

博士論文

論文題目 自然主義的意味論の研究

氏 名 次田 瞬

## 謝辞

本稿の執筆にあたり、多くの人々の協力を得た。学部の卒業論文から今日まで私の哲学研究を指導していただいている一ノ瀬正樹教授からは、初稿の段階から多くのアドバイスを頂戴した。口述試験の審査をつとめられた飯田隆教授（日本大学）・戸田山和久教授（名古屋大学）からも多くを学んだ。本稿に直接目を通されたわけではないが Richard Dietz 博士（東京電機大学）と Nikolaj Pedersen 教授（延世大学）には本稿で展開するいくつかのアイデアをめぐって議論に付き合ってください、論述をより厳密にすることができた。そして、佐藤暁博士（首都大学）、葛谷潤氏（専修大学）、宮園健吾教授（広島大学）には実際に原稿の一部を読んでいただき、有益なコメントを頂戴した。貴重な時間を割いて協力していただいたこれらの方々に深く感謝する。

## 目次

序章 .....	5
1 章 行動主義との比較.....	17
1.1 節 三つの問題 .....	17
1.2 節 方法論的問題.....	18
1.2.1 節 方法論的行動主義.....	18
1.2.2 節 方法論的行動主義の問題点.....	20
1.2.3 節 ホムンクルス機能主義.....	22
1.3 節 概念的問題 .....	25
1.3.1 節 分析的行動主義.....	25
1.3.2 節 意味のネットワーク説とラムジー化.....	28
1.3.3 節 ラムジー化に関する諸注意.....	31
1.4 節 心身問題 .....	34
1.4.1 節 存在論的行動主義.....	34
1.4.2 節 存在論的行動主義の諸問題.....	36
1.4.3 節 因果役割機能主義.....	38
1.4.4 節 傾向性と因果役割.....	41
1.5 節 本章のまとめ.....	45
2 章 同一説の再検討.....	47
2.1 節 同一説 .....	47
2.2 節 タイプ同一説とトークン同一説.....	49
2.3 節 機能主義と同一説.....	52
2.4 節 実現者機能主義：同一説の復権.....	54
2.4.1 節 性質の形而上学.....	54
2.4.2 節 実現者機能主義.....	56
2.5 節 同一説を支持する論証の批判的検討.....	58
2.5.1 節 狂人の痛みを想定する議論.....	59
2.5.2 節 理論的同一視にもとづく議論.....	60
2.5.3 節 役割性質の因果的無効力性にもとづく議論.....	62
2.6 節 本章のまとめ.....	66
3 章 機能主義の課題と限界.....	68
3.1 節 感覚経験の現象的性格とクオリアの反転.....	68
3.2 節 過少決定の問題.....	72
3.3 節 本章のまとめ.....	75

第二部の見通し .....	77
4 章 クワインの言語哲学.....	84
4.1 節 翻訳の不確定性と指示の不確定性.....	84
4.2 節 根源的翻訳と翻訳の不確定性.....	85
4.2.1 節 クワインの行動主義：言語獲得と言語習熟.....	86
4.2.2 節 翻訳の不確定性の論証.....	89
4.2.3 節 翻訳の不確定性をどう評価するか.....	93
4.3 節 指示の不確定性と意味理論.....	94
4.3.1 節 エヴァンズの議論.....	95
4.3.2 節 鏡映制約と話者の言語能力.....	98
4.4 節 指示の不確定性と命題的態度の帰属.....	101
4.5 節 本章のまとめ.....	103
5 章 ルイスの解釈理論.....	105
5.1 節 合理的な解釈.....	105
5.2 節 制約充足の問題を解くということ.....	106
5.3 節 解釈理論への不満.....	108
6 章 自然的記号と知覚的信念.....	111
6.1 節 心のレシピ .....	111
6.2 節 ドレッキの情報理論とその再検討.....	113
6.2.1 節 情報の確率的理論.....	113
6.2.2 節 情報の確率的理論への批判と応答.....	115
6.3 節 自然的記号から表象へ.....	118
6.4 節 ドレッキ型の目的意味論に対する批判.....	121
6.4.1 節 起源説、および起源説的な機能概念に対する批判.....	122
6.4.2 節 反応機能は起源説的な機能概念と両立するか.....	124
6.4.3 節 識別能力の限界.....	125
6.5 節 本章のまとめ.....	127
7 章 生得的な表象内容.....	129
7.1 節 欲求ベースの自然主義的意味論.....	129
7.2 節 ミリカン型の目的意味論.....	130
7.3 節 単純な生命活動と表象.....	133
7.4 節 比較研究 .....	134
7.4.1 節 ドレッキ型との比較：過剰警戒の事例をめぐって .....	134
7.4.2 節 動物行動学のコミュニケーション理論との比較.....	136
7.5 節 正常な条件について.....	136

7.5.1 節	カミンズの提案とその問題点.....	136
7.5.2 節	二通りの応答戦略.....	138
7.6 節	本章のまとめ.....	140
8 章	基本的な信念・欲求.....	141
8.1 節	導入 .....	141
8.2 節	原理 (R) .....	141
8.3 節	一切の障害はないという信念？ .....	143
8.4 節	条件付けの概念を用いた欲求内容の決定方法.....	145
8.5 節	前節の方法では扱えないさまざまな欲求.....	147
8.6 節	小休止 .....	149
8.7 節	二つの提案 .....	149
8.7.1 節	基本的欲求.....	150
8.7.2 節	基本的信念.....	152
8.8 節	結論 .....	154
文献表	.....	157

## 序章

“Contemporary philosophical theory of meaning is something of a black hole.”

Dennett 1984, p. ix

## 信念と欲求

本稿の主題は、一口で言えば、信念・欲求のもつ志向性という特徴である。志向性とは何か、どうしてそれが物議を醸すのかについてはまもなく触れるとして、まずは信念・欲求について簡単に考察することから始めよう。

信念と欲求は心的状態の一種である。信念は世界がしかじかのあり方をしている和我々が考えているときに我々がとっている状態である。例えば、バラク・オバマは第44代アメリカ合衆国大統領であると考えている私は、バラク・オバマは第44代アメリカ合衆国大統領であるという信念状態にある。欲求は、世界がしかじかのあり方をしていて欲しいと考えているときにわれわれがとっている状態である。例えば、次のアメリカ合衆国大統領はドナルド・トランプであって欲しいと考えている人は、トランプ氏が第45代アメリカ合衆国大統領であることを欲求している状態にある。

一般に、信念と欲求の内容は平叙文によって表現される。われわれは平叙文が表現する命題 (proposition) を内容とする信念・欲求を抱くのであり、それゆえ、信念・欲求は「命題的態度」と呼ばれる<sup>1</sup>。信念・欲求の他にも、予期、記憶、希望などさまざまな命題的態度がある。しかし、予期は未来についての信念、記憶は過去についての信念かもしれないし、希望は実現する蓋然性の低い欲求かもしれない。こうした分析が可能だとすれば、信念と欲求は他の態度よりも基本的な命題的態度とみなすことができる。

信念と欲求は行動と深く結びついている。世界がどのようなあり方をしているかを表象する信念は、地図と似ている。信念が地図と似ているとすれば、欲求は目的地にたとえられる。目的地にたどり着くために地図を利用するように、欲求を充実するための行動は信念によって導かれる。それゆえに、人がどのような信念・欲求を抱いているのかが分かれば、その人の行動を予測できるだろう。逆に、人のなす行動を調べるだけでも、その人がどんな信念・欲求を抱いているのかはある程度の見当がつく。人間は社会的動物なのでこの種のスキルはわれわれの日常生活において重要な役割を担っている。

---

<sup>1</sup> 命題の存在論的身分に関してはさまざまな見解がある。本稿の議論はほとんどの見解と両立すると思うが、私自身は命題を可能世界の集合と同一視するのを好む。たしかに、この命題観にはいろいろな弱点が指摘されている。例えば、可能世界の集合は命題と同一視するには肌理が粗すぎると言われる。可能世界の集合間には包含関係などが成立するが、ラッセルの命題のように構造化されておらず、必然的真理を表現する命題はすべて同一になるからである。また、すべての心的態度が命題の内容をもつかどうかも議論の余地がある。例えば、本質的に指標的な心的内容は命題ではない可能性がある。しかし、ストルネイカーはこれらの批判にうまく応答している。必然的真理を表現する命題に関しては Stalnaker 1976; 1984, chap. 5 を参照。本質的に指標的な内容に関しては Stalnaker 2014, chap. 5 を参照。

また、日常生活において重要な役割を担う信念・欲求は、学術的な研究の対象にもなりうる。このことは特に信念に当てはまる。信念に関する研究は、大まかにいえば、規範的側面に注目するものと、記述的側面に注目するものに分けられる。形式的認識論や計算機科学は、信念の規範的側面に注目する傾向にある。例えば、信念論理の研究者は、合理的な思考者の信念はKD4やKD45のような論理に従うべきだと言うかもしれない。ベイズ主義は、合理的な思考者がある仮説の正しさに関する信念の度合いを証拠に照らして修正する方法はベイズの規則に従うべきだと言うだろう。彼らにとって、現実のわれわれが信念を抱く仕方がこれらの論理や規則からたびたび逸脱することはあまり問題ではない。彼らは理想的な信念のあり方に注目しているからである<sup>2</sup>。これに対し、認知科学の研究は信念・欲求の記述的側面に注目する傾向にある。普通の人間が信念を抱く仕方は、理想から逸脱していたとしても、興味深いパターンをもつかもかもしれない。例えば、発達心理学における誤信念課題の実験によれば、平均的な4歳未満の幼児は他者の信念についての信念を抱くのが苦手である。人によって信じている事柄が異なりうる、という可能性に気づくまでに、われわれは平均して4-5年の歳月を要するらしい。

こうした簡単な考察だけでも、信念・欲求が社会的動物であるわれわれの日常生活において重要な役割を担っていること、そして、信念・欲求をめぐる認識論者や認知科学者がさかんに議論を積み重ね、成果を上げてきたことは確認できる。しかし、その一方で、哲学では、信念・欲求の概念は本当に現代人の抱いている科学的世界像と両立する概念なのか、というかなり根本的な疑惑にもさらされてきた。本稿で行う議論は、この根本的な疑惑と関わりを持っている。

### 問題：志向性は科学的世界像に適合するのか？

現代哲学の大きな潮流をなす考え方の一つに自然主義（naturalism）がある<sup>3</sup>。「自然主義」という表現は分野によって多義的に用いられるが、ここではおおまかに次のような意味で用いる。すなわち、自然主義者は、実在（reality）がほかならぬ自然科学の方法によって探求できる事柄ですべて尽くされていると主張する。したがって、もしも哲学が実在について何らかの知識をもたらしてくれる営みだとすれば、哲学もまた自然科学と連続した営みであるということになる。

現代の哲学者は、こうした帰結をすすんで受け入れる傾向にある。自然科学が発達し、われわれの日常生活にまで大きな影響を与えている現在、自然科学が原理的に扱うことのできない超自然的な現象や存在者を認めないという態度は、ほとんど常識的であるように

---

<sup>2</sup> ただし、規範的研究と記述的研究はそれほど厳格には区別できないかもしれない。現実の思考者が規範から逸脱した思考をすることがあるとしても、逸脱の程度があまりにも大きい思考者に対してはそもそも信念を帰属してよいのかどうか疑わしくなる。合理的な規範をある程度遵守していることは信念・欲求を帰属するための必要条件なのかもしれない。この路線に沿った信念・欲求の理論には5章で触れる。

<sup>3</sup> cf. Macdonald & Papineau 2006; Papineau 2015.

すら思われる。

それにも関わらず、自然主義はさまざまな挑戦を受けてきた。ここでいう自然主義に対する挑戦はおしなべて次のような形をもつ。自然科学が原理的にすら扱うことができないが、実在にとっての不可欠の部分であるような、そういう現象がたしかにある。反自然主義者によれば、われわれがさまざまな信念や欲求を抱いているという事実はそうした現象の一つである。しかし、信念・欲求という心的状態のどのような側面がそれほど物議を醸すのか？

伝統的に哲学では、心的状態は志向性という物的状態にはない特徴をもつと言われてきた。志向性とは何かについてのものである、という特徴である。志向性を持たない心的状態もあるかもしれないが、信念や欲求は志向性をもつ典型的な心的状態である。信念・欲求は何かについての信念・欲求であり、単に信じている・欲している、ということはない。例えば、冷蔵庫に向かって歩いている人は、はっきりとそう意識してはいなくても、冷蔵庫の中身についての思考をめぐらしている。冷蔵庫の中にはチョコレートが入っていると信じている。チョコレートを食べることを欲求している、などなど。

信念・欲求がもつ志向性は、一見するとどうということのない特徴だが、問題は、この「何かについてのものである」とか「何かに向けられている」という関係性は何なのか、である。いくつかの点で、志向性は物理的世界の中に見出されるような種類の相互作用とは異質であるように思われる。例えば、因果性は物理的世界の中に典型的に見出される相互作用だが、志向性は因果性とはしばしば対比される。冷蔵庫の中にはチョコレートが入っているという信念は、冷蔵庫の中にチョコレートが入っていることを原因として形成される必要はない。冷蔵庫の中にはチョコレートなどないかもしれないからである。そういう場合ですら、われわれは冷蔵庫の中のチョコレートについて、つまり、冷蔵庫の中にチョコレートがあると信じることがある。こうした信念は偽であり、存在していないものに関わっている。

偽な信念を抱くときのように、外界のあり方を誤って捉えることがあるという特徴は、志向性をもった心的状態に特有だと言われる。物的状態にはそもそも誤りというものがない。物体が自由落下するとき、電子と電子が反発するとき、そこには重力法則やクーロンの法則といった物理法則にしたがって物質が運動するだけで、そこに例外はない。もちろん、物理的世界の中には確率的な現象もある。例えば、喫煙が肺がんを引き起こすかどうかは、確率の問題と言える。しかし、誤りの可能性に開かれているかどうかは、確率の問題とは別である。例えば、冷蔵庫の中にチョコレートがあるという信念と、冷蔵庫の中にチョコレートがあることとの間には、強い相関関係があるかもしれないが、こうした相関関係のゆえに、この信念は冷蔵庫の中にチョコレートがあるという内容をもつわけではない。なぜなら、この信念は、冷蔵庫の中にチョコレートまたはミルクがあるという選言的な事態とより強い相関関係にあるかもしれないからである。もし信念の内容が相関の強さ依存するのであれば、当の信念の内容は、冷蔵庫の中にチョコレートまたはミルク



がある、ということになってしまうだろう。これを信念に関する正しい考え方とみなすことはできない。

以上の議論は、簡単ではあるが、志向性は物理的世界の中に見出されるような種類の相互作用とは異質であり、それゆえに、心的状態の志向性は自然主義に対する挑戦となるということを示唆する<sup>4</sup>。少なくとも、自然主義者の中にはそういう直観をもつ人が多い。ジェリー・フォーダーは次のように述べている。

私思うに、遅かれ早かれ物理学者は、物理学がこれまで物体の究極のかつ還元不可能な特徴について集めてきたもののカタログを完成させるだろう。そうしたカタログを完成させたときには、おそらく、スピン、チャーム、電荷などがそのリストに載るだろう。しかし、ついて性 (aboutness) は確実に載らないだろう。志向性はそれほど究極のものではない<sup>5</sup>。

もし志向性が世界の基本的な特徴ではないのだとすれば、志向性を科学的世界像に適合させるためには、志向性を世界のより基本的な特徴へと還元すべしという研究課題が立てられるだろう。これが本稿の研究課題である。

歴史的にみれば、志向性が心的状態に特有の特徴であるという見解は少なく見積もっても100年以上昔からある。しかし、本稿で取り上げる素材の大半は、20世紀後半の英米圏の哲学から取られている。古典を軽視するつもりはないのだが、ごく表面的に見ても、20世紀後半にさまざまな概念や道具立てが整備され、心の哲学が発展したことは否定できないと思われる。新たな概念や道具立てとしては、ラムジー文、多重実現可能性、因果役割、情報意味論、機能の起源説などを挙げることができる。たしかに、志向性を世界のより基本的な特徴へと還元するという課題がすでに達成されたというわけではないだろうし、私もそれほど楽観的ではない。しかし、これらの道具立てをフルに利用することで、心的状態に特有な特徴とされる志向性に関して、現代の心の哲学はどこまで前進できるのかを考えるのは価値がある。

本稿は、上で引用したフォーダーの要求に適合するような志向性の理論を構築するために提案されてきた従来の理論を比較検討し、利点と難点を冷静に見極める。そして、最も有望な選択肢を選びだし、擁護することを目指す。

---

<sup>4</sup> ここまでの数段落で提示した議論は、心の哲学の文献で「誤りの問題」、「選言問題」と呼ばれている話題を、私なりに翻案したものである。cf. Crane 2003, pp. 178–180. なお、物的状態にはない心的状態の特徴として、志向性の他に意識を挙げることでもある。志向性と意識は共に（ただし、異なる理由で）物理的世界の中に適切に位置づけるのが哲学的に困難であると考えられており、自然主義者にとって問題となっている。意識について3章でごく手短に触れる。

<sup>5</sup> Fodor 1987, p. 97.

## 本稿のアウトライン

すでに述べたように、本稿の主題は信念と欲求の志向性である。しかし、信念・欲求にたどり着くまでに少し遠回りをする。本稿は二部構成になっており、信念・欲求の志向性を論じるのは後半（4-8章）である。前半（1-3章）では心の機能主義を扱う。

なぜ機能主義を取り上げるのか。言語的意味と心的状態はともに志向性という特徴を持つものの、言語的意味のもつ志向性は信念などの心的態度に依存していると思われる。信念・欲求といった心的態度は一種の心的性質である。機能主義は現代の心の哲学において、心的性質についての標準的な理論とみなされている。それゆえ、心的なものに関心があるならば、まずは機能主義の検討から始めるのがよいだろう。

1, 2章では、機能主義とはどのような立場であるのかを詳細に検討する。「○○とは何か？」という問いに答える一つの方法は、○○が何でないかを述べることで、○○の本質的な特徴を浮かび上がらせることである。1章では、機能主義を行動主義と比較し、どのような点において機能主義が行動主義の弱点を克服したのかを考察する。2章では、機能主義と同一説の関係を考察する。機能主義には同一説と両立する実現者機能主義と、両立しない役割機能主義の二種類がある。ここでは役割機能主義を擁護し、実現者機能主義を支持する議論に反論を加える。

これら二つの章で本稿が擁護する立場は、心の哲学ではすでにオーソドックスな部類に入ると思う。読者の中には、すでに主流となっている見解をわざわざ擁護するという作業は退屈で魅力を感じない、という人もいるかもしれない。しかし、この作業は少なくとも二つの点で重要である。まず、同じ結論を導く複数の方法がありうるということを指摘したい。例えば、数学においてはすでに証明されている定理の別証明を探すことも一つの仕事として認められる。たしかに、新たな定理を証明することと比較すれば別証明を提示することはあまり評価されないかもしれないが、簡潔な別証明が与えられることで既に知られていた定理の理解が深まることもある。ここでの狙いは、数学でのそうした営みに類比させることができる。また、主流となっている立場に対しても異論ないし対立する立場がある以上、そうした立場との比較検討を丁寧に行うことにはやはり意義があると思われる。もっとも、オーソドックスな見解がオーソドックスであることにはそれなりの理由があるのだから、慎重な検討の結果として、やはりオーソドックスな見解に落ち着く、というのは自然なことである。例えば、2章ではタイプ同一説を否定する。このように、本稿の立場は全体的にやや保守的である。

3章では、行動主義と同一説を否定した機能主義に依然として残された課題について論じる。機能主義に残された課題はさまざまだが、その中には、命題的態度の本性は依然として不明瞭であるという問題が含まれる。命題的態度は志向性を持つとされる典型的な心的性質であるから、このことは、機能主義単独では志向性を科学的な世界観と折り合いをつけることができないことを示唆する。そこで、命題的態度の本性を明らかにするには機能

主義とは別立ての補完的な理論が必要であると結論づける。この新たな課題への取り組みが、本稿の第2部を形づくる。

第2部では、第1部の結論を踏まえて、命題的態度に関して自然主義者たちがこれまでに提出してきた諸見解を検討する。議論の見通しをつけるため、まず、命題的態度に関する自然主義者の諸見解を

- ・ 消去主義
- ・ 解釈理論
- ・ 信念ベースの理論
- ・ 欲求ベースの理論

の四種類に分類することを提案する。第2部の諸章では、この分類枠に沿って、それぞれの立場の典型的な議論を批判的に検討していく。

4章では、意味や命題的態度に関する消去主義として、クワインの言語哲学を考察する。クワインの言語哲学は行動主義的な前提を描いており、この前提から指示の不確定性のようなパラドキシカルな帰結を導いているので、クワインの議論は帰謬法なのだとよく言われる。私はこの診断に基本的に賛成だが、4章では、ここでいう「行動主義的な前提」の意味するところを明確にし、なぜその前提が指示の不確定性を帰結するのかを分析することに精力を傾けた。

5章では、測定理論との類比に基づくルイスの解釈理論を考察する。そして、ルイスの解釈理論には、高度な合理性を仮定しなければ命題的態度を帰属できず、また、動物にも備わるような内在的志向性を取り扱えないといった弱点があることを指摘する。この指摘はおそらく解釈理論にとって致命的というわけではないだろうが、それでも、信念・欲求に関するボトムアップの研究を行う動機付けにはなる。では、「信念・欲求に関するボトムアップの研究」とはどういう意味か。

20世紀の終盤にかけて、心理学は進化生物学と合流し、認知能力の進化論的な起源や機能が論じられるようになった。自然界を眺めると、生命活動にはさまざまな情報をやり取りが含まれている。外界の知覚能力はその典型的な例だが、蜜蜂のような昆虫でさえ、ダンスによって餌のある方位や距離を仲間に伝えることが知られている。信念・欲求に関するボトムアップの研究ということで私が念頭に置いているのは、人間以外の動物にも共有されている認知能力のあり方にヒントを得た、志向性の研究である。6-8章で取り上げる諸見解は、細部に違いはあっても、すべてボトムアップ型の研究である。

6章では、信念ベースの理論として、ドレッスキの議論を取り上げる。ドレッスキは情報意味論の創始者の一人として知られており、彼の情報意味論は生物学の哲学における機能の起源説と組み合わせることで、知覚的信念の内容を決定するという目的に応用できるかもしれない。このアイデアを「ドレッスキ型の目的意味論」と呼ぶ。しかし、6章後半では、ドレ

ツキ型の目的意味論に対する三種類の批判を挙げる。中でも三番目の批判が重要であり、ドレツキ型の目的意味論は、自然選択による進化が最適にデザインされた認知能力をもった生物を生み出すという非現実的な前提を抜きにしては成功しないと指摘する。この難点のゆえに、本稿はドレツキ型の目的意味論を退ける。

こうして、残される選択肢は欲求ベースの理論だけとなった。これまでに提案されている欲求ベースの理論は、生物学的機能を利用するタイプと、オペラント条件付けを利用するタイプに分けられる。

7章では、生物学的機能に注目するミリカンの理論を取り上げる。ただし、ここでは信念・欲求について論じる代わりに、蜜蜂のような比較的単純な動物にみられるコミュニケーション行動について考察している。一見単純な動物も複雑な行動を行なうことがあり、そうした複雑な行動を説明する上で、彼らは何らかの内容を伝達するコミュニケーションを行なっていると考えerことは有用であり、直観的にももっともらしい。ミリカンは動物のコミュニケーション行動において伝達される内容についての理論を提案しているのだが、この理論は彼女にとって信念・欲求についての自然主義的意味論を展開するための基礎部分として位置づけられている。

8章では、オペラント条件付けを利用するホワイトの理論を考察する。ホワイトの理論はミリカンとは一部異なるが欲求ベースの理論に分類できる。ホワイトの議論は細部で疑わしい仮定を用いているため、まずホワイトの議論を批判的に検討する。しかし、彼の議論には正しい方向を向いていると思われるいくつかの論点がある。8章の後半では、ホワイトの議論を踏まえて、改善案を提示する。

## 形而上学的な前提

本稿のアウトラインはおおよそ以上の通りである。このまま本論に入ってもよいのだが、その前に、本稿が（明示的にせよ暗黙のうちにせよ）採用している形而上学的な前提をできる限り明確にしておきたい。以下では心的／物的、出来事、状態、因果性、内在的／派生的志向性といった基本的概念を取り上げ、予備的な考察を行う。本稿が採用する形而上学的な前提は、現代哲学の中では標準的な部類に属しており、自然主義と問題なく両立すると私は考えている。

### (1) 心的なものと物的なもの

第一部では「心的 (mental)」と「物的 (material)」といった形容を頻繁に用いる<sup>6</sup>。これらの形容は「対象」や「性質」、「出来事」などに適用される。

「心的」と「物的」は伝統的には対比的に扱われてきた。本稿でもこの伝統をある程度踏襲している。ただし、以下の二点には注意が必要である。一つ目の論点は、心的性質と

---

<sup>6</sup> 本稿では「心的」は「心理的 (psychological)」と、「物的」は「物理的 (physical)」と同義的に用いる。

物的性質は排他的 (disjoint) とは限らないということである。2章の内容を先回りすることになるが、例えば、タイプ同一説によれば、心的性質は何らかの物的性質と同一であるとされる。また、たとえタイプ同一説を否定したとしても、心的性質が物的性質でないということにはならない。何らかの物的性質によって実現される機能的性質は別の物的性質であり<sup>7</sup>、心的性質は物的性質によって実現される機能的性質だとすれば、心的性質はやはり物的性質の一種ということが帰結するだろう。

二つ目の論点は、「物的」という形容は物理学だけではなく化学・生物学など自然科学全般に関わるという意味で、かなり広範囲の事柄に用いられるということである。すなわち、質量や電荷といった物理学において対象に帰属される性質だけではなく、脳のしかじかの領域が活性化しているといった神経学的な性質も物的性質に含まれる<sup>8</sup>。さらに厳密に言えば、「化学的」、「生物学的」、「神経学的」といった形容ですら、相当にラフな言い回しであることを認めなければならない。例えば、ニューロンに関する研究と一口に言っても、単一のニューロンに焦点を合わせた研究から嗅球や海馬といった単位に焦点を合わせた研究に至るまで様々なレベルが区別できるし、実際に区別されている<sup>9</sup>。以上の考察は、物的対象や物的性質は決して一枚岩ではないということを示している。しかし、本稿における議論の正否は、こうした事情には左右されないだろう<sup>10</sup>。

## (2) 出来事と状態

本稿では述語によって表現されるものを一括して「性質」と呼んでいる。性質の中には顕在的な性質とそうでない性質がある。顕在的な性質の中には、「殴る」「壊れる」「溶ける」など、日常的な語感からすれば「性質」と呼ぶのがためらわれるものもある。しかし、これらの述語も本稿の用語法では性質を表現する。これらの述語は出来事 (event) の性質を表現する。

しかし、誰かが誰かを殴ったり、あるいは、何かが壊れたり、溶けたりすると言うときに、出来事に対する述定がなされているように見えないと反論されるかもしれない。この反論に応答するには、デイヴィドソンによる出来事の理論を参照するのが有益である。デイヴィドソンは、「ピーターがジョンを昨日殴った」といった行為文には出来事に対する量が含まれていると主張した。デイヴィドソンによれば、この文は

$\exists e$  [殴る (ピーター、ジョン、 $e$ )、かつ、時点 (昨日、 $e$ )]

<sup>7</sup> Kim 1998, p. 114.

<sup>8</sup> 山口 2012, p. 12.

<sup>9</sup> Churchland 1986, pp. 359–360.

<sup>10</sup> 物的なものが一枚岩でないということを真剣に受け止めて、さらなる腑分けを行うこともできる。例えば、キムは実在を物的なものと心的なものに二分する古い見方に代えて、実在は順序付けられたさまざまなレベルから成り立つという階層的なモデルを提案している。Kim 1998, pp. 15–16; Schaffer 2003.

と分析される<sup>11</sup>。このように分析すれば、殴るということは誰かが誰かを殴ることであるという考えと、誰かが誰かを殴るときに出来事への述定がなされているようには見えないという考えを両立させることができる<sup>12</sup>。

デイヴィッドソンの分析には他にも多くの利点がある。例えば、この分析は、ピーターがジョンを一度しか殴らなかったとしても、あるいは何度も殴ったとしても、「ピーターがジョンを昨日殴った」という文が真であるという事実を上手く説明するだろう。さらに、この分析は、副詞修飾の除去が妥当な推論であるということをうまく説明する。例えば、「ピーターがジョンを昨日殴った」から「昨日」という副詞修飾を除去して「ピーターがジョンを殴った」を推論するのは妥当であろう。つまり、「ピーターがジョンを昨日殴った」を上のように分析するならば、副詞修飾の除去は連言除去の一例として説明がつく。

もしデイヴィッドソンが正しければ、出来事は行為文の分析において要請される対象である。だが、出来事は他の文脈でも登場する。例えば、因果関係の関係項は、一般的には出来事として理解されている<sup>13</sup>。このように多様な役割を担う出来事とはどのような対象なのだろうか。この問いに対するデイヴィッドソンの回答は、大まかに言って、出来事とは時間的・空間的な位置をもった何らかの具体的対象である、というものである。デイヴィッドソンの回答は有力な見解ではあるものの、これとは別の有力な見解もある。それは、出来事は性質例化であるという見解である。出来事存在論に関するこれら二つの代表的な選択肢は、タイプ同一説とトークン同一説の相違を検討する場面で重要になる（2章）。結論から言うと、本稿はトークン同一説を擁護するために、デイヴィッドソンの出来事存在論を擁護することになる。

ところで、出来事は状態（state）と比較されることがある。例えば、出来事が物体に生じる変化と関わるのに対して、状態は物体が一定の期間にわたる無変化と関わりがあるとされる。出来事は生じる（happen）のに対して、状態は持続する（persist）という仕方で時間

---

<sup>11</sup>  $e$  は出来事の上を動く変数である。したがって、この記号列は「ある出来事  $e$  が存在し、 $e$  はピーターがジョンを殴るという出来事であり、 $e$  は昨日生じた」と読む。

<sup>12</sup> デイヴィッドソンによる行為文の分析は、「「シーザーが死んだこと（that Caesar died）」という語句は、存在命題をあらわし、ある種の出来事存在を述べる」というラムジーの分析に由来する。Ramsey 1927, p. 37. 本稿はラムジーに由来するアイデアを数多く取り入れているが、これはその一例である。

<sup>13</sup> デイヴィッドソンは、出来事は単称名によって指示されるという意味で文によって表現される事実（fact）とは異なり、因果関係は出来事間の関係であるから、事実間の関係ではないと論じた。デイヴィッドソンの議論はスリングショット論法に基づく。スリングショット論法は、事実の同一性条件が

- ・ 二つの単称名 “a” と “b” の指示対象が同じならば “Fa” と “Fb” は同じ事実を表現する。
- ・ 論理的に同値な二つの文は同じ事実を表現する。

という制約を満たすならば、事実の一つしかない」と結論づける（真である文のみが事実を表現すると仮定した場合）。これらの前提はもっともらしいので、因果言明の論理形式は「…という事実が～という事実を引き起こす」ではない。

ただし、スリングショット論法は事実因果にとって必ずしも致命的ではない。スリングショット論法は妥当ではないかもしれないし、仮に妥当だとしても健全ではないかもしれない。スリングショット論法については Neale 1995 を参照。ここでは差し当たり、スリングショット論法は事実因果を拒否して出来事因果を採用する一つの動機にはなる、と考えておく。

と関係する、と言われることもある<sup>14</sup>。しかし、本稿ではこのような相違は表面的なものであり、出来事と状態をひとまとめにした統一的なカテゴリーについて語ることができると仮定する。こう仮定する一つの理由は、状態もまた因果関係の関係項として言及されるからである。例えば、信念や欲求といった心的状態は行動の原因とみなされうる。そこで、本稿ではむしろ広義の出来事と狭義の出来事とを区別して、状態は狭義の出来事とは区別されるが広義の出来事の一つではあるという方針を採用したい。すなわち、「殴る」や「壊れる」が出来事の性質を表現するのに対し、「楕円形である」や「乾いている」といった述語、あるいは信念述語などは状態の性質を表現する、と考える。

こうした方針に反対して、そもそも状態が原因であるという考えを否定すればよいと言う人もいるかもしれない<sup>15</sup>。しかし、状態も因果関係の関係項でありうるという考えは因果論では広く受け入れられている。例えば、マッチを擦って火をつけるという古典的な例を考えよう。たしかに、マッチを擦ったという出来事はマッチに火がついたという出来事の原因である。しかし、火がつく原因にはマッチを擦るという出来事だけでなく、マッチが乾いているとか周囲に十分な酸素があるといった状態も含まれると考えられる。なぜなら、これらの状態とマッチに火がつくことの間には反事実的な依存関係があるからである。例えば、マッチを湿らせることによって、私はマッチに火がつくという結果を生じさせないように操作することができるだろう。このように、状態を出来事と共に原因として語るの是一般的に認められるべきだと思われる<sup>16</sup>。

デイヴィッドソンの出来事存在論の下で、状態を出来事の一つとみなすならば、出来事と同様に状態もまた時空領域を占めると考えることになる。例えば、信念や欲求といった心的状態は行動の原因として言及されることがあるが、同一説を前提にするならば、そうした因果性は神経系ないし身体を含む時空領域を占める状態が行動を引き起こすことによって成立しているということになる<sup>17</sup>。

本稿の方針は出来事と状態を並行的に扱うというものであるから、行為文には出来事に対する量化が隠れていると分析したように、物体が静的に保有している性質を帰属する文にも出来事への量化が隠れていると分析したい。例えば、あるマッチに火がついた原因として、そのマッチの乾きについて語りたいことがある。これを可能にするには、「乾いている」という述語を、マッチと特定の状態を値とする二項述語として扱えばよいだろう。このとき「マッチが乾いている」という文は

$\exists e$  [乾いている (m, e)、かつ、マッチ (m)]

<sup>14</sup> Steward 1997.

<sup>15</sup> この反応は、デイヴィッドソンが命題的態度を行為の原因であると論じたときにありうる反論として検討しているものでもある。Davidson 1963.

<sup>16</sup> こうした見解は文献に広く見受けられる。Armstrong 1968; Kim 2006, p. 7.

<sup>17</sup> cf. Davidson 1969, p. 176.

