Yang (2019) "Judicial Politics and Sentencing Decisions," *American Economic Journal: Economic Policy* 11(1): 160–191.

Danzigera, Shai, Jonathan Levavb, and Liora Avnaim-Pesso (2011) "Extraneous Factors in Judicial Decisions," *Proceedings of the National Academy of Sciences* 108(17): 6889–6892.

Deakin, Simon and Christopher Markou (2020) "From Rule of Law to Legal Singularity," in Simon Deakin and Christopher Markou (eds.), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, Oxford: Hart, pp.1-29.

Dworkin, Ronald (1978) *Taking Rights Seriously,* with a New Appendix, Cambridge, MA: Harvard University Press.

Frankel, Marvin E. (1973) *Criminal Sentences: Law without Order,* New York: Hill and Wang.

Goldsworthy, Daniel (2019) "Dworkin's Dream: Towards a Singularity of Law" *Alternative Law Journal* 44(4): 286–290.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein (2021) *Noise: A Flaw in Human Judgment,* London: William Collins.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag (2010) "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review* 60(2): 295–411.

Sunstein, Cass R. (2022) "Governing by Algorithm? No Noise and (Potentially) Less Bias," *Duke Law Journal* 71(6): 1175–1205.

Topaz, Chad M., Shaoyang Ning, Maria-Veronica Ciocanel, and Shawn Bushway (2023) "Federal Criminal Sentencing: Race-based Disparate Impact and Differential Treatment in Judicial Districts," *Humanities and Social Sciences Communications* 10: 366.

Ulenaers, Jasper (2020) "The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?" *Asian Journal of Law and Economics* 11(2): 20200008.

U.S. Sentencing Commission (2023) *Annual Report 2023*.

Vilaça, Luiz Doria, Marco Morucci, and Victoria Paniagua (forthcoming) "Antipolitical Class Bias in Corruption Sentencing," *American Journal of Political Science*. DOI: 10.1111/ajps.12885

Volokh, Eugene (2019) "Chief Justice Robots," *Duke Law Journal* 68(6): 1135–1192.

Weber, Robert F. (2020) "Will the 'Legal Singularity' Hollow Out Law's Normative Core?" *Michigan Technology Law Review* 27: 97–165.

Yang, Crystal S. (2014) "Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from Booker," *New York University Law Review* 89(4): 1268–1342.

宇佐美誠（2020）「AI・技術的失業・分配的正義」宇佐美誠編『AI で変わる法と社会：近未来を深く考えるために』岩波書店、86-112 頁。

宇佐美誠（2023）「AI 裁判官：技術・機能・原理」田中成明・足立幸男編『政治における法と政策：公共政策学と法哲学の対話に向けて』勁草書房、181-206 頁。

宇佐美誠・児玉聡・井上彰・松元雅和（2019）『正義論：ベーシックスからフロンティアまで』法律文化社。

総務省（2024）『情報通信白書令和 6 年版』

恩師ヨセフ・アガシ（1927-2023 年）*

立花　希一

Joseph Agassi, My Mentor(1927-2023)

Kiichi TACHIBANA

年を取ると過去を振り返る傾向が強くなり、しかも自分の過去を恥も外聞もなく語りたくなるようだが、ご多分に漏れず私もそうである。1980 年から 1983 年までイスラエル政府国費留学生として、アガシのいるテルアヴィヴ大学に留学した。留学の主たる目的は二つで、その一つはポパーの科学哲学および社会哲学の研究を深めることであった。1980 年にはポパーは 78 歳で、大学をすでに定年退職していた。かれに直接師事することはできない相談だったが、ポパーの直弟子たちから学ぶ道が残されていた[1]。そのなかから誰にするかが問題になったのだが、私はアガシによる「ポパー以上にポパーリアンの立場をめざしている」という言葉[2]に惹かれて、かれの下への留学を決意した。当時、アガシはボストン大学とテルアヴィヴ大学の兼任教授だったので、留学先をアメリカにするかイスラエルにするかの選択肢があったのだが、イスラエルのテルアヴィヴ大学を選んだのには訳があった。私には、もう一つの目的、すなわちポパーのユダヤ的背景を探るという目的ももっていたからである。ユダヤ人の生活や思想で溢れているイスラエルがうってつけの場所で、テルアヴィヴ大学のアガシの下に留学することは一石二鳥に思われた。ポパーの『開かれた社会とその敵』なくして、私の学問・人生はないと言っても過言ではないが、アガシとの関わりなくしても、その後の教育・研究生活は成立しなかったであろう。高木勘弐（1919-1993 年）が日本における恩師だが[3]、アガシはそれと同等かそれ以上の恩師である[4]。

ヨセフ・アガシ（Joseph Agassi）は、1927 年、現在のイスラエル（当時は英国の委任統治領）に生まれた。宗教的シオニストであった父親は東欧からイスラエルへの移民の一人であった。宗教的な家庭に生まれたヨセフは、宗教学校に入れられたが、しばらくしてユダヤ教に疑問をもち、学校を退学する。私見では、この宗教学校退学という十代半ばの決断がアガシの人生にとって決定的だったと思われる。この決断の状況を詳しく述べてみたい。

1980 年に私が初めて会ったときのアガシは 53 歳で、当然のごとく哲学の先生であった。青年期あるいは少年期に哲学に興味をもったアガシは哲学をする目的のためには物理学が不可欠な手段だと考えて、ヘブライ大学の物理学科に入学したのはそれより 34 年前の 1946 年である。それ以後のアガシは、研究者志望の学生、哲学科の院生、助手、講師、助教授、教授へと昇進し、定年退職後、名誉教授となったが、こうした社会的地位とは関係なく、生涯にわたってひたすら哲学に関わる研究・教育一筋の生活を送った（かれは「ワーカホリック」だと自嘲気味に語っていた）。これが、アガシが実際に送った人生である。

しかし、アガシが大学入学前に、入学したものの退学した宗教学校「イェシヴァ」は、改革派や保守派のラビ養成ではなく、「正統派のラビ」養成学校であった。このイェシヴァは、正式名称「イェシヴァ・ハメルカジット・ハオラミット（中心的・世界的宗教学校）」（通称「メルカズ・ハラーヴ（ラビクック・センター）」）という学校で、アシュケナジ主席ラビを務め宗教的シオニストでもあったアブラハム・イツハク・クック（1865-1935 年）が 1924 年に設立した権威ある宗教学校である。歴史的事実かどうかはさておき、正統派ユダヤ教では、「バビロニア捕囚」からのエルサレム帰還民によって再建された第二神殿が 70 年に崩壊して以後、「神殿祭司中心のユダヤ教」

から移行した「ラビのユダヤ教」という伝統的なユダヤ教を正統に継承するものだと主張されてきた。正統派ユダヤ教の特徴は、613 の戒律からなる「モーセの律法」の厳格な遵守にある。正統派ユダヤ教徒たちは日常生活の隅々にいたるまでこの戒律を遵守した生活を送っている。このような生活の送り方は「俗生活を聖化」させるためだと言われる。戒律を正しく解釈し、正統派ユダヤ教徒に伝授する役割を果たす宗教的指導者が「ラビ」で、正統派ユダヤ教徒からの尊敬を受ける存在である。正統派ユダヤ教徒は常に「キパ」という小さな帽子を頭に載せているので「正統派ユダヤ教徒」であることは一目瞭然である。そのうえ、アガシの育った家庭は、正統派のなかでも「ハレディーム（神を畏れる者たち）」と呼ばれる超正統派ユダヤ教の家庭だった。髭をたくわえ、もみあげを伸ばし、黒いコートと黒い山高帽を身につけたユダヤ人たちをテレビなどで見たことがあるはずだ。「バル・ミツヴァ（13 歳になった男子を祝福する成人式）を終えたアガシは父親の期待を一身に受けて、「ラビ」になるためにイェシヴァに入学した。アガシの優秀さを鑑みると、果たせるかな「ラビ」の称号を取得し、前途有望な将来が約束された青年・成年へと成長し、さらにはクックのような主席ラビにまでなったとしても不思議ではない。かりにラビになれなかったとしても、超正統派ユダヤ教徒の一員として戒律遵守の生活を送り、かれらの居住する特別な地区から外にはほとんど出ることなくその生涯を終えていたかもしれない。アガシが尊敬する哲学者スピノザを読むこともけっしてなかっただろう。イスラエル留学中、北から南までいろいろな場所を訪問したが、超正統派ユダヤ教徒の居住地にはただの一度も足を踏み入れたことはなかったので、アガシと出会うこともけっしてなかっただろう。どちらもありえたアガシの二つの人生、生活様式の（程度問題ではありえない、まるっきり異なる質の）相違を比較してみれば、アガシの人生にとって、「宗教学校退学」という十代の選択がいかに決定的なものであったかがわかっていただけるであろう。しかもアガシはイェシヴァを退学しただけではなく、超正統派の衣服を脱ぎ棄て、「律法遵守」という正統派ユダヤ教の宗教的義務・束縛から自らを自らの手で解放することで、自律的で独立した「自由人」になったのである。こうした生活様式の劇的な変化およびそれに伴うパーソナリティの変化には、「批判精神」が決定的な役割を果たしたであろう。私が出会ったアガシはすでにこのような人物であった。

先に言及したように、宗教学校中退後、ヘブライ大学に入学したアガシは、哲学の理解を深める目的のために、物理学を選んだ。しかし、当時のかれの物理学の教師たちは、約束主義の影響を受け、物理学を応用数学の一種とみなし、哲学とは無関係の学問として扱っていた。こうした態度に失望したアガシは、カール・ポパーの論文「哲学的諸問題の性格と科学におけるその根源」（1952 年）を読んで感動し、ポパーの下で研究することを決心する[5]。このポパー論文との出会いが、アガシの研究対象や研究の方向性を決定づけたといえよう。

ヘブライ大学で修士号（M. Sc.）を取得（1951 年）後、ロンドン・スクール・オヴ・エコノミクスに留学する（1952 年）。ポパーの指導の下、The Function of Interpretation in Physics（「物理学における解釈の機能」）という論文で、博士号（Ph. D.）を取得する（1956 年）。この学位論文における「解釈」というのは、「形而上学的解釈」ないし、「形而上学」ということであり、当時、有力であった帰納主義と約束主義という二つの考え方では排除されていた形而上学が科学の研究に大きな役割を果たすことをファラデーの事例研究を中心に論

じたものである。この論文が基になって、1971年に *Faraday as a Natural Philosopher*[6]（『自然哲学者としてのファラデー』）として結実し、科学史家としての地位を確立することになる。

ポパーとの関係に注目しながら、博士号取得後のアガシの経歴をみていきたいと思う。というのも、アガシによると、アガシのあえて行った行為がきっかけで、二人の間は蜜月から断絶に変わったからである。

1960年にようやくアガシは、哲学の講師として香港大学に赴任する。それまでの数年間、アガシはポパーの助手やロンドン・スクール・オヴ・エコノミクスの研究所の助手を務めるなどしながらポパーの研究の手伝いを献身的に行っており、例えば、ポパーの『探求の論理』の英語版『科学的発見の論理』（1959年）の刊行に際しては、アガシが索引項目の収集および編集を任されており、また注では、確率理論などにおけるいくつかの研究成果が二人の協働作業の結果であることなどが記されている。香港大学赴任後もポパーとの良好な関係は続き、1963年刊行の『推測と反駁』でも、注のなかでアガシ博士に対する謝辞などが述べられている。1965年に発表されたアガシによる『推測と反駁』の書評では、『科学的発見の論理』においては、真理概念の使用にたいして控え目だったポパーが『推測と反駁』では真理概念を積極的に用いているだけではなく真理接近理論も展開している点に貴重な貢献があり、それだけでもこの書の刊行は有意義であると、高い賛辞が述べられていた。ところが、香港大学、イリノイ大学を歴任した後、テルアヴィヴ大学とボストン大学に移ってすでに数年たっていたアガシが、1972年に刊行された『客観的知識』の書評を、1974年に発表した途端、まったく音信不通となり、以後絶縁状態となったのである。ポパーと二度と会うことができなかったという。アガシはあえて行ったのだが、好意的な評価をするのが当たり前とみなされている「書評」の場で、アガシは大小合わせてかなり多くの批判を述べたのだ。ここではアガシが私に語った批判だけを紹介しよう。著者ポパーは、常日頃から主題ではなく問題に取り組むよう勧めているにもかかわらず、この書では、主題は明瞭だが取り組もうとしている問題が不明であると、ポパーの持論を用いてポパーの著作を公然と批判したのである。しかも、この批判なら、批判を尊重する批判的合理主義の提唱者ポパーなら歓迎とはいわないまでも容認はしてもらえるはずだと期待して。しかし、その期待は失望に変わっただけではなく、報復も待っていたそうだ。というのも、執筆時期と出版時期はずれがあるので、年が前後するが、1975年に出版されたアガシの『流動する科学』によると、同書に収められている論文「修正約束主義（Modified Conventionalism）。」は元々、1974年に刊行された、シルプ編『カール・ポパーの哲学』の寄稿論文として執筆されたものであったが、ポパーの要求によって、元の論文の十分の一にも満たない3ページ余りの抜粋が掲載されることになったという。しかも、シルプ編の哲学叢書では書くことが決まりごとになっている批判論文にたいする「返答」でポパーは、自分は哲学的見解・立場を「主義」として扱っているとよく言われるが、アガシは「主義」ではなく、「主義主義（ismism）」として論っているだけだとアガシの批判的考察を一蹴したのである（『流動する科学』所収の「修正約束主義」を是非読んでいただきたい）。

アガシによると、「批判的合理主義者」らしからぬポパーの振る舞いは今に始まったことではなかったようで、アガシからみると好意的でしかも適切と思われる批判を弟子のバートリーが、ポパーが取り組んだ重要な問題の一つである「境界設定の問題」の解決の試みにたいして行った際にも見ら

れたそうだ。ポパーにとって弟子はいつまでも弟子であり、その弟子は批判を自分に直接プライベートな形で行ってもよいが、パブリックな場で公然と行ってはならないというのが、ポパーの定めた暗黙のルールのようだとアガシは 2000 年に夫妻で来日した際、語ってくれた。身内による公の場での批判は自分の哲学を弱体化させ敵を利するだけになるとポパーは危惧していたようだとも話してくれた。アガシはこの考えに反対で、哲学者は相互に独立した人間で自律的でなければならず、「希一も私に追従せず自律的であるべきだ」と念をおされた。

ポパーは「批判的合理主義」の提唱者であっても実践者ではなかったようだ。逆にアガシは「批判的合理主義」の優れた理論家のひとりであると同時に実践者たろうと努力する人間であった。「言行一致」はまさに言うは易く行うは難しなのだが、この点で、アガシはポパー以上の哲学者だといえるかもしれない。真の「哲学者」という観点から、アガシの人物像をもう少しみてみよう。

アガシは、師のポパーと同等あるいはそれ以上の哲学者だと述べたが、「哲学者」とはいったいどんな人間なのだろうか？「哲学・哲学者」の語源からしても、ピタゴラスの有名な定義は適切だと思われる。それによると、「哲学者」は、地位・名誉・利得などではなく、「真理を追求する者」である。人間にとって重要な価値として、真理・善・美・正義などが挙げられるように、真理は価値の一つであるが、ポパーは、真理が価値のなかで高次の価値を占めると示唆した。例えば、「法律順守のほうが人命尊重より価値がある」は真であるのか？とか、「信仰のほうが客観的真理より価値がある」は真であるのか？と問えるように、価値に限らずどんな主張・見解にたいしてもその真理値が問えるからである。このような意味での真理をアガシも追求した。では、次の発言はだれの言葉だろうか？

> わたしとは、どんな人間であるかといえば、もしわたしの言っていることに何か間違いでもあれば、こころよく反駁を受けるし、他方また、ひとの言っていることに何か本当でない点があれば、よろこんで反駁するような、とはいってもしかし、反駁を受けることが、反駁することに比べて、少しも不愉快にはならないような、そういう人間なのです。なぜなら、反駁を受けることのほうが、より大きな善であるとわたしは考えているからです。

ポパーやアガシなら言いそうな言葉だが、実は『ゴルギアス』のなかのソクラテスの言葉である。ポパーは「批判的合理主義」の提唱者として知られているが、ポパー自身はソクラテスを「批判的合理主義の祖」と位置づけていた。ポパーの言葉を引用する[7]。

> 『ソクラテスの弁明』は、われわれはいかにほとんど何も知らないかを自覚すべきであり、そして、あらゆる理論や信念が受けるべき批判的議論によってわれわれは学ぶことができるという謙虚な見解が特徴の「批判的合理主義」の最初のもっとも偉大なマニフェストである。この批判的合理主義は、西洋思想（や西洋科学）に対するその影響において測り知れないものがある。

この意味でアガシもまた「ソクラテスの弟子」であり、批判・反駁を通して誤りから学び真理を探究しようとする哲学者としての道を生涯歩み続けた人物だと思われる。20 世紀以降の歴史を振り返ってみると、「哲学者」と呼ばれる人びとのなかで、「批判的合理主義」を標榜していたかどうかを

不問にしても、ソクラテス的哲学の実践者は残念ながらあまり見当たらないように思われる。

ここで、ポパーとアガシの科学哲学における立場の共通点・相違点も確認しておきたい。ポパーとアガシの反証主義は、帰納の問題の解決策としての帰納主義の否定・却下という点では共通しているが、両者の間には大きな相違がある。ポパーの反証主義は、科学と非科学の境界設定問題の解決策として提案されたものであり、その境界設定基準が反証可能性であった。他方、境界設定問題よりも知識の成長・進歩の問題を重要視するアガシは、科学以外にも反証可能な理論が存在することと科学理論にも反証不可能な形而上学的要素が存在すると主張し、ポパーによる境界設定基準としての反証可能性は成立しないとして却下する[8]。さらに、科学は経験的性格だけではなく、形而上学的性格も備えており、形而上学を非科学として排除することなどできないだけではなく、形而上学の軽視は科学的知識の成長の妨げにもなると主張する。アガシによれば、科学が「経験」科学と呼ばれる理由が、経験（実験・観察）による実証（verification）や確証（confirmation）、験証（corroboration）といった経験的支持（empirical support）にあるのではなく、まさに経験による反証可能性および実際の反証にあるのだと主張する。すなわち、アガシの反証主義は、科学の経験的性格を特徴づけるものであって、ポパーのように科学を包括的に特徴づけるものではない。しかも、アガシにとって反証可能性はそもそも境界設定の基準ではないので、アガシの場合には、科学以外にも、反証可能性という科学と同様の経験的性格があったとしてもまったく問題にはならないのだ。

アガシを懐かしむ思いから、かれの人柄を示すエピソードも述べさせていただきたい。テルアヴィヴ大学にはアガシの所属する「哲学科」とは別組織として「ユダヤ哲学科」があった。そこで、ユダヤ哲学の受講を計画し、この計画をアガシに伝えると、怒ったように「ユダヤ哲学科のスタッフは全員「キパ」をかぶっている連中で、批判的ではない。そんな授業は価値がない、受けても無駄だ」と言われてしまった。たまたまそばにいた同僚のゲルション・ヴァイローが「無駄かどうかは受講してみなければわからない。顔を出してみたらどうだ」と、取りなしてくれたおかげで、無事に、「マイモニデスの『迷える者への手引』[9]の講義を受講することができた。当時のアガシは激昂しやすい質であった。アガシの「哲学の諸問題」の授業は、受講者が自由に執筆したペーパーを読み、その後、議論する形式であった。私が用意した「独断主義・懐疑主義・批判主義」というペーパーを読む機会があったとき、読み始めてまもなく「ラカトシュ」の名前に言及するやいなや制止され、みるみる顔つきが変わっていき、「ラカトシュはポパーを懐疑主義者だと言ったり独断主義者だと言ったりするが、かれが「懐疑主義」で何を意味しているのかまったくわからない」と言われ、私がペーパーを読むのが打ち切られてしまった。授業の後で、受講していたかなり年配の学生に「よくあることだ。気にするな」と慰められた。アガシは学会の質疑応答でも、同様の振る舞いをする傾向があるらしく、日本の研究者で、ポパーの「批判的合理主義」を評価する哲学者からも、日本に帰国した後、「アガシはアグレッシヴ（攻撃的）だ」と言われたことがある。アガシの「怒りの爆発」は国際学会でも実際にあったようで、その場面が公刊された本で読むことができる（Discussion: On papers by I. Lakatos and Y. Elkana, *The Interaction between Science and Philosophy*, edited by Y. Elkana, Humanities Press, 1974, pp. 280-297.参照）。議論の内容とは切り離してであるが、アガ