と分析される。また、デイヴィドソンによる行為文の分析には副詞に関する推論を適切に説明できるという利点があったが、副詞修飾は行為文だけでなく状態に関する文にも生じるであろうから、出来事と状態を並行的に扱うことにすればこの点に関しても利点を引き継ぐことができる。

### (3) 内在的志向性／派生的志向性

志向性という特徴をもつのは心的状態だけではない。言語的意味も志向性をもつと言われる。例えば「冷蔵庫の中にチョコレートがある」という文は冷蔵庫の中にチョコレートがあることを意味している。実際に冷蔵庫の中にチョコレートが入っていても入ってなくても、どちらの場合であれ、この文は冷蔵庫の中にチョコレートがあるということを意味する。思考が実在を誤って捉えることがあるのと同様に、平叙文は誤っていることがありうる。この点で、言語的意味は信念・欲求のような心的状態がもつ志向性と似たような特徴を持っている。

それでは、言語的意味の志向性と心的状態の志向性はどちらが優先するのだろうか。この点に関して哲学者はまだ合意に達していないが<sup>18</sup>、本稿では言語的意味の志向性は心的状態の志向性に依存すると前提している<sup>19</sup>。志向性に関する限り、信念・欲求のもつ志向性にもっぱら注目するのはそのためである。この前提の下では、心的状態のもつ志向性は内在的 (intrinsic) であり、言語的意味のもつ志向性は派生的である。

言語的意味の志向性が心的状態の志向性に依存するという見解は、直観的に言っても理解しやすいと思う。例えば、本に書かれた文章が何かの表現・表象であるのは、その文章を書いた人が伝達意図をもって書いたからだだろう。ネコがタイプライターをデタラメに叩いたことで、偶然にも用紙の上に有意味に思われる文章ができたとしても、それは言語表現ではなく単なるインクの染みだという直観は広く共有されている。

たしかに、言語を使いこなすことによって、われわれが考えられる事柄の範囲が大きく広がった、という可能性は想像できるし、事実そうなのかもしれない。しかし、認知科学者たちが心的イメージによる推論などを研究しているように、すべての思考が言語的というわけではないと考えることはできる。言語を抜きにしても、われわれ人間は、あるいはより単純な動物であっても、信念・欲求を伴った思考ができるのではないか<sup>20</sup>。本稿は、こうした可能性に賭ける。

<sup>18</sup> 両陣営の対立は「ホメロスの闘争 (Homeric struggle)」と呼ばれている。Miller 2007, sec. 7.1.

<sup>19</sup> 古典的には、グライスの名が挙げられる。一般に、グライスのプログラムは言語的意味を話し手の意味という一種の心的状態へと還元することだと理解されている。グライスのプログラムと本稿の議論との関係は、第二部の見通しでもう少しコメントする。

<sup>20</sup> この話題に関しては Carruthers 2002 が参考になる。



## 第一部

### 心の機能主義：その利点と弱点

## 1 章 行動主義との比較

“psychologists don’t find behaviorism very *reinforcing* these days.”

Dennett 1978, p. 53

### 1.1 節 三つの問題

心の哲学において、20 世紀後半は機能主義が台頭してきた時代である。21 世紀に入った現在でも、機能主義はしっかりとオーソドックスの地位を確立しているように見える。それでは、機能主義の強みはどこにあるのだろうか。本章では、機能主義を行動主義と比較することを通して、この問題に回答する。

なぜ行動主義と比較するのか。それは、機能主義が 20 世紀の前半に隆盛を誇った行動主義に取って代わるようにして浮上してきた研究プログラムだからである。それゆえ、機能主義は、それに先行する研究プログラムの行動主義と比べて、どんな点が優れているのかと問うのは自然である。この疑問に答えることは、行動主義の問題点を挙げるだけの作業とは異なる。実際、心の哲学の歴史を解説する文献に少しあたれば、行動主義に多くの問題点が指摘されてきたことは分かる。しかし、ある論点が行動主義にとって問題だとしても、同じ論点は機能主義にも当てはまるかもしれない。いかなる意味で機能主義が行動主義よりも優れているのかを示すには、行動主義にとって都合が悪く、機能主義にとっては問題とはならないような論点を、慎重に一つ一つ特定しなければならない。

厄介なことに、行動主義と機能主義のどちらにもいくつかの変種がある。「行動主義」と「機能主義」という名称は共通していても、論者によって取り組んでいる問題が微妙に異なっているために、無印 (generic) の行動主義と機能主義を定式化して単純に比較することはできない。むしろ、行動主義と機能主義は互いに独立であるような複数の主張の束であるように思われる。そこで、次のような方策を採用することにしたい。本章では、心の哲学者たちが行動主義や機能主義といった立場を引き合いに出すことで取り組んできた問題として、以下のような三つの問題に注目する<sup>1</sup>。

- ・ **方法論的問題**：心理学の研究対象は何であるべきかという問題
- ・ **概念的問題**：心的語彙の意味を分析するという課題
- ・ **心身問題**：心的性質とは何かを特徴づけるという存在論的な問題

---

<sup>1</sup> この問題設定は Churchland 1988 を下敷きにしている。もちろん、これら三つで心の哲学の主要な問題がすべてカバーされるわけではない。チャーチランドは他にも認識論的問題を論じている。例えば、自分自身の心的性質に関して特権的なアクセス能力があるように思われるが、それは本当だろうか。もし本当だとすればなぜだろうか。こうした問題は本稿では取り上げない。その理由は、こうした問題が難しいからではなく、機能主義が行動主義に比して前進をみせたとは言いがたいからである。

行動主義と機能主義の無印の定式を得ることはできなくても、これらの問題に対する行動主義的な回答と機能主義的な回答を定式化することはできる。それらの回答を順に比較していけば、全体として、機能主義的な回答がどのような意味で画期的であるのかを見極めることもできるだろう。

本章の構成は以下のようなものである。上に挙げた三つの問題に対する行動主義と機能主義のスタンスの違いを明らかにする。1.2 節では、方法論的問題に関する行動主義的な回答と機能主義的な回答を比較する。1.3 節では、概念的問題に対する行動主義的な回答と機能主義的な回答を比較する。1.4 節では、心身問題に対する行動主義的な回答と機能主義的な回答を比較する。

## 1.2 節 方法論的問題

本節では、方法論的問題に関する機能主義の貢献が何であったかを考察する。1.2.1 節では方法論的問題に対する行動主義的な回答を概観する。1.2.2 節では、行動主義的な心理学理論の問題点を明らかにする。1.2.3 節では、方法論的問題に対する機能主義的な回答を提示した上で、行動主義の側から寄せられるであろう反論から擁護する。

### 1.2.1 節 方法論的行動主義

方法論的問題に対する行動主義的な回答は次のように纏められる。

**方法論的行動主義：**心理学は、皮膚より内側の内的状態ではなく、行動についての科学でなければならない<sup>2</sup>。

方法論的行動主義によれば、ある学科が行動についての科学でないならば、それは端的に心理学ではないことになる。そのため、たとえ「心理学者」と呼ばれている人々の研究方法が、方法論的行動主義に反するとしても、そのことによって方法論的行動主義が論駁されることはない。この意味で、方法論的行動主義の主張は規範的である<sup>3</sup>。

とはいえ、方法論的行動主義の指針によってもたらされた科学の生産性を評価することはできるかもしれない。そもそも、心理学が行動についての科学であるべきだという主張

---

<sup>2</sup> Graham 2015.

<sup>3</sup> 方法論的行動主義は何ゆえに心理学に対してこれほど厳格な制約を課すに至ったのだろうか。この疑問には、さしあたり次のように答えておけばよいだろう。19 世紀の内観心理学は、科学にふさわしい客観性を保持していないという批判を浴びていた。科学にふさわしい客観性は公共的に観察可能な対象や性質を扱うことによってもたらされると思われる。しかし、皮膚の下の内的状態を研究する分野はすでに神経生理学として確立されているので、心理学の研究対象はもはや皮膚の下の内的状態ではありえない。そこで、心理学が神経生理学とは異なる自律的な学科として成立するためには、その研究対象はむしろ皮膚の外にあらわれる行動に求められるべきである。McCulloch 1995, pp. 110–113.

は、生命体の行動の多くは、皮膚より内側の内的状態に訴えることなく記述・説明・予測できるだろう<sup>4</sup>、という期待の裏返しとも考えられる。こうした期待はどの程度かなえることができるのだろうか。

まず、内的状態を参照しないということは、生命体の内部状態をブラックボックスとして取り扱うことを意味する。行動主義者はブラックボックスに対する刺激とそこから出力される行動だけを参照し、それら刺激と行動がどのように関係しているのかを考察する。行動の中には、眼に対する光の明るさの変化によって瞳の大きさが変化する瞳孔反射のように、刺激と行動を容易に関係付けられるものもある。

しかし、瞳孔反射のような無条件反射は、行動の中でもかなり単純な部類に入る。生命体は同じ種類の刺激に対して同じ種類の行動を返すとは限らない。このことは行動パターンの違いや変化が生命体の皮膚下の内的状態の違いや変化に由来することを明確に示していないだろうか<sup>5</sup>。

たしかに、その通りである。しかし、行動主義者には次のように応答する道が残されている。なるほど、生命体の内部状態というものを描定したければしてもよい。ただし、行動パターンの違いや変化をもたらした内的状態の違いや変化は、さらにさかのぼれば、生命体がこれまでどのような条件付け（conditioning）の歴史を経てきたのかに由来するだろう。つまり、生命体の行動は刺激の関数ではなく、刺激と条件付けの歴史の関数である。したがって、生命体の行動を予測・説明するために、生命体の内部状態を参照する必要はない。問題となっている生命体に対して、条件付けによる学習が可能であるような内部構造を持つといった最小限の仮定をする必要はあるが、それでも、生命体の内部状態に関してそれ以上の仮定を措く必要はないと期待してもよい<sup>6</sup>。

ここでいう「条件付け」とは、もともとは関連のなかった刺激を別の刺激ないし行動と連合する（associate）プロセスをいう。条件付けには古典的条件付けとオペラント条件付けの二種類がある。古典的条件付けのパラダイムはパヴロフの犬である。餌の視覚的刺激は唾液の分泌を無条件に誘発するのに対し、ベルの聴覚的刺激は唾液の分泌を無条件には誘

<sup>4</sup> Graham 2015. cf. Kim 2006, p. 76. なお、生命体（organism）が有機物（organic matter）でできているとは前提していない。例えば、心的性質の多重実現可能性を論じる文脈では、生命体は炭素ベースでない可能性もある。cf. Kim 2006, p. 117.

<sup>5</sup> 例えば、チョムスキーは、同じ絵画を見せられたとしても人によって「オランダの絵画だ」「見たことのない絵だ」「君はもっと抽象的な絵画が好きだと思っていたのだが」などさまざまな反応がありうる、という例を挙げている。Chomsky 1959.

<sup>6</sup> こうした制約の存在はしばしば無視されるが、クワインは例外的にこの点に関して自覚的であり、次のように述べた。「行動主義は学習態勢の生得的なメカニズムに、それと知りながら喜んで、首までつかっている。行動主義にとって決定的に重要な反応の強化と消去そのものは、刺激をいわば質的な空間に布置する主体のやり方が生まれながらに不均等であることに依存している。[...] 生得的な偏向と傾向性が行動主義の礎石である」（Quine 1969b, p. 57）。このコメントは、学習可能な行動レパトリーが種によって異なるという明白な事実を踏まえている。例えば、人間以外の動物に言語行動を教えるという試みはすべて失敗してきた。おそらく、種ごとに学習可能な行動レパトリーがどのように異なるのかは、皮膚の下の内的状態へと言及しないことには説明できないだろう。しかし、行動主義者は種ごとに異なる行動レパトリーを確定することは生理学が解明すべき課題であり、心理学にとって行動レパトリーは所与だと考えていると思われる。

発しない。しかし、ベルを鳴らしながら餌を与えるようにすると、ベルを鳴らただけで唾液を分泌するようになる。つまり、犬はベルの刺激と餌の刺激との間の連合（いわゆる恒常的接続）を学習する。これに対し、オペラント条件付けでは行動とその結果として得られる刺激との間の連合が学習される。例えば、「キー」と呼ばれる照明された円板を備え付けた箱の中にハトを入れ、キーをつつくと餌が飛び出る仕組みの実験装置をつくる。ハトは箱の中でさまざまな行動をとるが、キーをつついた後に餌を与えられるということを繰り返すと、キーをつつく頻度が上がる。このように、たまたま何らかの行動をした後に報酬を受けるとその行動の頻度が上がる、という現象は多くの動物において観察される。この現象を「強化（reinforcement）」という。逆に、何らかの行動をしたあとに罰を受けるとその行動をとる頻度は下がる。例えば、牛が電線を張り巡らした柵に触るとショックを受けるならば、牛は柵に近づく頻度は下がるだろう。この現象を「強化」との対比で「弱化」という。弱化も多くの動物において観察される。

古典的条件付けとオペラント条件付けのどちらにおいても、行動パターンが変化するという意味での「学習」が生じている。古典的条件付けで強化できる行動は腺の分泌など不随意的行動が主であるのに対して、オペラント条件付けでは随意的な行動を強化できる。たまたまなされた随意的な行動の頻度を報酬によって上げることから、オペラント条件付けは試行錯誤（trial and error）による学習とも呼ばれる<sup>7</sup>。もっとも、自然界では、同じ行動をとったとしても常に報酬が手に入るとは限らない。そこで、行動主義の心理学者たちは行動の後に報酬を与えるさまざまなパターンを試す。例えば、 $n$ 回キーをつついた後に報酬を与えるようにハトを訓練する、などである。この $n$ の値をどこまで大きくできるのか、というのは興味深い問題である。この種の疑問に対して精緻な実験によって回答を与えることは、行動主義者にとって重要な研究課題である。

行動主義者たちは数多くの条件付けの事例を実験室の中で確認してきた。野生の状況でどのような条件付けが生じているのかを見通すのは難しいが、野生の状況で条件付けが自然に生じていると推測するのは自然である。そこで、野生であれ実験室で育てられたのであれ、生命体の内部状態は条件付けの歴史に依存している、と仮定しよう。そうすると、たしかに、内部状態が異なれば、同じ種類の刺激に対して生命体が異なる行動を返しても不思議ではない。しかし、もし生命体の内部状態が条件付けの歴史に依存しているならば、原理的には、行動を記述・説明・予測するために生命体の内部状態まで参照する必要はない、ということになる。

### 1.2.2 節 方法論的行動主義の問題点

行動主義は報酬や罰を用いることで、生命体の特定の行動を強化できることを見出した。

---

<sup>7</sup> ここで「行動の強化」を「形質の選択」と読み替えると、オペラント条件付けによる学習はダーウィンの自然選択と類比的な構造をもつことが分かる。

たしかに、方法論的行動主義にもとづく心理学は 20 世紀後半になって人気を落としたが、条件付けが動物による学習の重要な一過程であるという事実は揺るがない。行動療法は不安障害のような精神疾患に対して精神分析などよりも有効だとされる<sup>8</sup>。行動主義が疑似科学だと判明したわけではない。行動主義が哲学者によってしばしば軽蔑的に扱われていることを思うと、この点はどれほど強調してもしすぎることはない。

しかし、行動主義の発見が今日でも受け入れられているという主張は、生命体の多くの行動が内的状態に訴えることなく記述・説明・予測できるという主張とは別である。生命体の行動の多くが内的状態に言及せずに説明できるというのは疑わしい<sup>9</sup>。方法論的行動主義にもとづく心理学は、生命体の多くの行動を適切に取り扱えないのではないかと疑われる。

まず、報酬という概念は、欲求の充実といった形で、暗黙に心的状態への言及を含んでいるのではないかという疑念がある。このような疑念を退けるためには、例えば、報酬を動因低減 (drive reduction) として特徴づけるのがよいかもしれない。「動因」はそのままで「欲求」の単なる言い換えだが、空腹や渇きといった、比較的単純な生理的メカニズム (ホメオスタシス) に基づいている動因であれば、行動主義者は所与として認めることができるはずである。しかし、動因低減という報酬の特徴づけは上手くいかないだろう。なぜなら、動物は餌を与えられることがなくても特定の行動を学習しうるからである。例えば、十分に餌を与えられて満腹になったラットにゴールに餌のある複雑な迷路を探索させると、ゴールまでの道筋を学習させることができる。すなわち、学習がなされた時点ではラットは満腹なので餌を食べない、つまり、報酬は与えられないのだが、十分に時間がたって空腹になったラットを迷路の入口に置くと、ゴールまで短時間で行くことができる。こうしたタイプの学習を「潜在学習 (lateral learning)」という。さて、潜在学習を条件づけの特殊事例として説明するには、迷路を探索したラットには何らかの報酬が与えられたのだと考えなければならない。一つの方法は、空腹や渇きのような動因の他に、好奇動因 (curiosity drives) のような新種の動因を指定し、そうした新種の動因が低減することをもって行動を強化する報酬がもたらされたと考えることである。しかし、好奇動因の生理的基盤が明らかにならない限り、「好奇動因の低減」は「好奇心が満たされること」といった日常表現の言い換えでしかない。好奇心のような内的状態に訴えて行動を説明するなどという心理主義的な説明を行動主義者は認めないので、好奇動因の低減という概念の使用を認めることはできそうにない<sup>10</sup>。

報酬の概念を不問に付したとしても別のより重要な問題がある。オペラント条件付けに

---

<sup>8</sup> アイゼンク 1988.

<sup>9</sup> 以下の議論は Chomsky 1959; Rey 1997, sec. 4.1 を主に参照している。

<sup>10</sup> 公平のために言っておくと、「動因低減」による「報酬」の定義は現在では廃れており、「行動の直後に出現するとその行動の頻度を上げるような刺激」などと定義するのが現在では一般的である。杉山ほか 1998, p. 5 などを参照。この定義によれば、報酬とはオペラント条件付けをもたらす因子ということになる。チョムスキーはこの種の定義を空虚だと批判しているようが、この批判に私は同意しない。傾向性を効果によって特徴付けることは一般的に行われるからである。cf. MacCorquodale 1970, pp. 86–87.

よって、動物は特定の行動と特定の刺激との連合を学習する。たしかに、迷路を抜けることは行動の一種ではある。しかし、「迷路を抜ける」という記述は多様な身体挙動を包摂するので、迷路を抜けることと餌とを連合させるということは、迷路を抜けるためのある特定の脚の動かし方をラットが学習したということではありえない。例えば、複雑な迷路を探索したことのあるラットは、訓練されたときとは別の場所から出発しても出口までたどり着くことができるし、出口までの経路が塞がれている場合には別の経路を走るし、同じ迷路が水で満たされた場合には泳いで出口にたどり着けることが知られている。こうした即興的かつ柔軟な行動をラットがとれるという事実は、単に刺激と行動が連合されたというよりも、むしろ、迷路の空間的構造についての情報がラットの脳内に保持されていることを示唆する。

ラットですらこのような即興的に知的な行動をとることができる、ということは印象的である。人間の場合、即興的になされる行動の賢明さはラットの迷路学習よりも一層鮮明になる。デネットは次のような例を挙げる<sup>11</sup>。強盗に銃を突きつけられて「財布を出せ」と脅されたとする。多くの人は強盗に遭遇した経験などないので、こういう場合のいかなる行動も強化されていないはずである。それでも多くの人々は賢明な行動、つまり、財布を渡すことができるだろう。なぜなら、強盗にあったとき、賢明な人は「もし肌身を護りたいと思っていて、自分は脅されていると信じているならば、強盗が欲すると思しき事柄をすべきだ」などと考えるからである。信念や欲求を措定すれば、新奇かつ知的な行動を説明することができる。

以上の考察は、心理学の研究対象から生命体の内的状態（心的状態）を締め出すことがいかに難しいかを示唆する。多くの動物は条件付けのような試行錯誤による学習に基づくことなく、新しい状況に対しても即興的にうまく立ち回ることができる。この即興性を説明する一つの方法は、信念や欲求といった心的状態、あるいは、そこまでいかなくても外的環境についての情報処理を行う内的状態を措定するのを認めることだと思われる。

しかし、そのような内的状態を科学としての心理学が研究することなどできるのだろうか。行動主義はこの可能性を否定するが、機能主義は肯定する。

### 1.2.3 節 ホムンクルス機能主義

方法論的問題に対する機能主義的な回答は次のように纏められる。

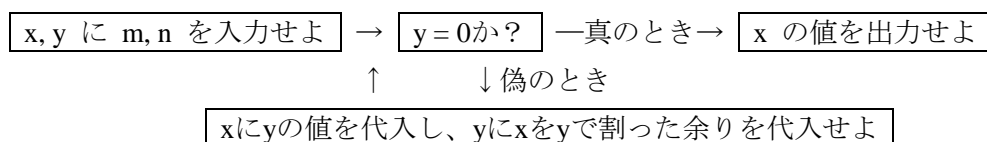
**ホムンクルス機能主義：**心理学の方法論は機能分析（functional analysis）である。

機能分析とは、研究対象となっているシステム的能力や傾向性を、当のシステムを構成する諸部分へと分解した上で、そうした構成部分のもつ能力や傾向性がどのように統合され

---

<sup>11</sup> Dennett 1978, p. 67.

ているのかを述べることで説明するという戦略をいう<sup>12</sup>。例えば、二つの自然数 $m, n$ に対して最大公約数を求める能力を、ユークリッドの互除法によって説明することを考えよう。互除法のアルゴリズムは以下の有向グラフで表現される<sup>13</sup>。



最大公約数を求める能力はこのように説明される。これは機能分析の一例である。

機能分析は心理学的行動主義とまったく無関係ではない<sup>14</sup>。例えば、動物の複雑な行動がどのように学習されるのかを、行動主義者は次のように説明する。ある刺激（弁別刺激）が提示されたときに限って何らかの行動をとるように条件付ける。そして、この行動の結果そのものが次の行動を強化する上での弁別刺激となるようにする。このようにして一連の行動を強化することができる（刺激反応連鎖）。歩く、箸を使うといった日常的な行動は刺激反応連鎖によって学習された行動の具体例である<sup>15</sup>。以上は、複雑な行動をより単純な行動へと分解することで複雑な行動がいかにして可能になるのかを説明するものであり、機能分析の方法論が使われていると言ってよい。

しかし、ホムンクルス機能主義は生命体の内部状態に言及することを許すという点で、方法論的行動主義と袂を分かť。この違いによってホムンクルス機能主義は生命体・システムの知的能力がいかにして可能になっているのかを説明する強力な武器となる。

現代の認知科学はシステムの認知能力が実現されるメカニズムを説明するために、機能分析を積極的に用いている。認知科学者たちは人間のもつ知的能力を、視覚・運動計画・カテゴリー化・帰納・言語理解などといった様々な種類の認知能力へと分解することを好む。こうした認知能力は別の認知能力を入れ子状に含むとされる。例えば、三次元の物体を識別する視覚能力は物体の輪郭を検出したり、その物体が投影する影を同定したりする能力が含まれる。カテゴリー化の能力には長期記憶に貯蔵された様々な心的表象を引き出し、知覚によって形成された表象と比較するといった能力が含まれていると考えられる<sup>16</sup>。人間の知的能力はこのような構成要素の能力によって説明される。

動物の知的活動に対しても機能分析の方法論は有効である。例えば、砂漠のアリは餌を求めて100メートルほども動き回った後で、自分の巣へと回帰する際には最短のルートをとることが知られている。こうしたアリの回帰行動は、自分がどの方向にどれほどの距離を移動したのかという情報を担う何らかの表象が脳内に形成されていると考えれば、自然な

<sup>12</sup> cf. Cummins 1983, pp. 29–30.

<sup>13</sup> 高橋 1991, p. 6.

<sup>14</sup> cf. Cummins 1983, pp. 128–129.

<sup>15</sup> 杉山ほか 1998, pp. 240–242.

<sup>16</sup> Machery 2009, sec. 1.1, 5.1.



形で説明される。自分がどの方向にどれほどの距離を移動したのかが分かれば、推測航法によって出発点に戻ることができる<sup>17</sup>。これはそれほど複雑なアルゴリズムではない。アリが脳内に蓄えていると想定されるこうした表象は「認知地図 (cognitive map)」と呼ばれている。迷路学習したラットは、脳内の海馬において認知地図を保持しているようだ。こうした認知地図は、原始的な信念とみなせるかもしれない。

ホムンクルス機能主義には典型的な批判が向けられてきた。その批判とは、システムの知的能力を、当の知的能力を所有しているホムンクルスをシステムの内部に措定することによって説明するのは悪循環ではないか、という批判である。デネットによれば、この種の批判は、スキナーが志向的語彙の使用に対して示した嫌悪感の中にその先駆けを見出すことができる<sup>18</sup>。

スキナーは、心理主義的な説明、つまり信念や欲求にもとづく行動の説明があまりにも容易に作れてしまうという事実を問題視した<sup>19</sup>。例えば、トムがアップタウン行きのバスに乗ったのは、メイシーズがアップタウンにあると信じており、メイシーズに行きたかったからだ、とある人が述べたとする。スキナーによれば、こうした説明は聞き手にほとんど情報を与えることがない。なぜなら、こうした説明は要求されていることをさっさと実現してくれるホムンクルスを持ち出しているからである。メイシーズがアップタウンにあると信じており、メイシーズに行きたいという前提からアップタウン行きのバスに乗るという行動をトムが導くには、何らかの合理的な計算過程が働いていなければならない。しかし、もし心理学の目標が人間の知性や合理性を説明することにあるとすれば、人間の中にあらかじめ合理性を帰属してしまうような心理主義は、人間の中にそれと同等の知能をもったホムンクルスを措定するのと変わらないのではないか。

デネットは、能力や傾向性を機能分析する際にホムンクルスを仮定することは、直ちに問題を引きおこすわけではないと論じる。なぜなら、仮定されているホムンクルスがシステムよりも、いわば一回り愚鈍でさえあれば、そのようなホムンクルスの系列を何重にも重ねていくことで、最終的には知的とは到底言えないような基本的な機構にたどり着くことがありうるからである。この段階にまでたどり着けば、仮定は落ちた (discharged) と言ってよい。したがって、スキナーが考えるのとは違って、機能分析において呼び出されているホムンクルスはその機能分析が説明すべき当の能力を所有しているわけではない。

実際、ユークリッドの互除法の例に立ち戻ってみれば、ホムンクルスを仮定することが悪循環でないというデネットの指摘が正しいことはかなり明白だと思われる。先のグラフにおいて、割り算の余りを求める能力はまだブラックボックスになっていた。これが、最大公約数を求める能力を説明するために仮定されたホムンクルスである。しかし、割り算の能力はより基礎的な能力（掛け算や引き算の能力）に訴えることで説明できる。ホムンクルスの中に入っているのは、あまり賢くないホムンクルスである。計算能力を説明する

<sup>17</sup> 推測航法 (dead reckoning) についての初歩的な解説は、ジョンソン＝レアード 1989, pp. 36–38 を参照。

<sup>18</sup> Dennett 1978.

<sup>19</sup> cf. Davidson 1995, p. 118.

ために仮定されるホムンクルスは徐々に愚かになっていく。そして、最終的には知性を必要としない単純な作業にたどりつく。

この例はごく初等的な関数の計算能力をホムンクルスによって説明しているだけだが、得られた結論は一般化できる。すなわち、システムの知性を説明するためにホムンクルスを仮定することが常に悪循環を引き起こすわけではない。たとえスキナーが言うように、人間の行動を説明するために「信念」などの志向的語彙を用いることが知的なホムンクルスを仮定することに繋がるとしても、ホムンクルスの仮定が上で述べた意味で最終的に落とされる限り、ホムンクルスの仮定それ自体は不当な手続きではない。

### 1.3 節 概念的問題

本節では、概念的問題に対する行動主義的な回答と機能主義的な回答を比較する。1.3.1 節では行動主義的な回答を提示し、どこに問題があるのかを診断する。それを受けて 1.3.2 節では、機能主義的な回答が行動主義が直面した問題をどうやって回避するのかを考察する。

#### 1.3.1 節 分析的行動主義

概念的問題、すなわち心的語彙の意味を分析するという課題に対する行動主義的な答えは以下のようである。

**分析的行動主義：**すべての心的語彙は、行動と物理現象に関する語彙だけを用いて明示的に定義される<sup>20</sup>。

「行動」ということで何を考えているのか、もう少し明確にしておこう<sup>21</sup>。行動には発汗、射精、心拍や血圧の増大といった生理的反応のほか、手を振る、レバーを押す、迷路を抜けるといった物理的な身体運動が含まれる。物理的な身体運動は、行動の形態や順序、運動に要する時間、運動に用いる身体部位、さらには、それが引き起こす結果の観点から分類することができる。例えば、レバーを押すという行動は、「レバーがしかじかの位置にあるということを引き起こす傾向のある身体運動」という風に、結果によって記述できるので、物理的な身体運動に含めてよい。

こうした「行動」の用法は行動科学の用法にのっとったものであり、日常的な語法よりもある意味では狭く、別の意味では広い。例えば、刺激に誘発されただけの生理的反応を「行動」と呼ぶのは不自然だが、行動科学では一般にこれらも「行動」に含める<sup>22</sup>。逆に、質問に対して応答するといった言語行動は、質問の理解や質問に答える意図などの心的概

---

<sup>20</sup> cf. Kim 2006, p. 60.

<sup>21</sup> 行動に関する以下の説明は Kim 2006, pp. 59, 63; 杉山ほか 1998, pp. 89–90 を参考にした。

<sup>22</sup> Dretske 1988, p. 2.

念を巻き込んでいるため、ここでは「行動」に含めない。

分析的行動主義を支える動機は何だろうか。しばしば挙げられるのは、文の意味内容はそれが成り立っていることが検証されるべき条件によって尽くされる、とする意味の検証主義（verificationism）による動機付けである。実際、分析的行動主義を支持していたのは、カルナップやヘンペルといった検証主義者たちであった<sup>23</sup>。しかし、言語哲学において、意味の検証主義には大きな困難があると指摘されてきた<sup>24</sup>。もし分析的行動主義の主要な動機が意味の検証主義に尽きるのだとすれば、分析的行動主義はそもそも問題のある前提によって動機付けられていることにならないだろうか。

キムによれば、我々は分析的行動主義に対してもう少しチャリティーを働かせることができる<sup>25</sup>。心的状態は各々の主体が私的にしかアクセスできないようにみえるが、他方で、心的状態に関する文の意味は伝達可能であるはずであり、何らかの客観性をもつべきだと思われる。そのためには、心的状態に関する文は、その真理をただちに保証するほど強い条件（＝検証条件）ではなくても、公共的に観察可能な基準を持つべきである。したがって、心的語彙の意味は、公共的に観察可能な物事を意味する語彙によって与えられなければならない。こうした考えが、分析的行動主義を動機付けていると思われる。

それでは、分析的行動主義はどのようにして心的語彙を物理的刺激や行動に関する表現へと翻訳するのだろうか。行動主義者・トールマンによる「予期」（expectation）の定義を考察しよう。餌が迷路のしかじかの位置にあるという予期を心理学者がラットに帰属するとき、彼らは何を言っていることになるのか。トールマンによれば

あるラットは餌が位置Lにあると予期していると主張するとき、我々は、もし（1）餌を奪われていて、（2）経路Pで訓練を受けていて、（3）Pに置かれているのだが、（4）Pがふさがれており、（5）Lに通じるPとは別の経路があるならば、そのラットはLに向かうその経路を走っていくであろうと主張している。あるラットは餌が位置Lにあると予期していないと主張するとき、我々は、そのラットは同じ条件下ではLに向かうその経路を走っていかないだろうと主張している<sup>26</sup>。

この複雑に入り組んだ定義は、ラットが報酬に行き着くための単純な反応を学習したのではなく、餌の置かれている場所をこそ学習したのだ、というアイデアを表現しようとしている。

トールマンの「予期」の定義は、科学哲学で「還元文」と呼ばれる文の一種である。還元文というアイデアは、傾向的（dispositional）な語彙を実質含意によって定義しようとする

<sup>23</sup> cf. Soames 2003, pp. 97–98.

<sup>24</sup> 飯田 1989, 1.3 節。

<sup>25</sup> Kim 2006, pp. 61–62.

<sup>26</sup> Rey 1997, p. 152 より引用した。ただし、この定義が還元文の一種であるという解釈は、レイが述べていることではなく、私自身の考えである。

る論理実証主義者の試みから生まれた。還元文の重要な特徴は被定義項（傾向的な語彙）を部分的にしか定義していないことである。還元文は“ $P \rightarrow (Q \leftrightarrow R)$ ”という形をしていることに注目しよう。傾向的な語彙は  $Q$  に出現するのだが、この文から  $Q$  を左辺にもつ双条件文は導出できない。したがって、 $Q$  に出現する傾向的な語彙を還元文によって消去することはできない。

トールマンが問題にした「予期」は傾向的な語彙の一種である。よく知られるように、傾向性の帰属は実質含意では表現できない。実際に、(1) から (5) の連言と「そのラットが  $L$  に向かうその経路を走っていく」を実質含意でつないだものを、餌が位置  $L$  にあると予期していることの定義とみなすと、(1) から (5) の連言が成り立たないラットは、前件が偽であるために、餌が位置  $L$  にあると予期しているということになってしまう。これは不合理である。そこで還元文の出番となる。トールマンの文章は (1) から (5) の文の連言が成立しているという条件の下で、「あるラットは餌が位置  $L$  にあると予期している」と「そのラットが  $L$  に向かうその経路を走っていく」が同値だと主張している、と解釈できる。記号化すると

トールマンの「予期」:  $(1) \ \& \ (2) \ \& \ (3) \ \& \ (4) \ \& \ (5) \rightarrow (\text{ラットは餌が位置 } L \text{ にあると予期している} \leftrightarrow \text{そのラットが } L \text{ に向かうその経路を走っていく})$

となる。

還元文はあくまでも実質含意のみを使用することで、傾向的な語彙を分析するものである。実質含意へのこうした拘りは、論理実証主義が反事実的条件法や因果性といった様相的な概念を使用することに対して及び腰だったことに由来する。様相論理や反事実的条件法の意味論が整備された現在では、様相的な概念の使用に関して多くの哲学者はそれほど禁欲的ではない。もし様相的な語彙の使用を認めるならば、ラットの「予期」を明示的に定義することもできるかもしれない。例えば、(1) から (5) の連言と「そのラットが  $L$  に向かうその経路を走っていく」を反事実的条件法の結合子でつないで、「餌が位置  $L$  にあると予期している」の定義とする。あるいは、「(1) から (5) の連言を刺激として  $L$  に向かう経路を走っていくような傾向性」と定義する<sup>27</sup>。これらはラットに餌の位置に関する予期を帰属するための悪くないやり方かもしれない。

しかし、多くの「信念」はこのやり方では定義できないだろう。一般に、信念は何らかの欲求と結びつかなければ行動を引き起こすことはない。何らかの刺激の下でしかじかの行動をするということから生命体が抱いている信念内容を決定するには、問題の生命体がどのような欲求を抱いているのかが与えられていなければならない。では、生命体がどのような欲求を抱いているのかを、刺激と行動のみに言及することで決定できるだろうか。たしかに、餌を奪われた状況では生命体（例えば、ラット）が餌を欲求していると仮定し

<sup>27</sup> これら二つの定義は同値ではない。この点は 1.4.4 節でより詳しく論じる。

でもあまり問題ないかもしれない。しかし、たとえこの仮定を認めたとしても、他の多くの「欲求」は同じように単純には定義できない。ある欲求を持つ人は、それを充実する手近な手段があると信じていればそうするだろう。しかし、その欲求を充実する手段が手近にあると信じていなければ、行動に移すこともないだろう。したがって、「信念」が「欲求」よりも先に定義されていなければ、「欲求」を定義することはできない。しかし、そもそも「欲求」に注目したのは「信念」を定義するためであった。

以上の議論から導かれる結論は、信念や欲求は相互に関連することで行動を引き起こすので、「信念」と「欲求」を個別に定義することはできないということである<sup>28</sup>。ある信念をもった主体が同じ種類の物理的刺激に対してどのように振舞うかは、欲求のあり方に応じて異なりうる。それゆえ、刺激や行動に関する語彙だけを用いて一方だけを定義することはできない。どれほど複雑なやり方を思案しても、心的語彙を物理的刺激や行動に関する表現へと翻訳することは原理的に成功しえないと思われる。

分析的行動主義が直面する以上のような困難を、ここでは「信念と欲求の循環の問題」、略して「循環の問題」と呼ぶことにしよう。循環の問題は、心的語彙を行動に関する言明によって定義することは原理的にできない、ということを示唆する。

### 1.3.2 節 意味のネットワーク説とラムジー化

概念的問題に回答するにあたって、機能主義は分析的行動主義の見解をできるだけ保持しようとする。その代わり、心的語彙の意味に関する考え方を改めることで、機能主義者は循環の問題を避ける。心的語彙の意味に関して機能主義者が典型的に採用する見解を、チャーチランドは「意味のネットワーク説 (network theory of meaning)」と呼ぶ<sup>29</sup>。

**意味のネットワーク説：**すべての心的語彙は心理学の理論によって陰伏的に定義される。

「信念」の意味は「欲求」の意味と相互に依存する関係にあるが、これら心的語彙の意味は心理学の理論全体によって規定されていると考えることはできる。とはいえ、意味のネットワーク説を採用するだけでは、まだ心的語彙を定義できたとは言えない。

しかし、この苦境は、理論語を観察語によって明示的に定義することができないという自然科学で広範に見られる現象と類比的なのかもしれない。20 世紀前半の科学哲学者たちは、科学理論の経験的内容、すなわち、観察語のみを使用して特徴づけられる内容を切り出すことに興味を持っていた。その際、観察語による翻訳を与えることのできない理論語

<sup>28</sup> Kim 2006, pp. 65–66. cf. Chisholm 1957, pp. 182–185; Rey 1997, p.154; Braddon-Mitchell & Jackson 2007, pp. 44–45. ただし、チザムのいう “expecting” は期待、つまり、一種の欲求であるが、本稿の「予期」は一種の信念である。

<sup>29</sup> Churchland 1988, p. 56.