シは自分が「かっとなった（lose temper）」ことを反省し素直に謝っている。おそらくアガシは自分の切れやすい性質を自覚し反省し、コントロールしようと努めてきたようである。イスラエル帰国後十数年ぶりに再会した2000年以後、2002年、2003年とアガシに会い、大小の学会や講演会でアガシの振る舞いを目にしたが、かれが激昂した場面を一度も見たことがない。「批判は敬意・尊重の表明だ」というのがアガシの口癖なので、かれの批判の鋭さ・厳しさは相変わらずだったが、激昂型というよりむしろ、穏やかで丁寧な口調でしかも礼儀正しく、いわば日本の美徳とされる「謙虚」な振る舞いが目立っていた。この事例は、先天的な性質も自己反省・自己批判的に変えようと努力すれば、なくしてしまうことはできないとしても後天的に変えていけることを示す模範例だと思われる。

イリノイ大学、テルアヴィヴ大学、ボストン大学、カナダのヨーク大学（トロント）を歴任し、定年退職後は、テルアヴィヴ大学・ヨーク大学名誉教授となった。かれの専門分野、関心は、自然科学・社会科学の方法論、科学史はもとより、物理学、形而上学、技術論、精神病理学、人間学、教育理論、宗教思想、民族問題など多岐にわたっている（ジャーヴィとラオールはアガシを「オールラウンダー」と評している[10]）。各方面における著書、編著書、共著書、論文など膨大な数にのぼるが、その詳細については、https://www.tau.ac.il/~agass/pub.html を参照していただきたい。ここでは英文の主要著書（単著）に言及しながら、かれの思想を紹介することにしよう。

香港大学に赴任してまもなく、1961年に *Towards an Historiography of Science*（『科学の歴史記述に向けて』）を書き上げる（出版されたのは1963年）。この本のなかで、従来、科学史家が科学の歴史を書く際に意識的あるいは無意識的に採用している科学哲学として帰納主義と約束主義を挙げ、その問題点を指摘しながら、ポパーの反証主義という第三の立場による歴史記述の方法を提示したものである。出版はクーンの『科学革命の構造』の一年後であるが、シルプ編、『カール・ポパーの哲学』のなかで、寄稿者の一人、J. O. ウィズダムが、アガシの『科学の歴史記述に向けて』とクーンのそれに言及して「この二書が最も斬新かつ重要なこの分野における貢献である」と高く評価している[11]。

次に出版されたのは *The Continuing Revolution: A History of Physics from the Greeks to Einstein*,1968である[12]（拙訳、ヨセフ・アガシ、『科学の大発見はなぜ生まれたか』、講談社ブルーバックス、2002年）。かれの歴史記述の立場は、ポパーの反証主義と、E．A．バートやA．コイレの立場---科学は形而上学の枠組のなかで発展していくものとして理解するもの---との総合をめざすものであり、科学的研究によって批判可能ないくつかの形而上学的枠組が競合する歴史として物理学の歴史を捉えようとする立場である。本書はかれの息子との対話という体裁をとっており、中身を読まないと子供向けの通俗書として受け取られてしまう恐れがある。しかし、学術的にも十分批判に耐えるものになっている。むしろ、高度に学術的な作品が、青少年にも手が届くように工夫されていると考えた方が適切であろう。

ボストン大学で10年間教鞭をとった後、かれの主要論文が *Science in Flux*（『流動する科学』）というタイトルで1975年に、Boston Studies in the Philosophy of Scienceの一書として出版された。因みに、Boston Studies in the Philosophy of Scienceは科学哲学において世界をリードする書物を多く刊行しているシリーズであるが、この著作

はそのなかでもとりわけ成功した作品の一つである。科学を高度に批判的で創造的な営みとみなす点、形而上学の強調に特徴があるが、他方、科学の社会的役割、責任にも目を配っている点にも特徴がある。

イスラエルの Van Leer Jerusalem Foundation での講演を基にまとめられた *Towards a Rational Philosophical Anthropology*(『合理的な哲学的人間学に向けて』) は 1977 年に出版された[13]。本書は、ギリシャ以来の二分法的発想をオール・オア・ナッシングの見方として退け、それに代わって、モア・オア・レスから出発してモアをめざしていくという積極的な見方を、物理学、生物学、哲学、社会科学、神学などといった多分野において提唱するという大胆な、画期的な企てである。あまりに大胆な企てのため直ちに理解されるものではなく、今後、次第に論議が高まっていくものと期待される。

1981 年に Boston Studies in the Philosophy of Science から出版されたかれの二つ目の論文集である、*Science and Society: Studies in the Sociology of Science* (『科学と社会:科学社会学における研究』)[14]は、科学の社会的側面をさらに考察し、学問の世界の運営が権威主義的ではなく、より民主主義的なものになれば、科学はさらにより良きものになり、人道的になるということを論じたものである。

Boston Studies in the Philosophy of Science から刊行された第三作目が、*Science and Culture*, 2003 (『科学と文化』) である[15]。これまで執筆してきた数多くの論文のなかから、自律、寛容、理性、哲学、責任というキーワードの下に精選された論文で構成されている。知性主義 (反経験主義)、経験主義 (帰納主義)、道具主義の科学観を退け、ポパー流の批判主義的科学観を採用したうえで、個人の自律、リベラリズム、多元主義、デモクラシーと科学との関わりを考察している。

Boston Studies in the Philosophy of Science から刊行された第四作目として、上述の *Towards an Historiography of Science*, 1963 を再掲すると共に歴史記述に関わる論文が付加された *Science and Its History: A Reassessment of the Historiography of Science*, 2008、(『科学とその歴史:科学の歴史記述再考』) があり、第五作目には、*The Very Idea of Modern Science: Francis Bacon and Robert Boyle*, 2013 (『近代科学の観念:フランシス・ベーコンとロバート・ボイル』) がある。

アガシは学界にデビューするのに、書評の分野において書評を学問にまで高めるという貢献が大きく寄与したが、かれのこれまでの書評のうち重要なものを集めて一書となった極めて稀な書が、1988 年に出版された、*The Gentle Art of Philosophical Polemics*, (『哲学的論争の作法』) である[16]。このなかでかれは模範的な哲学的議論を示している。

最新の著書、*Academic Agonies and How to Avoid Them*, Social Epistemology Review and Reply Collective, 2020 (『学術上の苦悩とその回避策』) からもわかるように、アガシは特に近年顕著に、さまざまな事象をモア・オア・レスという程度問題として把握しようとする従来からの立場に依拠し、エリートと大衆とか知識人と素人とかといった二分法を退け、学問の世界と社会一般の垣根を可能な限り取り払って、民主社会において、いわゆる文系・理系を問わずさまざまな知的活動を相互交流させ、活性化させるため、ともすれば権威主義的で閉鎖的になりがちな学界をさらにリベラルで開かれた社会へと改革するための方策を研究・提言している。同書は、こうした実践的活動の一環として、興味のあるひとなら誰でもオンラインで読めるように提供されている (https://social-

epistemology.com/2020/10/19/table-of-contents-academic-agonies-and-how-to-avoid-them-joseph-agassi/)。因みに、拙訳『科学の大発見はなぜ生まれたか』の「日本のみなさまへ」でも、ガリレオやアインシュタインを引き合いに出して「ポップ・サイエンス」の意義を強調することによって、アガシのアカデミズム打破の姿勢が滲み出ている[17]。

その他、科学と技術の峻別、公害等環境問題の科学に基づく技術的解決の可能性、環境問題のグローバルな視点からの把握・対策を訴えている *Technology: Philosophical and Social Aspects*, 1985（『テクノロジー：哲学的・社会的側面』）[18]、哲学入門書としては個性的な *The Siblinghood of Humanity: Introduction to Philosophy*, 1990（『人類皆同胞：哲学入門』）[19]、量子力学の展開を扱った、*Radiation Theory and the Quantum Revolution*, 1993（『放射理論と量子革命』）[20]、ポパーの伝記的作品である、*A philosopher's Apprentice*, 1993（『哲学者の弟子』）[21]、民族問題を扱った、*Liberal Nationalism for Israel: Towards an Israeli National Identity*, 1999（『イスラエルにとってのリベラルなナショナリズム』）[22]、ポパーの反証主義に対する批判を批判的に検討する、*Popper and his Popular critics: Thomas Kuhn, Paul Feyerabend, Imre Lakatos*, 2014（『ポパーと通俗的なポパー批判者たち：トマス・クーン、ポール・ファイヤアーベント、イムレ・ラカトシュ』）[23]、教育論に関する論考等を収録した、*The Hazard Called Education by Joseph Agassi: Essays, Reviews, and Dialogues on Education from Forty-five Years*, 2014（『教育に潜む危険要素：45年にわたるアガシの教育に関するエッセイ、書評、対話』）[24]、批判的合理主義の立場からヴィトゲンシュタインの思想を考察した、*Ludwig Wittgenstein's Philosophical Investigations: An Attempt at a Critical Rationalist Appraisal*, 2018（『ルートヴィヒ・ヴィトゲンシュタインの哲学探究：批判的合理主義的評価の試み』）[25]等がある。

2017年にはアガシ生誕90年を記念した論文集が出版され[26]、拙稿、How can we attain both democracy and constitutionalism? [27]も収録されているが、この年は私の定年退職の年度でもあり、私にとってもいい記念になる論文集の刊行であった。

2018年以降もアガシは精力的に研究を続け、共著を含めると20本以上の論文を執筆するなど、研究・執筆意欲は最期まで衰えなかったようである。

恩師ヨセフ・アガシが2023年1月22日、95歳で逝去したとの訃報を同年3月にアガシの弟子で知的財産相続人となったアブラハム・メイダンからの電子メールで知った。先に言及したアガシの著書、*The Continuing Revolution* は、2002年に『科学の大発見はなぜ生まれたか』という書名で、講談社ブルーバックスの一書として邦訳出版されていたが、2011年の第三刷を最後に長い間品切れとなっていた。奇しくも十数年ぶりに復刊の動きがあり、2024年7月、『父が子に語る科学の話：親子の対話から生まれた感動の科学入門』と題する新装改訂版として復刊された。残念ながら、生前には間に合わなかったが、復刊を機に新しい命を吹き込まれたこの拙訳書を恩師ヨセフ・アガシに捧げる。

---

* 本稿は、恩師ヨセフ・アガシの逝去（2023年1月22日）を受けて、アガシについてすでに、日本ポパー哲学研究会ウェッブサイトに掲載された「世界の批判的合理主義者たち」第四回にあたる、「ヨセフ・アガシ」に関する拙稿を大幅に加筆・修正した最終版である。