の存在が問題となった。

ラムジーは、理論語を観察語へと翻訳することなしに、科学理論の経験的内容を切り出す方法を提示した。公理化された科学理論が与えられたとすると、その理論中に出現するすべての理論語に対して存在量化を施す。こうして得られる文は「ラムジー文」と呼ばれる。ラムジー文は、観察語だけを含む文を演繹する能力に関して言えば、存在量化を施す前の科学理論と同等の表現力をもつ<sup>30</sup>。したがって、ラムジー文は科学理論の経験的内容を余すところなく掬い取っていると考えられる。

ルイスはラムジーの理論語に関する仕事を発展させれば、理論全体によって陰伏的に定義されている理論語を明示的に定義にしてやることができると主張した<sup>31</sup>。ルイスが用いた手法は、今日では「ラムジー化 (ramification)」と呼ばれている。現代の機能主義者は心的語彙を定義するためにラムジー化の手法を好んで利用するようになった。

それでは、ラムジー化とはどのような手法なのか。ルイスはまず、科学理論とは関係のない架空の事例を例に持ち出すことで、ラムジー化の基本的なアイデアを説明する。例えば、ある家で起きた殺人事件を解決しようとしている探偵が関係者を前にして自身の推理を披露する、という場面を考えよう。探偵は以下のように語る。

ふたりの人物をここでは「X」と「Y」と呼んでおこう。このふたりは二週間まえ、あるテニスクラブで会い、スミス殺害の件について話し合った。Xはその会員だがYはそうではない。翌日、Xはストリキニーネを買い、Yに手渡した。Yはそれをスミスのマティーニに混ぜた<sup>32</sup>。

このシナリオは、一つの理論とみなすことができる。X と Y は、この理論における理論語である。聞き手は、X と Y が誰を指示しているのか知らないかもしれない。しかし、探偵の語りが詳細に関して十分に明確であるならば、特定の人物によってしか X と Y の役割は満たされない。それゆえ、我々はある人物の名前を「上の探偵の語りにおいて X の役割を満たす唯一の人物」という確定記述によって明示的に定義できる。

この例において X と Y は、人物という観察可能な対象を指示している。また、X と Y の指示対象が何であるのかは探偵のシナリオとは全く別個のルートで理解されうる。その限りで、探偵のシナリオは科学哲学者が問題にするような科学理論とは性格が異なる。しかし、こうした違いに関わらず、ルイスのアイデアは科学理論にも応用できるほど一般的なものであるのは明らかである。というのも、彼のアイデアは理論語と観察語という対比に依拠していない。そして、理論に出現する理論語は、たとえ独立には定義できないとしても、理論の全体によって陰伏的に定義されているならば、それを明示的な定義へと変換することができるからである。

<sup>30</sup> Ramsey 1929.

<sup>31</sup> Lewis 1970; 1972.

<sup>32</sup> この文章はルイスの具体例をエヴニンが簡略化したものを引用した。エヴニン 1996, p. 246.

心的語彙が何らかの心理学理論において陰伏的に定義された理論語の一種なのだとすれば、ラムジー化によって心的語彙の定義を得ることができる。ラムジー化によって心的語彙を定義する手順は以下のようである<sup>33</sup>。

### ラムジー化による心的語彙の定義

心的状態（痛み、怒り、信念など）と入出力（刺激・行動）の相互関係を適切に記述している心理学理論をThとする。Thは有限公理化可能と仮定する<sup>34</sup>。よって、Thに含まれるすべての文を導出するような単一の式 $\Phi$ が存在する。

Thに出現する心的語彙を $M_1, \dots, M_n$ とする。これら語彙は $\Phi$ にも出現するだろう。そこで、 $\Phi$ の代わりに $\Phi(M_1, \dots, M_n)$ と表記しよう。また、個々の心的語彙 $M_i$ は単称名ではなく「 $x$ は $t$ において $M_i$ である」といった二項述語であると仮定する（ $x$ は生命体を値にとる変数、 $t$ は時点を値にとる変数）。心的状態が性質であるとすれば、このように仮定するのは自然である。

ここで、 $\Phi(M_1, \dots, M_n)$ に出現している任意の心的語彙 $M_i$ を述語変項 $F_i$ で置き換えて存在量化を施すと

$$\exists F_1, \dots, \exists F_n \Phi(F_1, \dots, F_n)$$

が得られる。この存在量化文を「ラムジー文（Ramsey sentence）」という。

ラムジー文には心的語彙は出現しない。だが、まだ心的語彙の明示的な定義にはなっていない（存在量化が現れていることは、ラムジー文が現実の世界に関する実質的な主張を含むことを示唆する）。心的語彙 $M_i$ の明示的な定義を得るには、次のようにする。

$$\forall x \forall t [M_i(x, t) \leftrightarrow \exists F_1, \dots, \exists F_n (\Phi(F_1, \dots, F_n) \& F_i(x, t))]$$

$M_i$ は任意なので、他の心的語彙も同様に定義できる。

ラムジー化の手法は何らかの心理学理論が与えられてはじめて利用可能になる。現在の心理学には形式化された理論は存在しないので、ラムジー化によって心的語彙を実際に定義した、といった業績は存在しない。その意味では、概念的問題に関して意味のネットワーク説がもたらした前進は思弁的な段階にとどまる。

<sup>33</sup> 以下に示した手続きはKim 2006を参考に行っている。この手続きは(i)二階量化を用いており、(ii)一意的な存在量化を用いない、という二つの点で、ルイス自身の手続きとは異なる。二番目の変更点については2.4.2節の註でやや詳しい説明を与える。

<sup>34</sup> ある理論Thが有限公理化可能(finitely axiomatizable)であるとは、論理的帰結の集合がThと一致する何らかの有限の文集合が存在するということである。

しかし、意味のネットワーク説はラムジー化の手法を概念的問題に持ち込んだことで、循環の問題を避けて心的語彙を明示的に定義するのは原理的には可能であることを示している。循環の問題は、心的語彙の意味を孤立させて分析しようとする場合に限って問題となる。心的語彙は物理的刺激と行動に関する語彙だけを用いて明示的に定義できるという分析的行動主義の基本的な発想は、心的語彙の意味に関する意味のネットワーク説と、陰伏的定義を明示的定義に変換するラムジー化の手法をたずさえた意味のネットワーク説によって、救い出すことができる<sup>35</sup>。

### 1.3.3 節 ラムジー化に関する諸注意

心身問題の検討に移る前に、ラムジー化に関して注意点を二つほど補足しておく。

#### (1) 分子論か全体論か<sup>36</sup>

意味のネットワーク説は、信念のような語彙を定義しようとするのと別の心的語彙を用いることが避けられないという循環の問題から抜け出すために提案された。つまり、心的語彙は心理学理論の全体によって陰伏的に定義されていると考えることが許されるならば、その後はラムジー化によって明示的な定義を得ることができる、というのがここでの大筋であった。

しかし、 $M_1, \dots, M_n$  がすべての心的語彙だと仮定した場合には、意味のネットワーク説は不合理な帰結をもたらすように思われる。ラムジー化の手続きによれば、任意の  $M_i$  に関して、 $x$  が  $t$  において  $M_i$  であるのは

$$\exists F_1, \dots, \exists F_n (\Phi(F_1, \dots, F_n) \& F_i(x, t))$$

のときであった。 $x$  が  $t$  において  $M_i$  であることは、 $x$  が  $t$  において  $M_j$  ( $M_j \neq M_i$ ) であることは含意しないものの、 $x$  が  $t$  において  $M_i$  であることは、 $x$  が  $t$  において  $M_j$  でありうるということを含意する。この帰結が不合理だと思われる理由は、特定の心的性質に限っては例化しえない生命体を想定することができるからである。例えば、視覚と聴覚を失ったヘレン・ケラーは、いくつかの心的性質（色の視覚経験など）を例化できない。しかし、だからといって、彼女は信念や怒りの感情といった心的性質すらも例化しえない、と考えるべきではないだろう。なお、ヘレン・ケラーの障害は後天的なものだが、その点はここでは重要でない。最初から眼を欠いている生命体は色の視覚経験をもたないだろうが、もしヘレン・ケラーが信念や怒りの感情を持つならば、その生命体も信念や怒りの感情を持ちうるというてよいだろうからである。したがって、いくつかの心的性質を例化する生命体が、特定

<sup>35</sup> cf. Rey 1997, p.154.

<sup>36</sup> この小節は、山田竹志氏（東京大学）からの示唆に多くを負っている。



の心的性質に限っては例化しえないということは十分に考えられる。このことは、ある心的語彙の意味内容が別の心的語彙と相互依存関係にあるにせよ、すべての心的語彙が相互依存関係にあるとは想定すべきではない、ということを示している。

では、どのように考えればよいのだろうか。一つの提案は、心的語彙の全体をいくつかのクラスに分割して、個々のクラスに属する心的語彙を段階的に定義する、というものである。例えば、信念や欲求は感覚入力からかなりの程度自律しているので、まずはこれらの心的語彙を相互定義する。その後で、知覚能力を信念を用いて定義し、意図を欲求によって定義する、という風に心的語彙を定義することが考えられる。このように心的語彙を段階的に定義するという戦略をとる場合には、 $M_1, \dots, M_n$  を各段階における原始的な (primitive) 心的語彙とみなして、上に示したラムジー化の手続きを複数回行う必要がある。定義の各段階において、理論  $Th$  は  $M_1, \dots, M_n$  以外の心的語彙を含んでいてもよいが、それらは心的でない語彙によってあらかじめ定義されている必要がある。

## (2) 心理学理論 $Th$ とは何か？

心的語彙を陰伏的に定義している心理学理論  $Th$  とは、具体的にはどのような理論だろうか。具体的な公理系が提案されている熱力学などとは違い、心理学に対してそのように形式化された理論を与えることは当座のところ期待できない。そのため、心の哲学においてラムジー化を適用しようとする論者は  $Th$  の正体に関してはかなり投機的にならざるをえないところがある。しかし、投機的ではあってもおおまかな提案がなされているのも事実である。いくつかの提案をみておこう。

心の哲学にラムジー化の方法を適用したルイスは、 $Th$  として常識心理学 (folk psychology) を念頭に置いていた。ルイスによれば、常識心理学とは心的状態と感覚刺激と運動反応の間の因果的関係に関する自明の理 (platitude) を取り集めてくることで形成されるような理論である<sup>37</sup>。例えば、

任意の人について、事態  $s$  を欲求しており、 $p$  することが  $s$  をもたらす手段であると信じており、 $p$  ことができると信じており、実際に  $p$  ことができ、 $s$  と競合する欲求や  $p$  の代案がなければ、その人は  $p$  するであろう<sup>38</sup>。

といった文は、常識心理学の理論に含まれる自明の理の候補である。たしかに、われわれが実際に常識心理学の理論をつかって他者の心を理解・予測しているのかどうかをめぐる論争があるのは事実である<sup>39</sup>。しかし、常識心理学は原理的には人の行動を予測・説明する

<sup>37</sup> Lewis 1972, pp. 257–258.

<sup>38</sup> Churchland 1986, p. 300

<sup>39</sup> 発達心理学では、他人の心を理解する心的過程のあり方をめぐって、理論説 (theory theory) とシミュレーション説 (simulationism) が対立している。この対立については、朴 2011 を参照。

ことができる、とこことでは仮定する<sup>40</sup>。

概念分析のために、常識として我々が所有している自明の理を収集するという発想は、今日では「キャンベラ計画 (Canberra Plan)」と呼ばれる哲学上の方法論に発展している<sup>41</sup>。キャンベラ計画によれば、哲学は分析される概念についての自明の理を収集する第一段階の作業と、自明の理を実現 (realize) するものを探す第二段階の作業という二つのステップを通じて遂行される。第一段階の作業とは、具体的には、自由意志・因果性・道徳的価値などの語彙について、それらの意味を構成する自明の理を我々の日常的な信念から選りわけ作業であり、キャンベラ派の人々によれば、この作業は純粋に概念的な探求である。第一段階の作業により、どのようなものが自由意志や道徳的価値とされると日常的には考えられているのかが明確にされる。その成果を踏まえて、第二段階では、哲学的諸概念に課せられた役割がいかんして実際に担われているのかが明らかにされる。第二段階の作業では、アポステリオリな科学的知見を持ち込むことが許されている。例えば、痛みの因果役割を実際に果たしている人間の身体組織が何であるかを追究する作業に携わるのは、科学者の仕事である、という風にである。

しかし、第一段階で集められる自明の理は、たんに誰でも知っているような事柄であればいいというわけではない。たんに市井の人々に尋ねるだけで心的語彙を定義するに足るまともな理論が得られるというのは、見込み薄だからである。一つの懸念は、市井の人々は人間の心に関して哲学者の眼から見るとかなり粗野な信念を抱いているかもしれない、ということである<sup>42</sup>。例えば、人間の心は非物質的であるといった心身二元論や、人間の心は神的な存在者によって創造されたとか、人間は永遠の生命をもった不死の心をもつ、といった信念などである。また別の懸念としては、市井の人々の信念を集めるだけで心的状態を実現する物的状態を一意的に特定できるという保証がない、ということも挙げられる<sup>43</sup>。こうした懸念が正当であるならば、ルイスの提案とは別にThを手に入れる方法を考える必要が生じてくる。

そこで考えられる代案は、心的語彙を定義するための心理学理論に経験科学としての心理学 (科学的心理学) の知見を取り入れることである。だが、この提案は科学的心理学の役割を、一般人が常識として知っている心理学的内容に単純に補完するだけだと前提していないだろうか。この前提は疑わしい。科学的心理学において一般人が日常的に用いている心的語彙 (とりわけ、「信念」や「欲求」といった志向的語彙) が居場所をまったくもたない、という可能性もある。つまり、信念や欲求といった概念は人間の行動を科学的に予測したり説明したりする上で有用ではないことが判明するかもしれない。その場合でも、

<sup>40</sup> Rey 1997, p. 201.

<sup>41</sup> cf. Papineau 2015, sec. 2.3. 「キャンベラ計画」という名称は、この方法論の支持者 (フランク・ジャクソン、マイケル・スミス、フィリップ・ペティットなど) がキャンベラに位置するオーストラリア国立大学に所属していたことに由来する。キャンベラ計画については、次の論文集が参考になる。Braddon-Mitchell & Nola 2009.

<sup>42</sup> Rey 1997, pp. 185–186.

<sup>43</sup> Field 1978, p. 56n21a; Williamson 2007, p. 121n34.

科学的心理学の理論に対してラムジー化を施すことはできるが、われわれが日常的に用いる心的語彙が定義されることはないだろう。

これはもっともな疑念である。しかし、たとえ科学的心理学の中に日常的な心的語彙が出現しないことが判明したとしても、日常的な心的語彙に類似した語彙が代わりに生き残る可能性はある。例えば、日常的な生物種 (species) の概念は生物学からは追い出されていても、進化論的な生物種概念はなお生き残っている。それと同じようなことが、日常的な心的語彙と科学的な心的語彙の間にも成り立つかもしれない<sup>44</sup>。その場合には、日常的な心的語彙を定義することはできなくも、われわれは科学的心理学の理論を援用することで、心的語彙の意味をいくらか明確にできるだろう。

## 1.4 節 心身問題

本節では、心身問題に対する行動主義的な回答と機能主義的な回答を比較する。1.4.1 節では存在論的行動主義を定式化する。1.4.2 節では存在論的行動主義に対して二つの問題点を指摘する。1.4.3 節では因果役割機能主義を定式化し、この立場が存在論的行動主義に対する二つの問題に対してどのように対処できるのかを検討する。1.4.4 節では存在論的行動主義の別の定式化を検討する。

### 1.4.1 節 存在論的行動主義

心身問題に対する行動主義的な回答は、以下のように定式化される。

**存在論的行動主義：**すべての心的性質は現実の物理的行動、あるいは、可能的な物理的行動のパターンである<sup>45</sup>。

この立場は心的性質を現実的な行動ないしは可能的な行動のパターンと同一視し、心的性質が神経学的な性質のように皮膚の内部に関わる性質ではないと考える。なお、前節と同様に、ここでも「物理的行動」は生理的反応と身体運動に限られる。

可能的な行動への言及は、信念や欲求のような心的状態と行動とのつながりをつけるために必要となる。心的状態の異なる二人は、現実の行動において違いがなくても、何らかの可能な状況における行動には違いが表面化すると思われるからである。例えば、鹿を撃ってみたいという欲求を持つ人は、適切な機会が得られなければ、そうした欲求を持たない人と現実の行動は変わらないかもしれない。しかし、可能的な状況のもとでは、両者の行動には違いが出るだろう。なぜなら、鹿を撃ってみたいという欲求を持つ人は、もし機

---

<sup>44</sup> Sterelny 2003, p. ix.

<sup>45</sup> cf. Kim 2006, p. 71; 金杉 2002, p. 98.

会があれば鹿を撃つような人だからである。そうした機会が現実 to 得られるときにもしかるべき行動をおこすだろうが、たとえそうした機会が現実 to 成り立っていない場合でも、もし機会があればしかるべき行動をおこすという点で、彼女は鹿を撃ててみたいという欲求を持たない人とは異なる。

存在論的行動主義からの重要な帰結は、心的性質を現実的な行動ないしは可能的な行動のパターンと同一視することで、心的性質の例化が行動を引きおこすという考え、すなわち、心身因果を否定するということである。鹿を撃ててみたいという人が鹿を撃ったとしても、その行動は鹿を撃ててみたいという欲求によって引きおこされたと言うことは認められないということである。その理由は、心的性質を帰属する言明は法則言明に類似しているという考えに基づいている<sup>46</sup>。決定論的な自然法則は、時点 $t$ における世界のスナップショットと合わせて時点 $t'$ における世界のスナップショットを論理的に帰結する<sup>47</sup>。しかし、自然法則は世界の状態遷移を支配するパターンを与えるものではあっても、時点 $t'$ における世界のあり方の原因であるとは言えない。さて、もしも傾向性の帰属が法則言明と似ているならば、傾向性を例化していることがその傾向性が発現したことの原因とはやはり言えないだろう。例えば、「 $x$ は鹿を撃ててみたい」という欲求の帰属は

$x$  は適切な機会を得ている  $\square \rightarrow x$  は鹿を撃つ

という反事実的条件法によって分析されると仮定しよう。たしかに、この条件文は特定の個体に関わるという意味で、自然法則のような普遍性を持ち合わせてはいない。しかし、それでもこの条件文からは「 $x$ は適切な機会を得ている」という前提のもとで「 $x$ は鹿を撃つ」が前件肯定式によって帰結する<sup>48</sup>。したがって、法則の場合と同様に、 $x$ が鹿を撃ててみたいという欲求をもつことは、 $x$ が鹿を撃ったということの原因ではないだろう。このように、心的性質を帰属する言明は法則言明に類似しているために、心的状態は行動の原因ではない。以上が存在論的行動主義からの帰結である<sup>49</sup>。

ただし、存在論的行動主義は、鹿を撃ててみたいという欲求を持つ人とそうでない人が物的性質に関して異なるということまで否定する必要はない<sup>50</sup>。両者の現実の行動が異なる場合、そこに何らかの原因があると考えるのは理に適っている。例えば、それは両者の脳状態の違いに由来しているのかもしれない。存在論的行動主義はこうした仮説を十分に認めることができる。存在論的行動主義が認められないのは、鹿を撃ててみたいといった心的性質の例化が行動を引きおこすという考えである。心的性質は行動のパターンのことであって、行動を引きおこすようなものではない、というのが存在論的行動主義の立場であ

<sup>46</sup> cf. Ryle 2009, pp. 101–109; Mumford 1998, pp. 133–136.

<sup>47</sup> 「時点 $t$ における世界のスナップショット」とは、時点 $t$ における世界全体の状態記述を意味する。

<sup>48</sup> 本稿で反事実的条件法の論理を話題にするときは、Lewis 1973 の VC を標準的な体系として扱う。VC では、反事実的条件法の前件肯定式は妥当な推論である。

<sup>49</sup> cf. Morton 1997, p. 329.

<sup>50</sup> Mumford 1998, pp. 16–17.

る。この立場は現実の行動に何らかの原因があることまで否定するわけではない。

本節を終える前に、存在論的行動主義と 1.3 節で取り上げた分析的行動主義の関係についても少し述べておこう<sup>51</sup>。結論からいうと、存在論的行動主義は分析的行動主義とは独立である。まず、分析的行動主義は存在論的行動主義を含意しない。なぜなら、すでに述べたように、存在論的行動主義は痛みや欲求といった心的状態をあくまでも行動ないし行動のパターンとして扱おうとするので、心的状態を行動の内的な原因とみなすことはできないのに対して、分析的行動主義は心的語彙を行動に関する語法へと翻訳することを目標としているので、「うめき声をあげる原因」といった言い回しを用いること自体は問題にならないからである。他方で、存在論的行動主義は分析的行動主義を含意しない。分析的行動主義とは異なり、存在論的行動主義は心的語彙の意味に関する概念分析には関わりをもたない。上の条件文における「適切な機会を得ているならば…」のような表現は、存在論的行動主義が心的語彙の意味を明らかにするという目的にとっては使用すべきものではないだろう。このことは、心的語彙を行動に関する語法へと還元することが不可能であるとしても、存在論的行動主義は生き残るかもしれないということを示している<sup>52</sup>。

#### 1.4.2 節 存在論的行動主義の諸問題

本節では、存在論的行動主義の問題として

- (1) 心身因果を否定することは適切か？
- (2) 身体運動を伴わない、顕在的な心的性質をどう扱うのか？

という二つの問題を取り上げる。

##### (1) 心身因果の否定

存在論的行動主義からの帰結は、心的性質は行動に付随するものであって、行動の内的な原因ではないということであった。この帰結は心身因果の否定に相当する。

しかし、帰結は受け入れがたい。たしかに、心身因果が成立するということを決定的に証明するのは難しい。しかし、心身因果が成立するという見解が我々の常識にかなりよく適合している、ということを示すことならばできる。クレインは心身因果を擁護する二つの論証を提示している<sup>53</sup>。

一つ目の論証は、行為論に見出される。この論証は、思考実験によって紹介することができる。ボレスラフという男は弟を強く憎んでおり、自分の人生がうまくいかないのは弟が妨害しているからだと信じているとする。このとき、ボレスラフは弟を殺す理由をもつ

---

<sup>51</sup> Kim 2006, p. 71.

<sup>52</sup> cf. 金杉 2002.

<sup>53</sup> Crane 2003, pp. 59–62.

ている。さて、ここで二つのシナリオを考える。一つ目のシナリオはこうである。ボレスラフは、ある夜、ある酒場でケンカに巻き込まれたのだが、偶然にも変装してその場にいた自分の弟をそれと知らずに殺してしまった。この場合、ボレスラフは弟を殺す理由を持っており、そして実際に弟を殺したわけだが、その理由に基づいて弟を殺したわけではない、と思われる。二つ目のシナリオはこうである。ボレスラフは、ある夜、ある酒場でケンカに巻き込まれ、偶然にもその場にいた弟を見つけて、どさくさにまぎれて殺した。この場合、ボレスラフは弟を殺す理由を持っており、その理由に基づいて弟を殺した、と思われる。これら二つのシナリオを、あるいは、行為の理由と行為がそれに基づいてなされる理由を区別するものは何か。この問いに対する最も自然な答えは、因果性の有無である。行為の理由が信念や欲求だとすれば、この思考実験は、心身因果が成立するというを示している。

二つ目の論証は、実際に我々は心的状態が物理的な原因によって影響されることを認めている、というものである。興奮剤やプロザックなどの化学物質を服用することで、心的状態が変化するということはほとんど自明の理であるように思われる。このように、心的出来事や心的状態が何らかの物的出来事の結果であるならば、心的出来事や心的状態が行動を引き起こすことはありえないと考える理由はない。

これら二つの論証は説得的であり、たしかに心身因果は成立するという印象を与える。したがって、心身因果を否定する存在論的行動主義を受け入れることはできない。

以上の議論に一つ気がかりな点があるとすれば、現代の形而上学において、心身因果がいかんにして成立しうるのかという問題をめぐる論争がある、ということである<sup>54</sup>。しかし、この論争は心身因果などそもそも成立しないと主張することで解消されるべきではないと思う。心身因果が成立すると考える理由は十分にあるのであって、むしろ心身因果がいかんにして成立しうるのかを考察すべきである。

## (2) 身体運動を伴わない、顕在的な心的性質

存在論的行動主義の主張は、すべての心的性質は現実ないし可能的な物理的行動のパターンであるというものであった。しかし、この主張は、そもそも身体運動を伴わない心的性質には当てはまらない。例えば、痛みなどの感覚の生起がそれである。たしかに、痛みを感じる人は、うめき声をあげるなどといった物理的行動をとるかもしれない。1.4.1 節で痛みをうめき声を上げるといった行動と同一視することを考えたのはそのためであった。しかし、痛みを感じていることをおくびにも出さないということもありうる以上、痛みを行動と同一視するのはやはりもっともらしくないのではないか。むしろ、通常の条件では、痛みを感じるのがうめき声を引きこすと考えるのが自然に思われる。

ライルは、感覚を持つことは概して心理学と無関係だとして、痛みなどの感覚の生起は心的性質ではないと提案した。彼によれば、感覚能力について述べることは、その人のオ

---

<sup>54</sup> 例えば、本稿 2.5.2 節の議論を参照のこと。

能とか人柄について述べることではない。耳が聞こえないとか乱視であることは、愚かさとか卑劣であることとは関係がない。これらを治すことは耳鼻科医や眼科医の仕事であり、道徳家や精神科医の仕事ではない。感覚をもつことは知性などの性質を発揮するのとは異なるので、「我々は爬虫類に対して感覚を認められないほど誇り高くはない」<sup>55</sup>。この議論は、爬虫類に対しては心的性質を帰属できないように「心的」という語の適用範囲が制限されることを含意する。このように制限された「心的」という語の用法は、絶対に採用できないというわけではない。とはいえ、たとえ感覚の生起は心的性質ではないという提案を受け入れるとしても、「それでは、人間や爬虫類に共通する痛みとは一体何なのか」という問題は相変わらず残るのではないだろうか。

プレイスやスマートなど初期の同一説論者は、感覚の生起を存在論的行動主義にとって致命的な事例とみなし、存在論的行動主義をすべての心的性質に適用することを諦めた。ただし、彼らは信念や欲求などの心的状態に関しては存在論的行動主義を採用しつつけており、行動主義は同一説によって補完されるべきだと論じた。行動主義と同一説の二本立てというわけである。これに対し、アームストロングは、プレイスらの考え方は心的性質一般に関する統一的な説明を与えるのを安易に放棄していると批判した<sup>56</sup>。アームストロングの批判はもっともであるように思われる。1.4.3 節では、因果役割機能主義のもとで心的性質一般に統一的な見解を与えられる見込みがあることを示す。

ちなみに、存在論的行動主義が扱えそうにない心的性質は痛みなどの感覚経験に限らない。例えば、判断、推論、暗算、黙読といった心的行為（*mental act*）について考えてみよう。これらの心的行為、とりわけ黙読や暗算には心的イメージを抱くという心的性質が関わっているが、存在論的行動主義は「行動」を生理的反応と身体運動に限るため、心的行為を行動に含めることはできない<sup>57</sup>。したがって、心的行為は身体運動を伴わない心的性質とみなさざるを得ないと思われる。

#### 1.4.3 節 因果役割機能主義

心身問題に対する標準的な機能主義の回答は、次のように定式化される。

**因果役割機能主義：**すべての心的性質は、それぞれの心的性質に特有の因果役割

---

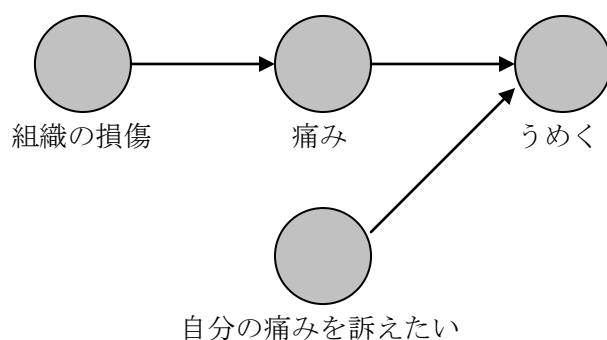
<sup>55</sup> Ryle 2009, p. 184.

<sup>56</sup> Armstrong 1968, p. 81.

<sup>57</sup> ライルはしばしば行動主義者とみなされるが、彼が「行動」として記述するものは、しばしば心的行為を巻き込んでいる。例えば、ライルは、地球は円いと信じるとはしかじかの推論や想像をするといった傾向をもつことだと述べるが（Ryle 2009, p. 32）、推論や想像は心的行為であって身体行動ではない。では、このことはライルが行動主義の擁護に失敗したことを意味するのだろうか。それとも、ライルはそもそも行動主義者ではなかったということの意味するのだろうか。これは解釈上の問題として興味深いが、本稿はこれ以上のライル解釈には踏み込まない。ライルが行動主義者かどうかに関しては、ソームズとハッカーの間で論争があった。Soames 2003, chap.4; Hacker 2006, p. 124; Soames 2006, sec. IV. cf. Tanney 2009.

の担い手を例化しているという性質である<sup>58</sup>。

性質の因果役割 (causal role) とは何か。管見に触れた限りでは、因果役割の明確な定義は、この用語を心の哲学に導入したルイスを含めてどの文献を探しても見当たらない。しかし、性質の因果役割ということで、哲学者たちはおおよそ、その性質への言及を含む因果的説明ないし因果モデルにおいてその性質が占めている役割のことを考えているようである。例えば、単純なケースとして、ある人がうめき声を上げたという出来事を、その人が腕を怪我したために痛みが生じ、その痛みがうめき声を引き起こした、と説明する場合を考えよう (以下のニューロン図を参照)。



このとき、痛み特有の因果役割とは、怪我によって引き起こされ、うめき声を引き起こすことである<sup>59</sup>。この因果役割の担い手となる性質が何であるかは生命体によって異なりうる。例えば、人間の場合には、C繊維発火がこの因果役割の担い手であり、イルカの場合には、D繊維発火が同じ因果役割の担い手である、といった可能性が考えられる。因果役割機能主義によれば、C繊維発火にせよD繊維発火にせよ、この因果役割の担い手となる何らかの性質を例化しているならば、その生命体は痛みを例化しているとされる。

もっとも、ここで述べた因果役割は議論のために簡略化したものであって、痛み特有の因果役割が実際にこのようなものであるとは期待できない。現段階で言えるのはかなり一般的なことである。すなわち、個々の心的性質に特有の因果役割とは、感覚入力と運動出力、そして他の心的性質に言及する複雑な因果的説明によって特徴づけられる、という程度のことである。なお、ここで他の心的性質に言及する必要があるのは、ある心的状態がどのような行動を引き起こすかは主体が他にどのような心的状態にあるかに依存する、

<sup>58</sup> 2章の内容を先取りすることになるが、実は、この定式は役割機能主義を前提している。

<sup>59</sup> cf. Kim 2006, p. 280. この図によると、組織の損傷によって引き起こされることは痛みにとって本質的であるかのようだ。しかし、組織の損傷によって生じた状態でなければ痛みではないのだろうか。例えば、スワンプマンの思考実験に対する一般的な直観は、様々な心的状態にある主体が突如として創造されることがありうると告げている。このように、心的状態にとってそれがどのように生じたのかは本質的ではない、という考えもありうる。アームストロングはこの点に関して慎重であり、次のように述べている。「心的状態の概念は何よりもまずある種の行動を引き起こしそうな人の状態の概念である。…いくつかの心的状態の場合においてのみ、それらはある種の刺激によって引き起こされそうなその人の状態でもある。しかし後者の定式は二次的なものである」。Armstrong 1968, p. 81.



という 1.3.1 節でも確認した論点を踏まえてのことである。

行動の因果的説明においてどんな種類の信念と欲求が必要となるのかを完全に解明することはできないかもしれない。しかし、信念・欲求に言及する行動の説明をニューロン図で表現できないと考える原理的な理由は、さしあたり特に見当たらない。

この点に関連して、因果役割機能主義は心的語彙の意味に関する概念分析には関わらないという点を指摘しておこう。例えば、心理学の理論が公理化不可能であるために心的語彙を物理的刺激や行動に関する語法に翻訳できないとしても、心的性質に特有の因果役割というものがある（きわめて複雑であろうが）あるのかもしれない。その場合には、1.3 節で論じた意味での機能主義（分析的行動主義＋意味のネットワーク説）は成り立たないが、因果役割機能主義は生き残る。

逆に、1.3 節で論じた意味での機能主義は、心的語彙の指示対象の存在論的身分について特に制限を設けていない。そのため、かつて「エクトプラズム」と表現されていた非物質的で幽霊のような対象が何らかの心的語彙によって述定されるという可能性は原理的には排除されていない<sup>60</sup>。他方で、因果役割機能主義はこのような可能性を許容するとは思えない。因果役割の担い手は物的性質しかありえないと仮定するならば、因果役割の担い手をもつという性質を例化するものは物的対象しかありえない。そして、この仮定はもっともらしいので、心的性質を例化する対象はエクトプラズムではありえない<sup>61</sup>。したがって、分析的行動主義が存在論的行動主義と独立だったように、因果役割機能主義は 1.3 節で論じた機能主義からは独立している、ということになる。

以上の考察を踏まえて、因果役割機能主義が 1.4.2 節で取り上げた存在論的行動主義の諸問題をどのように回避・克服しているのかを順に考察する。

### (1) 心身因果

存在論的行動主義とは違い、因果役割機能主義者は心的性質が行動に対して何らかの因果的効力をもつという考えを否定しない。ただし、人間が心と身体という二種類の実体から構成され、心は物理的世界には属さず、神秘的な仕方では身体と因果関係に立つと主張する実体二元論とは異なり、因果役割機能主義者は心的性質を例化するものはあくまでも物的対象であるとする<sup>62</sup>。この意味で、因果役割機能主義は実体一元論に属する。

しかし、1.4.2 節でも述べたように、多くの哲学者がいかんして心身因果が可能なのかを説明するという課題に取り組んできた。因果役割機能主義それ自体は、心身因果がいかん

<sup>60</sup> Lewis 1966, p. 102. 非物的対象の候補としては、他に靈魂やセンスデータなどが考えられる。実体一元論はこうした対象が存在することを否定する。

<sup>61</sup> 一ノ瀬正樹教授からは、裁判所の決定は「因果的役割を果たせ」るが、「それが物的対象だというのは、きわめて納得しにくい」という批判をいただいた。正直なところ、私自身は、裁判所の決定が何らかの因果役割と結びついているのかどうか、よく分からない。また、仮に裁判所の決定が何らかの因果役割と結びついているとしても、この役割の担い手が何らかの物的性質ではありえないというのもよく分からない。決定を下した人々の脳状態は、この役割の担い手となる性質の候補かもしれない。

<sup>62</sup> ここでの実体二元論の定式はソームズによる。Soames 2003, pp. 93–94.

して可能であるのかを説明するわけではない。因果役割機能主義が心身因果の可能性と両立可能であることは別個に示さなくてはならない。この点は 2.5.3 節でさらに議論する。

## (2) 身体運動を伴わない、顕在的な心的性質

存在論的行動主義は、心的行為や感覚の生起などの顕在的な心的性質について、まったく扱えないか、同一説に訴えるなどして顕在的でない心的性質（心的状態）とは別建ての説明を与える必要があった。しかし、因果役割機能主義はどちらの可能性も免れている。心的状態は何らかの物的状態と同一であり、顕在的な心的性質は何らかの顕在的な物的性質と同一である、という風に並行的に扱うことができる。

### 1.4.4 節 傾向性と因果役割

1.4.1 節では、存在論的行動主義を次のように定式化した。

**存在論的行動主義：**すべての心的性質は現実の物理的行動、あるいは、可能的な物理的行動のパターンである。

文献では、「可能的な行動のパターン」の代わりに、「行動の傾向性」という表現が用いられることがある。つまり、行動主義とは「すべての心的性質は現実の物理的行動、あるいは、物理的行動の傾向性である」といった立場である、と。

一般に、傾向性の概念は可能的な状況における物体のふるまいと関わりを持っている。例えば、割れやすいグラスは適度な衝撃を受けたという状況の下で割れる。また、物体がある傾向性をもつことは、その傾向性が現実が発現していることを含意しない。割れやすいグラスは現実には割れているとは限らないだろう。これらの論点は心的状態にも当てはまるように思われる。例えば、鹿を撃つてみたいという欲求を持つ人は現実には鹿を撃つとは限らず、しかるべき機会を得たときに鹿を撃つだろう。もちろん、割れやすさのように単純な傾向性と比べると、心的状態ははるかに複雑である。割れやすい物体が画一的な反応を示すのと比べ、鹿を撃つてみたいという欲求の持ち主は多様な仕方その欲求を表出する。しかし、割れやすさとの類比が成り立つ以上は、心的状態を何らかの行動の傾向性と考えることは意味をなしているように思われる<sup>63</sup>。

「可能的な行動のパターン」を「行動の傾向性」と言い換えることは、傾向性の反事実的分析（counterfactual analysis）によってさらに動機づけられる。反事実的分析によれば、刺激  $s$  に対して反応  $r$  を返すといった単純な傾向性  $D$  は

$x$  は時点  $t$  で傾向性  $D$  をもつ  $\Leftrightarrow$  ( $x$  は  $t$  で  $s$  を受ける  $\square \rightarrow x$  は  $t'$  で  $r$  する)

<sup>63</sup> cf. Mumford 1998, sec. 1.2.

と分析される<sup>64</sup>。例えば、「このガラスは時点 $t$ で割れやすい」という傾向性帰属は「このガラスは時点 $t$ で適度な衝撃を受けたならば $t'$ で割れるだろう」と分析される。傾向性の反事実的分析を受け入れるならば、物体に対する傾向性の帰属は見かけ上は物体の性質を記述してはいるが、実際には、その物体に生じうる出来事のパターンを簡潔に表現するための言い回しだと考えることが正当化されるだろう。

傾向性の中には多様な刺激に対して多様な反応をするものがある。例えば、電荷  $q$  を帯びているといった性質がそうである。電荷を帯びている物体は条件に応じて決まった反応を返すが、そうした条件と反応のペアはクーロンの法則によって記述される。すなわち、何かが電荷  $q$  を帯びているとは、それが電荷  $Q$  を帯びている対象と距離  $r$  におかれるとお互いに  $F (= \epsilon Qq/r^2)$  の力で引き寄せ／反発しあうということである。この種の性質は、多様な仕方で「顕在化する傾向性 (multi-track dispositions)」と呼ばれる。

多様な仕方で顕在化する傾向性の存在は、傾向性の反事実的分析に対する反例にはならない。多様な仕方で顕在化する傾向性は、反事実的条件法の連言によって定義できるかもしれないからである。例えば、刺激  $s_1$  に対して反応  $r_1$  を、刺激  $s_2$  に対して反応  $r_2$  を…刺激  $s_n$  に対して反応  $r_n$  を示す、という性質  $F$  は

$$Fx \leftrightarrow [(S_1x \square \rightarrow M_1x) \& (S_2x \square \rightarrow M_2x) \& \dots \& (S_nx \square \rightarrow M_nx)]$$

と分析できるかもしれない<sup>65</sup>。

しかし、傾向性の反事実的分析には、性質の形而上学で二つの問題が指摘されている。一つは異論の余地があるが、もう一つはより説得的である。

傾向性の反事実的分析の一つの問題は「マスク」と呼ばれるタイプの反例の存在である。割れやすさを例にとろう。反事実的分析によれば、割れやすいという傾向性は、強い衝撃を受けるならば割れるだろう、ということである。しかし、少し複雑な状況を考えることができる。例えば、あるガラスが強い衝撃を受けるならば、そのガラスにはラップがしてあるであろう、という反事実条件文が成り立つと仮定する。この場合には、そのガラスは割れやすいという傾向性をもち続けるにも関わらず、強い衝撃を受けるならば割れるであろうという反事実条件法は成立しない。よって、反事実的条件文が真であることは、傾向性帰属の必要条件ではなく、傾向性の反事実的分析は成功していない。

この議論はそれほど説得的ではないかもしれない。例えば、ラップしてあるガラスには衝撃が届かないのだから、そのガラスは強い衝撃を受けたにも関わらず割れなかったということにはならない、よって、上の事例が示しているのは、せいぜい「そのガラスの周囲が強い衝撃を受けたならば割れるだろう」という反事実条件文は成り立たないということ

<sup>64</sup> 鉄が磁性を獲得したり、ゴムが弾力性を失ったりするように、傾向性の帰属は時点に相対化される。この論点はシンプルだが見過ごされることもある。Mellor 1974, sec. 3 を参照。

<sup>65</sup> Bird 2010, pp. 21–24.

に過ぎないのではないか、などと反論することができよう。もしこの反論が正しければ、傾向性の反事実的分析はマスクの事例に基づく反論からは擁護できる。

しかし、反事実的分析にはさらに別の問題が指摘されている。それは「フィンク (fink)」と呼ばれるタイプの反例の存在である<sup>66</sup>。フィンクとは、傾向性を特徴づける刺激が、まさにその傾向性を消滅させてしまうようなケースの総称である。次のような例を考えよう。割れやすいグラスが机の上に置かれている。そのグラスの近くには常に魔法使いが控えており、グラスに強い衝撃が加えられるときには神秘的な力によってグラスの材質をプラスチックに変えてしまうのだとしよう。このとき、「そのグラスは強い衝撃を受けたならば割れるだろう」という反事実的条件文は偽である。しかし、強い衝撃を加えられていない限りでは、そのグラスは割れやすさを有していると思われる。したがって、反事実的条件文が真であることは、傾向性帰属の必要条件ではない<sup>67</sup>。

この議論に対しては、反事実的条件法の論理は前件強化 (antecedent strengthening) を妥当な推論とは認めないのだから、上の事例は反例になってないと言われるかもしれない。要するに「そのグラスは強い衝撃を受けたならば割れるだろう」という反事実的条件文は真であり、「そのグラスは (強い衝撃をうけ、かつ、魔法をかけられた) ならば、割れるだろう」という反事実的条件文こそが偽なのである、と。

この反論は、「そのグラスは (強い衝撃をうけ、かつ、魔法をかけられた) ならば、割れるだろう」という反事実条件文が偽であると主張する限りではまったく正しい。しかし、このことは「そのグラスは強い衝撃を受けたならば割れるだろう」という反事実条件文が真であることを確立するわけではない。フィンクという反例のポイントは、「そのグラスは強い衝撃を受けたならば割れるだろう」という発話が、「そのグラスが強い衝撃を受けたならば魔法をかけられるだろう」という反事実的条件文が真であるという状況下でなされている、と想定しているところにある。反事実的条件法の論理では

$$A \Box \rightarrow B, A \Box \rightarrow C \models (A \& C) \Box \rightarrow B$$

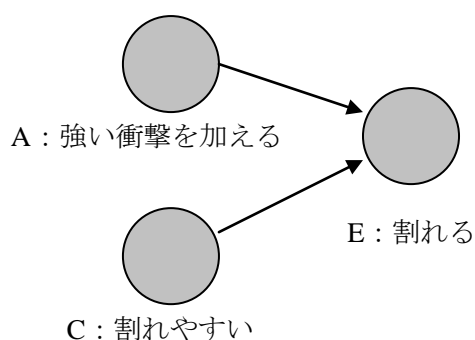
という推論図式は妥当であるため、A に「そのグラスは強い衝撃を受ける」、B に「そのグラスは割れる」、C に「そのグラスに魔法がかけられる」を代入すると、結論は偽なので、二つのうちどちらかの前提が偽となる。仮定により  $A \Box \rightarrow C$  は正しいので、結局  $A \Box \rightarrow B$  が偽となる。つまり、「そのグラスは強い衝撃を受けたならば割れるだろう」は偽である。

<sup>66</sup> Lewis 1997. このタイプの反例は C. B. マーティンに由来する。“fink”はスパイや裏切り者を意味する英単語だが、哲学の邦語文献では「フィンク」で流通しているので、その慣習に従う。

<sup>67</sup> 同じような論法で、十分条件に対する反例を作ることでもできる。プラスチックのグラスがあるとしよう。そのグラスの近くには常に魔法使いが控えており、グラスに強い衝撃が加えられるときには神秘的な力によってグラスの材質を薄いガラスに変えてしまう。このとき、「そのグラスは強い衝撃を受けたならば割れるだろう」という条件文は真である。しかし、強い衝撃を加えられない限りでは、そのグラスはプラスチックのままであるので割れやすさを有しているとは思われない。したがって、反事実的条件文が真であることは傾向性帰属の十分条件ではない。

さて、以上の議論が正しければ、物体に対する傾向性の帰属は、まさに見かけどおりにその物体の性質を記述しているのものであって、出来事のパターンを表現しているわけではないことになる。したがって「可能的な行動のパターン」を「行動の傾向性」と言い換えることには賛成できない。実のところ、このように考えるからこそ、本稿では、一般に「行動主義」と結び付けられている「行動の傾向性」という表現を本節に至るまでは使用せず、「行動のパターン」と表現してきたのであった<sup>68</sup>。

しかし、傾向性の帰属が可能的な状況において当の物体に生じうる出来事のパターンを記述することでないとするれば、それは一体何なのか。標準的な見解によれば、傾向性は性質の一種であるので、傾向性帰属は何らかの性質の帰属である。しかし、ここでより踏み込んで述べるならば、傾向性は因果役割によって特定される性質の一種とみなしてもよいと思われる。一般に、傾向性は特定の刺激条件の下でどのような出来事が典型的に生じるのかを記述することで特定される。割れやすさの場合には、それは、強い衝撃を加えるという出来事を刺激条件として割れるという出来事を典型的に引き起こすような状態のことである。個々の出来事・状態と因果関係をニューロン図で示すと以下のようなになる。



この図はA, Cのニューロンが共に発火するときEが発火することを示す<sup>69</sup>。

傾向性は一般に、このニューロン図と同様の因果役割を持つと思われる。マスクの事例

<sup>68</sup> 心の哲学で「傾向性」という表現を明示的に用いはじめたのは、一般に行動主義者として認識されているライルである。しかし、彼は「傾向性」を行動のパターンという意味で用いている。「傾向的性質を持つことは、特定の状態にあることや、特定の変化を被ることではない。それは、特定の条件が実現した暁には、特定の状態にあったり、特定の変化を被ったりしがちだということである」(Ryle 2009, p. 31)。これに関連して、愛煙家 (smoker) や (職業としての) 運転手であることをライルが「傾向性」に分類していることが注目される。たしかに、常に喫煙していたり車を運転していなくても、人は愛煙家であったり運転手であったりすることができるという意味で、これらは傾向性に似ている。しかし、一度も喫煙したり車を運転したことのない人が愛煙家であったり運転手であることはできないだろう。ところが、ここで定義した意味での「傾向性」をある物体が有するためには、その傾向性は顕在化する必要はないのである。例えば、グラスは一度も割れたことがなくても、それどころか現実世界では割れることがないとしても割れやすいグラスでありうる。cf. Mumford 1998, p. 21.

<sup>69</sup> ただし、これと同じニューロン図はEがAとCによって過剰決定されるケースにも用いることができる。この違いを明確にするには、変数の値が別の変数の値にどのように依存するのかを示す構造方程式 (structural equation) を利用する必要がある。A, C, E それぞれの変数に 0 か 1 のどちらかの値をとらせて、1 はノードの発火を示すことにしよう。すると、目下の文脈では  $E := A \cdot C$  という式が成り立つ。A, C の一方のニューロンだけが発火しても E は発火しない cf. Woodward 2003, pp. 42–45.

では、ガラスがラップで包まれているために「強い衝撃を加える」というノードが発火しないので、ガラスは割れない。フィンクの事例では、ガラスは魔法使いによって割れやすすくないガラスへと変化させられたために、強い衝撃を加えられても割れなくなっている。フィンクの事例は割れやすさに関する上図の分析に対して問題を引きおこすことはない。また、電荷のように、多様な刺激に対して多様な反応を返す傾向性を扱うには、刺激と反応を表すニューロンが二つより多くの値をとっていると仮定すればよい。

傾向性の概念をこのように因果役割の概念によって説明するならば、心的状態は行動の傾向性であるという主張は、因果役割機能主義に限りなく近づく<sup>70</sup>。もっとも、両者は完全に同値というわけではない。問題は「行動の傾向性」というところにある。というのはも、心的状態の中には刺激条件や反応が物理的刺激や物理的行動ではないようなタイプの傾向性があるからである。無制限の全称量化を含む一般的信念 (general belief) を取り上げよう。一般的信念の一例は、2 より大きな偶数は素数の和であるという信念である。この信念は

$$\forall x [x \text{ は } 2 \text{ より大きな偶数である} \rightarrow x \text{ は二つの素数の和である}]$$

という内容（ゴルトバッハ予想）をもち、無制限の全称量化を含んでいる。さて、アームストロングによれば、この一般的信念は何か 2 より大きな偶数であるという信念の獲得を刺激として、それは素数の和であるという信念を生じさせるという傾向性である<sup>71</sup>。しかし、この傾向性の刺激条件と反応は信念の獲得という心的出来事であって、物理的刺激や物理的行動ではない。したがって、一般的信念のこのような分析が正しいとすれば、この種の傾向性の存在は「すべての心的状態は物理的刺激を刺激条件とする何らかの行動の傾向性である」という主張に対する反例となる。とはいえ、これらの論点を別にすれば「すべての心的状態は何らかの行動の傾向性である」という（しばしば行動主義に帰属される）主張は、因果役割機能主義の主張にかなり近いといえることができるだろう。

## 1.5 節 本章のまとめ

「行動主義」と「機能主義」は、厳密に言えば独立であるさまざまな立場に対して用いられてきた。本章では、方法論的問題・概念的問題・心身問題という三つの問題に対してどのような回答を与えるのかという観点から、行動主義と機能主義の双方に関して、三種類のアイデアを切り出し、定式化した。その結果は以下の表にまとめられる。

<sup>70</sup> cf. Armstrong 1968, p. 85; Kim 2006, pp. 121–124.

<sup>71</sup> Armstrong 1973. cf. Stalnaker 1984, chap. 6.

	行動主義	機能主義
方法論的問題	方法論的行動主義	ホムンクルス機能主義
概念的問題	分析的行動主義	意味のネットワーク説＋ラムジー化
心身問題	存在論的行動主義	因果役割機能主義

こうした分類表を作成することで、行動主義と機能主義を有意味な仕方で比較することが可能になる。本章では、三つの問題すべてに関して機能主義の優位を主張した。

今後、機能主義が何らかの形で挑戦を受けたときには、上の分類表を参照することで、その挑戦が機能主義のどの部分に対して差し向けられているのかをチェックできる。この意味でも、上の分類表は有益である。とはいえ、特に区別する必要がある限りは、単純化のために、「機能主義」はこれら三つの問題に対する機能主義的な回答の連言を意味し、「機能主義者」はその連言を支持する人のことを意味する、と約定してもよいだろう。以下では、この方針で議論を進めていく。

## 2 章 同一説の再検討

“functionalism is not so profoundly different from the identity theory as was first made out.”

Churchland 1988, p. 42

機能主義を行動主義と比較した前章につづいて、本章では、同一説 (identity theory) の批判を通して機能主義の形而上学的な基礎を明確にすることを目指す。

多くの機能主義者は、心的性質が多重実現可能 (multiply realizable) であるという前提に基づいて、同一説を否定する傾向にある<sup>1</sup>。痛み状態にある人間と他の生物は、異なる神経学的状態にあるであろうから、痛みを何らかの神経学的状態と同一視することはできない、といった直観がこの議論を支えている。しかし、心的性質の多重実現可能性に基づいて同一説を拒否する議論は、二つの方面から批判を浴びてきた。一方に、同一説は多重実現可能性とわずかな調整によって両立させることができると論じる立場がある。ルイスやジャクソン、パーゲッター、プライアーといった機能主義者がこの立場を支持している。他方で、そもそも心的性質の多重実現可能性に懐疑的な立場もある。近年、こちらの立場は支持者を増やしている<sup>2</sup>。同一説をめぐる論争はまだ決着がつかそうにない。

本章では、同一説と心的性質の多重実現可能性が両立可能であるという立場を退ける。多重実現可能性に関して懐疑的であるような同一説は、本章では検討しない。この意味で、本章の同一説批判は限定的なものであることをあらかじめ断っておく。

本章の構成は以下のようである。2.1節では、同一説を定式化する。そこで定式化される立場は、多くの文献で「タイプ同一説」と呼ばれている立場に相当する。一般に、同一説には二種類あり、タイプ同一説よりも弱いトークン同一説を考えると可以说われている。2.2節では、心の哲学におけるこうした「タイプ」と「トークン」の区別が何を意味するのかを明確にする。その上で、「同一説」はタイプ同一説を意味すると考えるのが自然であり、トークン同一説は「同一説」の名に値しないと論じる。2.3節では、機能主義はタイプ同一説の否定を帰結するというオーソドックスな議論を提示する。2.4節では、性質の形而上学の基本的な論点を整理しながら、同一説と機能主義を両立させようとするルイスやジャクソンらの試みを説明する。2.5節では、同一説を支持する三つの議論を批判的に検討する。

### 2.1 節 同一説

同一説を批判するためにも、まず同一説とは何かを明確にしよう。同一説は心脳同一説とも呼ばれるが、この心脳同一説は「心とは脳のことである (The mind is the brain)」という

<sup>1</sup> Kim 2006, p. 113. なお、イギリスの哲学者は“variably realizable”と表現する傾向がある。cf. Blackburn 1991; Mumford 1998.

<sup>2</sup> cf. Gozzano & Hill 2012.



風に定式化されることがある。しかし、このように定式化すると、心脳同一説は「宵の明星は明けの明星である」といった同一性言明と類比的にみえる。これはミスリーディングな定式である。まず、死体は脳を持っていても心をもたないと考えるのが自然である。また、心という実体について現代の心の哲学者たちが語ることは稀であり、彼らはむしろ、心的性質が脳の性質とどのような関係にあるかに関心をよせている。そこで、心脳同一説は次のように定式化するのがよいかもしれない。

すべての心的性質は何らかの脳の性質と同一である。

このように定式化された心脳同一説は「宵の明星は明けの明星である」といった同一性言明よりも、「熱は分子運動である」といったタイプの同一性言明と類比的に解釈されるべきものである。熱は個体（実体）ではないので、この言明は個体同士の間の同一性ではなく、性質同士の同一性を主張していると思われる。

細かいことをいえば、心的性質は人がもつことのできる性質なので、それが何らかの脳の性質と同一であるという考えは、そのままでは不合理にみえる。だが、このパズルは、例えば、脈拍が100であるという人の性質はその人の心臓が100回鼓動するという性質と同一である、といったケースとの類比によって答えられると思う。

また、すべての心的性質が脳（特に大脳）の性質と関わりをもつという意見には異論があるかもしれない。例えば、頭に小さな穴をあけて、そこから高圧の水を流し込んで大脳を破壊された犬が、8か月近く生き延び、歩き回ったり刺激に反応して目覚めることができたといわれている<sup>3</sup>。こうした除脳動物（decerebrate animal）に関する研究から、動物は大脳中枢なしでも刺激に対して反応したり、オペラント条件付けによって学習するということが明らかになった。この種の事例は、脳の性質とは同一視されない心的性質があるという可能性を示唆するが、その一方で、この種の事例は心脳同一説にとってそれほど深刻な反例とも思われない。例えば、心脳同一説の定式化に使われる「脳の性質」は、脳以外の神経系の性質も含むと考えておけば、この種の事例には対処できそうである。この点を明確にするため、本稿では

**心脳同一説：**すべての心的性質は何らかの神経学的な性質と同一である。

という定式化を正式なものとして採用する。

心脳同一説が何であるかは、以上の考察によってだいぶ明確になった。以下では、単純さのために「同一説」という略称を用いる<sup>4</sup>。

<sup>3</sup> 下條 1994, pp. 245–248.

<sup>4</sup> 同一説には様々な批判がある。例えば、同一説はライプニッツの法則と緊張関係にあると言われるが、この批判にもさらに多くの変種がある。cf. Rey 1997, pp. 48–62; Braddon-Mitchell & Jackson 2007, pp. 99–100. 本章では同一説と機能主義の関係や、タイプ同一説とトークン同一説の対立に議論を集中するため、同一

## 2.2 節 タイプ同一説とトークン同一説

多くの文献はタイプ同一説とトークン同一説という二種類の同一説を区別している。前節で定式化した同一説はタイプ同一説に対応する。それを確認するため、タイプとトークンの区別が心の哲学において何を意味するのかを明確にしよう。

形而上学では、タイプとトークンの区別は言語表現や芸術作品などの人工物の存在論を論じる際によく利用される。例えば、「スタインの “Rose is a rose is a rose” という表現にはいくつの単語が出現しているのか？」という質問は明らかに多義的である。ある意味では3個だが、ある意味では7個である。タイプとトークンの区別は、こうした多義性を解消するために利用される。つまり、「単語」をタイプの意味で解釈すれば答えは3個だが、トークンの意味で解釈すれば答えは7個である、というわけである。

実のところ、この例においてタイプとトークンとして区別されているものが本質的には何なのか、という問題はそれほど簡単には答えられない。しかし、一見すると、タイプとトークンの区別は普遍とそれを例化する個別の対象という区別に重なるように見えるのは確かである<sup>5</sup>。そのため、心の哲学では「タイプ」と「トークン」は単純に、普遍とそれを例化する個別の対象という意味で用いられる<sup>6</sup>。心の哲学者たちはタイプとトークンの区別をこのように理解した上で、この区別を心的性質とは何かという問題に持ち込んだ。タイプは普遍のことであり、普遍が性質なのだとすれば、前節で定式化した同一説はタイプ同一説を意味する。これに対し、タイプ同一説と区別される同一説のバージョンは次のように定式化される。

**トークン同一説：**すべての心的性質のトークンは何らかの神経学的な性質のトークンと同一である。

心的性質や物的性質のトークンとは、それらの性質を例化する対象である。例えば、私が歯に痛みを感じるとき、私は歯が痛いという性質を例化している対象であるが、私は同時に、自分の神経系がしかじかの状態にあるという性質を例化している対象でもある。ただし、厳密にいうと、痛みはある時点で生じたり消えたりするので性質の帰属は時点に相對

---

説に対する反論を取り上げることもしない。

<sup>5</sup> この同一視は安易すぎるという異論もある。例えば、目の前にある金属片は酸化銅(II) CuO というタイプと相対的にはトークンだが、CuO は酸化銅というタイプと相対的にはトークンであり、酸化銅は酸化物というタイプと相対的にはトークンである…という風に、タイプとトークンの区別は相対的な区別であるとステュワートは論じている。Steward 1997, pp. 123–124. 酸化銅と CuO の間には確定可能 (determinable) と確定的 (determinate) という区別は成り立つにせよ、酸化銅と CuO はどちらも個別の対象ではなく普遍に分類されるであろうから、タイプとトークンの区別は普遍とそれを例化する対象の区別とは重ならないだろう。また、ウェッツェルは普遍には成り立たないかもしれないタイプの特徴を他にもいくつも挙げている。Wetzel 2009, p. xii.

<sup>6</sup> Braddon-Mitchell & Jackson, 2007, p. 100.

化する必要がある。例えば「歯が痛い」が正確には「 $x$ は $t$ において歯が痛い」というように対象と時点を引数にとる二項述語として扱う必要がある。この場合には、対象と時点の順序対を心的性質のトークンとみなせばよいだろう<sup>7</sup>。

タイプ同一説とトークン同一説はどういう関係にあるのか。二つの状態が同一であるなら、それらは例化する個体に関しても一致するので、タイプ同一説はトークン同一説を含意する。他方、トークン同一説がタイプ同一説を含意しない。このことは、「色を持つ対象は何らかの形をもつ」が成り立つからといって「すべての色は何らかの形と同一である」とはいえない、というアナロジーを用いると理解しやすい<sup>8</sup>。これと同様に、心的性質を例化するものは何らかの物的性質を例化する、ということから、個々の心的性質が何らかの物的性質と同一であると推論することはできない。したがって、タイプ同一説はトークン同一説よりも真に強い主張である。

実のところ、トークン同一説はトリビアルではないにせよかなり弱い主張である。トークン同一説の主張は、色をもつすべての対象は形をもつという主張と似ている。しかし、色をもつすべての対象は何らかの形をもつということが成り立つからといって、特定の色と特定の形の間に法則的な関連性があると指摘したことにはならない。同様に、トークン同一説は、特定の心的性質と特定の物的性質との間に法則的な関連性があると述べているわけではない。トークン同一説が述べているのは、すべての心的性質を例化する個体は、同時に何らかの物的性質を例化しているということに過ぎない。トークン同一説が偽であるとすれば、それは、デカルト主義的な靈魂などが心的性質の担い手であるときに限られる。そう考えると、トークン同一説は実体一元論の別名に過ぎない<sup>9</sup>。本章の冒頭で、トークン同一説は同一説の名に値しないと述べたのはこのためである。

さて、ここまでは同一説を個体のもつ性質に関する学説として議論してきた。しかし、同一説は出来事に関する学説として定式化されることもある。出来事に関する学説として解釈される場合、タイプ同一説は、心的出来事タイプは何らかの物的出来事タイプと同一であると主張するのに対して、トークン同一説は、すべての心的出来事トークンは何らかの物的出来事トークンと同一であると主張する。

出来事に関するタイプ同一説とトークン同一説の区別は、一見するとこれまでの議論とパラレルに理解することができる。デイヴィッドソンによる行為文の分析を参考にして、個々の心的述語 $M$ を主体と出来事とを関係づける二項述語とし、 $M$ は心的出来事タイプを表現すると考えよう。そうすると、例えば「私は $t$ において歯が痛い」という文は

$\exists e$  [歯が痛い (私、 $e$ )、かつ、時点 ( $e, t$ )]

<sup>7</sup> ここでは三次元主義の枠組を前提している。四次元主義の枠組では「左腕が痛い」は私の時間的部分において端的に例化されるので、述語を時点に相対化する必要はない。

<sup>8</sup> Kim 2006, pp. 102–102.

<sup>9</sup> cf. Crane 2001, p. 57.

と分析される。出来事に関するタイプ同一説が正しければ、歯痛という心的出来事タイプと同一であるような何らかの物的出来事タイプPが存在することになる。これに対し、出来事に関するトークン同一説は、歯痛という心的出来事タイプと同一の物的出来事タイプの存在を要求する代わりに、時点tにおける私の歯痛のようなすべての心的出来事トークンに対して、それと同一である物的出来事トークンが存在することを要求する。すなわち

$$ie \text{ [歯が痛い (私、} e \text{) 、かつ、時点 (} e, t \text{) ]} = ie \text{ [P (私、} e \text{) 、かつ、時点 (} e, t \text{) ]}$$

といった同一性が成立するような物的出来事Pが存在することを主張する。この同一性が成り立つには、心的語彙を用いて記述された特定の心的出来事が物的語彙のみを用いて再記述できなければならない。第一部の冒頭で紹介したデイヴィドソンの出来事存在論によれば、そうした再記述は可能である。

しかし、出来事に関するタイプ同一説とトークン同一説を以上のように区別する道は、どんな出来事存在論に対しても開かれているわけではない。例えば、出来事のある時点における性質例化 (exemplifications of properties at times: EPT) とみなスキムの理論について考えてみよう<sup>10</sup>。EPTは出来事の主体と例化する性質とその時点との順序対として記述される構造化された対象である。EPTは以下のような同一性条件をもつ。

$x$  が性質  $P$  を時刻  $t$  に例化するという出来事が、 $y$  が性質  $Q$  を時刻  $t'$  に例化するという出来事 と同一であるのは、 $x = y$  かつ  $P = Q$  かつ  $t = t'$  のときそのときに限る。

EPT の出来事観によれば、心的出来事が物的出来事と同一であるためには、心的出来事を構成する性質が何らかの物的出来事を構成する性質と同一でなければならない。例えば「私が時刻  $t$  に痛みを感じる」と「私の C 繊維が時刻  $t$  に発火すること」と同一の出来事を表現するには、痛みが C 繊維発火すると同一の性質でなければならない。このように EPT の出来事観では、同一説はタイプ同一説に帰着してしまう。

トークン同一説は理解可能な学説なのだから、タイプ同一説とトークン同一説の区別を表現できないような出来事存在論には問題がある、としてステュワートはEPTの出来事観を批判している<sup>11</sup>。この批判は完全にフェアというわけでない<sup>12</sup>。先に示したように、「出来事」という語を使わなくても、タイプ同一説とトークン同一説の区別が捉えようとしている論点は押さえることができる。実際、キムも「本質的には、それ [=トークン同一説] が言っているのは、心的性質と物的性質は同一の存在者によって例化されるということだ

<sup>10</sup> Kim 1976, Kim 2006, p. 103.

<sup>11</sup> Steward 1997, sec. 1.2.

<sup>12</sup> cf. Hall 2001.