[1] 日本人でポパーに直接学んだ哲学者には、市井三郎（1922-1989年）、神野慧一郎（1932年- ）がいる。因みに、ポパーの京

都賞受賞記念として開催されたワークショップ（1992 年 11 月 12 日）で論文を読み上げた神野氏は、ポパーから、「神野の論文は極めて優れている（extremely good）」と評された。市井に関するエピソードとしては、私がアガシに初めて面会した際、アガシから「市井を知っているか」と尋ねられ「はい」と答えると、にっこりとし懐かしそうな表情を見せたのを覚えている。かれらは同じ時期にポパーの学生であった。二人とも理学部出身だったので、共通の話が多くあったのだろう。その際、市井の執筆した本を紹介できれば良かったのだが、少なくとも題名を翻訳して伝える機敏すら私は持ち合わせていなかった。市井については、小林傳司、「日本におけるポパー哲学受容の一形態---市井三郎の創造的受容」、小河原誠編、『批判と挑戦』、未来社、2000 年、202-244 ページ、参照。

その後の世代の研究者は、実にいろいろなポパーの弟子に学んだ。若干、例を挙げると、小河原誠（バートリー、アメリカ）、萩原能久（アルバート、ドイツ）、渡部直樹（ワトキンズ、イギリス）、吉田敬（ジャーヴィ、カナダ）、そして私（アガシ、イスラエル）である。しかも興味深いのは、私も含めそれぞれが、ポパーのそれぞれの弟子の影響を大いに受けてポパー哲学を理解しており、見解・立場に相違が生じていることである。

[2] Joseph Agassi, Modified Conventionalism is More Comprehensive than Modified Essentialism, P.A. Schilpp, ed., *The Philosophy of Karl Popper*, Open Court, 1974, p. 693. 因みに、この論文は、Modified Conventionalism の極端な圧縮版で、元の論文は一年後の 1975 年に、本文でも紹介した *Science in Flux* に収録され刊行された。Joseph Agassi, Science in Flux, R. S. Cohen and M. W. Wartofsky eds., Boston Studies in the Philosophy of Science, 28, Reidel, 1975, pp. 365-403.「ポパー以上のポパーリアンの立場」とは、反証主義の徹底である。アガシによれば、1959 年の『科学的発見の論理』ではなく 1963 年の『推測と反駁』において、ポパーは反証主義者（falsificationist）から験証主義者（corroborationist）への逡巡的移行を示しており、験証された科学理論の受容が合理的である、という主張に傾斜したという。そもそもの反証主義の立場からは、合理的受容（rational acceptance）は問題にならず、むしろ理論と経験（実験・観察結果に関する言明）とが矛盾するという意味での反証が具体的に示された理論はそのまま真理と認めることはできないとして拒否するのが合理的であるとみなす、合理的拒否（rational rejection）の見方のほうが重要だという。アガシによれば、「矛盾の認知が合理性の要」なのである。Joseph Agassi, *Science in Flux*, p.138. ポパーは、アガシが異を唱える自分の立場に「実証主義の匂い（whiff of verificationism）」があると認めている。K. R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge and Kegan Paul ,1963, p. 248, note 31.率直に言って、私はイスラエルでアガシの見解を理解するまで、験証された科学理論の受容が合理的であるという見解がポパーの反証主義の立場だと思い込んでいた。アガシから学んだこの新たな知見に基づいて執筆したのが、次の注 3 でも言及した拙稿、FALSIFICATIONISM VERSUS CORROBORATIONISM、1982 年、筑波大学哲学・思想学会、『哲学思想論叢』、第 1 巻, pp. 67-78.である。typo 修正版は、以下のアドレスで閲覧可能である。。
https://www.kiichiposition.com/_files/ugd/070434_7eccad9ad61e4fbdb43b8cf5a518a1b7.pdf

J. Agassi, Science in Flux： Footnotes to

Popper, *Science in Flux*, 1975, pp. 9-50. 同じ著書に収録されている、Replies to Diane: Popper on Learning from Experience も参照（pp. 81-91）。アガシとは対照的に、験証主義をさらに推し進め、科学的研究プログラムの方法論（MSRP）を唱えたのが、ラカトシュである。ラカトシュは自分の方法論がポパーの反証主義を洗練させたものだと主張したが、ラカトシュによるデュエム＝クワイン・テーゼの扱い方に注目してラカトシュの MSRP が反証主義ではまったくないことを明らかにしたのが、拙稿、「デュエム＝クワイン・テーゼと反証主義」、『批判と挑戦』、2000 年、141-178 ページ、である。クーンやファイヤアーベントの立場がアンチ反証主義であるのは明白だが、実はラカトシュの立場もアンチ反証主義であることを浮き彫りにしたのが、注 23 のアガシの著書である。

[3] ポパー哲学との最初の出会いをつくってくださったのが高木先生である。私が学部３年次だった 1973 年、（後に私の卒論の指導教官となる）高木先生が、「これまで私はテキストを使って授業をしたことは一度もないのだが、今回はポパーの科学哲学の主著『科学的発見の論理』をテキストにする」と言われて始めた講義を受講したとき、ポパーの名前を初めて知ったからである。イスラエル留学を後押ししてくださったのも高木先生だし、後年知った事実なのだが、私の最初の本格的論文とも言うべき、注 2 で言及した、拙稿、FALSIFICATIONISM VERSUS CORROBORATIONISM を、英語論文でありながら、筑波大学哲学・思想学会の学術雑誌、『哲学思想論叢』への掲載を認めるよう支援してくださったのも高木先生であった。私が論文投稿した『哲学思想論叢』創刊号には、英語論文の投稿規定がなかったのだ。

[4] テルアヴィヴ大学には 3 年在籍したが、先にも述べた通り、アガシは半年ボストン大学、半年テルアヴィヴ大学だったので、アガシから指導を受けたのは実質 1 年半で、半年毎の飛び飛びでもあった。しかも、イスラエルの院生たちの間では、当たり前のようにアガシを愛称・渾名の「ヨスケ」で呼び、相互にフランクに付き合っていたのだが、25 歳も年長でしかも指導教授であるかれを名前で呼ぶことなど、私にはまったくできなかったし、ましてや渾名で呼ぶことなどできるわけがなかった。私の 3 年間の留学生活は思ったほど成果を挙げることができなかったのは、哲学の問題に取り組むうえでは、先生も学生もない対等な関係を築くことができなかったことが大きな原因ではなかったかと思えてならない。

因みに、ソクラテスは、議論に参加する意思のあるひとなら誰とでも無償の議論を行ったが、そのなかには、プラトンなど多くの若者もいた（ソクラテスとプラトンの年齢差は 47 歳である）。プラトンの作品を通してみる限り、ソクラテスと若者たちとの対話では、互いに丁寧で明晰な言葉は使われているが、当然、尊敬語や謙譲語は使用されていない。しかも、意見が異なる同士の間でも、つねにではないが、相互の尊重や信頼は醸成されている。ところが、『プラトン全集』の翻訳では、不思議なことにまるで日本人同士の会話のように、年少者はソクラテスに敬語で話しているのだ。

日本では敬語（尊敬語・謙譲語）の使用を当然視する伝統的規範が私の足を引っ張ったと言いたくなってしまう。学問をめざす子どもたちのために、日本語の翻訳でも、敬語表現など用いずに、古代アテネにおけるかれらの実際の対話を活き活きと再現する工夫がなされてしかるべきかもしれない。

2000 年に香港大学から招聘されたアガシ夫妻が、そのついでとして 10 月 7 日に来日

し、17 日の離日まで講演活動等で東京、京都、東京、秋田、東京を旅行した。その期間中、私はずっとかれらに同行したのだが、その間にようやく、「ヨセフ」、「ユディット」と呼べるようになった。「郷に入れば郷に従え」と口で言うのは易しいが、行うのは私にはとてつもなく難しかったのである。因みに、日本ポパー哲学研究会では、例年夏季に開催している研究大会をアガシ夫妻の来日に合わせて秋季に開催（10 月 14 日、慶応義塾大学三田キャンパス）し、夫妻がそれぞれ特別講演を行った。

イスラエルでは当たり前の師弟関係に至るのに実に 20 年も要してしまった。これをきっかけにその後は、主として電子メールによるものだが、アガシとのコミュニケーションを驚くほどスムーズに行えるようになった(2002 年にウィーンで開催されたポパー生誕記念国際学会（ポパー・コングレス 2002）で再会でき、しかも私の研究発表をわざわざ聞きに来てももらえたし、さらに 2003 年 10 月 22 日から 25 日までの台湾でのアガシの講演活動にも同行し、哲学上のいろいろな話をすることができた)。2000 年は私にとって決定的な年であった。ポパー・コングレス 2002 で発表した原稿に加筆・修正し完成した拙稿が、Can the Japanese Learn to Welcome Criticism Openly?, Ian Jarvie, Karl Milford and David Miller eds., *Karl Popper A Centenary Assessment*, Ashgate, Vol. I, pp. 203-215. である（因みに、この論文名もアガシが提案してくれた複数の候補のなかから私が選んだものである)。

[5] Karl R. Popper, The Nature of Philosophical Problems and their Roots in Science, *Conjectures and Refutations*, pp. 66-96. アガシは後（1964 年）に、The Nature of Scientific Problems and their Roots in Metaphysics（「科学的諸問題の性格と形而上学におけるその根源」）という論文を書いている。J. Agassi, The Nature of Scientific Problems and their Roots in Metaphysics, *Science in Flux*, Boston Studies in the Philosophy of Science, Vol. 28, 1975, pp. 208-239. これはアガシによるポパーの下での研究を決断させたポパーの論文、「哲学的諸問題の性格と科学におけるその根源」を捻ったもので、一見するとポパー論文とはまさに対照的な主張に思えるが、私見では、対立というより科学と形而上学（哲学）のどちらに焦点をあてるかという重点の相違であるように思われる。

[6] J. Agassi, *Faraday as a Natural Philosopher*, Chicago University Press, 1971.

[7] K. R. Popper, *After the Open Society: Selected Social and Political Writings*, edited by Jeremy Shearmur and Piers Norris Turner, Routledge, 2008, p. 222.

[8] Joseph Agassi, Popper's Demarcation of Science Refuted, *Methodology and Science*, 24, 1991, 1-7.

[9] この『迷える者への手引』は、哲学に目覚めたばかりのスピノザやソロモン・マイモンにとって読むことのできた唯一の「哲学的書物」であった。最初はマイモニデスの生涯や著作についての講義だったが、次第に演習形式になり、ヘブライ語を始めたばかりの私は授業についていくことができなかった。それでも、シュロモー・ピネスの英語版があったのでそれと対照させながらテキストを読むことができた。この読書から生まれたのが、拙稿「ポパーの反証主義の背景としてのマイモニデスの否定神学」である。社会思想史学会、『社会思想史研究』、第 11 巻、1987 年、93-103 ページ。修正版は、ポパー哲学会編、『批判的合理主義』第 1 巻、2001 年 9 月、未来社、所収。科学哲学の分野で、論文名に「マイモニデス」の名前が用いられている論文はおそらく後にも先にも私のだけだと思う。因みに、テルアヴィヴ大学のユダヤ哲学科では、マイモニデスに関する講義は学部 1 年次の必

修である。イスラエル内外のユダヤ社会では、マイモニデスは今日でも重要不可欠な思想家である。因みに、マイモニデスは、カントの第一のアンティノミーに先んじた議論を行っており、哲学史でもっと論じられてよい思想家である。

[10] I. C. Jarvie and Nathaniel Laor, The Philosopher as All-Rounder, I. C. Jarvie and Nathaniel Laor eds., *Critical Rationalism, Metaphysics and Science*, Kluwer Academic Publishers, 1995, pp. xi-xxii.