けである」と述べている<sup>13</sup>。したがって、出来事存在論に中立を保ったままでも、タイプ同一説とトークン同一説の区別を立てることはできるはずである。

それでは、同一説を出来事に関する学説として定式化することには何の眼目もないのだろうか。いや、そうではない。たしかに、あらゆる文脈において、同一説を出来事に関する学説として定式化する必要はないと思う。このことは、ちょうど論理学において、動詞を含むような日常言語の文を一階言語の文に翻訳するときに、述語に対して出来事の項を増やすことが必ずしも求められないのと似ている。しかし、心的因果を考慮すべき文脈では、心的出来事を話題にすることに眼目がある。なぜなら、第一部の冒頭で述べたように、出来事は因果関係の関係項となる存在者だからである。デイヴィドソンは、心的出来事が何らかの物的出来事とトークン同一でさえあれば、心身の因果関係が確保できると考えている。これに対して、出来事を EPT とみなすキムは、デイヴィドソンのような仕方で出来事のタイプとトークンを区別することができない。キムにとっては、同一説はタイプ同一説でしかない。これにより、キムは、タイプ同一説が成り立たない場合に心身の因果関係がいかんにして成立しうるのかに関して難問を抱えることになる。2.5.2 節では、タイプ同一説が成り立たない場合に、いかんにして心的因果が成り立つのかという問題をより詳しく検討する。

## 2.3 節 機能主義と同一説

ここまでは同一説の定式化について議論をしてきた。本章の残りの部分では、機能主義と同一説の関係を考察する。本節ではまず私自身の考えを率直に述べる。それに対してなされる反論の検討は、次節以降で行うこととする。

因果役割機能主義によれば、心的性質とは何らかの因果役割を担い手をもつという性質のことであるから、心的性質を例化する対象は何らかの物的性質を例化する。したがって、因果役割機能主義はトークン同一説を含意する。

これに対し、トークン同一説は因果役割機能主義を含意しない。トークン同一説は心的性質Mと物的性質Pの間に何らの法則的な関係がなくても成り立つが、因果役割機能主義は心的性質と物的性質の間の法則的な (nomic) 関係があることを要求するからである。因果役割機能主義が心的性質と物的性質の間の法則的な関係を要求する仕組みは次のようである。まず、物理的に可能な世界では現実世界で成り立つすべての自然法則が成り立っている。そして、現実世界と同じ自然法則が成り立つ可能世界では、物的性質は同じ因果役割を担うはずである。したがって、心的性質Mに結び付けられている因果役割を物的性質Pが担うとすれば、物理的に可能なすべての世界では、Pを例化する対象はMも例化しているは

---

<sup>13</sup> Kim 2006, p. 104. 興味深いことに、キム『心の哲学』の第3版では、第2版には含まれていたタイプ同一説とトークン同一説についての節が削除されているようである。この変更点を教えていただいた Wolfgang Ertl 教授（慶應義塾大学）に感謝する。

ずである<sup>14</sup>。記号化すると、

$$\Box \forall x \forall t [P(x, t) \rightarrow M(x, t)]$$

が成り立つ。必然性のオペレータは物理的な必然性を表す。これこそが、心的性質と物的性質との間に法則的な関係があるということの表現である<sup>15</sup>。

次に、因果役割機能主義はタイプ同一説とは両立しないことを示す。そのための議論は二段階に分かれる。まず、それぞれの心的性質に特有の因果役割は様々な担い手をもちうるように思われる。例えば、人間にとっての痛みはC繊維の発火であっても、イルカにとっての痛みはD繊維の発火かもしれず、火星人間にとってはさらに別のO繊維の発火が痛みの因果役割を担うのかもしれない。しかし、C繊維の発火とD繊維の発火は、相異なる二つの状態である。すると、同一性の推移性により、痛みをC繊維の発火といった特定の神経生理学的な性質とは同一視できないことになる。以上の考察は、他の心的性質にも当てはまるだろう。心的性質に関わる因果役割は様々な担い手をもつので、因果役割を担い手をもつという性質は、因果役割の担い手となる神経学的な性質と同一ではない。

ここまでの議論を「多重実現論法」と呼んでおく<sup>16</sup>。次に、多重実現論法の結論を含めて以下の三つの仮定からタイプ同一説の否定を導く<sup>17</sup>。

A1：心的性質はその性質に特有の因果役割の担い手をもつという性質と同一である。

A2：心的性質に特有の因果役割の担い手をもつという性質はその担い手となる神経学的な性質と同一ではない。

A3：心的性質に特有の因果役割の担い手をもつという性質は、その担い手となる神経学的な性質と同一でなければ、そもそも神経学的な性質ではない。

A1は因果役割機能主義である。A2は多重実現論法の結論である。A1とA2からは、心的性質がその性質に特有の因果役割の担い手となる神経学的な性質と同一ではないことが帰結する。しかし、このことから、問題の因果役割の担い手をもつという性質が何かまったく別の神経学的な性質であるという可能性は排除されない。こうした可能性を排除するにはA3の仮定が必要になる。たしかに、A3はいささかアドホックだが、この仮定はほとんどの心の哲学者によって暗黙に認められていると考えられるので、議論を単純化するために、ここでは問題のない仮定とみなす。ともあれ、以上の三つの仮定からは、心的性質がいか

<sup>14</sup> cf. Kim 1998, pp. 23–24

<sup>15</sup> cf. Blackburn 1984, p.184. この表現はいわゆる「強い付随性」にほぼ対応する。

<sup>16</sup> 2.4.1 節では、多重実現論法をより一般的な形で再び取り上げる。

<sup>17</sup> 心的性質の多重実現可能性から同一説の否定を導く論証の定式化は一通りではない。例えば、Gozzano & Hill 2012, p. 8 などを参照。

なる神経学的な性質とも同一ではないことが帰結する。したがって、因果役割機能主義はタイプ同一説と両立しないことが明らかになった。

本稿はこの論証を支持する。しかし、この論証に抗して、同一説はある種の機能主義と両立させることが可能なのだと論じる論者もいる。彼らの考えでは、A2とA3の仮定は受け入れられるが、A1を受け入れるべきではない<sup>18</sup>。ここで興味深いのは、彼らがA1を否定するやり方である。次節では、そのやり方を考察する。

## 2.4 節 実現者機能主義：同一説の復権

本節では、同一説と機能主義を両立させるために、ルイスとジャクソン、パーゲッター、プライアーが機能主義をどのように定式化したのかを考察する。ただし、ルイスやジャクソンらの機能主義は、性質の形而上学における議論を背景にしているので、まずは、目下の議論に関係する範囲内でいくつかの論点を整理しておく。

### 2.4.1 節 性質の形而上学

1.4.3 節で確認したように、ある状態の因果役割とはその状態を生じさせる典型的な原因、もしくは、特定の条件の下でその状態から生じる典型的な結果のことであった。例えば、一定量の青酸カリを服用すると、生命体は死に至る<sup>19</sup>。したがって、青酸カリの因果役割とは、それを一定量服用した生命体に死をもたらすことである、とすることができる。そして、青酸カリはこの因果役割の担い手 (occupant) である。役割とその担い手というこの対比は、ちょうど「バラク・オバマは (2015 年において) 合衆国大統領という役割の担い手である」といった言い回しと類比的である。われわれは、オバマが何の役割を担っているのかを問うのと同様に、さまざまな薬品に関して、どのような因果役割を担うのかを問うことができる。

これとは逆に、任意の因果役割に関して、その役割の担い手をもつという性質を考えることもできる。例えば、ある対象が毒性をもつとは、それを一定量服用した生命体に死をもたらすという因果役割の担い手をもつことだと言えよう。このように、毒性をはじめとする傾向性一般は、それが引き起こす典型的な結果に言及することで定義することができる。それどころか、因果役割に言及することで定義されるのは毒性のように日常的な概念だけではない。例えば、ワトソンとクリックによって DNA の分子構造が発見されるまで、メンデル遺伝学では対立遺伝子 (allele) を、その物理的構造によってではなく、表現型に及ぼす影響によって同定していた。自然科学においても因果役割によって理論的概念を定義することは珍しくない。

---

<sup>18</sup> 例えば、Lewis 1966 における同一説の論証は A1 と矛盾する前提を措いている。

<sup>19</sup> ここでいう「一定量」はある程度の幅を許容する。

ある因果役割の担い手をもつという性質は、その因果役割の担い手とは別個だと思われる<sup>20</sup>。例えば、毒性と青酸カリを比較しよう。それを一定量服用した生命体に死をもたらす化学物質は青酸カリとは限らない。他の化学物質もそれを一定量服用した生命体に死をもたらすだろう。その限りで、毒性の外延は青酸カリの外延よりも広い。外延が異なる性質は別個なので、毒性は青酸カリとは別個の性質であるに違いない。多くの場合、ある因果役割を担い手をもつという性質とその因果役割の担い手はこのように外延が異なる<sup>21</sup>。たしかに、因果役割の中には現実には多重実現していないものもあるかもしれない。しかし、その因果役割が多重実現しているのであれば、つまり、他の可能世界では多重実現しているのであれば、その役割の担い手をもつという性質と担い手の間には同一性が成り立つことはないと思われる。こうした議論は「多重実現論法」と呼ぶことができる。

さて、ある因果役割の担い手をもつという性質をその因果役割の担い手から区別したところで、議論の構図の整理も兼ねていくつかの用語をまとめて導入する（以下の表も参照のこと）。まず、ある因果役割の担い手は、その因果役割を担い手をもつという性質とは別個ではあり、前者は後者を「実現する (realize)」という関係に立つ。例えば、青酸カリは毒性を実現する性質である。そして、因果役割の担い手をもつという性質を「役割性質」または「機能的性質」といい、因果役割の担い手となる性質を役割性質の「実現者 (realizer)」という。

因果役割、役割性質、実現者の三者関係は、論理学の表記法と概念を利用することで、もう少し明確にできる。性質の形而上学では、因果役割の担い手をもつという性質は「二階の性質」と呼ばれる。二階の性質とは何か。

**二階の性質**：何らかの性質の集合Bを一階の性質の集合として与えられたとき、ある性質Pが二階の性質であるのは、Pがこの性質集合Bの要素であって、一定の条件Dを満たすような性質をもつという性質 (property of having a property) であるときそのときに限る<sup>22</sup>。

<sup>20</sup> この主張は Prior, Pargetter & Jackson 1982 の「区別性テーゼ (distinctness thesis)」に相当する。ただし、区別性テーゼには批判もある。Mumford 1998, chap. 5 を参照。ホーソンとマンリーは、マンフォードのブライヤー批判は二階の性質という概念に関する無理解に基づいていると指摘している。Hawthorne & Manley 2005, pp.186–189 を参照。この点に関して、私自身はホーソンとマンリーに同意する。

<sup>21</sup> 本章の冒頭で、現代の同一説論者の中には心的性質の多重実現に懐疑的な人々がいると述べた。傾向性に関してこれに相当する立場も理論的にはありうる。マンフォードは、傾向性の多重実現はそれほど広範には成立しないのではないかと疑っている。例えば、彼は、衣類の吸水性は繊維と繊維の間に水が入りこみやすいということと、繊維そのものの間に水が入りやすいという二つの仕方で実現されるというマッキーの例に反論して、二つ目のケースは単により小さな繊維と繊維の間に水が入り込むことなのではないかと述べている。Mumford 1998, p. 102. 傾向性の多重実現がどのくらい広範に成り立つのかは経験的な問題であり慎重な扱いを必要とするという点までは、私もマンフォードに同意する。しかし、彼のマッキーに対する反論は失敗していると思う。繊維そのものの間に水が入るというのは、セルロースの構成単位であるグルコースがもつヒドロキシル基 (-OH) が水素結合によって水を引きつけるということであって、繊維と繊維の間に水が入ることではないと思われるからである。

<sup>22</sup> Kim 1998, p. 20. ただし、直観的に分かりやすくなるように、ここでは表現を変えている。



このままでは分かりづらいので、具体例を通してこの定義の意味を確認しよう。一階の性質の集合 **B** として化学物質の種類の集合を考える。青酸カリは **B** の要素だが、毒性は **B** の要素ではない。しかし、毒性はそれを服用した生命体に死をもたらすという因果役割によって、以下のように定義できる。

$x$  は毒性をもつ  $\Leftrightarrow \exists F \in B [\forall y (Fy \rightarrow y \text{ は } y \text{ を一定量服用した生命体に死をもたらす}) \& Fx]$

この定義は次のように理解すればよい。 $x$  が青酸カリであるならば  $x$  は毒性をもつと思われる。なぜだろうか。それは、変数  $F$  の値として青酸カリをとるならば、 $F$  はそれを一定量服用した生命体に死をもたらすようなものであり、 $x$  は  $F$  である、ということが成り立つからである、と。毒性を以上のように定義するならば、青酸カリが一階の性質であるときに毒性が二階の性質であるのはもはや明らかである。毒性とは、化学物質の種類の集合の要素であって一定の条件（それを服用した生命体に死をもたらす）を満たすような性質をもつという性質だからである。二階の性質の定義はこのように用いられる。

なお、青酸カリと毒性は性質の性質ではなく、あくまで対象の性質であるから、論理学の標準的な用法では本来どちらも一階の性質に分類されるべきではある。標準的な意味で二階の性質とすべきなのは、むしろ、それを服用した生命体が死に至るという因果役割であり、実際、これは諸性質の性質（property of properties）である。それに関わらず毒性が「二階の性質」と呼ばれるのは、一階の性質上の存在量化を含むために、ラッセルとホワイトヘッドの用法ではオーダーが 2 であるという事情による<sup>23</sup>。

いささか用語法が煩雑になってしまった。本節を終える前に、リマインダーとして、ここまで導入した用語の関係を表にまとめておこう。

	具体例	オーダー
因果役割	一定量服用した生命体に死をもたらす	（標準的な用法で）二階の性質
役割性質（＝役割の担い手をもつという性質）	毒性	（ラッセルとホワイトヘッドの用法で）二階の性質
実現者（＝役割の担い手）	青酸カリ、ヒ素 etc.	一階の性質

## 2.4.2 節 実現者機能主義

傾向性をはじめとする二階の性質に関する性質の形而上学の議論は、心の哲学における心

<sup>23</sup> Putnam 1970, p. 313. ラッセルとホワイトヘッドの用法では、オーダー 2 の述語とはオーダー 1 の述語への量化を含む述語である。ラッセルによる、タイプ 1 でオーダー 2 の述語の具体例は「 $x$  はすべての偉大な将軍がもちあわせている性質をもちあわせている」である。記号化すると、 $\exists P [Px \& \forall y (y \text{ は偉大な将軍である} \rightarrow Py)]$  となる。

的性質についての機能主義と類比的な関係にある。なぜなら、1.4.3 節で定式化した因果役割機能主義によれば、心的性質には何らかの因果役割が結び付けられているとされるからである。痛みを例にとろう。議論のために、痛みの因果役割は、組織の損傷によって引き起こされ、うめき声や顔をしかめるといった行動を引き起こすことであるとしよう。このとき、因果役割機能主義によれば、痛みとはこうした因果役割の担い手をもつという性質に他ならない。他の心的性質についても同様である。要するに、因果役割機能主義は、心的性質を傾向性と同様に二階の性質とみなす。

これに対して、ルイスやジャクソンらは、心的性質が傾向性のような二階の性質に対応するとは考えない。たしかに、彼らは、心的性質が何らかの因果役割と本質的に結びついていることを認めるという点で、因果役割機能主義に近いところにいる。しかし、彼らは心的性質がむしろ傾向性のような二階の性質の実現者に相当すると考える<sup>24</sup>。

因果役割機能主義とルイスやジャクソンらの機能主義との対立は、心的性質を役割性質と同一視するか、それとも実現者と同一視するかにかかっている。1.4.3 節では、因果役割機能主義を次のように定式化しておいた。

**因果役割機能主義：**すべての心的性質はそれぞれの心的性質に特有の因果役割の担い手をもつという性質である。

この定式化は心的性質を役割性質と同一視している。これに対し、心的性質を因果役割の実現者と同一視するならば、因果役割機能主義は次のように修正されることになる<sup>25</sup>。

すべての心的性質はそれぞれの心的性質に特有の因果役割の担い手と同一である。

因果役割機能主義から区別されるこの立場は「実現者機能主義 (realizer functionalism)」と呼ばれる。

一見すると、実現者機能主義はかなり奇妙な立場である。痛みには特有の因果役割の担い手となる性質が複数ありうるとすれば、痛みをそれら諸性質のうちのどれと同一視すればよいのだろうか。むしろ、どの性質とも同一視できないのではないか。なぜなら、例えば、痛みと C 繊維発火を同一視するならば、痛みには特有の因果役割を担う別の性質、例えば D 繊維発火をもはや痛みとはみなせなくなってしまう。

こうした疑問に答えるには、実現者機能主義者が心的性質を何らかの因果役割の担い手と同一視するとき、彼らが問題にしている心的性質は集団 (population) に相対化されている、ということを念頭におく必要がある。まず、個々の心的性質に特有の因果役割の担い手は、人間やイルカや火星人といった集団と相対的には一意に定まると仮定しよう。この

<sup>24</sup> Jackson, Pargetter & Prior 1982; Lewis 1980; 1994; Braddon-Michell & Jackson 2007.

<sup>25</sup> cf. Braddon-Michell & Jackson 2007, p. 48.

仮定の下では、人間という集団と相対的には、痛み特有の因果役割の担い手は C 繊維発火であり、イルカという集団と相対的には、痛み特有の因果役割の担い手は D 繊維発火ということになる。この帰結を受けて、実現者機能主義は、人間の痛み=C 繊維発火、イルカの痛み=D 繊維発火といった具合に、心的性質をそれと結びつけた因果役割の担い手とを同一視する。

集団へと相対化してまで実現者機能主義が同一視にこだわる動機は、自然科学における還元とはそのようなものだ、という前提に由来するのだろう。例えば、メンデル遺伝学では親から子へと遺伝して表現型に影響する何らかの因子を「遺伝子」と呼んでいた。そして、地球上の生物に関していえば遺伝子の正体は DNA に他ならないことが後に判明した。つまり、この発見は、遺伝子の役割を担うメカニズムが少なくとも地球上の生物に関しては DNA 以外ではありえないということの発見であった。実現者機能主義者は、心的性質に関しても同様の発見がなされることを期待している。つまり、地球上の生物、などといった集団と相対的には、心的性質を実現しているメカニズムが唯一つ存在することを彼らは期待している。

心的性質を因果役割の担い手と同一視する以上、実現者機能主義者は因果役割機能主義者のように「心的性質は多重実現可能である」とは言えない。それでも、彼らは心的性質に特有の因果役割が（集団に対する相対化をやめるならば）さまざまな担い手を持ちうるということは認めるはずである。この意味で、実現者機能主義は「心的性質の多重実現可能性」という考えを実質的には受け入れている。しかし、心的性質と結びつけた因果役割の担い手は物的性質なので、実現者機能主義は心的性質を物的性質と同一視することになる。こうして実現者機能主義は同一説とも両立する<sup>26</sup>。

## 2.5 節 同一説を支持する論証の批判的検討

実現者機能主義が因果役割機能主義よりも好ましいとされる理由は何だろうか。あるいは、実現者機能主義が形式的には問題ないのだとしても、同一説の何が（あるいは、機能主義を同一説と両立させることの何が）それほど望ましいのだろうか。

本節では、同一説（実現者機能主義）を擁護する方法として三つの論証を取り上げる。

<sup>26</sup> ここで、1.3.2 節で与えたラムジー化の手続きは、因果役割機能主義と折り合いがつくようにルイスによるオリジナルの定式化に変更を加えている、という点を補足しておく。実現者機能主義により忠実であるならば、心的語彙  $M_i$  の明示的定義は以下ようになる。

$$\forall x \forall t [M_i(x, t) \leftrightarrow \exists! \langle F_1, \dots, F_n \rangle (\Phi(F_1, \dots, F_n) \& F_i(x, t))]$$

“ $\langle F_1, \dots, F_n \rangle$ ” は性質の列をあらわす。よって、“ $\exists! \langle F_1, \dots, F_n \rangle$ ” は何らかの条件を満たす性質の列がただ一つだけ存在すると述べる。こうした表記法は通常の二階言語では用いられないが、定義によって問題なく導入できる。この存在量化は“ $\exists! F_1 \dots \exists! F_n$ ”とは異なることに注意すべきである。Lewis 1972, p. 254n7. 例示のため、ここでは一階量化を用いた例を与えておく。量化のドメインを  $\{a, b, c\}$  とし、述語 R の外延を  $\{\langle a, b \rangle, \langle a, c \rangle, \langle b, c \rangle\}$  とすると、“ $\exists! x \exists! y Rxy$ ” は真だが“ $\exists! \langle x, y \rangle Rxy$ ” は真でない。よって、“ $\exists! x \exists! y Rxy$ ” と“ $\exists! \langle x, y \rangle Rxy$ ” は真理条件が異なる。ほぼ同じことが二階量化の場合にも言える。

私の考えでは、最初の論証はそれほど説得的でないが、残り二つの論証は重要であり強力である。しかし、いずれも反駁することは不可能ではないと思う。順に見ていこう。

### 2.5.1 節 狂人の痛みを想定する議論

ルイスは火星人の痛み (Martian pain) と狂人の痛み (Mad pain) という二つの思考実験を提示している<sup>27</sup>。痛み状態にある火星人は、我々に痛みを引き起こすのと同じ種類の物事によって引き起こされ、同じような反応を返す。例えば、火星人の皮膚を強くつねると、彼はそれをやめさせようとする。しかし、火星人の身体組成は人間とは大きく異なる。例えば、炭素ベースの人間に対し、火星人は珪素ベースだと想定する。要するに、火星人の痛みという思考実験は、同じ因果役割が異なる担い手を持ちうるということを示す。

これに対し、狂人の痛みの思考実験は、同じ物的性質が異なる因果役割の担い手になるかもしれないことを示唆する。まず、正常な人間の場合、C 繊維の発火が痛みの因果役割を担うと仮定する。そして、狂人は人間であるから、その身体組成は正常者とほとんど変わるところがないとする。ところが、痛みに関わる狂人の振舞い方は常識的な人間とはまるで異なる。狂人の場合、C 繊維の発火は身体の損傷によってではなく、体操によって引き起こされる。狂人は組織の損傷を避けようとはしない。また、C 繊維の発火は、うめき声ではなく指を鳴らす反応を引き起こす。このように、狂人の振舞いは他の人間の典型的な振舞いと大きく異なるため、狂人の C 繊維の発火は痛みの因果役割の担い手とは言いがたい。しかし、ルイスは、狂人もかつては彼の同類である人間と変わらぬ振舞いをしていたのだから、通常の人間の場合に痛み状態であるところの C 繊維の発火を、狂人の場合にも「痛み」と呼んではいけないことはない、と主張する。そして、ルイスはこの直観を説明するには同一説に訴えなければならず、したがって、痛みは役割性質ではなく実現者と同一視すべきであると結論する。

これはかなりトリッキーな議論である。組織の損傷を避けようとししない狂人に対して痛みを帰属できるというルイスの直観には疑問の余地があるからだ<sup>28</sup>。

また、たとえ狂人の痛みは真正の痛みの事例であるというルイスの直観を受け入れたとしても、因果役割機能主義は、狂人の痛みの事例を少なくともルイスと同程度にうまく扱うことができると思われる。そのことを示すには、ルイスが実際に狂人の痛みをどのように扱っているのかを確認するのがよい。

ルイスは、痛み状態を、痛みの因果役割を典型的に担っている物的状態と同一視する。健常者にとって痛みの因果役割を担うのは C 繊維の発火なので、痛みの因果役割を典型的に担う物的状態とは C 繊維の発火である。したがって、痛み状態は C 繊維の発火と同一である。ところで、狂人は、振舞い方こそ違えども健常者と同じく人間である。そこで、狂

<sup>27</sup> Lewis 1980.

<sup>28</sup> Soames 2015, pp. 86–87. なお、狂人の痛みに関するルイスの直観はルイスと同様にタイプ同一説を支持するジャクソンらによってすら共有されていない。Jackson, Pargetter & Prior 1982, pp. 214–215.

人における C 繊維の発火も、痛み状態と同一視することができる。

以上がルイスの戦略である。この戦略は「典型的に」という表現を使用しているが、機能主義者がこの表現を問題なく使用できるのか、疑問の余地がある<sup>29</sup>。しかし、仮に「典型的に」という表現を使用することが許されるのなら、因果役割機能主義者は次のように応じるはずである。すなわち、痛みを、痛みの因果役割を典型的に担う性質をもつという役割性質と同一視すればよい、と。そうすれば、狂人もこの状態を例化することが可能になる。というのも、狂人は、痛みの因果役割を典型的に担う C 繊維の発火という物的状態にあるので、痛みの因果役割を典型的に担う性質をもつという性質を例化しているからである。したがって、狂人の痛みの思考実験は実現者機能主義を支持しない。

## 2.5.2 節 理論的同一視にもとづく議論

実現者機能主義を支持する二つ目の議論は、科学哲学において理論的同一視の例証として論じられる具体例との類比にもとづく議論である。2.3 節や 2.4.1 節で論じたように、役割性質を実現者と同一視できないことの一つの論拠は、心的性質は多重実現されうるという見解にあった。しかし、科学哲学において理論的同一視とみなされる事例を調べると、多重実現可能性は同一説にとって必ずしも障害にならないことが示唆される。これが、理論的同一視にもとづく議論のモチベーションである。

理論的同一視は特定の領域に相対化されるならば多重実現可能性と両立する。2.4.2 節では、実現者機能主義を定式化するにあたって、痛みは人間・イルカ・火星人間…といった領域と相対的には何らかの物的性質と同一視できると述べた。こうした考え方はかなり広範囲に応用しうる。例えば、系の温度はその系を構成する分子の平均運動エネルギーとしばしば同一視される<sup>30</sup>。しかし、エンクやチャーチランド夫妻によれば、この理論的同一視は気体からなる系という特定の領域に相対化されているということに気づくのが重要である<sup>31</sup>。なぜなら、この同一性は固体の温度、プラズマの温度、真空の温度などには適用できないからである。固体の分子は気体分子のように自由に運動できない。プラズマや真空の場合には、そもそも分子が含まれていない。よって、固体やプラズマや真空の温度に関しては、それらがどのように実現されるのかを個別に探求する必要がある。

しかし、温度が領域に相対化されるとすれば、気体や固体が同じ温度（例えば、摂氏 100 度）の値をとるということが理解不能になってしまうのではないか。同じことは心的性質にも言える。痛み特有の因果役割は複数の担い手をもつが、それらは同一の因果役割を

<sup>29</sup> この点は宮園健吾博士によって指摘された。

<sup>30</sup> 物理の教科書では、気体の温度は気体分子の平均運動エネルギーに「比例」する、と表現されることが多いが、哲学者と物理学者の間に本質的な対立があるわけではないと思われる。気体の温度と気体分子の平均運動エネルギーが比例の関係にあるならば、比例定数を 1 として両者を同一視するのは原理的に可能だろう。標準的には比例定数を 1 に設定しないというだけのことである。ファインマン 1968, p. 193.

<sup>31</sup> Enç 1983; Churchland 1986; pp. 356–357; Churchland 1988, pp. 41–42. ただし、よく知られるように、チャーチランド夫妻は機能主義には批判的である。

担うからこそ、物的性質としては異なるにも関わらず、より抽象的な観点から痛みとして一まとめにする、ということに機能主義のポイントがあったのではないか<sup>32</sup>。

実現者機能主義はこの懸念をある程度払拭することができる。ルイスは痛みに関して、「痛み概念 (concept of pain)」と「痛み状態」という区別を導入しており、この区別がここで参考になる<sup>33</sup>。痛み概念とは痛み特有の因果役割の役割性質であり、因果役割機能主義者がまさに痛みとして同定するところのものである。これに対し、痛み状態とはその実現者（例えば、C繊維の発火）と同一視される。より一般的には、ある心的語彙 $M_i$ がラムジー化によって

$$M_i(x, t) \leftrightarrow \exists! \langle F_1, \dots, F_n \rangle (\Phi(F_1, \dots, F_n) \& F_i(x, t))$$

と定義されたとすると、 $M_i$ に対応する性質が $M_i$ 概念であり、変数 $F_i$ の値となるのが $M_i$ 状態である<sup>34</sup>。温度に関しても同じような区別を持ち込むならば、同一の因果役割の担い手たちを一まとめにできないという懸念には答えられそうである。

とはいえ、このように考えると、実現者機能主義と因果役割機能主義の違いはかなり微妙ではないか。結局のところ、実現者機能主義も  $M_i$  概念という因果役割機能主義が  $M_i$  という心的語彙に対応する心的性質として措定するものを必要とする。実現者機能主義は言葉上の巧妙な操作によって機能主義をタイプ同一説と両立させただけなのではないか。

そうかもしれない。しかし、実現者機能主義と因果役割機能主義の間にはもう少し実質的な対立もあると思う。ルイスは、「グルー」のように奇妙な述語も含めた任意の述語について、それに対応する性質を割り当てることができるとしつつも、他方で、実際には少数の述語だけが自然的な性質に対応する、といった性質の形而上学を採用している<sup>35</sup>。彼は、痛み概念のような性質を措定する一方で、実際には、痛み状態こそが自然的な性質であると考えている。そうであるからこそ、上でみたような言葉上の操作によって心的性質と物的性質の同一性を確保しようとする動機が生じる。しかも、こうした見解の適用範囲は、心的なものに限られない。ルイスは、役割性質はおしなべて自然的な性質ではないと考えているように思われる。

因果役割機能主義者は、役割性質が自然的な性質でないという厳格な考えには抵抗するべきだと思われる。温度について考えてみよう。温度の同値関係は熱力学の第 0 法則によって定義される。すなわち、エネルギーの流入と流出が正味でつりあっているとき、二つの系は同じ温度とみなされる。こうして温度に関して系の同値類をとることができる。気体の系であれば平均運動エネルギーがこの結果に関わっているが、固体であれば分子の振

<sup>32</sup> Sterelny 1990, pp. 200–201.

<sup>33</sup> Lewis 1980, pp. 124–126.

<sup>34</sup> ルイスは、心的語彙は集団に応じて実現者が異なりうるという意味で非固定的だと述べている。「一等賞」や「優勝者」といった表現がイベントに応じて指示対象を変えるように、「(人間の) 痛み」と「(火星人の) 痛み」は異なる性質を指示しうる。cf. Lewis 1980, p. 125.

<sup>35</sup> Lewis 1983.

動エネルギーが関わっているかもしれない。しかし、こうしたメカニズムが判明したところで、温度に関する際立った同値関係を自然でないとする理由はない。ブラックバーンによれば、固体から気体へといった相転移にも関わらず、同一の種類の性質（温度やエントロピー）を系に対して適用できることに熱力学の強みがある<sup>36</sup>。より一般的に言って、物理学的思考（physical thinking）は、本質的に、多くの実現を包摂するような状態を見出すことに存する。ブラックバーンは力学的エネルギーについても論じている。ある系の力学的エネルギーは、系の諸対象の位置と速度によって構成される高次の性質といえる。しかし、古典力学は、重力場において様々な高さや速度で運動している諸対象からなる閉鎖系に関して、運動エネルギーと位置エネルギーの和が一定であることを見出した。これにより、実現者としてみた場合には異なる状態にある様々な系を一まとめにして取り扱うことが可能になる。

ブラックバーンの指摘はかなりの的を射ているように思われる。実現者機能主義は心的性質を役割性質ではなくその実現者と同一視することに拘るが、因果役割機能主義を反駁するにはさらに論証を与える必要がある。ここで鍵となりうる概念は因果性である。実現者機能主義を擁護する三つ目の議論に目を向けよう。

### 2.5.3 節 役割性質の因果的無効力性にもとづく議論

実現者機能主義を擁護する三つ目の議論は次のようである<sup>37</sup>。役割性質は因果的効力を持たないので、心的性質が役割性質であるとすれば、心的性質には因果的効力がないことになってしまうが、これは不合理である。よって、心的性質は役割性質ではない。心的性質に因果的効力をもたせるためにも、むしろ、心的性質は何らかの因果役割の実現者と同一視されるべきである。以上の議論のコアとなるのは前半部分である：

心的性質はすべて因果的効力をもつ。

役割性質はすべて因果的効力をもたない。

よって、心的性質はすべて役割性質ではない。

この三段論法は明らかに妥当なので、因果役割機能主義はどちらかの前提を拒否する必要がある。

一番目の前提は、我々が例化する心的性質が因果的効力をもつと主張している。これはどういう意味だろうか。ある性質が因果的効力をもつ（causally efficacious）とは、その性質

<sup>36</sup> Blackburn 1991, pp. 236–241.

<sup>37</sup> Braddon-Mitchell & Jackson 2007, pp. 103–104. cf. Lewis 1983, pp. 43–45. なお、キムはジャクソンが心的性質の因果的効力を否定していると述べているが（Kim 1998, p. 72）、少しミスリーディングだと思う。たしかに、ジャクソンは1990年代の前半まで性質二元論を擁護していたが、質的でない心的性質に関して言えば、彼は一貫してそれらに因果的効力をもたせるためにタイプ同一説を擁護してきた。

が何らかの因果的説明に出現するということだと理解することができる。そうすると、心的性質が因果的効力をもたないということは、心的性質に言及することで我々の行動を因果的に説明することはできないということを意味する。これは心身因果を否定することなので、できることならば受け入れたくはない。

二番目の前提は、役割性質は因果的効力をもたないと主張している。なぜだろうか。考えられる理由はおおまかには次のようなものである<sup>38</sup>。まず、役割性質とは何らかの因果役割の担い手をもつという性質のことであつた。そして、ある性質の因果役割とは因果的説明においてその性質が占める位置・役割のことであつた。このことは、因果役割を特徴づける因果的説明において実際に出現するのは当の因果役割の担い手の方であつて、役割性質そのものは出現しないということを示唆する。例えば、毒性とは「それを服用したために死んだ」といった因果的説明において「それ」の位置に当てはまるような性質を担うということだが、この因果的説明において「それ」の位置に本来出現すべき性質は毒性ではなく、例えば、青酸カリのような役割の実現者であろう。このように、毒性という役割性質は因果的説明において役割を果たすわけではない。したがって、毒性は因果的効力をもたないことになる。以上を一般化すると、役割性質というものはおしなべて因果的効力をもたない、と考えられる。もし心的性質が役割性質であるなら、心的性質もまた因果的効力をもたないことになるだろう。これは一番目の前提と矛盾する。こうして、心的性質は役割性質ではないと結論される。

しかし、この議論の運び方には二つの点で疑問がある。まず、たとえ役割性質が前段落の意味で因果的効力をもたないとしても、その意味で心的性質が因果的効力をもたないということは、次のように考えれば、そこまで耐え難いわけではない。例えば、ある物体が毒性をもつということは、その物体がそれを一定量服用すると死に至るという因果役割を実現する何らかの性質をもつということである。毒物を服用したので死んだ、という説明は真正の因果的説明、例えば、青酸カリを一定量服用したので死んだ、といった説明を準備するようなものではある。このように、ある説明が真正の因果的説明を準備するような場合、その説明において言及される性質は因果的関連性（causal relevance）をもつ、などと言われる。因果的効力をもつのは因果役割の実現者のみだが、因果的関連性は役割性質と実現者の両方に帰属されうる。そして、因果的関連性をもつ役割性質は、その実現者に言及することで因果的に説明される出来事と法則的に切り離されているわけではない。物理的に可能なすべての世界において、役割性質を例化する対象はその実現者を例化しており、実現者は因果的効力をもつからである。もし心的性質が役割性質だとすれば、心的性質は行動と切り離されているわけではない。こうして因果役割機能主義は、心的性質が因果的に無効であるという考えを問題なく受け入れることができる。

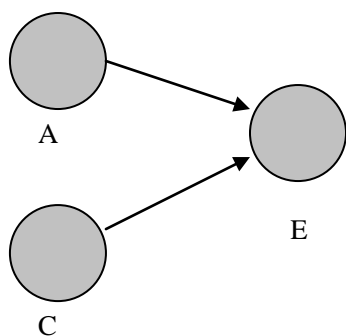
また、そもそも役割性質は因果的効力を持たないという先の議論にも疑わしいところがある。「毒物（毒性をもつ物質）を服用したために死んだ」といった言明もまた真正の因果

<sup>38</sup> cf. Prior, Pargetter, Jackson 1982, pp. 255–256.



的説明である、と論じる余地があるのではないだろうか。たしかに、この言明は「青酸カリを服用したために死んだ」という説明よりも素朴で情報量が少ないかもしれない。しかし、異なる出来事間に反事実的依存があることは、因果性の反事実的分析を支持する理論家によって、因果関係の十分条件とみなされてきた。この基準を目下のケースに適用すると、もし服用した錠剤が毒物でなかったならばそれを服用したとしても死ななかつただろう、という反事実的依存は成立することから、毒性に因果的効力を認めるのは自然だということになる。むしろ、「毒物（毒性をもった物質）を服用したために死んだ」という説明が因果的説明でないと言われてしまうとすれば、それはなぜなのか。

この疑問に対してありうる回答は過剰決定の不可能性に訴えるものである<sup>39</sup>。「過剰決定 (overdetermination)」は、同一の出来事を引き起こすのに十分な原因が複数存在するような因果関係を一般に指す。過剰決定の事例は以下のようなニューロン図で描かれる<sup>40</sup>。



このニューロン図が表しているのは、出来事 A と C が E を引き起こしたが、たとえ出来事 C が生じなかったとしても、出来事 A 単独で出来事 E を引き起こしたであろうし、逆に、出来事 A が生じなかったとしても、出来事 C 単独で出来事 E を引き起こしたであろう、という事態である。いま、図の A に青酸カリの錠剤を服用すること、C に毒性をもった錠剤を服用すること、E に死ぬことを代入することができるのだとすれば、これは過剰決定の事例ということになる。もし過剰決定が不可能だとすれば、青酸カリの錠剤を服用することか、毒性をもった錠剤を服用することのどちらかは死ぬことの原因ではない。原因としての身分が否定されるべきであるのは毒性をもった錠剤を服用することの方であり、したがって、毒性は因果的効力をもつとは言えない。

しかし、この議論には色々と疑わしいところがある。まず、なぜ過剰決定は不可能だと前提されなければならないのか。むしろ、過剰決定の事例は十分に想像できるのではないか。例えば、二人の処刑人が同時に発砲して死刑囚を殺害するといったケースがそうである。銃弾が同時に着弾して同時に死刑囚の死を引き起こすなどということは極めて稀かも

<sup>39</sup> cf. Crane 2001, pp. 49–50, 59.