[11] J. O. Wisdom, The Nature of 'Normal' Science, P.A. Schilpp, ed., The Philosophy of Karl Popper, Open Court, 1974, p. 820.

[12] J. Agassi, *The Continuing Revolution: A History of Physics from The Greeks to Einstein*, McGraw Hill, 1968.

[13] J. Agassi, *Towards a Rational Philosophical Anthropology*, Kluwer, 1977.

[14] J. Agassi, *Science and Society: Studies in the Sociology of Science*, Boston Studies in the Philosophy of Science, Vol. 65, 1981.

[15] J. Agassi, *Science and Culture*, Boston Studies in the Philosophy of Science, Vol. 231. 2003.

[16] J. Agassi, *The Gentle Art of Philosophical Polemics: Selected Reviews and Comments*, Open Court, 1988.

[17] 因みに、「哲学は、講壇哲学ではなく市井の哲学でなければならない」が高木先生の口癖で、かれもまた権威主義的で閉じたアカデミズムに反対であった。

[18] J. Agassi, *Technology: Philosophical and Social Aspects*. Kluwer, 1985.

[19] J. Agassi, *The Siblinghood of Humanity: Introduction to Philosophy*, Caravan Press, 1990.

[20] J. Agassi, *Radiation Theory and the Quantum Revolution*, Birkhäuser, 1993.

[21] J. Agassi, *A Philosopher's Apprentice: In Karl Popper's Workshop*, Series in the Philosophy of Karl R. Popper and Critical Rationalism, Rodopi, 1993.

[22] J. Agassi, *Liberal Nationalism for Israel: Towards an Israeli National Identity*, Gefen, 1999. 拙稿、「民族主義の倫理的一考察---K. R. ポパ-の民族問題に関する発言を手掛かりとして---」、『秋田大学教育学部研究紀要』、第 38 集、1988 年、23-32 ページも参照。因みに、この拙稿は、1987 年イスラエル大使館エッセイコンテスト入賞作品の加筆・修正版である。アガシは、イスラエルをユダヤ人国家にするのではなく、イスラエルがより民主的でリベラルなナショナリズムをめざすよう提案している。同様に、私は、拙稿で、ポパーの開かれた社会と両立可能な民族主義を、両立不可能な「全体主義的民族主義（Totalitarian Nationalism）」と対比させて、「個人主義的民族主義（Individualistic Nationalism）」と呼び、後者が、ポパーが看過したリベラルな民族主義であると主張した。2000 年に来日した際、アガシは秋田大学も訪れたが、私の研究室で、テーブルに置いてあった拙稿をたまたま見つけて直ちに Abstract に目を通し、ナショナリズムについて「われわれは同様の見解のようだ」というコメントをしてくれた。拙稿は、以下のアドレスからアクセスして読むことができる。
https://air.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=1175&item_no=1&page_id=13&block_id=21

[23] J. Agassi, *Popper and his Popular critics: Thomas Kuhn, Paul Feyerabend, Imre Lakatos,* Springer Briefs in Philosophy, 2014. 本書には、本題と同名の項目の下にいくつかの論文が掲載されているが、その前に置かれた第 4 章の論文の元論文の邦訳がすでにある。拙訳、J. アガシ、「通俗的なポパー批判者たち：クーン、ファイヤアーベント、ラカトシュ」日本ポパー哲学研究会、

『批判的合理主義研究』、Vol. 2, No. 1, 2010年6月、7-23ページ。

[24] J. Agassi, *The Hazard Called Education by Joseph Agassi: Essays, Reviews, and Dialogues on Education from Forty-five Years*, edited by Ronald Swartz and Sheldon Richmond, Springer, 2014.

[25] J. Agassi, *Ludwig Wittgenstein's Philosophical Investigations: An Attempt at a Critical Rationalist Appraisal*, Synthese Library, 401, Springer, 2018.

[26] Stefano Gattei and Bar-am Nimrod eds., *Encouraging Openness: Essays for Joseph Agassi on the Occasion of His 90th Birthday*, Boston Studies in the Philosophy and History of Science, Springer, 2017.

[27] Kiichi Tachibana, How can we attain both democracy and constitutionalism?, *Encouraging Openness: Essays for Joseph Agassi on the Occasion of His 90th Birthday*, 2017, chapter 26, pp.305-318. この拙稿の私訳「民主主義と立憲主義」は以下のアドレスからアクセスして読むことができる。https://f3cf63f9-a5fa-47a7-8b5b-859966dd92ef.filesusr.com/ugd/070434_b64ba648fec04c009264d7cdece17cff.pdf

ポパーの相互作用主義における存在論：心の主観性と実体性を超えて

池田健人（大阪大学）*

1. 序論

本論文では、心の主観性と実体性をめぐる批判に対して、ポパーの相互作用主義にみられる存在論を明示することで応答する。ポパーが心身問題へ本格的に着手し始めたのは、のちにポパーと共同研究を実施することとなる神経生理学者のエクルズが、オーストラリアから数回にわたりポパーのもとを訪れた 1952 年のことである。翌年には、ポパーの手になる心身問題についての最初の論文「言語と身心問題：相互作用主義の再説」（1953）が発表された。自身の仕事を振り返るなかで、「『開かれた社会とその敵』と『歴史主義の貧困』という合間ののち、わたしは他の諸問題、まずは論理（自然演繹の論理）、そして合理主義の歴史とその今日的な諸脅威の問題へと向かった。その少しあとには、1953 年、わたしは心身問題を論じた」（Popper 1974: 1068）と、ポパーは回想している。ポパーにとって、公式には、これが相互作用主義の出発点である。

もっとも、1952 年より前から、ポパーは心身問題に興味をもってはいた。しかし、「心身問題との最初の遭遇は、わたしをして長年のあいだ、これはとても解決の望みのない問題だという感じを抱かせてきた」（Popper 1976 [2002]: 219）。ところが、時代は 20 世紀も中葉へとさしかかり、ポパーの言語論が進化論に結びつくと、それ以来、状況は一変する。たちまち、ポパーは物理主義へと反旗を翻し、二元論を擁護する徹底的な相互作用主義者となった。社会哲学と科学哲学につづき、心の哲学において、ポパーは第三の大きな貢献を果たした、とニーマン（Niemann 2012）は述べている。もっとも広範で詳細な相互作用主義の説明は『自我と脳』（1977）で試みられた、とみなされるのが一般的である（cf. Århem 2021; Puccetti 1985）。そのあとも、1994 年の 9 月に逝去する間際まで、ポパーは相互作用主義の立場から心身問題を論じつづけた。

物理主義を筆頭に、ポパーの相互作用主義は、これまでさまざまな方面から批判されてきた。たとえば、その後景をなす形而上学が深刻な欠陥を抱えていることや、科学の成果が真摯に受け止められていないこと、あるいは正しく理解されていないことなどが、ポパーの相互作用主義には指摘されている（Albert 1977; Keuth 2000; Rooijen 1987）。しかし、そうした批判のなかには、妥当なものばかりでなく、ポパーの相互作用主義を十分に検討することなく提出されているものも多い。本論文では、とりわけ存在論の観点から、心の主観性と実体性にかかわる批判が、ポパーの意図を十分に汲んでいないということを、ポパーのテクストに即しながら論証する。

まずは、ポパーの相互作用主義を簡単に説明しておこう。そこで主張される心身相互作用論は、ポパーに独自の進化論（能動的ダーウィン主義）と言語論（言語の四機能説）から導出されたものである。ポパーは、言語における意味内容の進化、および意味内容がもつ生存価に着目し、意味内容と身体の関係は相互作用だと帰結したうえで、それらが相互作用するためのシステムが心だとみなした。言語論的な意味内容の次元までをも包摂して進化論的に考察されるとき、はじめて心と身体の関係は理解しうるものとなる、というのがポパーの見立てである（第 2 節）。

心の主観性が定義されないかぎり、いかなる心身相互作用論も成立しえず、よってポパーの心身相互作用論も無効だとする批判がある。この批判によれば、心の主観性を定義

* 本論文に助言をくださった 2 名の匿名査読者に感謝いたします。

することは、あらゆる形態の心身相互作用論に求められる要件だが、その要件は本質的に充足されえない。そうだとすれば、すべての心身相互作用論は不可能であり、ポパーの心身相互作用論も幻想である。しかし、その批判は、あくまでも相互作用の認識論に困難を指摘するものであるにすぎず、すべての心身相互作用論を完全に放逐できるほど強力なものではない。本論文では、そのような批判が、相互作用の存在論的な性格をもつポパーの心身相互作用論に向けられるのは誤りだと主張する（第３節）。

身体からは独立の心を論じていながら、心の実体性を語らないポパーの心身相互作用論は空疎だとする批判もある。この批判によれば、およそ状態なるものはすべて実体の状態であり、ある何らかの状態を調査するためには、その状態にある実体が吟味されなければならない。それにもかかわらず、心的実体には言及することなく、ひたすら心的状態の力説にばかり集中するポパーの心身相互作用論は脆弱である。しかし、その批判は、ただ過程のみが存在するというポパーの定立を等閑にしているため、ポパーの心身相互作用論を適切に反駁できていない。本論文では、そのような批判にとっては、プロセス存在論を追及することが何よりも先決だと主張する（第４節）。

このことから、本論文では、ポパーの心身相互作用論には相互作用の存在論、および過程の存在論が組み込まれているために、心の主観性や実体性を重視する批判は、ポパーの心身相互作用論を転覆させられるほど決定的な批判ではないと結論する。ただし、これはポパーの心身相互作用論に積極的な支持を与えるものと誤解されてはならない。あくまでも、本論文の目的は、心の主観性や実体性にもとづくポパーへの批判における過言を指摘することを通じて、ポパーの心身相互作用論を正しく把握することにある。なぜなら、それはポパーの哲学を批判するうえで不可欠の前提だからである。ポパーの哲学を擁護するのではなく批判するにしても、その企ては、つねにポパーの意図をできるだけ忠実に捉えることを原点とするものでなければならない。以下では、ポパーが構想していた心身相互作用論の大略を確認することから始めよう。