<sup>40</sup> ただし、これと同じニューロン図は E が A と C が同時に発火したときそのときに限って発火する場合にも適用されうる。1.4.4 節の註でも確認したように、この多義性を解消するには、変数の値が別の変数の値にどのように依存するかを示す構造方程式を利用する必要がある。

しれないが<sup>41</sup>、そういうケースが成立することは十分に想像できる。

過剰決定は不可能だと言う人は、純粋な過剰決定に注目してそう言っているのだと思われる<sup>42</sup>。同時に死刑囚の死を引き起こすことは過剰決定のケースではあるが、純粋な過剰決定ではない。処刑人の一方が発砲した場合と、処刑人が二人とも発砲した場合とでは、死刑囚の死に方（肉片の飛び散り方など）が多少異なるだろう<sup>43</sup>。もしこれが純粋な過剰決定だとすれば、一方の出来事が生じなかったとしても、結果が寸分たがわずに生じるのでなければならない。これはたしかに奇妙であり、純粋な過剰決定などというものは物理的に不可能なのだと考えるのが自然である。というのも、出来事A単独でも出来事Cを引き起こすのに十分な原因であるならば、AとCが共同で作用する場合には、Eの発火する強さは二倍になるはずなのに、単独であろうと共同であろうと発火の強さが一定というのは不可解だからである。これをポールとホールは「加法性の問題（problem of additivity）」と呼んでいる<sup>44</sup>。加法性の問題こそが、純粋な過剰決定を物理的に不可能にさせる元凶である。純粋な過剰決定は現実の自然法則という制約を取り払ったときに仮想的にのみ成立するような因果性である。

よって、もし青酸カリと毒性が死をもたらすことが純粋な過剰決定の事例だとすれば、どちらかの性質には因果的効力がないと論じてもよいだろう。しかし、そもそも青酸カリ（毒性）の錠剤を服用して死ぬということは過剰決定の事例なのだろうか。そうではないと思われる。純粋であろうと不純であろうと、過剰決定では原因と結果となる出来事間に反事実的依存が成り立たないが、目下のケースでは反事実的依存が問題なく成立するからである。このような違いが生じる理由は、青酸カリと毒性の間には緊密なつながりがあるからである。過剰決定の事例では、原因となる二つの出来事は独立しているが、青酸カリの錠剤を服用するという出来事と毒性の錠剤を服用するという出来事は、論理的には独立だが、物理的に可能なすべて世界において、前者が生じるならば後者も生じるという意味で、物理的には独立していない。

それどころか、トークン同一説の下では、青酸カリの錠剤を服用するという出来事と毒性の錠剤を服用するという出来事は同一の出来事である、と論じうる<sup>45</sup>。すなわち

ie [服用する (a, b, e) 、かつ、時点 (e, t) 、かつ、青酸カリ (b, t) ]

<sup>41</sup> シャッフアーは（不純な）過剰決定の事例は実際には稀どころか極めてありふれていると論じている。例えば、大きな石を投げて窓を割った場合、石の上の半球と下の半球はどちらも窓を割るのに十分なだけの運動量をもっているかもしれない。Schaffer 2003, p. 28.

<sup>42</sup> 「純粋な過剰決定」という用語はポールとホールによる。Paul & Hall 2013, pp. 113–114.

<sup>43</sup> 過剰決定の事例は因果性の反事実的分析に問題を引き起こすと指摘されてきた。一見すると、二つの原因の一方が生じなかった場合には結果が生じる仕方が変わるという観察は、反事実的分析にとって好都合である。しかし、結果となる出来事の同一性条件をあまりにきめ細かくしてしまうと、常識的にはある出来事の原因とみなされないような出来事ですら原因になってしまう恐れがある。因果性の反事実的分析の支持者は、過剰決定や遅い先回り（late preemption）といった厄介なケースに対処する際にはこの点に十分注意する必要がある。cf. Paul & Hall 2013.

<sup>44</sup> Paul & Hall 2013, p. 148. cf. Schaffer 2003, pp. 26–27.

<sup>45</sup> cf. Mumford 1998, pp. 139–141.

ie [服用する (a, b, e) 、かつ、時点 (e, t) 、かつ、有毒 (b, t) ]

という記述句は同一の出来事を指示するのではないか。それゆえに、過剰決定を不可能にしている加法性の問題も目下のケースでは生じない。

心的性質が役割性質だとすれば、ここで述べたことは心的性質にもすべて当てはまる。何らかの物的性質 **P** が心的性質 **M** に結び付けられている因果役割を担うとすると、物理的に可能なすべての世界において、**P** を例化する対象は **M** も例化するという意味で、**P** と **M** は物理的に独立ではない。また、2.2 節で定式化した出来事に関するトークン同一説によれば、心的性質はいかなる物的性質と同一ではないとしても、出来事トークンとしては同一であるがゆえに、加法性の問題も生じない。したがって、心的性質が因果的効力をもつと考えた場合でも、心的性質の例化は行動を過剰決定するわけではない。

以上の議論が正しければ、過剰決定の不可能性に基づいて、役割性質は因果的効力をもたないという考えを正当化することはできない。また、たとえ役割性質に因果的効力を認めないとしても、先に論じたように、因果的関連性を認めるという余地もある。実現者機能主義を支持する三つ目の議論も決定的なものではない。

## 2.6 節 本章のまとめ

本章では、同一説の再検討を行った。本章の議論内容を序盤・中盤・終盤の三つの部分に分けて整理しておこう。

まず、本章の序盤 (2.1-2.2 節) では、同一説の定式化に関わる諸問題を取り上げた。いくつかの主要な論点を箇条書きにしておく。

- ・ 同一説 (=心脳同一説) は心的性質と神経学的性質との同一性を主張する経験的な仮説である。
- ・ タイプ同一説とトークン同一説という二種類の同一説がしばしば区別される。本章では「同一説」の名に値するのはタイプ同一説であると論じた。
- ・ タイプ同一説とトークン同一説の区別は個体のもつ性質に関してつけられる場合と、出来事に関してつけられる場合の二通りがある。出来事に関するタイプ同一説とトークン同一説を区別する眼目は、心的因果を論じる文脈において浮上してくる。

本章の中盤 (2.3-2.4 節) では、機能主義と同一説の両立可能性をめぐって二つの立場を提示した。まず、因果役割機能主義の下では、個々の心的性質に特有の因果役割は多重実現可能であるため、同一説は否定されることになる。これに対し、機能主義と同一説の両立可能性を擁護する人々は個々の心的性質をそれ特有の因果役割の担い手と同一視する。この立場 (実現者機能主義) によれば、因果役割ないしそれを担うという状態は多重実現

可能だが、因果役割の担い手そのものは何らかの物的性質と同一視される。

本章の終盤（2.5 節）では、実現者機能主義を擁護する論証を三つ取り上げ、それらすべてに対して問題点を指摘した。狂人の痛みの事例は同一説においてのみ取り扱うことができる、という第一の論証に対しては、まず、狂人の痛みという事例が思念可能かどうかに関して問題があること、また、仮に思念可能だとしても、それは因果役割機能主義の下で取り扱うことができると論じた。理論的同一視に基づく第二の論証に対しては、理論的同一視の本性に関する誤解があるのではないかという指摘を行った。役割性質の因果的無効力性に基づく第三の論証に対しては、まず、役割性質が因果的に無効力であるのと同じ意味で心的性質が無効力であるとしても特にコストが大きいとは思えないこと、また、そもそも役割性質が因果的に無効力であるという議論にも欠陥があると指摘した。以上の応答により、同一説による批判から因果役割機能主義は免れることができた。

### 3 章 機能主義の限界

1 章と 2 章では、機能主義を行動主義や同一説と比較することで機能主義という立場の強みがどこにあるのかを明確にする作業を行った。しかし、1 章と 2 章の議論が正しかったとしても、機能主義によって行動主義や同一説の難点がすべて解決したと楽観することはできない。行動主義と同一説のどちらも、ここまで論じていない二種類の問題に手を焼いてきた<sup>1</sup>。感覚経験の現象的性格（phenomenal character）と志向性である。機能主義もまたこれらの問題に巻き込まれる。

本章ではこれら二つの問題が機能主義にどのようなトラブルを引き起こすのかを確認する。3.1 節では、機能主義者が感覚経験の現象的性格にまつわるパズルの一例として、クオリアの反転を取り上げ、このパズルに対処する方法の一例をスケッチする。3.2 節では、志向性は物的状態へと還元できない心的状態の特徴であるというブレンターノの見解を取り上げ、因果役割機能主義はこの見解を退けることができないだろうと論じる。

#### 3.1 節 感覚経験の現象的性格とクオリアの反転

機能主義に残された課題の一つとして、経験の現象的性格に関する諸問題がよく知られている。感覚経験の現象的性格は、多くの心の哲学者たちにとって悩みの種であり、機能主義者にとっても自説を揺るがしうる問題である。しかし、感覚経験の現象的性格に由来するパズルに対して、機能主義は全く手が付けられないわけではないと思う。

まず、感覚経験の現象的性格とは何かを大まかに確認しておこう。心的状態は現象学的な側面、すなわち、そのような状態であるとはどのような感じであるか（what it is like）という側面から分類される。心的状態の中でも、とりわけ五種類の感覚経験の場合には<sup>2</sup>、そうした経験をするとどのような感じであるのかが明瞭であるように思われる。例えば、熟れたりんごは赤いわけだが、このように対象（りんご）に帰属される赤さとは区別された、感覚経験それ自体の性質に注目することもできるように思われる。そうした感覚経験には、赤いものをみるときに特有の感じ、としか表現できないような性質を帰属することができるかもしれない。感覚経験それ自体に帰属されるこうした主観的な性質を「感覚経験の現象的性格」という。言い換えると、感覚経験の現象的性格とは、どのような感じかという側面から感覚経験を分類するような、経験それ自体に備わる性質のことである。ここで「感覚経験の」という点を強調するのは、信念や欲求などの命題的態度に関しては、感覚経験の場合に比肩しうるような特有の主観的経験が伴うのかどうか疑わしいからで

<sup>1</sup> 行動主義者は志向性に関して手を焼いたと言うよりは、ほとんど何も述べなかったと言うほうが実情に近いかもしれない。Armstrong 1968, p. 57.

<sup>2</sup> 感覚様相を五種類に分類するにあたって、ここでは触覚（皮膚感覚）のうちに、痛覚や温度感覚や圧力感覚なども含めているが、これらの感覚様相を統一するような特徴があるのかどうかはよく分からない。「人間の感覚様相はいくつあるのか？」という問題は意外に厄介である。cf. de Vignemont & Massin 2015.

ある。

さて、感覚経験の現象的性格に関する問題は、現象的性格が脳状態や心的性質に固有の役割性質とは独立であるように思われることから生じる。次のようなシナリオを考えよう。なお、ここでいう「クオリア」は「感覚経験の現象的性格」の単なる別名である。

**クオリアの逆転 (inverted qualia) :** 私が赤いトマトを見るとき感覚経験は、むしろ、あなたが緑のホウレン草を見るとき感覚経験とよく似ている。

このシナリオは、そのままでは同一説と機能主義のどちらにも含意をもたない。しかし、シナリオの詳細によっては、現象的性格は脳状態や心的性質に特有の因果役割とは独立であるという帰結をもつ。例えば、

**クオリアの逆転 ( $\alpha$ ) :** 私とあなたは物的な複製 (duplicate) の関係にあるが、私が赤いトマトを見るとき感覚経験は、むしろ、あなたが緑のホウレン草を見るとき感覚経験とよく似ている。

同一説によれば、すべての心的性質は何らかの物的性質と同一である。しかし、もしこのシナリオが可能だとすれば、私とあなたは同じ物的性質を例化しているにも関わらず、異なる心的性質を例化していることになる。したがって、同一説は偽である。

この批判に対して、同一説の支持者は、あるシナリオが思念可能 (conceivable) であることは、そのシナリオが物理的に可能であるということを含意しないと応じることができるだろう。時間の向きと宇宙のエントロピー増大が食い違うことは思念可能だが、こうした食い違いが生じることはエントロピー増大則と矛盾するので物理的に可能ではない<sup>3</sup>。同じように、クオリアの逆転 ( $\alpha$ ) が思念可能だとしても、それが物理的に可能とは限らない。直観によって思念可能な事柄まで考慮しなくてはならないという正当な理由がなければ、クオリアの逆転 ( $\alpha$ ) が同一説に対して反例を突きつけたことにはならない。

とはいえ、時間の向きと宇宙のエントロピー増大が食い違うことが物理的に不可能であることは、エントロピー増大則が物理法則であることから帰結するのに対して、クオリアの逆転 ( $\alpha$ ) が本当に物理的に不可能であるということはそれほど明らかでない。その意味で、この応答は不十分だと言われる余地を残している。

クオリアの逆転 ( $\alpha$ ) は物理的に不可能だということはできるだろうか。感覚経験の現象的性格はそれを実現する物的状態に付随する、と多くの論者は考えているように思われる<sup>4</sup>。実際、感覚経験の現象的性格は (少なくとも部分的には) 生理学的に説明可能なはずであ

<sup>3</sup> cf. Paul & Hall 2013, pp. 30–31.

<sup>4</sup> 付随性 (supervenience) には強い付随性と弱い付随性の二種類がある。ここでは強い付随性を念頭に置いている。心身付随性を弱い付随性として解釈することへの批判は Blackburn 1984, pp. 184–185 などを参照。なお、心身付随性には異論がないわけではない。Fish 2009, chap. 5 などを参照。

る。例えば、色盲について考えてみよう<sup>5</sup>。

色盲には様々な種類があることが知られており、色盲の人すべてが色のない白黒の世界に住んでいるわけではない。色盲の人の多くは、二色型色覚（特に、第一色盲と第二色盲）に分類される。正常者においては、すべての色が三色の混合で作れるが、第一色盲と第二色盲は二色の混合によって作れる色だけを識別できる。第一色盲と第二色盲の場合、赤・緑光の混色の大部分は、正常者にとっての黄色光に対応してしまう。この事実は、眼の網膜にある三種類の錐体細胞（cone）のうち、一種類（第一色盲では赤を感受する錐体、第二色盲では緑を感受する錐体）が機能していないと考えるとうまく説明がつくという。

しかし、ここまで述べたことは、公共的に観察可能な色の識別能力だけに基づいている。問題は、色盲の人が色をどのように経験しているのかにかかっている。結局のところ色の経験は主観的であって、色盲の人が色をどのように経験しているのかは正常者には知りえないように思える。この問題に取り組む上での重要な手がかりは、片目は正常だが片目が色覚異常という事例である。こうした人物からの内観報告は、正常者と色覚異常者を橋渡しするという意味で貴重である。驚くべきことに、その報告は前の段落で述べた生理学的な推定を裏付けるものであったという。

ポール・チャーチランドは、生理学的な説明を視覚とそれ以外の感覚様相に応用している<sup>6</sup>。まず、視覚に関しては、すでに述べたように三種類の錐体細胞があるので、それらが感受する三種類の波長の光それぞれの強度を三つ組にして、個々の色をベクトル表示する。すると、正常者が経験するすべての色は、三次元の状態空間（state space）のどこかの座標に割り当てられる。上で取り上げたような二色方色覚の人は、一つの次元に対応する情報が欠落していると考ええる。味覚に関しては、チャーチランドは四つの基本味（甘味、苦味、酸味、辛味）に対応する受容器があると仮定して、人間が感じられる任意の味を四次元の状態空間の中に位置づける<sup>7</sup>。色の場合と同様に、任意の味はこれら四種の基本味の割合によって表現される。その他の感覚様相についても同様に考えられる。

チャーチランドは状態空間を調べれば、他の動物であるとはどのようなことかについての手がかりが得られるとも主張する。ネコは、味覚に関しては人間と同様に四種類の基本味を有するが、苦味に対する感度が異なる。つまり、ネコは苦味に対する識別に優れている。そのため、人間にとっては砂糖とそれほど大きな違いのないサッカリンは、ネコにとってはより苦く感じられるに違いない、などと推測することができる。

このように、感覚経験の現象的性格が神経生理学的な基盤に付随している、という仮説はもっともらしい。したがって、クオリアの逆転（ $\alpha$ ）から同一説を否定するという議論はそれほど説得的ではない。同一説を否定するには、前章で検討した、心的性質の多重実現

<sup>5</sup> 大山 1994, 3 章。以下で言及した片目は正常だが片目は色覚異常という事例はかなり興味深いと思われるのだが、寡聞にして哲学文献で取り上げられているのを見たことはない。

<sup>6</sup> Churchland 1992, pp. 102–108. 状態空間を用いた感覚のコード化はニューラルネットワークによって効率よく達成される。なお、感覚器官で多くの並列処理が見られることは、古典的計算主義者ですら認める事実である。Simon 1996, p. 81.

<sup>7</sup> 実際には次元の数はもっと大きいと思われるが、ここでの議論にとっては正確な数は重要ではない。

可能性にもとづく議論の方がより説得的だと言える。

しかし、(α)とは異なるバージョンのクオリアの逆転シナリオは、機能主義に対して問題を引きおこすかもしれない。物的な複製関係にある二人の間でのクオリアの逆転は物理的には不可能だったとしても、そのことが機能主義を救うとは限らないからである。人間の痛みと火星人の痛みに関する前章の議論を参照しよう。人間と火星人は身体の組成は大きく異なるにも関わらず、人間の痛みと火星人の痛みの因果役割は同じである。このとき、因果役割機能主義によれば、両者は同じ心的性質をもつが、それを実現する物的性質において異なる。因果役割機能主義のもとでは、人間と火星人が同じ痛み状態にあることを保証するのは、両者が同じ物的性質を例化していることではない。したがって、感覚経験の現象的性格はそれを実現する物的状態に付随するにせよ、人間と火星人には身体の組成が異なる以上、人間と火星人の間でのクオリアの逆転は物理的に可能かもしれない。すると、次のようなシナリオも可能かもしれない<sup>8</sup>。

**クオリアの逆転 (β) :** 私とあなたは機能的に同値 (functionally equivalent) の関係にあるが、私が赤いトマトを見るときに感覚経験は、むしろ、あなたが緑のほうれん草を見るときに感覚経験とよく似ている。

機能的に同値とは、例化している機能的性質（役割性質）が同じということである。そのため、例えば、私とあなたは熟したトマトに対して「これは赤い」と言い、ほうれん草に対して「これは緑だ」などと言うのであり、刺激に対する反応に注目する限りでは大きな違いは見当たらないと仮定する。

もしこのようなシナリオ (β) が可能だとすれば、因果役割機能主義は偽である。こちらの議論に対しても、先ほどと同じように、シナリオの思念可能性はそれが物理的に可能であることを含意しない、という方針で応答したい。しかし、この方針を貫くことはシナリオ (α) の場合ほど容易ではないという点に注意する必要がある。なぜなら、シナリオ (β) の私とあなたは、シナリオ (α) の私とあなたとは異なり、物的組成が同種であるとは仮定されていないからである。たしかに、因果役割機能主義は心をもつ個体の物的組成に対して何の制約もしないわけではない。とはいえ、人間とは物的組成の大きく異なる存在者が、しかるべき機能的性質を例化することで人間と同じ心的性質を例化する可能性を認めるところに、因果役割機能主義と同一説との違いがある。

したがって、シナリオ (β) が物理的に不可能であると主張することは、シナリオ (α) の場合ほど簡単ではない。しかし、ここで一つの注目に値する論点がある。それは、赤さの感覚には緑の感覚とは違って鮮やかさ (vibrancy) がある、という論点である。赤いものは鮮やかさのゆえに緑のものよりも眼を引きやすい。そのため、赤いものを見るときに緑のものを見るときのような感覚経験をもつ人と、そのような感覚経験をもたない人の行動

---

<sup>8</sup> cf. Fodor & Block 1972, pp. 53–55.



を比べると、違いが出てくるのではないかと考えられる<sup>9</sup>。この論点は必ずしも赤と緑の対比に限定されない。例えば、青は寒色 (cool color) であり、それゆえ青の感覚は黄色の感覚とは異なる。安全なエリアであることを示すために青色が使われるのは、そのためである。このような具合で、色の感覚に関しても機能的性質に付随すると論じることができるかもしれない。そうだとすれば、シナリオ (β) における二つの心的状態は、厳密には機能的に同値ではありえないということになる。

この論点は、クオリアの反転を機能主義に対する批判として構成するには、シナリオの細部をかなり詰めなければ必ずしも説得的にならないことを示唆する。クオリアの反転シナリオはしばしば「スペクトルの反転」とも呼ばれるように、色の感覚経験を記述することで提示される。しかし、色相環 (hue circle) を単純に 180 度回転するだけでは、鮮やかさの有無や暖色と寒色の区別など、行動上の違いをもたらす性質が保存されないだろう。クオリアの反転 (β) は、機能主義の批判としては決定的ではないと考えられる。

### 3.2 節 過少決定の問題

ブレンターノによれば、心的状態と違って物的状態には志向性がない。志向性は心的状態の本質的な特徴であり、物的状態には還元できない特徴であるとされる。しかし、唯物論者 (機能主義者を含む) はこうした見解に反対して、心的状態が志向性をもつのだとすれば、志向性がいかにして物理的な世界において可能なのか、という問題はむしろ解決されるべき課題とみなすだろう。本稿は後者の立場を採用する。

「信念」や「欲求」といった語彙を、心的語彙を含まないで定義することができれば、問題は解消するかもしれない。何らかの心理学理論 Th があって、心的語彙は Th によって陰伏的に定義されているとすれば、ラムジー化の手続きによって個々の心的語彙に対して明示的な定義を与えられるはずである (1.3 節)。

しかし、ここにはちょっとした技術的な問題がある。ラムジー化によって理論語を明示的に定義するには、理論は公理化可能でなければならない。この条件は、内容を持たないか、限られた範囲の内容しか持たない心的状態に限れば成立していてもおかしくないが、信念や欲求などの命題的態度の場合には、そもそも心的語彙の数が有限であるということすら容易には前提できない。なぜなら、命題的態度の内容の数には原理的に上限がないからである。おそらく、我々は無限に多くのことを思考しうる。そのため、もし無数の信念述語を「空は青いという信念 (believe-that-the-sky-is-blue)」とか、「草は緑であるという信念 (believe-that-grass-is-green)」といった、それ以上分節化できない原始語として導入するとすれば、心的語彙のリストは有限では済まないだろう<sup>10</sup>。そして、語彙のリストが有限でなければ理論は公理化可能ではないので、ラムジー化を用いて命題的態度をあらわす心的

<sup>9</sup> Braddon-Mitchell & Jackson 2007, p. 133.

<sup>10</sup> Block 1997, p. 21.

語彙を定義することはできないだろう。

しかし、この問題は、個々の命題的態度をそれ以上分節化できないような主体の性質とみなしたことから生じているように見える。そこで、この仮定を放棄してしまえばよい、という解決策がただちに思いつく。命題的態度は主体と命題との間に成り立つ二項関係であり、「信念」や「欲求」を含む言い回しには命題変数に対する全称量化が隠れている、と考えてみよう<sup>11</sup>。この考え方のもとでは、例えば、信念は、「Pと信じる」、「Qと信じる」などなどに共通する関係とみなされるだろう。すると、心理学理論Thは「空は青いという信念」といった表現ではなく、「信念」や「欲求」といった態度一般を陰伏的に定義しているのだと言えるようになる。心的語彙のリストが有限個では収まらないという問題は、こうして回避される。

では、心的語彙の分析において命題変数を用いるという上の提案を採用すれば、志向性がいかにして物理的な世界において可能なかを解明する道筋は立つのだろうか。そうではないと考える理由がある。上の提案にしたがって、「信念」を定義した場合にはどのような結果になるかを見てみよう。まず、1.3.2節で示したように、心的語彙のリストを  $M_1, \dots, M_n$  とし、「痛み」に対応する語彙を  $M_i$  とする。 $M_j$  の定義は

$$M_i(x, t) \leftrightarrow \exists F_1, \dots, \exists F_n (\Phi(F_1, \dots, F_n) \& F_i(x, t))$$

となる。それでは、「信念」に対応する語彙を  $M_j$  として同様の定義を試みよう。すると  $M_j$  の定義は、任意の命題  $p$  に関して

$$M_j(x, t, p) \leftrightarrow \exists F_1, \dots, \exists F_n (\Phi(F_1, \dots, F_n) \& F_j(x, t, p))$$

という形をとるだろう。 $M_i$  の定義とは違って、ここでは命題変数が  $M_j$  と  $F_j$  の関係項に現われている。 $F_i$  の値、すなわち、具体的対象である主体と何らかの抽象的対象である命題（と時点）とを結びつける関係は、感覚入力と運動出力によって特定されるような関係に限定されるはずである。

ここで問題が生じる。こうした限定の下では個々の命題的態度の内容を確定できそうにないのである。次のようなシナリオを考えてみよう<sup>12</sup>。フレッドが苦痛を感じることを欲しているメアリーは、午前 3 時にチェロをかき鳴らせば隣人のフレッドが苦痛を感じると信じているので、午前 3 時にチェロをかき鳴らした。これら二つの信念と欲求をメアリーに帰属すれば、午前 3 時にチェロを演奏するという彼女の行動を説明することができる。しかし、同じ行動は別の信念と欲求のペアによっても説明することができる。例えば、メアリーは自分の隣人がフレッドではなくアルバートだと信じているとする。アルバートが苦

<sup>11</sup> Lewis 1972, p. 258n13.

<sup>12</sup> Stalnaker 1984, p. 17. cf. Field 1986, p. 440.

痛を感じることを欲しているメアリーは、午前 3 時にチェロをかき鳴らせばアルバートが苦痛を感じると信じているので、午前 3 時にチェロをかき鳴らした。このように、フレッドに関するメアリーの命題的態度をすべてアルバートに置き換えてしまえば、メアリーの行動はまったく同じように説明される。

もちろん、メアリーがアルバートに会ったこともないのだとすれば、二番目の説明は明らかに倒錯している。しかし、ここで問題なのは、メアリーがどちらの命題的態度を抱いているのかはこれまでの道具立てからは決定できないということである。チェロをかき鳴らすという行動に対して何らかの欲求状態 **D** と信念状態 **B** が因果的に関与しているということまでは、はっきりしている。しかし、**D** と **B** がどのような内容をもつのかは、それらがどんな行動を引き起こすのかということだけからは定まらない。**D** の内容はフレッドが苦痛を感じることもかもしれないし、アルバートが苦痛を感じることもかもしれない。一方の信念・欲求の体系においてフレッドが苦痛を感じることの欲求に結び付けられた因果役割と、他方の信念・欲求の体系においてアルバートが苦痛を感じることの欲求に結び付けられた因果役割は同じということになるだろう。したがって、心的性質は因果役割によって個別化されるという因果役割機能主義の考え方では、命題的態度の内容を個別化することまではできそうにない。科学哲学の用語を使うならば、因果役割は命題的態度の内容を過少決定 (underdetermine) している。

いまのシナリオでは、命題的態度の内容は単称的だった。しかし、この点は本質的ではない。別のシナリオを考えよう<sup>13</sup>。デイヴィドは一杯のビールを飲むことを欲しており、パブに行けばビールを飲めると信じている。これら二つの欲求と信念により、デイヴィドがパブに足を運んだことは説明できる。だが、同じ行動は別の欲求と信念のペアによっても説明できる。デイヴィドは泥の皿を欲しており、パブに行けば泥の皿が手に入ると信じている。これら二つの欲求と信念により、デイヴィドがパブに脚を運んだことは先と同じように説明できる。

したがって、表象状態の内容を物理的刺激や行動に関わる語彙を用いて、あるいは、感覚入力や運動出力によって特定されるような因果役割によって確定させることは、一般にはできそうにない。こうした結果に直面して、多くの機能主義者は次のような見解にまで後退することになった<sup>14</sup>。すなわち、ある物的性質が与えられたとき、その性質が担っている因果役割によって定まるのは態度 (attitude) に関わる部分まで、つまり、その物的性質は信念を実現しているのか、欲求を実現しているのか、といった問いに対する答えまでであり、そうした心的態度の内容 (content) が何であるのかを決定するには、さらに多くの要因が必要になる、と。喩えて言うなら、この状況は文章中の見知らぬ単語について、その語が出現している文脈から品詞や格を決定することはできても、それが何を意味しているのかまでは決定できない、といった状況に似ている。

---

<sup>13</sup> Lewis 1986, p. 107.

<sup>14</sup> cf. Loewer 1997, p. 110.

### 3.3 節 本章のまとめ

本章では、機能主義が行動主義や同一説から引き継いだ二種類の課題を取り上げた。一つ目の課題は感覚経験の現象的性格に関わる。3.1 節では、感覚経験の現象的性格が機能主義対してもたらす問題の一例として、クオリアの反転のパズルを取り上げ、機能主義者は感覚経験の現象的性格は機能的性質に付随すると述べることでこのパズルに対処できるのではないかと述べた。よって、クオリアの反転は機能主義批判として決定的ではない。

二つ目の課題は、志向性は物的状態に還元できない心的状態の本質的な特徴であるというブレンターノの見解に関わる。3.2 節では、ストルネイカーやルイスのシナリオに基づいて、ラムジー化や因果役割といった道具立てを用いて、信念や欲求といった心的態度の内容を確定することはできそうにないことを示した。こうして機能主義者に残された問題を本稿では「過少決定の問題」と呼ぶことにする。過少決定の問題のゆえに、現在では機能主義者の多くは、命題的態度に関する別立ての理論が必要だと考えている。こうした見立てのもと、本稿の第二部では、機能主義を補完すべき命題的態度に関する別立ての理論を探っていく。

## 第二部

### 自然主義の意味論と命題の内容

## 第二部の見通し

1980年代以降、ドレツキやフォーダー、ミリカンといった哲学者たちによって、「自然主義的意味論 (naturalistic semantics)」と総称される諸理論が提案されてきた。自然主義的意味論は、主体が抱えている信念・欲求の内容を決定する物理的事実とはどのようなものか、という問題に回答を与えることで、機能主義を補完することが意図されている。

第二部では、自然主義的意味論に属する諸理論を検討する。ここではその準備作業として、自然主義的意味論の立ち位置を明確にするため、5つの論点について議論する。

- (1) 記述的意味論と基礎的意味論
- (2) 自然主義的意味論とグライスのプログラム
- (3) 意味論的自然主義と反自然主義
- (4) 自然主義的意味論の諸理論
- (5) 自然主義的意味論と心の表象理論

### (1) 記述的意味論と基礎的意味論<sup>1</sup>

論理学では、「意味論」という表現は「構文論 (syntax)」と対にして導入される。品詞や文法について考察する分野としての構文論に対し、言語中の任意の表現に対して、それが何を意味するのかを特定する分野として意味論を考える。具体的には、意味論は言語表現に何らかの対象を意味論的値として割り当てことで、ある特定の言語の文に対して真理条件を特定したり、推論の妥当性を説明する。

こうした課題に答える理論は「記述的意味論 (descriptive semantics)」と呼ぶことができる。しかし、「意味論」は記述的意味論とは別の取り組みを指す場合もある。指示や命題的内容といった意味論的な概念を意味論的でない概念によっていかにして説明するのかという問題に答える理論は「基礎的意味論 (foundational semantics)」と呼ばれる。基礎的意味論は記述的意味論の成果を背景としつつも、個々の言語表現が何かを意味するという事実がどのようにして構成されるのかを明らかにしようとする。

固有名を具体例にとって、記述的意味論と基礎的意味論の対比を明らかにしよう。記述的意味論が扱うのは、固有名が含まれる文の真理条件に対して固有名はどのような寄与をするのか、言い換えると、固有名の意味論的値は何か、という問題である。この問題に対する代表的な回答は、大まかにミル主義とフレーゲ主義の二種類に分けられる。ミル主義の記述的意味論によれば、固有名の意味論的値はその指示対象である。フレーゲ主義の記述的意味論によれば、固有名の意味論的値は記述的内容や、あるいは、内包（可能世界から対象への関数）である。次に、固有名がその意味論的値をもつという事実がどのように

---

<sup>1</sup> この区別は Stalnaker 1997, p. 535 で提示されている。ストルネイカーによれば、この区別を明確にしたところが、『名指しと必然性』(Kripke 1980) の大きな功績である。

構成されているのかという基礎的意味論の問題に目を向けよう。この問題に対するフレーゲ主義の回答はそれほど明らかではない。他方、ミル主義の記述的意味論を支持する人はここで指示の因果説 (causal theory of reference) に訴えるかもしれない。すなわち、現在における固有名の使用から指示対象に対する命名にまで遡りうる因果連鎖によって、固有名が特定の対象を意味するという事実が構成される、と。

自然主義的意味論は、記述的意味論よりも基礎的意味論という意味での「意味論」に関わる。たしかに、自然主義的意味論を論じる哲学者の多くは、言語哲学よりも心の哲学という領域で研究している。しかし、「真である」といった意味論的評価の述語は自然言語の文表現だけでなく信念にも適用される。自然主義的意味論を論じる哲学者たちは、個々の信念が意味論的に評価できる内容をもつという事実がいかにして成立しているのかを問題にしている。この意味で、自然主義的意味論は基礎的意味論に関わる<sup>2</sup>。

## (2) 自然主義的意味論とグライスのプログラム

言語表現と心的状態との間に志向性をもつという共通点があることは、両者がどのような関係にあるのか、あるいはどちらが優先するのか、といった問題を引き起こした。この問題に対し、多くの自然主義者は、言語表現の志向性が心的状態の志向性に依存するという可能性を探ってきた。さらにいえば、多くの自然主義者は、心的状態の志向性を何らかの自然的性質に依存すると考え、心的状態が何かについてのものであることはいかにして可能か、という問いに向かった。

実際、自然主義的意味論はグライスのプログラムと対比されることがある。グライスは話し手の意味（非自然的意味）を、信念や意図を前提にして定義しようとした。これに対し、自然主義的意味論は、信念などの心的状態が何かを表象するための（必要）十分条件を、因果性のような概念のみによって与えようとする。ただし、これら二つの研究プログラムは必ずしも競合するわけではない。例えば、意味に関する次のような素朴な描像の下では、グライスのプログラムと自然主義的意味論は補完的な関係におかれる。すなわち、言語的意味は話し手の意味によって説明され、話し手の意味は話し手の信念や意図といった心的状態によって説明され、信念は因果性のような概念によって説明される、といった描像である。

本稿では、この素朴な描像を包括的に論じること、ましてや包括的に擁護することはできない。本稿第二部は、上の素朴な描像に照らしていえば、命題的態度を因果性のような概念によって説明するという最後のステップだけを検討する。

<sup>2</sup> 心の哲学において意味論の自然主義者は、指示の因果説に類似したアイデアをしばしば利用する。指示の因果説が回答しようとしている問題は、言語表現が現に意味するものを意味するのはどのようなメカニズムによるのか、といった種類の問いであった。これに対して、自然主義者たちは、信念のような心的状態が何かについてのものであるのはどのようにしてか、という問題に答えるためにクリプキのアイデアを流用する。ここには重要な共通性がある。Godfrey-Smith 1992, Sterelny 1990, pp. 114–118. ただし、クリプキ自身は、指示の因果説によって指示の概念を消去することを目指していたわけではない (Kripke 1980, p. 97)。

### (3) 意味論的自然主義と反自然主義

自然主義的意味論の諸理論は、意味論的性質が何らかの物的性質によって決定されると考える点で一致する。つまり、本稿の5章以降で個別に検討する自然主義的意味論の諸理論は、意味論的自然主義 (semantic naturalism) にコミットしている。

**意味論的自然主義**：志向性をもつ性質は何らかの物的性質Pによって法則必然的に含意され、かつ、この含意関係は明瞭に特徴づけることが可能である<sup>3</sup>。

このテーゼの前半を否定する哲学者はあまりいないと思う。しかし、このテーゼの後半を受け入れられるかどうかはかなり微妙であり、それゆえ、意味論的自然主義にコミットする人々には温度差がある<sup>4</sup>。極めて徹底した自然主義者（例えば、ミリカン）は、言語表現の意味から信念・欲求といった命題的態度に至るまで、志向性をもつすべての性質とそれを必然化する物的性質との関係を明瞭に特徴づけることが可能だと論じる。やや穏健な自然主義者（例えば、ストルネイカー）は、信念・欲求といった命題的態度を必然化する物的性質との関係を明瞭に特徴づけることは可能だと論じる。さらに穏健な自然主義者（例えば、ステレルニー）は、志向性をもつ性質のなかには明瞭な特徴づけが可能なものもある、と主張する。

本稿では最終的に、もっとも穏健な自然主義の一種を擁護する。すなわち、信念・欲求の一部のクラスに関して、それを必然化する物理的性質との関係を明瞭に特徴づけることは可能だという見解を擁護する。この見解は穏健であるがトリビアルではない。以下の諸章を読んでもらえれば、たとえ志向性をもつ性質のごく一部であっても、それを必然化する物理的性質との関係を明瞭に特徴づけるという作業が、そう単純ではないことは少なくとも理解してもらえると期待する。

穏健な自然主義の外側には、志向性をもついかなる性質もそれを必然化する物的性質との関係を明瞭に特徴づけることは不可能である、という悲観的な立場が位置する。この立場を「意味論的反自然主義」と呼んでおく。レーアによれば、意味論的反自然主義が正しいとすれば哲学者には二つの道しか残されていない<sup>5</sup>。一つは志向性をもつ性質などそもそも実在せず、したがって例化されることもないという立場（消去主義）である。4章で取り上げるクワインの哲学は消去主義の立場に分類されうる。もう一つの選択肢は、志向性をもつあらゆる性質は、物的性質に付随しないか、あるいは、それを必然化する物的性質との関係を明瞭な特徴づけは不可能なほどに二つの性質は独立している、という二元論を採用することである。これを意味論的二元論と呼んでおく。ただし、本稿では、意味論的二元論を扱ってはいない。

---

<sup>3</sup> cf. Loewer 1998, pp. 108–109.