2. ポパーの心身相互作用論

ダーウィン主義的な進化論は、無作為な突然変異と自然淘汰を通じて生存競争に従事する生物の理論として知られている。生物を取り巻く環境は、その環境に適合しない変異を淘汰することによって、いわば生物を彫刻する。このとき、生物に対する環境の作用は、ほとんど一方的なものである。進化の主要な原動力は環境による淘汰であり、変異の偶成という消極的な関与を除いては、生物が進化に貢献することはない。すべての生物は、ただ環境に翻弄されるのみである。このようなダーウィン主義を、ポパーは受動的ダーウィン主義と呼んでいる。

ポパーによれば、受動的ダーウィン主義は不十分である。「わたしが敵対的な環境の理論、または受動的ダーウィン主義と呼んでいるものに、わたしが友好的な環境の理論、または探索的ダーウィン主義、能動的ダーウィン主義と呼んでいるところの、第二の補完的な理論を加えなければならない」（Popper 1982: 39）。能動的ダーウィン主義とは、新たな生態学的ニッチや生活条件、行動様式などを求めて、自身の置かれた環境を積極的に探索する生物のダーウィン主義である。生物の進化を正しく分析するためには、能動的ダーウィン主義が考慮されなければならない、とポパーは考えている。

能動的ダーウィン主義を人間に適用するとき、そこには言語の叙述機能と論証機能が見出される。これら２つの機能は、ポパーが提唱した言語の四機能説に由来している。言語の四機能説によれば、人間言語には４つの

主要な機能がある（e.g. Popper 1972 [1979]）。それは、もっとも基礎的なものから順に表出機能、信号機能、叙述機能、論証機能である。表出機能と信号機能は動物言語と人間言語に共通の機能（言語の低次機能）であり、叙述機能と論証機能は人間言語のみに固有の機能（言語の高次機能）である。人間は、動物とは異なり、言語の高次機能を利用して環境の探索に従事している。

言語の高次機能を備えた高次言語は、意味内容と結びついている。いいかえれば、高次言語を手に入れたことにより、「わたしたちは、理論を定式化したり表現したりするさまざまな様式を抽象化し、その不変の内容または意味に注意を払うすべを学んだ」（Popper 1972 [1979]: 240、強調は原著者）。意味内容は物理学へ還元されえないので、「人間言語についての物理主義的な因果理論は成り立ちえない」（Popper 1963 [2002]: 395）。それゆえ、高次言語と人間（身体）の関係は、物理主義の枠内では決して解明されえない、というのがポパーの見立てである（cf. Searle 1992）。では、高次言語と身体の関係は、いったいどのようなものなのだろうか。

高次言語は、目や耳、足などの身体内的な器官と同じように、ダーウィン主義的な進化の産物である。ポパーの構想は、「言語の叙述機能と論証機能は、他の動物におけるより単純な表出機能と伝達機能から、人間において進化した」（Boyd 2016: 229）というものであった。このことについて、ポパーは高次言語としての理論を例に説明している。「理論はわたしたちが身体外に発達させる器官であり、器官はわたしたちが身体内に発達させる理論である」（Popper 1968: 163）。生存競争に従事するなかで、人間は言語という器官を身体外に発達させる道を選んだ。「わたしたちは、人間の理論的構成物 …… を身体外における突然変異のようなものとして、あるいは「身体外的突然変異」と呼ばれるものとしてみなすことができる」（Popper 1994: 13、強調は原著者）。ポパーにとって、あくまでも高次言語はダーウィン主義的な原理の埒内にある。

してみれば、意味内容、それゆえ高次言語と身体の関係は相互作用である。ダーウィン主義的な進化の産物であるかぎり、「言語のより高度な次元は、２つの事柄――わたしたちの言語のより低度な次元と、わたしたちの環境への適応――をよりよく制御する必要の圧力のもとで …… 進化した」（Popper 1972 [1979]: 240、強調は原著者）。そのうえで、言語のより高度な次元（高次言語）は自然淘汰の脅威に耐えている。これは、高次言語が人間の生存に対する有用性を備えているということである。すなわち、高次言語が人間に対して非効果的だということはありえない（cf. Richards 1987）。このように、ダーウィン主義的な進化論の標準的な理路に照らすとき、生存の見地から身体を制御する仕方で、高次言語は人間へと有益に反作用している。

身体は高次言語を創造し、高次言語は身体を制御する。こうした相互作用の関係を仲介する機構（システム）として要請、あるいは発見されたのが、ポパーのいう心である。ポパーは、心そのものを探究したのではなかったが、高次言語と身体が相互作用の関係にあるのなら、それらの中間には何らかのシステムが存在している必要があり、それが心だと判断するに至った。高次言語からの作用を受けて、「意識の状態（「心」）は身体を制御し、身体と相互作用する」（Popper 1972 [1979]: 251、強調は原著者）。ポパーによれば、心とは高次言語と身体を連繋するシステムである。

はたして、環境の人間による能動的な探索を考察することから、ポパーに独自の心身相互作用論が導かれる。身体（物理運動）が心（心的状態）を創造するとともに、「わたしたちの心的状態はわたしたちの物理運動（のあるもの）を制御するものであり、心的活動と

有機体の他の諸機能とのあいだにはあるやりとり、あるフィードバック、それゆえある相互作用がある」(Popper 1972 [1979]: 251–2、強調は原著者)。このような洞察から、ポパーは心身相互作用論を標榜している。

3. 相互作用の存在論

ムンツ(Munz 2007)によれば、主観性を定義できないかぎり一般に心身相互作用論は不可能である。なぜなら、心は本質的に主観性を伴うものだからである。すなわち、主観性を定義することなしに心は解明されえない。心身相互作用論は、心と身体の作用関係についての理論である。心と身体の関係が明示されるためには、その両者が十分に精査できるものでなければならない。バンジも述べているように、心と身体の「相互作用は具体的な事物に対してのみよく定義される」(Bunge 1980: 83)。してみれば、主観性の定義を通じた心の解明なしに心身相互作用論はありえない。心身相互作用論を提案するための条件は、事前に主観性を定義することだとムンツはいう。

ところが、主観性を定義することはできない。なぜなら、定義とは、ふつう言語を手段として対象となる事物に適用されるものであるが、主観性は本質的に言語化されえないからである。「身体内の主観性は主観的なものであり、それゆえ純粋に私的な言語であらわされることができないので、明確に定義できるものではない」(Munz 2007: 308)。公的な言語どころか、その前提となる私的な言語によってさえ、主観性は表現されえない。これは、どのような方法をもってしても、主観性は非言語性を本性とする対象なので、主観性が定義されることは決してありえない、ということである。

定義とは、主観性を客観性へと変容させる、あるいは主観性を客観性に代替する操作である。いかなる形であれ、ともかくも言語的に表現された瞬間、主観性は主観性であることをやめる(cf. Svensson 1994; Munz 1997)。主観性とは、ムンツによれば、形のない感覚の流動性、または生の感覚がもつ属性である。内観するとき、わたしたちは生の感覚に接触することができる。ただし、さらに進んで、生の感覚を理解し、捕捉しようとするとき、わたしたちは生の感覚に名前をつけなければならない。つまり、いわば言語の鋳型へと生の感覚を流し込み、定義することなしには、内観を通じて、わたしたちが内面の何かを把握できるようになることはないのである。このとき、その把握される内面の何かは、ある特定の形を言語により与えられ、定義された客観性であり、もはや生の感覚、それゆえ主観的対象ではなく、そこに主観性はない。

したがって、ムンツによれば、ポパーの心身相互作用論は誤りである。ムンツにいわせれば、主観性が定義できないのにもかかわらず、ポパーが心身相互作用論を提案できるのは奇妙である。ポパーが心と身体の相互作用を宣言できたのは、たんにポパーが心(意識)の問題を片手間に扱っており、真剣に考えていなかったからだとムンツはみている。「意識についてのポパー自身の考えは思いつきのもので、ほとんど偶然的なものであり…… 修正されなければならない」(Munz 2007: 307)。主観性が特定できないのなら、心の全容を解明することはできない。してみれば、ムンツにとって、心身相互作用論は成立しえないと判断するのが妥当である。だが、その批判では、心身相互作用の認識論と存在論が混同されている、と本論文では主張する。

たとえ主観性は定義できないという指摘が正しかったとしても、そこから言えるのは、心身相互作用論は世界を適切に記述できていないということではなく、せいぜい心身相互作用の機序についての知識を得ることはできない(認識論)、ということまでである。主観性なしの心はありえないので、主観性が定義できないということは、そのまま、心を三人称的な概念として認識することはでき

ない、ということである。いいかえれば、心は、ただ一人称的な知覚としてのみ認識されうる。そうすると、心身相互作用の機序を解明することは原理的に不可能である。

なぜなら、心と身体が相互作用する機序の解明は、概念的な営為だからである。このことは、心身相互作用の機序を解明する試みにとって、あらかじめ心の概念を認識しておくことは避けられない、ということを意味している。心の概念が認識されえないのなら、心身相互作用の機序は解明されえない。もちろん、心身相互作用の機序に主観性はまったく関与していない可能性もある。しかし、たとえそうだとしても、何らかの機序の解明をもって心身相互作用の機序が解明されたと宣言するためには、やはり心の概念的な認識を看過して済ますことはできない。さもなければ、ただ適当な対象に心というラベルを貼り付け、その対象と身体の相互作用を分析するだけで、心身相互作用を説明したと言い張る議論を許容するということにもなってしまいかねない。心の概念を認識することができないかぎり、いつまでも心身相互作用の機序は解明されないままである。

それゆえ、ポパーが唱える心身相互作用論のなかでも、もしあるのなら、心身相互作用の機序へ言及する側面に対して、ムンツの批判は有効でありうる。たとえば、心は意識（心理的な力）と無意識（物理的な力、とくに電磁力の場）から構成されており、それらは注意を起点として相互作用することができる、とポパーは述べている（Popper 1982; 1989 [1979]; Popper et al. 1993）。一見したところ、ここでは心身相互作用の機序が言及されているようにみえる。ポパーは、注意という主観的対象の機能を推測しているだけであり、注意が相互作用を引き起こすように機能する仕組み、つまり心身相互作用が生じるための機序までは推測していない、という見方もある。しかし、もしこれが心身相互作用の機序に言及するものだと認められうるのなら、ムンツの批判は、そうしたポパーの主張を退けることはできるかもしれない。