<sup>4</sup> Godfrey-Smith 1996, p. 176. cf. Nanay 2013, p. 156.

<sup>5</sup> Loewer 1998, p. 110.



#### (4) 自然主義的意味論の諸理論

3.2 節で論じたように、因果役割機能主義の限界は、心的性質と結び付けられた因果役割に注目する限り、主体が抱いている信念と欲求の内容は一意に定まらないという点にあった。私見によれば、この難点を克服する手段としてこれまでに提案されてきた理論は、以下の三種類に分けられる。

- ・ 信念と欲求の内容を一挙に定めることを可能にする合理性の制約を導入する。
- ・ 信念の内容を決定する物理的事実を特定し、判明した信念の内容を所与として、欲求の内容を決定する。
- ・ 欲求の内容を決定する物理的事実を特定し、判明した欲求の内容を所与として、信念の内容を決定する。

これら三つの選択肢に共通する前提は、志向性をもつ心的性質は信念・欲求の他にもいろいろあるかもしれないが、信念と欲求の二つの体系は比較的自律しており、内容に関する限りは信念と欲求の相互依存関係のみに注目してもよい、というものである。三つの選択肢の違いは、この相互依存関係をどのように解きほぐすかにかかっている。

一つ目の選択肢は、5 章で取り上げるルイスの解釈理論に代表される。ルイスによれば、ある生命体がどのような内容の信念と欲求を抱いているのかという問題は制約充足問題である。そして、生命体の行動を因果的に説明できて、なおかつ、合理性の諸制約を満たすような信念・欲求の体系は一意に定まる。こうして、過少決定の問題は解消する。

二つ目の選択肢と三つ目の選択肢は、信念と欲求のどちらか一方の内容が定まれば、他方の内容も定まると仮定する。例えば、信念の内容を所与として欲求の内容を決定する原理は次のように定式化できるかもしれない。

p を欲求するとは、当人の信念（何であれ）が真である世界で p をもたらす（bring it about that p）であろう仕方で行為する傾向にあるということである<sup>6</sup>。

欲求の内容を所与として信念の内容を決定する原理については

p を信じるとは、p が真であるような世界では、当人の欲求（何であれ）を満たすような仕方で行為する傾向にあるということである<sup>7</sup>。

と定式化できるかもしれない。これらの原理が本当に正しいかどうかは、それ自体として吟味に値する<sup>8</sup>。しかし、仮にこの種の原理が正しいのだとしても、二つの原理は明らかに

---

<sup>6</sup> Stalnaker 1984, p. 15.

<sup>7</sup> *ibid.*

<sup>8</sup> 8 章では、欲求内容を所与として信念内容を決定するのにふさわしい原理が何かを考察する。

循環している。何らかの方法で循環の中に入ることができなければ、信念内容と欲求内容のどちらも決定することはできないだろう。そこで、二つ目の選択肢は信念の内容を欲求の内容とは独立に決定する方法を模索する<sup>9</sup>。これと対照的に、三つ目の選択肢は欲求の内容を信念の内容とは独立に決定する方法を模索する<sup>10</sup>。ストルネイカーやドレッキは二つ目の選択肢を、ミリカンやパピノー、ホワイトは三つ目の選択肢を支持している。6章では二つ目の選択肢を、7,8章では三つ目の選択肢について検討を加える。

信念・欲求の内容を因果役割だけでは決定できないという機能主義の苦境と、三つの選択肢がこの苦境をどう克服しようとしているのかを、比喩的に述べてみよう。いま、二つの変数を含む一次方程式  $ax+by=c$  が与えられたとする。このままでは解が無数にある。この状況を解消する一つの方法は、独立の一次方程式  $a'x+b'y=c'$  を持つてくることである。そうすれば連立方程式を立てて解を見つけられる。この方法は一つ目の選択肢に相当する。これに対して、まず  $x, y$  のどちらかの値を何らかの方法で決定し、しかる後に  $ax+by=c$  を解いて他方の変数の値を決定するという方法もありうる。こちらの方法は二つ目と三つ目の選択肢に相当する。

以上の説明だけでは三つの選択肢が何を意味しているのか、まだ判然としないかもしれない。だが、詳しい内容は第二部の本論の中で次第に明らかになるはずである。ここでは差し当たり、5章以降で検討していく諸々の立場は、三種類の選択肢のいずれかに分類されるということだけ了解してもらえればよい。参考までに、(3) と (4) の小節で導入した諸見解の関係を表にまとめておく。

反自然主義	消去主義 (4章)
	意味論的二元論 (本稿では扱わず)
自然主義	信念・欲求の内容を一挙に定める (5章)
	信念の内容を第一に定める (6章)
	欲求の内容を第一に定める (8章)

#### (5) 自然主義的意味論と心の表象理論

心的性質を特定の因果役割の担い手をもつという性質と同一視する機能主義に従うと、命題的態度  $A(x, t, p)$  が成り立つのは、 $A$  に特有の因果役割を担う何らかの物的性質  $P$  があって  $P(x, t, p)$  が成り立つということである。しかし、3.2節で論じたように、因果役割だけでは命題的態度を個別化することができない。そのため、機能主義を補完するような理論を別個に立てる必要があるというのが第一部の結論であった。

ここでいう機能主義の「補完」の内実に関しては、「心の表象理論 (representational theory of mind: RTM)」と呼ばれる見解に基づいて、より具体的なイメージを与えることができる

<sup>9</sup> Stalnaker 1984, p. 19.

<sup>10</sup> Whyte 1993.

かもしれない。RTMは次のように定式化される<sup>11</sup>。

**心の表象理論**：任意の主体  $x$ 、時点  $t$ 、命題  $p$ 、態度  $A$  に関して、 $A$  を実現する物的関係  $P$  があって、 $A(x, t, p) \leftrightarrow \exists r [P(x, t, r)$ 、かつ、 $r$  は  $p$  の内的表象である]。

内的表象は何らかの物的状態、具体的には何らかの脳状態である。RTMによれば、任意の命題的態度  $A(x, t, p)$  は命題 $p$ の内的表象と主体（と時点）との関係によって実現される。RTMを支持する一つの理由は、我々は同じ内容に対してさまざまな態度をとることができるという心的事実がどのような物理的メカニズムによって実現できるのかを考えた場合、それは同じ内容を意味している記号（内的表象）に対して心的態度の種類に応じて異なるさまざまな関係を結ぶからだと考えれば一応の説明がつく、といったものである<sup>12</sup>。例えば、脳の中には「信念箱（belief box）」とでも呼びうる領域があり、信念箱にある内的表象が蓄えられているということは、その内的表象の命題内容が主体によって信じられているという状態を実現しているのかもしれない。あるいは、脳の中には「欲求箱」とでも呼びうる領域があり、そこにある内的表象が蓄えられていることは、その表象の内容が主体によって欲求されているという状態を実現しているのかもしれない。主体と内的表象との間に結ばれるそうした関係は、個々の心的態度に特有の因果役割の担い手として特徴づけられる。

しかし、RTMの定式化にも出現している「 $r$ は $p$ の内的表象である」という部分に関しては、3.2節で論じたような理由で、いまだ何も明らかになってはいない。もしこの部分に対して意味論的な概念を抜きにして説明を与えることができれば、機能主義を補完したといえるだろう。これが自然主義的意味論の行おうとしていることである。

一見すると、以上はなかなか魅力的な考え方である。だが、内的表象とは何であるのか具体的に特定されないままに、「 $r$ は $p$ の内的表象である」という表現に説明を与えることなどできるのだろうか。たしかに、RTMの支持者たちは内的表象に関してもう少し具体的な提案もしている。RTMの支持者の多くは、内的表象は文のような（sentence-like）何かであると主張する傾向にある<sup>13</sup>。しかし、文と類比的とされる内的表象がいかなる文法構造を

<sup>11</sup> cf. Rey 1997, p. 210.

<sup>12</sup> cf. Rey 1997, pp. 231–232.

<sup>13</sup> この見解は「思考の言語説（language of thought hypothesis）」と呼ばれ、我々が何らかの思考をはたかせるとき、我々の脳内では命題を表現する何らかの言語的対象についての記号操作がなされている、と主張する。例えば、 $p$ と信じているということは、理論  $Th$  が特徴付ける信念態度を実現する脳内の信念ボックスに  $p$  を表現する文が含まれるというのである。この  $p$  を表現する文というのが、思考の言語説が解釈するところの内的表象の正体である。なお、「言語的対象」と言っても、脳内に（例えば、日本語の）文が文字通りの意味で書き込まれている必要はない。日本語の文トークンは音声でもインクの染みでもありうることから明らかなように、言語表現は媒体に関して中立である。したがって、日本語の文トークンが神経細胞の配置され方によって実現されるといった可能性はアプリアリには排除できない。

思考の言語説は Fodor 1975 や Field 1978 において提案された。教科書的な解説は Fodor 1987, Appendix; Sterelny 1990, chap. 2–4; Rey 1997, chap. 8 など参照。ただし、ここで名前を挙げた哲学者が今でも思考の言語説を支持している保証はない。フィールドとステレルニーはすでに懐疑的である。Field 1986; Devitt &

もった言語を構成しているのかは論者の間で合意がない<sup>14</sup>。また、意味論の自然化を目指してこれまでに提案されてきた理論の多くは、個々の心的態度とは独立に存在するとされる内的表象に対して、それがどのような内容をもつのかを明らかにする、といった形式をとってはいない。

こうした状況を踏まえると、自然主義的意味論と心の表象主義は、無関係ではないにせよ、かなりの程度独立していると言わざるを得ない<sup>15</sup>。本稿で論じる自然主義的意味論は、内的表象の内容を決定するというよりは、個別の心的態度Aに関して「xはpをAするのは…ときそのときに限る」といった双条件による説明を与えることを目指す。心的態度の種類に応じて、主体と命題内容との関わり方は大きく異なりうるだろう。

---

Sterelny, 1999, p. xi などを参照。

<sup>14</sup> ピンカーはフォーダーが思考の言語の詳細をまったく明らかにしていないと批判している。Pinker 2005, pp. 5–8.

<sup>15</sup> シェアによれば、フォーダーは思考の言語説を 1970 年代に提示したが、内的表象がいかにして意味をもつのかという問題に取り組む必要性を 1980 年代に入るまで認識していなかった。Shea 2004, pp. 42–43.

## 4 章 クワインの言語哲学

“Quine’s philosophy is a large continent” “I do not know how all the pieces of it can be reconciled”

Putnam 1983, p. 240

### 4.1 節 翻訳の不確定性と指示の不確定性

自然主義的意味論は、意味論的性質が自然的性質に付随するという主張にコミットしている。これに対し、クワインは、意味論的性質に関する消去主義を主張した代表的な論者である。意味論の自然化を考察する上でクワインの言語哲学は無視できない。そこで、本章ではクワインの言語哲学を批判的に検討する。

意味論的性質に対するクワインの批判は、根源的翻訳 (radical translation) の思考実験に基づいていると考えられる。クワインは、根源的翻訳の思考実験から二つのパラドキシカルな主張を導き出した。一つは翻訳の不確定性 (indeterminacy of translation) テーゼである。

**翻訳の不確定性テーゼ：**二組の言語学者がある言語の表現を母国語の表現へと写像する翻訳マニュアルを作成するとき、単純性などのプラグマティックな制約を無視するならば、科学的に入手可能な証拠すべてを考慮に入れたとしても、互いに競合するような二つの翻訳マニュアルを作ることができる。

もう一つの主張は指示の不確定性 (indeterminacy of reference) テーゼである<sup>1</sup>。

**指示の不確定性テーゼ：**ある特定の言語の表現に指示対象を割り当てるとき、単純性などのプラグマティックな制約を無視するならば、科学的に入手可能な証拠すべてを考慮に入れたとしても、「言語表現 ‘ $\alpha$ ’ は  $\beta$  を指示する」という形をした意味論的な主張の真理値はすべて不確定である。

二つのテーゼのどちらがより重要だろうか。エヴァンズは「翻訳の企てに対する制約はごくわずか (slight) であるのに、翻訳の不確定性に関してクワインと些細なことで論争するのは口うるさい人だ」と述べている<sup>2</sup>。私はエヴァンズのこの見解に同意する。そのため、本章では主として指示の不確定性テーゼの方を取り上げることになる。

指示の不確定性テーゼはかなり受け入れがたい主張である。たしかに、いくつかの言語

<sup>1</sup> 『ことばと対象』(Quine 1960) では「指示の不可測性 (inscrutability of reference)」と言われるが、後年の『真理を追って』(Quine 1992) では「指示の不確定性」と言い直されている。丹治が指摘するように、「不可測」という表現は、我々には知ることができないだけで事実としては確定している、といったニュアンスをもつことが変更の理由なのかもしれない。丹治 2009, p. 237.

<sup>2</sup> Evans 1975, p. 26.

表現に関しては、指示対象の割り当てが不確定であってもそれほど驚きはない。例えば、「赤い」のような曖昧述語 (vague predicate) は境界が不明瞭に思われる。そのため「この斑点」という直示的表現によって指示されている斑点が、赤色とオレンジ色の境界事例 (Borderline case) である場合では、文「この斑点は赤色である」は真か偽のどちらかを決定する手段がない<sup>3</sup>。そのため、何人かの哲学者たちは、指示の不確定性を反映した意味論を曖昧述語に対して適用することを実際に試みてきたのであった<sup>4</sup>。しかし、クワインによる指示の不確定性テーゼは、曖昧述語の事例から予測されるよりも、はるかに広範囲にわたって言語表現の指示が不確定であることを主張している。

本章の構成は以下のようなものである。4.2 節では、翻訳の不確定性テーゼの論証を三段論法として定式化し、二つの前提が根源的翻訳の思考実験によってどのように擁護されているかを考察する。4.3 節では、根源的翻訳の思考実験に基づいて指示の不確定性を導くことはできない、というエヴァンズの議論を検討する。4.4 節では、指示の不確定性が受け入れられない理由を命題的態度の帰属という実践と両立しないことに求める。

## 4.2 節 根源的翻訳と翻訳の不確定性

本節では、『ことばと対象』における翻訳の不確定性テーゼの論証を提示する。あらかじめ大枠を示しておく、クワインによる翻訳の不確定性テーゼの論証は以下のような三段論法として定式化できる<sup>5</sup>。

翻訳作業において利用できる証拠は言語行動の傾向性に限られる (クワインの行動主義  $\alpha$ )。

もし翻訳作業において利用できる証拠が言語行動の傾向性に限られるならば、互いに競合する複数の翻訳マニュアルが存在する。

よって、互いに競合する複数の翻訳マニュアルが存在する (翻訳の不確定性テーゼ)。

この論証は妥当なので、問題は二つの前提が成り立つかどうかにかかっている。クワインは根源的翻訳の思考実験によって、これらの前提、特に二つ目を動機づけている。

<sup>3</sup> ただし、曖昧性から意味論の変更という結論にただちに飛びつく必要はない。例えば、「この斑点は赤色である」の真理値は実際には確定しているのだが、曖昧述語の意味についての我々の認識能力に限界があるために境界事例では真理値が不確定にみえるにすぎない、という可能性はある。曖昧性についてのこうした考え方は「認識説 (epistemicism)」と呼ばれ、二値原理を擁護する哲学者 (古くはストア派の哲学者たち) によって支持されてきた。cf. Williamson 1994.

<sup>4</sup> 例えば、超付値 (supervaluation) の意味論によれば、ある文が真であるのは、すべての許容可能な (admissible) 解釈のもとでその文が真であるときそのときに限る。例えば、「この斑点は赤色である」という文は、指示されている斑点が赤色の典型例である場合には、すべての許容可能な解釈のもとで真となるので、真である。他方で、指示されている斑点が赤色とオレンジ色の境界事例である場合には、真になる許容可能な解釈もあれば偽になる許容可能な解釈もあるので、真でも偽でもない。

<sup>5</sup> 以下の定式化とは微妙に異なるが、バーンも翻訳の不確定性の論証を三段論法として定式化している。Byrne 2007.

本節の前半では、クワインが翻訳作業をどのように特徴づけているのかを考察することで一番目の前提が何を意味するのかを明確にし、本節の後半では、なぜクワインは二つ目の前提が成り立つと考えたのかを考察する。

#### 4.2.1 節 クワインの行動主義

根源的翻訳とはどのような思考実験なのか。根源的翻訳とは、我々とこれまで一切の接触をもってこなかった人々とコミュニケーションする必要にせまられた言語学者が、現地に赴いて、そこで使用されている未知の言語を我々の言語（例えば、日本語）に翻訳するために行う仮想的な作業のことである。このように極端な条件下で翻訳が行われることは、現実にはありそうにない。通常の翻訳作業では、言語学者は翻訳のための手がかりとして様々なものを利用できてしまうからである。「親族関係にある言語間、例えば、フリースランド語と英語の翻訳では、語源を同じくする語の形態が類似していることが助けとなる」であろうし、「ハンガリー語と英語のように無関係な言語間の翻訳では、文化の共有に伴って進化しながら伝統的に与えられてきた同値関係がその助けとなるかもしれない」<sup>6</sup>。しかし、根源的翻訳の眼目は、最も純粋な翻訳作業を想定することで言語的意味やその理解の本質的な側面を明らかにすることにある。つまり、こうした極限状態において言語学者が利用できる翻訳の手がかりは何かを想像すれば、言語的意味や意味理解が何に存するのかもおのずと見えてくるだろう、と期待される。こうした眼目に基づくなら、翻訳の状況が非現実的なこと自体はあまり重要でない。

根源的翻訳に携わるフィールド言語学者は、まず、現地人による言語行動の傾向性に注目するだろう。言語行動の傾向性は、眼の色彩照射パターンのように感覚器官に与えられる近刺激（proximal stimuli）と、観察可能な言語行動の組によって特定される。根源的翻訳を行う言語学者はそうした言語行動の傾向性を参照しながら、現地語の表現から日本語の表現への写像である翻訳マニュアルの作成を試みることができる。ここまでは特に異論もないだろう。

しかし、クワインは、根源的翻訳に携わる言語学者が利用できる手がかりは現地人による言語行動の傾向性に限られる、と言う。ある翻訳マニュアルが現地人の言語行動の傾向性と合致していれば、それは正しい翻訳マニュアルである。翻訳の正しさに関するこうした見解を、ここでは「クワインの行動主義」と呼ぶことにしよう<sup>7</sup>。

ここでいう「行動主義」の用法はややトリッキーである。一見したところ、根源的翻訳の思考実験は行動主義心理学の学習理論に依拠しているわけではない。根源的翻訳の思考実験において翻訳作業にたずさわる言語学者が、すでに母語に習熟した大人の言語学者と

<sup>6</sup> Quine 1960, pp. 28, 76. 英語とフリースランド語がともにインド・ヨーロッパ語族に属するのに対して、ハンガリー語はウラル語族に属する。

<sup>7</sup> 翻訳の不確定性は行動主義からの帰結であると言われる。クワイン本人もその解釈を受け入れているので、この呼び名は正当だと考えられる。Quine 1987, p. 5. cf. Quine 1992, sec. 14.

仮定されていることから明らかなように、クワインは言語獲得 (language acquisition) の発生論的な分析を行っているわけではないからである。

ちなみに、言語の個体発生に関して付け加えておくと、そもそも、言語獲得は条件付けの学習メカニズムによって説明されるべきだという主張にクワインは全面的には与していなかった可能性がある。クワインは『ことばと対象』3章で指示表現の使用に関する発生論的ともいえる考察を行っている。そこで彼は、条件付けのメカニズムはせいぜい語の習得の初期段階でしか支配的でないかもしれず、複雑な構文に習熟する過程では「さらに別の力が働いている余地がある」と述べている<sup>8</sup>。実際、現在の認知科学において言語獲得を説明する最も有力な学説はおそらくチョムスキー派の生得主義 (nativism) である。一般に、生得主義の妥当性は最終的には経験的な問題だと考えられている<sup>9</sup>。そうだとすれば、普遍文法 (universal grammar) の規則が心の中で表象されているといったチョムスキーの主張の中でも特に物議を醸している主張を除けば<sup>10</sup>、クワインは生得主義にあまり強くは反対しないかもしれない。

いずれにせよ、クワインの立場を「行動主義」と呼ぶからといって、1.2.1 節で取り上げた行動主義心理学への批判をここで持ち出しても、翻訳の不確定性テーゼをただちに否定することはできない。現実の言語獲得のあり方がどのようなものと判明するにせよ、クワインの関心は別のところに向けられていたと思われる<sup>11</sup>。それは、言語に習熟しているということ (language mastery) は何に存するののかという問題である。この問題は、幼い時期に始まり急速に進む人間の言語獲得がいかにして達成されるのか、という問題とは異なる。現実には達成されている言語学習を説明するのに行動主義的心理学が指定している学習メカニズムだけでは不十分かもしれないが、言語習熟という心的状態をある種の行動の傾向性として特徴づけられる、という可能性はまだ残っている。つまり、言語行動の傾向性であり、これが翻訳マニュアルの正しさを決定する手がかりのすべてとなる。

翻訳の正しさに関する以上のような見解は、一般的な通念とははっきり異なる。常識的には、正しい翻訳マニュアルとは意味を保存するマニュアルであるはずだが、クワインはこうした通念を「意味」に言及しているという理由で退ける。なぜなら、意味の概念は不明瞭で科学的に取り扱える代物ではないからである。意味について語るとき、われわれは二つの語の意味が同じかどうか (同義性)、ある文が意味によって真かどうか (分析性)、

---

<sup>8</sup> Quine 1960, p. 82.

<sup>9</sup> 言語獲得の生得説に関しては Pinker 1994; Laurence & Margolis 2001 を、言語獲得の経験主義に関しては Cowie 1999; 2008 を参照。

<sup>10</sup> この主張を疑う理由は様々なものが考えられる。まず、普遍文法の規則は素人には極めて難解であるため、すべての話者が聞いても理解できないような規則を表象しているとは考えにくい。また、文法規則は外界のあり方と対応するようなものではないため、表象の対象としてふさわしくないとも考えられる。

<sup>11</sup> デイヴィッドソンは次のように述べている。「もし言語およびその獲得や使用に関するすべてを知りたければ、どんな証拠が関連しうるのかに関してアプリアリな制約は存在しない。しかし、話者が発話によって意味していることについての我々の理解は、話者が行うことの知覚を通じて習得したり選び出すことができるものに、直接的であれ間接的であれ部分的に基づいているという確信を、私はクワインと共有している。どれだけ多くの文法を揺りかごから備えてこようと、特定の言語の言葉が何を意味するかは習得しなければならない」 Davidson 1995, p. 132.



といった事柄を問題にする。クワインは、こうした語りが科学的に厳密なものではありえないと考えた。クワインが分析性の概念に哲学的重要性を認めることを経験主義のドグマとして批判した際に、分析性は不明瞭な概念であり、同義性のように分析性と同じ程度に不明瞭な概念によってしか特徴付けられないと論じた。

ただし、クワインは言語的意味についての科学的な探求を完全に諦めたわけではない。たしかに、クワインは、言語的意味は同一性条件が不明瞭であることを問題視しているのだが、他方で、より明確な同一性条件をもった擬似的な対象としての言語的意味というもの的人工的に導入し、それについての科学的な探求を行うことは可能だとも考えた。例えば、クワインは日常的な意味の概念の代わりに「刺激意味 (stimulus meaning)」の概念を導入した。ある文 $s$ の刺激意味とは、 $s$ の同意を促すであろう刺激の集合（肯定的刺激意味）と不同意を促すであろう刺激の集合（否定的刺激意味）の順序対として定義される<sup>12</sup>。なお、ある刺激が同意を促すであろうということは、その刺激が提示されたとすれば話者は同意したであろう、という反事実的条件法によって定義される<sup>13</sup>。

文の刺激意味は文の真理条件とは別である。例えば「これはウサギだ」という文は目の前にウサギがいるときそのときに限って真だが、本物そっくりのウサギの模造品を目にした人はこの文を肯定するだろうし、典型的なウサギの姿には似ていないウサギを目にした人はこの文を否定するだろう<sup>14</sup>。また、刺激意味は話者に相対的ですからある。例えば、人によって誰が独身者であるかに関する信念が異なるので、「彼は独身者だ」という文の刺激意味は話者によって異なる。このように、同一の言語共同体に属するメンバーの間で同一の文が同一の刺激意味をもつわけでもないはずである。

だが、こうした違いにも関わらず、刺激意味を真理条件的な意味に近づけていくことがある程度まで可能である。例えば、話者によって「彼は独身者だ」の刺激意味が異なるとしても、言語共同体のどのメンバーに関しても「彼は独身者だ」の刺激意味は「彼は非婚者だ」の刺激意味と同一である。そこで、クワインはこれら二つの文は社会的に刺激同義であると言う。社会的な刺激同義性は、クワインがその不明瞭さを厳しく批判した同義性の概念に似ている。また、言語共同体のすべての話者によって常に同意されるような文もあろう。クワインはそうした文を社会的に刺激分析的と呼ぶ。「雪は白い」のような文も社会的に刺激分析的に分類されてしまうという限界もあるが、社会的な刺激分析性は、クワインがその不明瞭さを批判した分析性と似ている。かくして、刺激意味は意味の単なる代用品ではなく、「行動主義的な代用品 (behavioristic ersatz)」と言われる<sup>15</sup>。意味の概念は不明瞭であるとするクワインにとって、言語表現の意味を科学的に研究するためにできる

<sup>12</sup> ここでクワインは、現実世界では生じないが生じたかもしれない刺激も刺激意味の構成要素に含めるために、刺激を個別の出来事ではなく普遍 (universal) として扱っている。Quine 1960, p. 34. クワインが普遍に関しては唯名論者であることを考慮すると、この措置は少し意外にも思えるのだが、ここでのポイントは普遍の存在にコミットしても翻訳の不確定性や指示の不確定性が成り立つということにあるのかもしれない。

<sup>13</sup> Quine 1960, p. 33.

<sup>14</sup> Quine 1960, p. 31.

<sup>15</sup> Quine 1960, p. 66.

ことは、せいぜい文の刺激意味を特定することまでだ、というわけである<sup>16</sup>。

本節では、言語的意味や言語習熟に関するクワインの行動主義的な立場について解説した。次節では、クワインの行動主義を仮定した場合には、なぜ根源的翻訳に携わる複数の言語学者が複数の互いに競合する翻訳マニュアルにたどり着いてしまうという可能性を原理的に排除できないのかを検討しよう。

#### 4.2.2 節 翻訳の不確定性の論証

根源的翻訳の思考実験において、フィールド言語学者はある未知の言語を日本語へと翻訳しようと試みる。未知の言語への入口は、その発話が特定の刺激と結びついて、しかも現地語の話者すべての反応が一致しているような文にちがいない。クワインはこうした文を「観察文 (observation sentence)」と呼ぶ<sup>17</sup>。例えば、狩りの場面においてウサギに出くわした現地人は

ガヴァガイ！

と言うとする。言語学者は、この文は日本語の一語文「ウサギ！」ないしは「ウサギがいる」に対応するのではないかと推測する<sup>18</sup>。そして、この推測が正しいかどうかを確かめるため、実際に様々な状況の中でその現地語の文を使ってみる。ウサギが見えるところで「ガヴァガイ！」と言い、それに対して現地人の同意が得られ、また、ウサギのいない場所でも同じように発話し、それに対しては不同意が認められたならば、推測の信憑性は高くなるといった具合である。この要領で、言語学者は現地語の情報提供者から様々なデータを引き出しつつ翻訳作業をつづけていく。

すでに述べたように、クワインは、正しい翻訳は意味を保存するという通念を意味概念の不明瞭さを論拠にして退けている。ただし、観察文の翻訳に関しては、この通念が擬似的に成立する。すなわち、観察文の正しい翻訳とは刺激意味が同じ文へと写すことである、と。

しかし、観察文は現地語の全体からすれば小さな言語断片に過ぎない。フィールド言語学者は観察文以外の現地語の文も日本語へと翻訳しなければならない。問題は、上の方法で翻訳できるのは観察文だけだということである。例えば、未知の社会のメンバー全員が刺激を与えられなくても同意（あるいは、不同意）するような種類の文があるはずだ。こういった種類の文を、クワインは「定常文 (standing sentences)」と呼ぶ。定常文は特定の刺

<sup>16</sup> Gibson 2004 は「科学的意味論 (scientific semantics)」と表現している。

<sup>17</sup> クワインは、刺激に応じて同意・不同意が異なる文を、一般に「場面文 (occasion sentence)」と呼ぶ。観察文は場面文の一種だが、観察文でない場面文もある。クワインの例は「独身者！」である。これが観察文でないのは、誰が独身者であるかについての知識・信念が人それぞれだからである。

<sup>18</sup> ここで「ウサギ！」と訳した表現は ‘Rabbit!’ である。クワインは以下で述べる理由のために、一語文 ‘Rabbit!’ を単語 ‘rabbit’ から厳格に区別するのだが、日本語でこの区別をつけるのはやや難しい。

激と結びついていないため、翻訳作業にとってきわめて厄介である。

常に同意（あるいは、不同意）されてしまう定常文の翻訳を進めるには、文の内部構造に立ち入った分析を行わなければならない。つまり、現地語の単語が日本語のどの単語に相当するのかが推測しなければならない。いくつかの論理語、特に、真理関数的結合子については翻訳を確定させられる見込みがあることをクワインは認める。例えば、現地語のある表現 $e$ が日本語の否定「ない」に翻訳されるのは、現地語の任意の平叙文 $s$ に $e$ を付加してできる文の刺激意味が、 $s$ の肯定を促す刺激と否定を促す刺激をちょうど逆にしたペアであるときかつそのときに限る、などとすればよい<sup>19</sup>。

しかし、クワインは論理外の表現に関しては困難が生じると言う。例えば、先ほど「ガヴァガイ！」を「ウサギ！」という一語文、ないしは「ウサギがいる」と訳するのが適切だとみなした。これらの文の刺激意味は同一だと推測できるからである。ここで「ガヴァガイ」は現地語の単語であると仮定しよう。たとえそうだととしても、「ガヴァガイ」と日本語の語「ウサギ」が外延を等しくするとは限らない。なぜなら、「ガヴァガイ」と「ウサギ」は同じ種類の語ではないかもしれないからである。どういうことか。

論理学では一般に、単語は単称名と一般名（述語）に分類される。「ウサギ」や「テーブル」は分割された指示対象をもつ述語である。これに対して、「雪」や「水」といった物質名詞は指示対象を分割しない。「赤さ」や「球形性」のような抽象名詞も指示対象を分割しない。これらは単称名の一種とみなすことができよう。

さて、「ガヴァガイ」が一つの語だとしても、それが単称名なのか一般名なのかはただちに明らかでない、とクワインは言う。「ガヴァガイ」は、ウサギどもからなる不連続の時空領域（rabbit fusion）を指す単称名かもしれないし、ウサギ性（rabbithood）という普遍を指示する抽象名詞かもしれない。「ガヴァガイ」が一般名だとしても、それがどんな外延をもつのかは判然としない。語「ガヴァガイ」が適用される対象は、個体としてのウサギかもしれないが、ウサギの分離されていない諸部分（undetached parts of rabbit）を指示するという可能性もある<sup>20</sup>。「ガヴァガイ！」の刺激意味はこれらの互いに両立しない可能性すべてに対して開かれているとして、クワインは次のように述べる。

一般名と単称名の区別ですら、刺激意味から独立である。[...] 一般名と単称名の区別と同様、具体的対象と抽象的对象という区別も、刺激意味から独立である<sup>21</sup>。

以上を踏まえて「ガヴァガイ」の翻訳である日本語表現の候補を表にすると、以下のようになる<sup>22</sup>。

<sup>19</sup> 連言や選言の翻訳はこれほど単純ではないが、その問題はここでは無視する。cf. Quine 1969a, pp. 103–104.

<sup>20</sup> Quine 1960, pp. 51–52. ウサギ部分とはウサギの耳や足などである。

<sup>21</sup> Quine 1960, p. 52.