とはいえ、明らかに、それは心身相互作用の現象が事実ではない（存在論）、ということではない。ポパーの心身相互作用論は、必然的に主観性を伴う心それ自体ではなく、相互作用する心、あるいは心と身体の相互作用という客観的な事態の存在論を基礎にしている。ムンツの批判は、もしそれが存在論の否定を意図しているものであるのなら、実際そうであるのだが、まったくの筋違いである。それにもかかわらず、ムンツは心身相互作用の認識論（心身相互作用の機序にかんする知識の理論）から存在論（心身相互作用の現象にかんする事実の理論）へと飛躍し、そのうえダマシオの身体信号仮説（e.g. Damasio et al. 1991; Damasio 1994）を代案に支持することで、ついにはポパーが主張した心身相互作用論を放棄するに至っている。

心身相互作用の機序にかんする知識は、心身相互作用の現象にかんする事実から分離されなければならない。なぜなら、すべての対象は存在しているのなら必ず認識されるという保証はどこにもないからである。心身相互作用の現象が事実であり、存在しているとき、その現象、とりわけ心身相互作用の機序にかんする知識を手に入れられるかどうかは、心を概念として認識することはできるか、したがって主観性を言語化、それゆえ定義することはできるかに依存している。たとえ心身相互作用が存在しているとしても、心が一人称的な主観性を本質にするものであるのなら、それを三人称的な概念として認識することはできず、心身相互作用の機序を特定することもできない。心身相互作用は、その機序が原理的に解明されえない場合でさえ、現象している可能性がある。

ムンツの誤りは、「相互作用が考えられるものであるとともに生じるためには、ある人の身体がもつ主観的状態は、自分が心とどのように相互作用するかをその人が決定でき

るように、定義できるほど明確でなければならない」（Munz 2007: 308）という発言に集約されている。たしかに、心身相互作用を考えるうえで主観的状態の定義は障害になりうる。とはいえ、心身相互作用が生じるために主観的状態が明確である必要はない。すなわち、心身相互作用は現象しているが、それを概念として把握することはできない、ということはありうる。わたしたち人間が心身相互作用について考えられることと、心身相互作用が実際に生じていることは別である。

ポパーが主張しているのは、心身相互作用の存在論である。つまり、心と身体は事実として相互作用している、というのがポパーの仮説である。もちろん、その事実が現象する機序の究明は、心身相互作用論の課題である。しかし、その課題が依然として残されてはいるのだとしても、あるいはムンツがいうように、その課題を解決することは原理的にできないということが真であるのだとしても、だからといって、ただちに心身相互作用論はまったく不可能だということにはならない。たしかに、こうした消極的な主張は、ただ感情や願望にのみもとづくものであったり、とくに偶然の思いつきであったりするのなら、ほとんど学術的には価値のないものということになるだろう。だが、ポパーの心身相互作用論は、決して偶然の思いつきではない。

そこには、ダーウィン主義という要石がある。ダーウィン主義的な進化論の標準的な理路に従えば、その高い生存価を示唆する高次言語の歴史的な状況からして、心と身体の関係を相互作用だと推定するのは合理的である。ワトキンズも指摘しているように、心（意識）それ自体ではなく、「意識的制御の出現に対する説明を提供」（Watkins 1974: 397）することによって、ポパーは心と身体の相互作用を議論しようとした。ポパーの心身相互作用論において、相互作用する心の問題は「心あるいは意識の進化について、またそれによって心の機能について」（Popper 1972 [1979]: 250）記述することで解決される。心それ自体がもつ主観性を特定することはできなくても、ダーウィン主義的な進化論に依拠することから、言語と身体の分析を通じて、相互作用する心、あるいは心と身体の相互作用という客観的な事態の存在を推察することはできる。

主観性が特定できないことを根拠にするだけでは、心身相互作用論を全面的に却下することはできない。ポパーの主張を偶然の思いつきだというムンツの誤りは、心身相互作用の認識論を心身相互作用論の要件だとする謬見に起因している。なるほど、主観性に着目するムンツの指摘は、心身相互作用論の難点を的確に示してはいる。しかし、それはポパーの心身相互作用論を謬説と断定するための根拠にはなりえない。認識できない心身相互作用は存在しない、というのは間違いである。ポパーの心身相互作用論を拒否するためには、ダーウィン主義的な進化論や機能主義的な言語論の観点から、心身相互作用の存在論が批判されなければならない。

4. 過程の存在論

バンジ（Bunge 1980）によれば、すべての状態は実体の状態である。いいかえれば、およそ状態と呼ばれるものは、抽象的なもの（抽象的対象）ではなく、何であれつねに具体的なもの（具体的対象）の性質である。その意味で、具体的対象（実体）の性質、すなわち実体的性質が状態である、とバンジはいう。たとえば、水の状態は水分子という実体、光の状態は波動ないし粒子という実体の性質として存在している。バンジにとって、それらは状態それ自体として存在しているわけではない。「物理学から生物学、社会学に至るまで、すべての科学において、性質は具体的なもの ..... に備わっており、事象とはある性質の変化である（わたしたちはもちろん、抽象的対象の性質ではなく、実体的性質とその変化について書いている）」（Bunge 1980:

20)。ここで述べられているのは、事象は性質（状態）の変化であり、状態の変化は実体の変化だということである。

ポパーの心身相互作用論は、心的状態（心）と物的状態（身体）の相互作用を主張する理論である。これは、ダーウィン主義的な進化論からの帰結であった。「自然淘汰の理論が心と身体 …… より適切には心的状態と物的状態のあいだの相互作用の教説に賛成する強力な論証を供与する」（Popper 1978: 350、強調は原著者）。ポパーは、心的状態という表現を自覚的に選択している。「物理的な対象や状態のほかに心的状態というものがあり、その状態はわたしたちの身体と互いに相互作用する」（Popper & Eccles 1977 [1983]: 36、強調は原著者）。ポパーにとって、心的状態が存在していることには積極的な意義がある。「わたしは主観的経験、心的状態、知性、精神の存在を否定するものではない。これらのものはきわめて重要なものだとさえ、わたしは信じている」（Popper 1976 [2002]: 159–60）。非物理的な心的状態は進んで容認されている。

そうすると、バンジによれば、ポパーは心的実体を詳解できるのでなければならない。あらゆる状態が実体の状態であるのなら、心的状態は、物的実体の状態ではないかぎり、それ以外の何らかの実体、いわば心的実体の状態だということになる。事実上、「心は身体と相互作用するにもかかわらず身体からは独立の対象または実体である、という古代のテーゼに賛成する議論をポパーは模索している」（Bunge 1981: 140）。してみれば、その実体を明示することのないままに、いくら心的状態が存在するといっても、ただそれを繰り返すばかりでは空虚である。「実体のない心は存在しない」（Bunge 1977: 508）。バンジにとって、身体（物的実体）の状態でない心的状態を論じるためには、その基底をなす心的実体についての説明が不可欠である。

しかし、その批判は、ポパーの実体論を適切に踏まえたものではない、と本論文では主張する。ポパーが心の実体説を拒否することは、バンジがポパーの心身相互作用論を批判する以前から、これまで幾度となく宣伝されてきた。たとえば、ある場所では「わたしは（「実体」について語るべきだとは決して考えなかったけれども）つねにデカルト的な二元論者であった」（Popper 1976 [2002]: 218）とか、また別の場所では「もちろん、相互作用の教説がまったく古めかしいものであることは重々承知している。それにもかかわらず、わたしは相互作用と古めかしい二元論を（ただし、わたしは「実体」と呼ばれるものの存在は拒絶するけれども）擁護することを提案する」（Popper 1978: 351、強調は原著者）などと、ポパーは発言している。

心的実体のみならず、一般に実体の概念を放棄するのがポパーの立場である。物的実体もまた、ポパーにとっては存在しない。心だけが特別に実体をもたないのではなく、「実体という観念そのものが誤りにもとづいている」（Popper & Eccles 1977 [1983]: 105）。長いあいだ科学、とりわけ唯物論において中枢的な地位を占めてきた実体の概念は、物理学の発展を通じて克服されてしまった、とポパーはいう。

> 物質は保存されないゆえに「実体」ではない。それは破壊されうるし、また創造されうるのである。…… 物質は高度につまったエネルギーで、エネルギーの他の形式に変換可能であり、したがって光、そしてもちろん運動、熱などのような他の過程に変換できるゆえに、過程という本性をもつものである …… ／このように現代物理学の成果は、実体や本質の観念を捨て去るべきであることを示唆しているといえる。それらの成果は、（いくらかの物質は「日常的な」状態のもとでもそうであるが）時間上のあらゆる変化を通じて持続する自己同一的な

> ものが存在しないことを、そして対象の性質や属性を持続して担い、保持する本質は存在しないことを示唆している。宇宙はいまや対象の集まりではなく、事象や過程の相互作用する集まりのように思われる …… ／それゆえ、現代物理学者が物理的対象――物体、物質――は原子構造をもっていると述べるのはもっともなことであろうが、原子はそれ自身の構造をもっており、その構造を「物質的なもの」として描くことはほとんどできないし、まして「実体的なもの」としてはまったく描けない。結局、物質の構造を説明するプログラムによって、物理学は唯物論を超越しなければならなかったのである（Popper & Eccles 1977 [1983]: 7、強調は原著者）。

実体を却下したうえで、ポパーが存在するというのは過程である。物質でさえ実体ではない。光や運動、熱などと同じように、物質もまたエネルギーの形式である。従来、すべての対象は、いずれ何らかの究極的なモノ（実体）へと還元されうると考えられてきたが、そのような対象の見方は素朴な物理学の遺物であるにすぎず、じつのところ物質とは、その微視的な構造までをも精密に分析してみると、さまざまな形式へと絶えず変化しつづけるエネルギーの局面的なプロセス（過程）だということが、現代物理学の成果により判明した。いみじくも、ニーニルオートは、ポパーの哲学は「常識や科学と両立する動的なプロセス存在論に相当する」（Niiniluoto 2006: 60、強調は原著者）と断言している。もはや、実体に固執する理由は何もない。ポパーによれば、心について、心的実体などのようなものはまったく存在せず、ただ心的状態という過程のみが存在している。

すべての存在はプロセスだという洞察のもと、その心身相互作用論において、ポパーは物心プロセス二元論を採用している。宇宙をモノ、つまり実体の集合だと考える場合、その集合が物質のみで構成されているとみなすなら物的実体一元論（唯物論）、精神（心）のみで構成されているとみなすなら心的実体一元論（唯心論、ないし観念論）、あるいはデカルトのように物質と精神により構成されているとみなすなら物心実体二元論へと、それぞれ分類される。このとき、ポパーは、宇宙をプロセス、つまり過程の集合だと考えたうえで、その集合は物質と精神からなるとみなしているため、その立場は物心プロセス二元論だといえる。過程の一元性が真だとする存在論の定立を背景に、そこへ物的過程と心的過程の二元性を組み込むのが、ポパーの立場である。

プロセス二元論が認められるのなら、実体の概念は幻想だったということになる。人間もまたエネルギーの形式、つまりエネルギーが四次元的な変化の過程（プロセス）でとりうるひとつの部分だということが自覚されなければならない。「生命はモノではなくプロセスである。わたしたちは、わたしたち自身をモノとして考えすぎている。わたしたちはモノではなくプロセスである。…… 生命は高分子ではない。生命は明らかにプロセスである。それは進行中のプロセスである」（Popper et al. 1993: 169）。とりわけ、人間のように意識的な生命は、心理過程（あるいは状態）と物理過程（あるいは状態）が統合したエネルギーの形式であるということが推測される。「わたしたちは実体であるというよりは、むしろ心理物理過程である」（Popper & Eccles 1977 [1983]: 105）。このように、ポパーは、ひとえに過程のみからなる宇宙の描像を支持している。

いかなる状態も実体の状態ではない。というのも、実体は存在しないからである。バンジは具体的なもの（実体）とその状態（過程）を区別していたが、ポパーにいわせれば、世界に充実するすべてのものはみな過程、それゆえ状態であり、実体などのようなものはま

ったく存在しない。心的実体が説明されない点に異議を申し立てるのなら、まずバンジはポパーの実体論、それゆえプロセス存在論を批判する必要がある。すでにポパーが反論した実体説を、相変わらず復唱しつづけるだけでは、バンジの主張にポパーの心身相互作用論への貢献を見出すのは難しい。ポパーのプロセス存在論がなぜ誤っているのかを指摘できたとき、はじめてバンジの主張は有意義なものとなる。

5. 結論

本論文では、心の主観性と実体性を足場に構成される批判は、ポパーの意図を誤解したものであり、ポパーの心身相互作用論を失脚させる試みとしては不十分だということを示した。まず、心の主観性が定義できないという批判は、相互作用の存在論を否定しうるものではない。それゆえ、心の主観性だけを根拠にすべての心身相互作用論を拒絶し、よってポパーの心身相互作用論も受け入れられないとする議論は間違いである。つぎに、心の実体性が説明されていないという批判は、それが実体の概念を素朴に前提しているだけのものであるかぎり、ポパーの心身相互作用論を効果的に糾弾しうるものではない。ただ過程のみが存在するというプロセス存在論に立脚しているという事情を考慮することなく、実体の観点からポパーの心身相互作用論を徒事だとして追及する議論は不毛である。ポパーの心身相互作用論に対する心の主観性や実体性にもとづいた批判は、いずれも失敗している。

一次文献

Popper, K. 1963 (2002). *Conjectures and Refutations: The Growth of Scientific Knowledge*, London and New York: Routledge Classics.（『推測と反駁：科学的知識の発展』、藤本隆志・石垣壽郎・森博訳、法政大学出版局、1980 年）。

Popper, K. 1968. 'Is There an Epistemological Problem of Perception?', *Problems in the Philosophy of Science*, ed. Lakatos, I. and Musgrave, A., pp. 163–4, Amsterdam: North-Holland Publishing Company.

Popper, K. 1972 (1979). *Objective Knowledge: An Evolutionary Approach*, Oxford: Clarendon Press.（『客観的知識：進化論的アプローチ』、森博訳、木鐸社、1972 年）。

Popper, K. 1974. 'Replies to My Critics', *The philosophy of Karl Popper*, ed. Schilpp, P. A., pp. 961–1197, La Salle, Illinois: The Open Court.

Popper, K. 1976 (2002). *Unended Quest: An Intellectual Autobiography*, London and New York: Routledge Classics.（『果てしなき探求：知的自伝』、森博訳、岩波書店、1978 年）。

Popper, K. 1978. 'Natural Selection and the Emergence of Mind', *Dialectica* 32(3/4): 339–55.

Popper, K. 1982. 'The Place of Mind in Nature', *Mind in Nature: Nobel Conference XVII Gustavus Adolphus College, St. Peter, Minnesota*, ed. Elvee, R. Q., pp. 31–59, Harper & Row.

Popper, K. 1989 (1992). *In Search of a Better World: Lectures and Essays from Thirty Years*, tr. Bennett, L. London and New York: Routledge.（『よりよき世界を求めて』、小河原誠・蔭山泰之訳、未來社、1995 年）。

Popper, K. 1994. *Knowledge and the Body-Mind Problem: In Defence of Interaction*, ed. Notturno, M. A., London and New York: Routledge.

Popper, K. & Eccles, J. C. 1977 (1983). *The Self and Its Brain*, London and New York: Routledge.（『自我と脳（上・下）』、

西脇与作・大村裕訳、思索社、1986 年)。
Popper, K., Lindahl, B. I. B. & Århem, P. 1993. 'A Discussion of the Mind-Brain Problem', *Theoretical Medicine*, 14(2): 167–80.

二次文献

Albert, H. 1977. *Kritische Vernunft und menschliche Praxis: Mit einer Autobiographischen Einleitung*, Stuttgart: Philipp Reclam.

Århem, P. 2021. 'Popper on the Mind-Brain Relation', *Karl Popper's Science and Philosophy*, ed. Parusniková, Z. and Merritt, D., pp. 279–94, Springer.

Boyd, B. 2016. 'Popper's World 3: Origins, Progress, and Import', *Philosophy of the Social Sciences* 46(3): 221–41.

Bunge, M. 1977. 'Emergence and the Mind', *Neuroscience* 2: 501–9.

Bunge, M. 1980. *The Mind-Body Problem: A Psychobiological Approach*, Pergamon Press.

Bunge, M. 1981. *Scientific materialism*, D. Reidel.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*, Vintage. (『デカルトの誤り : 情動、理性、人間の脳』、田中三彦訳、筑摩書房、2010 年)。

Damasio, A. R., Tranel, D. & Damasio, H. C. 1991. 'Somatic Markers and the Guidance of Behavior: Theory and Preliminary Testing', Frontal Lobe Function and Dysfunction, ed. Levin, H. S., Eisenberg, H. M. & Benton, A. L., pp. 217–29, New York: Oxford University Press. (「ソマティック・マーカーと行動指針 : 理論と予備的検証」、北川玲訳、『感情』、梅田聡・小嶋祥三監修、pp. 285–306、岩波書店、2020 年)。

Keuth, H. 2000. *Die Philosophie Karl Poppers*, Tübingen: Mohr Siebeck.

Munz, P. 1997. 'The Evolution of Consciousness: Silent Neurons and Eloquent Mind', *Journal of Social and Evolutionary Systems* 20(4): vii–xxviii.

Munz, P. 2007. 'The Phenomenon of Consciousness from a Popperian Perspective', *Consciousness Transitions: Phylogenetic, Ontogenetic, and Physiological Aspects*, ed. Århem, P. and Liljenström, H., pp. 307–26, Elsevier Science & Technology.

Niemann, H. 2012. 'Geist als ein Kraftfeld: Bemerkungen zu Karl Poppers Ideen von 1991', *Aufklärung und Kritik* 44: 12–22.

Niiniluoto, I. 2006. 'World 3: A Critical Defence', *Karl Popper: A Centenary Assessment, Volume II: Metaphysics and Epistemology*, ed. Jarvie, I., Milford, K. and Miller, D., pp. 59–69, College Publications.

Puccetti, R. 1985. 'Popper and the Mind-Body Problem', *Popper and the Human Sciences*, ed. Currie, G. and Musgrave, A., pp. 45–55, Martinus Nijhoff Pblishers.

Richards, R. J. 1987. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, University of Chicago Press.

Rooijen, J. V. 1987. 'Interactionism and Evolution: A Critique of Popper', *The British Journal for the Philosophy of Science*, 38(1): 87–92.

Searle, J. R. 1992. *The Rediscovery of the Mind*, MIT Press.

Svensson, G. 1994. 'Reflections on the Problem of Identifying Mind and Brain', *Journal of Theoretical Biology* 171(1):

93–100.

Watkins. J. W. N. 1974. 'The Unity of Popper's Thought', *The philosophy of Karl Popper*, ed. Schilpp, P. A., pp. 371–412, The Open Court.

＜会員活動報告＞

小河原誠会員が『政治のロジックを問いただす』の PDF ファイルを 12 月 1 日から有料頒布（￥1260）されています。PayPay もしくは auPay にて代価 1260 円を氏の口座 ID の 08067417262 に送金されると折り返し指定のメールアドレスに上記ファイルが添付ファイルとして返送されてくるというかなり画期的なシステムで運営されています。(問い合わせ先：Makogawara47@outlook.jp）ご著書の内容は次の通りです。

まえがき
第一章　理由といっても、どんな理由なの――「なぜ」と問うのではなく、「どうなるか」を問え
第二章　両論併記をこえる途?報道は反証主義の立場で――報道は中立・公平でなければいけないのか?
第三章　「――すべし」と「――できない」、そして「――するつもり」
第四章　パラドックスまがいのことでも避けましょう
第五章　陰謀論を思想史の文脈で見る
第六章　不正選挙論はどこが間違っているのか
第七章　弁証法をとらえ直す
第八章　定義はものごとを明確化するか
第九章　賛成の数が多いといっても何ひとつ価値のある証拠にはならない――数の論理への批判――
第一〇章　「批判する」とはどういうことか
あとがき

編集後記

今回は 2024 年 8 月に行われた研究大会シンポジウム発表の完成稿を掲載しました。

研究会のホームページも随時更新されていきますので、またご確認いただければ幸いです。

本号についてのご意見等につきましては、編集委員（現在は、志村 昌司 shojishimura@gmail.com）までご連絡いただければ幸いです。

批判的合理主義研究（通巻 28 号）

2024 年 12 月発行

本誌は、『ポパーレター』（1989～2008,通巻 38 号）を改題し、継承したものです。

発行人 志村 昌司

編集・発行 日本ポパー哲学研究会事務局機関誌編集部

〒600-8018 京都府京都市下京区市之町251－2 壽ビルディング 2F

℡. 090-3842-9002

Email:shojishimura@gmail.com

入退会・名簿変更、会費徴収・会計管理に関しては、「日本ポパー哲学研究会事務局組織・会計部」にお願いいたします。

〒162-8473 東京都新宿区市谷本村町 42-8 中央大学大学院法務研究科 冨塚研究室 2826 号

tel. 03-5368-3661

fax. 03-5368-3630