<sup>22</sup> このリストは網羅的ではない。例えば、クワインは「ウサギ段階」という述語も訳語の候補として検討しているが、現代形而上学における三次元主義と四次元主義の対立にまで議論が不必要に拡散してしまう

「ガヴァガイ」の訳語の候補	
単称名	「ウサギ融合体」、「ウサギ性」
一般名（述語）	「ウサギ」、「分離されていないウサギ部分」

クワインはここで、実際に根源的翻訳をする状況におかれたならば、誰もが「ガヴァガイ」をとりあえず「ウサギ」と翻訳するだろう、ということまで否定してはいるわけではない。我々はおそらく暗黙のうちに「持続的で相対的に一様な対象は、短めの表現に対応する」という格率を採用している<sup>23</sup>。しかし、クワインは、この格率はせいぜい翻訳する側からの投影の産物であって客観的な根拠はない、と主張する。実際、ウサギを直示しようとしても、それは同時に、ウサギの分離されていない部分でもあるし、あるいはウサギ性という普遍の例化であるかもしれない。このようにして複数の翻訳マニュアルが得られる可能性は、どうやっても排除できないというわけである。

ここで、単語の翻訳は多様になされうるというクワインの主張は、個々の発話とその刺激意味を孤立させて扱うことに由来しているのではないか、という疑問が生じるかもしれない。発話の中に現れる語句は他の発話の中にも現れるということを考慮するならば、翻訳の候補は自ずと制限されるだろう。例えば、ウサギの耳と足を指差しながら「これとこれは同じガヴァガイか？」と現地人に尋ねて同意が得られたならば、この観察は「ガヴァガイ」を「分離されていないウサギ部分」と翻訳する可能性を排除するだろう。このようにして、翻訳作業の初期段階における誤りは、後であぶりだせるはずである。ちょうどクロスワードパズルが最終的には唯一の解答を許容するように、翻訳も最終的には唯一に定まると期待できるのではないか<sup>24</sup>。

しかし、この反論は上手くいかないだろう。たしかに、クワインは現地語の表現中から対象の個別化に関わる文法的な基準、クワインの言い方では「対象化を行うような装置の全体（冠詞、代名詞、さらに同一性、複数性、述定、あるいは正準的記法における量化、といった語法）」を見つけ出すことができれば<sup>25</sup>、ある語が単称名なのか述語なのか、述語である場合にはどのように指示対象を分割するのが判明するということを認める。例えば、ある表現Fが「三つのFがある（There are three Fs）」とか「あのFは昨日我々が見たFと同一である（That F is the same as the one we saw yesterday）」といった文脈で現れるならば、Fを物質名詞や抽象名詞と解釈することはできない。問題は、個別化に関わる文法的な基準を手に入れられる保証がないことである。ウサギの耳と足を指差しながら「これとこれは同じガヴァガイか？」といった質問ができるためには、「同一である」の正しい現地語訳が確定している必要がある。しかし、「同一である」と訳した現地語の表現は「ガヴァガイ」

恐れがあるため、ここでは「ウサギ段階」を除外した。

<sup>23</sup> Quine 1969a, p. 34.

<sup>24</sup> cf. Blackburn 1984, p. 285.

<sup>25</sup> Quine 1960, p. 53.

を「切り離されていないウサギ部分」と訳す翻訳マニュアルでは「共にある (belong with)」に対応させているかもしれない<sup>26</sup>。クワインは次のように述べる。

〔表現上の〕装置全体は相互依存的であり、名辞という概念は、関連する装置と同様、我々の文化に特有である。現地人は非常に異なった言語構造を通じて実質的に同じ効果を挙げるかもしれないので、我々の装置を現地語でたまたまうまく解釈したり、あるいはその逆をすると、そうした解釈は不自然で大幅に恣意的なものであることがわかるかもしれない<sup>27</sup>。

したがって、翻訳マニュアルに対して体系的な調整を行うならば「ウサギ」以外の訳語を割り当てられる可能性は残りつづけるように思われる。クロスワードパズルの正解が一つしかないということを証明するのは容易ではないのだ<sup>28</sup>。

ここまでの議論が正しければ、現地語に対して二つ以上の競合する翻訳マニュアルがあることになる。つまり、一つの翻訳マニュアルによれば「ガヴァガイ」は「ウサギ」と翻訳され、別の翻訳マニュアルによれば「ガヴァガイ」は「分離されていないウサギ部分」と翻訳される。翻訳の正しさを決定する事実をすべて考慮した上で、なおこのような決定不全性残る以上、どれが正しい翻訳マニュアルなのかを決定するような事実は存在しない。それゆえ

(A) 日本語の任意の表現「 $\alpha$ 」に関して、「ガヴァガイ」は表現「 $\alpha$ 」に翻訳されるという主張は真理を表現しない<sup>29</sup>。

こうして翻訳の不確定性の一例が手に入る。

ところで、根源的翻訳の思考実験は遠い未知の言語にだけ関係しているわけではない、という点にクワインは注意を促している<sup>30</sup>。「ガヴァガイ」に関する翻訳の不確定性と類比的なケースは、自国語においても原理的には生じうる。自国語の翻訳に関してここまで議論してきた類の不確定性が表面化しないのは、同音 (homophonic) の翻訳マニュアルを使用するのが便利であるというだけのことである。同音翻訳を意図的に止めるならば、不確定性が表面化しうる。もちろん、同音翻訳に大きな効用があることは否定できない。言語を習得する上で大人の発話を模倣することは有効であり、同音翻訳からむやみに逸脱することは混乱を招きやすい。しかし、相手の発話のうちに特異性があるときなど、同音翻訳を諦める場合がないわけではない。クワインは同音翻訳を自国語の翻訳マニュアルが採用すべき必須の制約とは考えない。かくして

<sup>26</sup> Quine 1969a, p. 33. cf. Quine 1960, p. 72.

<sup>27</sup> Quine 1960, p. 53.

<sup>28</sup> cf. 丹治 2009, p. 118.

<sup>29</sup> cf. Soames 2003, p. 260.

<sup>30</sup> Quine 1969a, p. 46.

(B) 日本語の任意の表現「 $\alpha$ 」に関して、「ウサギ」は表現「 $\alpha$ 」に翻訳されると  
いう主張は真理を表現しない。

翻訳の不確定性は自国語でも再生産される。

厳密に言えば、(A) と (B) は翻訳の不確定性の一例に過ぎない。しかし、ここまでの議論は「ガヴァガイ」や「ウサギ」にのみ当てはまる事柄には依拠していない。よって、もしここまでの議論が成り立つならば、(A) と (B) を論理外の任意の言語表現へと一般化することも許されるだろう。こうして翻訳の不確定テーゼが帰結する。

#### 4.2.3 節 翻訳の不確定性をどう評価するか

『ことばと対象』が出版されて以来、翻訳の不確定性テーゼは多くの疑問にさらされた<sup>31</sup>。しかし、すべての論者が翻訳の不確定性テーゼに対して懐疑的というわけではない。例えば、4.1 節で引用したように、エヴァンズは翻訳の不確定性をトリビアルな事実として問題視しない<sup>32</sup>。いま私もエヴァンズとほぼ同じスタンスを採ろうとしている。なぜ翻訳の不確定性に私が同意するのか、手短かに述べておこう。

翻訳の不確定性というパラドキシカルな考えを前にしてひるまないためには、「翻訳」という作業に関して一種の冷めた態度をとる必要がある。「ガヴァガイ」に対して「ウサギ」と「切り離されていないウサギ部分」という二つの訳語のどちらを使っても現地人の言語行動を整合的に記述することができるならば、どちらを訳語に選ぼうと、現地人と対話する際に同じくらいの成果を挙げることができるのだから、それでよい、といった見解である。例えば、「ガヴァガイ」を含む現地のある命令文を「ウサギを焼こう」と「切り離されていないウサギ部分を焼こう」のどちらに翻訳しようと、その現地人の命令を実行するためには結局のところ同じような行動をすればよい。そうすることで現地人を満足させられるなら、現地人との対話は十分に円滑であり、二つの翻訳マニュアルは同じ程度に上手くいっている、と言ってよいのではないだろうか。

私はすべての人が「翻訳」という語をこのような意味で理解すべきだとはまでは主張しない。上のような仕方で翻訳や対話の円滑さといった観念について理解することも可能だ、

<sup>31</sup> 本章で触れる翻訳の不確定性の議論は、『ことばと対象』2章に由来する「下からの論証 (argument from below)」に依拠している。下からの論証は、不確定性の具体例を与えているという意味で構成的である。しかし、クワインは論文「翻訳の不確定性の諸理由について」で新たな論証として「上からの論証 (argument from above)」を提示している (Quine 1970)。大まかに言うと、上からの論証は、翻訳マニュアルは一切の物理的事実を考慮したとしても一意には定まらなると論じるもので、下からの論証よりも強い結論を導いている。しかし、私自身は、上からの論証がそれほど興味深いのかについて懐疑的である。クワインにとって、翻訳の際に利用すべき証拠は行動上の証拠ですべてが網羅されているのだから、それ以外の物理的事実は証拠にはならないということは容易に想像がつく。cf. Soames 2003, pp. 243–258.

<sup>32</sup> 鈴木 2000, p. 121 はエヴァンズの論文に触れつつ、エヴァンズが翻訳の不確定性を否定していると解釈している。だが、以下で詳細に論じるように、私はこの解釈には同意できない。

というのがここでのポイントである。また、私はクワインが意味に関して主張したことにすべて同意しているわけでもない。クワインの結論の中には明らかに受け入れがたい要素が含まれている。ただし、その受け入れがたい要素が何であるのかを特定するには、翻訳の不確定性よりもむしろ指示の不確定性へと戦場を移すのがよいと思う。

#### 4.3 節 指示の不確定性と意味理論

一見すると、前節で提示した翻訳の不確定性テーゼの論証は、意味理論における不確定性をも含意する。(B) が示すように翻訳の不確性が自国語でも再生産されうるとすれば、次の主張も導くことができるように思われる。

(C) 日本語の任意の表現「 $\alpha$ 」に関して、「ウサギ」は $\alpha$ を指示するという主張は真理を表現しない<sup>33</sup>。

厳密に言えば (C) は指示の不確定性の一例に過ぎない。しかし、前節の議論は「ウサギ」にのみ当てはまる事柄には依拠していなかった。もし前節の議論が成り立つならば、(C) を論理外の任意の言語表現へと一般化することも許されるだろう。そこで、(C) を一般化すると、任意の言語表現に関して、その表現の指示対象を特定するような意味論的仮説は真ではありえないことになる。こうして、指示の不確定性が、つまり、意味理論の不確定性が導かれる。この論証は前節の冒頭で提示した論証とパラレルに定式化することができる。

意味理論を作る際に利用できる証拠は言語行動の傾向性に限られる (クワインの行動主義 $\beta$ )<sup>34</sup>。

もし意味理論を作る際に利用できる証拠が言語行動の傾向性に限られるならば、互いに競合する複数の意味理論が存在する。

よって、互いに競合する複数の意味理論が存在する (指示の不確定性テーゼ)。

この論証は妥当であるから、結論を受け入れたくなければ、二つの前提のどちらかを否定しなければならない。では、どちらを否定すべきだろうか。

<sup>33</sup> cf. Soames 2003, p. 261.

<sup>34</sup> 一ノ瀬正樹教授から、言語行動の傾向性を同定するには意味理解が先取りされていなければならないはずであり、一番目の前提では循環が生じているのではないか、という疑問をいただいた。「言語行動の傾向性」を、意味理解を先取りしなければ同定できない性質と解釈するならその通りである。しかし、4.2.1 節で論じたように、ここでいう「言語行動の傾向性」は、物理的刺激とともに提示された言語形式に対して肯定ないし否定という行動をする傾向性の全体と解釈している。言語形式は意味抜きで言語表現のことであり、意味は前提されていない。そもそも、「言語行動の傾向性」をこのように貧弱なものとして解するからこそ、クワインの行動主義には問題があると論じる余地が生じる。

以下では、互いに競合する複数の意味理論が存在するというクワインに反対して、利用できる証拠から、意味理論は一つに定まるというエヴァンズの議論を参照する（4.3.1 節）。私はエヴァンズの議論は説得的だと考える。ただし、エヴァンズは上のような仕方でクワインの論証を定式化しているわけではないので、もし二つの前提のどちらを否定するのかと尋ねられたらエヴァンズが何と言うかは解釈を要する。私はエヴァンズの論文に対するフックウェイとミラーのコメントを参照することで、エヴァンズは一番目の前提を拒否するであろうと解釈する（4.3.2 節）。

#### 4.3.1 節 エヴァンズの議論

エヴァンズによれば、クワインの議論は翻訳マニュアルではなく意味理論における不確定性を含意するときになってはじめて哲学的な意義を獲得する。クワインはこの含意を示すことに失敗している、というのがエヴァンズの診断である。

まず、翻訳マニュアルと意味理論の違いを明確にしよう。翻訳マニュアルは二つの言語に属する表現同士の対応をつけるだけなので、表現が何を意味しているのかを教えるものではない。例えば、「ドイツ語の“schnee”は英語の“snow”である」という文を知るために、英語の“snow”が何を意味するのかを知っている必要はない。この文は、ドイツ語のある語が英語のある語に対応するとしか言っていないからである。これに対して、意味理論は問題となっている言語（ないし言語断片）に属する表現が何を意味するのかを特定することが意図されている。例えば、エヴァンズが取り上げるような真理条件的な意味理論の場合、それは問題となっている言語に属するすべての文に対して、その真理条件が何であることを述べる定理を帰結する。そのために、文より小さな単位の表現、つまり名前や述語に対しては、その指示対象や適用条件を割り当てる。翻訳マニュアルと違って、こうした意味理論では指示や適用や真理といった意味論的概念が用いられる。

4.2.2 節の議論を踏まえると、日本語の「ウサギ」には四つの双条件のいずれかを公理として含む日本語断片の意味理論がありうる。

- (1)  $\forall x$  [「ウサギ」は  $x$  を指示する  $\leftrightarrow x$  = ウサギ融合体]
- (2)  $\forall x$  [「ウサギ」は  $x$  を指示する  $\leftrightarrow x$  = ウサギ性]
- (3)  $\forall x$  [「ウサギ」は  $x$  に適用される  $\leftrightarrow x$  は分離されていないウサギ部分である]
- (4)  $\forall x$  [「ウサギ」は  $x$  に適用される  $\leftrightarrow x$  はウサギである]

クワインによれば、これらの内のどの理論が適切な意味理論であるかを決定するような事実は存在しない。これに対して、エヴァンズは、(4) 以外の双条件を公理とする意味理論は日本語断片の適切な真理条件を与えない、と主張する。たしかに、対象の個別化に関わる表現装置（同一性や冠詞など）の翻訳を調整すれば、「ガヴァガイ」や「ウサギ」が単称



名なのか述語なのかも定まらない、とクワインは述べていた（4.2.2 節を参照）。しかし、エヴァンズによれば、「ウサギ」の指示が何であるかを定めるという目的にとって個別化に関わる表現装置は二次的な道具に過ぎず、別の方法を使えば、たかだか有限個の文しか持たない貧弱な言語であっても語の指示は確定しうる。

エヴァンズが考察しているのは次のような言語である<sup>35</sup>。まず、この言語に物体の種類をあらわすと考えられる有限個の語彙（G語）がある。G語の具体例は「ウサギ」、「ヒト」、「小屋」などである。G語とは別に、物体の種類とは関係なく環境の何らかの特徴をあらわすと考えられる有限個の語彙（F語）もある。F語の具体例は「白い」、「温かい」、「フワフワしている」などである。さらに否定表現「ない」もある。これら三種類の原始的な語彙に加えて、F語とG語と否定表現を連結した複合表現を形成することもできる。その際には、否定表現のスコープの広さも区別する。つまり、「(F G) でない」と「(Fでない) G」は別の言語表現として取り扱う。そして、この日本語断片に属する表現 $\phi$ は「 $\phi$ !」という主張、または「 $\phi$ ?」という疑問文で使用される。例えば、「ウサギ!」という主張は近くにウサギがいるときになされ、「ウサギ?」という疑問は近くにウサギがいるときになされたならば、聞き手から同意を引き出せるだろう。

エヴァンズは、以上の想定のもとで（1）から（3）はこの日本語断片の意味理論の公理として適切でないと主張する。各々のケースを順に見ていこう。

#### （1）ウサギ融合体として解釈

この解釈のもとで「白いウサギ!」の刺激意味を適切に捉えるためには、「白い」の適用条件は

$$\forall x \text{ [「白い」は } x \text{ に適用される} \leftrightarrow x \text{ の適切な部分は白い}]$$

となるはずである。問題は「適切な部分」がどの部分かである。単なる部分ではありえない。なぜなら、しっぽが白い茶色いウサギを目にした日本語話者は「白いウサギ!」には同意しないからである。ところが、白いウサギである部分と考えるわけにはいかない。そうしてしまうと、「白い犬!」のケースで上手くいかない。

#### （2）ウサギ性として解釈

この解釈のもとで「白いウサギ!」の刺激意味が適切に捉えるためには、「白い」の適用条件は

$$\forall x \text{ [「白い」は } x \text{ に適用される} \leftrightarrow x \text{ は白い事例をもつ}]$$

---

<sup>35</sup> Evans 1975, pp. 33–34.

となるはずである。そうすると「白くないウサギ！」が真であるのは、ウサギ性が白い事例をもつのではないときそのときに限られる。これは、ウサギ性が白い事例を持たないときである。しかし、「白くないウサギ！」という部分否定がほしいのは、むしろウサギ性が白でない事例をもつということでなければならないはずである。

### (3) 分離されていないウサギ部分として解釈

この解釈もまた、「白い」の適用条件をどうすればいいのかという問題に直面する。「白い」の適用条件を

$$\forall x \text{ [「白い」は } x \text{ に適用される} \leftrightarrow x \text{ は白い}]$$

とすると、「白いウサギ！」はしっぽだけが白い茶色のうさぎがいるときにも真になってしまう。そういう状況では、話者は「白いウサギ？」を否定するだろう。他方で

$$\forall x \text{ [「白い」は } x \text{ に適用される} \leftrightarrow x \text{ は白い動物の部分である}]$$

とすると、「白いハンカチ！」のような文の刺激意味を正しく捉えられない。白いハンカチを見せられた話者は「白いハンカチ？」に同意するが、そういう状況では「白いハンカチ！」は偽になってしまう。

こうして (1) から (3) の公理はすべて退けられる。

しかし、エヴァンズの議論を追っていると、「ウサギ」や「ガヴァガイ」を翻訳するときには不確定性がつきまとうのに、なぜ意味理論に関してはこれほどあっさりと「ウサギ」の指示が確定するのだろうか、と思わずにはいられない。翻訳マニュアルを作成するときには使えなかった何らかの手がかりを、エヴァンズは意味理論を作る際には利用しているのではないだろうか。この疑問を掘り下げてみよう。

エヴァンズは翻訳マニュアルが多様な選択肢に開かれているということを、自国語翻訳を例にとって次のように説明している。ある翻訳家が、日本語の「ウサギ」を「ウサギ融合体」に対応づけるとする。整合性を守るために、彼女は、「白い」を「白い」に対応づけつつ、「白い」^「ウサギ」を「白いウサギ融合体」に対応づけようとするだろう。「ウサギ」を「ウサギ性」に対応づける別の翻訳家は、「白い」を「白い」に対応づけつつ、「白い」^「ウサギ」を「白いウサギ性」に対応づける<sup>36</sup>。そして、「ウサギ」を「ウサギの分離されていない部分」に対応づけるさらに別の翻訳家は、「白い」を「白い」に対応づけつつ、「白い」^「ウサギ」を「白いウサギの分離されていない部分」に対応させることができる。これに対して、意味理論の場合には上で見たように、翻訳マニュアルのように多様な選択

<sup>36</sup> 抽象名詞をつくるオペレータ (性-hood) のスコープを明示するなら「(白いウサギ) 性」である。

肢に向かって開かれてなどいない。

#### 4.3.2 節 鏡映制約と話者の言語能力

エヴァンズの論文に対するコメントで、意味理論が翻訳マニュアルのように多様な選択肢に開かれていないということは明らかではない、とフックウェイは論じている<sup>37</sup>。フックウェイは

(3)  $\forall x$  [「ウサギ」は  $x$  に適用される  $\leftrightarrow x$  は分離されていないウサギ部分である]

という公理を守るために、「白い」に関する公理を次のように手直しすることもできる、と論じる。

$\forall x$  [「白い」は  $x$  に適用される  $\leftrightarrow$  「白い」が「ウサギ」と並んで出現しているならば  $x$  は白い動物の部分である、それ以外の文脈では  $x$  は白い]

「白い」の公理をこのように手直しするならば、エヴァンズが考慮した二つのケース、つまり、しっぽだけが白い茶色のうさぎがいるときや、白いハンカチがあるときなどは適切に処理することができる。

しかし、おそらく、エヴァンズは「白い」に関するフックウェイの公理を認めないだろう。意味理論には翻訳マニュアルにはない制約があると考えており、フックウェイの公理は適切な意味理論ならば採用できないような形をしている、というのがその理由である。適切な意味理論はいくつかの制約を満たさなければならない。例えば、意味理論の公理は有限個でなければならない、というのはそうした制約の一つである。もっとも、上で提示したエヴァンズの言語は整式がたかだか有限個の言語なので、この場合には有限性の制約は容易に満たされる。適切な意味理論に課せられた、フックウェイの公理を排除するべき制約は他にあるはずである。それは何か。

エヴァンズによる (1) から (3) の可能性の排除は、「白い」と「ウサギ」に割り当てられた意味論的性質から、「白いウサギ」の意味論的性質を割り出せないことを根拠としている。「白いウサギ」に対して、日本語断片の意味理論は、直接に意味論的性質を割り当てべきではないという考慮がここでは働いている。その理由は、日本語の表現が構成要素である語への分割を許すならば、その複合的な表現の意味は構成要素の意味によって与えられるべきだと考えられるからである。ある言語の意味理論は、たんにその言語に属するすべての文に対して意味（例えば、真理条件）を与えればよいというものではない。それだけであれば、任意の言語の同音的な意味理論は代入量化を用いて一言で述べることができ

---

<sup>37</sup> Hookway 1988, sec. 9.3.

る。すなわち

$$\prod p \text{ [「} p \text{」は真である} \leftrightarrow p]$$

である。しかし、この種の意味理論は意味の合成性を捉えていないために、望ましい意味理論とは言えない。

意味の合成性はなぜ重要なのか。その理由は、おそらく、意味理論は現実の言語話者の能力（competence）を構成している心理的メカニズムを説明しなければならないと考えられるからである<sup>38</sup>。現実の言語話者はまだ見聞きしたことのない文をそれまでに見聞きした文の理解を基礎にして理解することができる。代入量化を用いた上のような文では、話者の言語能力を説明することなどできそうにない。適切な意味理論はむしろ、複合表現の意味を、その表現の部分表現の意味に基づいて与えるのでなければならない。したがって、意味の合成性は言語表現の理解から切り離して定式化することはできない。実際、合成原理の標準的な定式化は次のようになされる。

**鏡映制約（Mirror Constraint）：**文Sの真理条件を特定する定理がSの部分文表現 $s_1, s_2, \dots, s_n$ の意味論的性質を特定することから導出可能であるのは、その言語の現実の話者がS以外の文において出現した $s_1, s_2, \dots, s_n$ に接して訓練を受けたことでSを理解できるようになるときそのときに限る<sup>39</sup>。

鏡映制約を満たす意味理論の諸公理は、原始的な表現の意味を与える。そこで、もし話者が鏡映制約を満たす意味理論の諸公理を暗黙に知っているとすれば、この話者はそれまでに見聞きした文の理解を基礎としながら、まだ見聞きしたことのない文を理解できるということにも説明がつく。

意味の合成性について考察したところで、フックウェイの公理に戻ろう。「白いハンカチ！」と「ウサギ！」という発話を理解する言語話者は、「白いウサギ！」という発話を耳にしたことがなくても、仮にこの発話がなされたとすれば理解できると思われる。よって、鏡映制約を満たす意味理論では、「白いウサギ」の意味論的性質は「白い」と「ウサギ」の意味論的性質から導かれなければならない。しかし、「白い」に関するフックウェイの公理はそうっていない。「ハンカチ！」「白いハンカチ！」「ウサギ！」しか聞いたことのない人物が、なぜ「白いウサギ！」を理解できるのかを説明できない。エヴァンズの観点からすれば、フックウェイの公理はまだ見聞きしたことのない文を理解できるという話者の能力を説明できず、それゆえ否定される。

鏡映制約は意味理論の制約として、もっともらしい。しかし、意味理論に鏡映制約を課

<sup>38</sup> Evans 1975, pp. 25–26. cf. Evans 1981.

<sup>39</sup> Miller 2007, p. 151.

すということは、クワインの行動主義 $\beta$ からは帰結しないし、むしろクワインの行動主義 $\beta$ とは両立しないだろう。問題は、意味理論の公理に関する暗黙的知識とは何かという点にある。複合表現の意味理解は部分表現の意味理解に依存しており、究極的には原始的な表現の意味理解に依存している。こうした原始的な表現の意味理解が、意味理論の公理の暗黙的知識である。エヴァンズの考えでは、ここでいう暗黙的知識は何らかの神経生理学的な状態として実現されている機能的性質である<sup>40</sup>。どういうことか。ある意味理論の公理系から一つの公理を取り除いたら、証明できなくなる定理があるだろう。ちょうどそれと平行に、意味理論の公理の暗黙的知識を実現している脳状態に損傷が生じたら、証明できなくなる定理に相当する複合表現の意味理解に障害が生じる、といったケースを想像すればよい<sup>41</sup>。もちろん、現段階の神経生理学はそのような暗黙的知識の実現者を同定するところまで到達していないが、これはあくまで原理的な話である。言語表現の意味理解に関するエヴァンズの描像は、話者の行動だけでなく神経生理学にまで踏み込むことを要求しており、クワインの行動主義 $\beta$ と両立しない。

しかし、だからといってクワインの行動主義  $\beta$  が正しくてエヴァンズは論点先取を犯しているということにはならない。そもそも、なぜそこまでして意味理論をつくる上で利用可能な証拠を行動の傾向性に限定しなければならないのかは判然としない。クワインの行動主義を支持する積極的な理由は何だろうか。さしあたり二つの可能性が考えられるが、どちらもあまり説得的ではない。

クワインによれば、心理学において人は行動主義者であつたりなかつたりしうが、言語哲学において行動主義は選択の余地のない義務である。われわれは他人の言語行動を観察し、自身のまごつく言語行動を他人に観察されながら言語を獲得する。それゆえ、観察可能な状況における明示的な振舞いから取り出される事柄を超えた言語的意味など存在しない、とクワインは言う<sup>42</sup>。しかし、この議論は次のように批判されうる<sup>43</sup>。われわれは電子に関する事実を観察可能な日常的対象の振舞いから学ぶしかないが、そのことは電子に関する事実が観察可能な事実を超えていないということは示していない。それゆえ、われわれが他人の言語行動を観察しながら言語を学ぶということは、言語的意味に関する事実が刺激と行動を超えていないことを示す証拠としては弱すぎる。クワインは、言語的意味に関する事実が神経科学によって解明されうるような類の事実にも依存しているという（エヴァンズが示唆しているような）可能性を否定できていない。

丹治は、言葉の意味を心の中のイメージと同一視する素朴な心理主義的言語観を退ける

---

<sup>40</sup> エヴァンズの見解は意味理論家の間で完全に同意を得られているわけではない。例えば、デイヴィッドソンにとって、意味理論はそれを用いれば話者を合理的に解釈するのに十分な内容を持っていれば適切であり、現実の話者が意味理論の公理について暗黙的知識を持っていることまでは要求されない。Davidson 1973. cf. 飯田 2002, 2.4 節。

<sup>41</sup> Blackburn 1984, p. 32.

<sup>42</sup> Quine 1987, p. 5. cf. Quine 1960, p. ix.

<sup>43</sup> Friedman 1975, pp. 365–369. cf. Soames 2003, p. 245.

ためには、言語習熟は行動主義的に説明されるほかないと論じている<sup>44</sup>。しかし、ありうる選択肢は素朴な心理主義と行動主義の二択だけではないだろう。鏡映制約を意味理論の制約として重視するエヴァンズの立場を第三の選択肢とみなしてもよいはずだ。

以上の議論が正しければ、クワインの行動主義  $\beta$  を受け入れる積極的な理由はない。クワインの行動主義  $\beta$  は指示の不確定性テーゼの論証の前提なので、指示の不確定性テーゼを受け入れる積極的な理由もない。

次節ではさらに、指示の不確定性テーゼは不合理な帰結をもたらすと論じる。つまり、指示の不確定性テーゼを否定する積極的な理由がある。それは、クワインの行動主義を否定する積極的な理由も与えることになる。

#### 4.4 節 指示の不確定性と命題的態度の帰属

4.1 節で述べたように、いくつかの言語表現に関しては、指示対象の割り当てが不確定だったとしてもそれほど驚くべきことではない。4.1 節では曖昧述語を例に挙げたが、もう少し興味深い例もある。フィールドは、科学史に現れる理論語の多くは指示に関して不確定だと主張する<sup>45</sup>。注意しなければならないが、彼は過去の科学者たちの用いた様々な語に関して、その指示対象が何であるかを知ることができない、と言っているわけではない。何を指示していたのかに関する事実問題は存在しないので、知るべき事実がそもそもない、と言う。

フィールドは一例として「質量」を挙げる。ニュートン力学に現れる「質量」という語を含む多くの文、例えば、「質量は、運動エネルギーを二倍したものを速度の二乗で割った値に等しい」などといった文は、特殊相対論では放棄された。その結果、現代物理学では固有質量 (proper mass) と相対論的質量 (relativistic mass) という二つの質量概念が区別されるようになった。この区別を踏まえると、我々は

$N_1$  : ニュートンの「質量」は相対論的質量を指示する。

$N_2$  : ニュートンの「質量」は固有質量を指示する。

という二つの仮説を立てることができる<sup>46</sup>。これらの仮説は両立しない。そして、ニュートン力学が近似的には十分に正確であることを考慮すれば、否定連言 (not- $N_1$  & not- $N_2$ ) は受け入れがたい。そうすると、我々には  $N_1$  か  $N_2$  の二択しか残されていない。しかし、ニュートンに関するどんな心理学的な考察を加えようとも、どちらの仮説が正しいのかを決定

<sup>44</sup> 丹治 2009, pp. 154–156. クワインはこうした素朴な心理主義的言語観を、言葉と意味の関係が展示品とそのラベルの関係に似ていることから「博物館の神話」と呼んでいる。

<sup>45</sup> Field 1973. 丹治も翻訳の不確定性テーゼを理解するための一例として、似たような具体例を提示している。タイムマシンでデモクリトスと話すことができた場合、デモクリトスのいう「アトモン」を「原子」と訳しても、「クォーク」と訳しても対話は円滑に進むかもしれない。丹治 2009, pp. 164–167.

<sup>46</sup> フィールドは「表示する (denote)」と言うが、ここでは「指示する (refer)」で統一する。

するだけの材料など存在しないのであって、ニュートンが本当の意味で用いていた質量概念はどちらなのかと問うのは無駄だと思われる。ここに現れた指示の不確定性は、決着をつけるべき事実問題の存在しない不確定性である。

フィールドの議論の肝は、誤っていることが明らかになった理論に対して翻訳を試みている、というところにあると思われる。デュエムとクワインによる確証の全体論によれば、理論は全体として経験の裁きを受けるため、理論が反証される場合には改訂の仕方は一通りではない。もちろん、改訂の際に理論に含まれる全ての文が同じように改訂の対象となるわけではなく、理論体系が全体として単純になるように考慮すべきであるのは間違いないが、そうした考慮を加えたとしてもなお、改訂のあり方が一通りでないということは、充分ありそうなことである。この場合、指示に関する不確定性が生じるのは、以上の例が示すように避けられないかもしれない。とはいえ、この例は同時に、指示の不確定性は実際には大した問題は引き起こさないのではないか、という印象も与えてくれる。現在の理論を用いて過去の誤った理論を解釈する際に、現在の理論によって措定される対象が過去の理論の中には明瞭に見出せないとしても、それは驚くことではない。

しかし、たとえ指示の不確定性という観念それ自体には大きな問題はないとしても、我々が4.3節で示したのは、クワインが正しければ指示の不確定性は言語表現のほぼ全てに対して適用されるということであった。この結論はいろいろな仕方で批判されてきたが、ここでは指示の不確定性テーゼと命題的態度の関係に注目したい。

フックウェイは、指示の不確定性テーゼを受け入れるならば、人々に対して非常に奇妙な心的態度を帰属することになると示唆している<sup>47</sup>。フックウェイの示唆するところを論証として再構成してみよう。まず、ある日本語話者が日本語の文sを誠実に肯定したとする。日本語断片の意味理論はsについて

sは真である  $\leftrightarrow$  p

という定理をもたらすはずだが、意味理論がもたらす定理に限っては

[sは真である  $\leftrightarrow$  p] ならば sはpを意味する

という含意関係を認めることができるだろう。ここで

[話者Sがsを誠実に肯定し、sがpを意味する] ならば、Sはpと信じている

という言語的意味と命題的態度を結びつける含意関係を前提すれば、話者がsを誠実に肯定するということから、その話者に対してpという信念を帰属できるだろう。さて、ここで、

---

<sup>47</sup> Hookway 1988, p. 158. cf. Soames 2003, pp. 264–270.

述語「ウサギ」は分離されていないウサギ部分に適用されると仮定しよう。そうすると、ある人が「ウサギ！」などと言うときには、分離されていないウサギ部分がそこにあるといった信念を帰属することになる。このような信念は、その人が分離されていないウサギ部分に対して知覚的に鋭敏であるといった奇妙な前提を措かない限りは成立しない。したがって、指示の不確定性テーゼは話者に対して心理学的に不適切な信念を帰属することを許容する。話者は、ウサギがそこにいると信じていると言ってもよいし、分離されていないウサギ部分がそこにあると信じていると言ってもよい。二つの信念帰属のどちらが正解ということはない。

クワインの行動主義から信念帰属の不確定性が導かれるというのは興味深い。このことは、言語表現の意味理解と信念という二つの心的状態の間に密接な繋がりがあることを示唆する。

また、ここで述べたような信念帰属の不確定性は、3章の末尾で浮上してきた論点といくらか似ている。個々の信念や欲求に結び付けられた因果役割だけでは、それぞれの信念や欲求の内容を決定できない、というのが因果役割機能主義からの帰結だったからである。しかし、機能主義者とクワインの間にこうした一致がみられるとしても、信念に関する両者の見方は大きく異なる。機能主義者の多くは、人々がどのような内容の信念や欲求を抱いているのかに関して、正解があると考えている。だからこそ、因果役割だけで信念や欲求の内容を個別化できないのであれば別の物理的事実によって個別化できるのではないかと考えて、自然主義的意味論の理論を立案しようとする。これに対し、クワインは、人々がどのような命題的態度を抱いているのかに関して唯一の正解は存在しないと考えている。「ウサギ！」という人に対して、ウサギがそこにいるという信念を帰属して、分離されていないウサギ部分がそこにいるといった信念を帰属しないのは、単にこれまでそうしてきたからであって、原理的には正解などない。

人々が何を信じ、何を欲しているのかに関して正解があると考えてる哲学者は、指示の不確定性テーゼを拒否するべきである。実際、そう考えるのはかなり常識的と言える。人々が何を信じ、何を欲しているのかは、人々の行動を説明する際に言及されるべき重要な指定物だからである<sup>48</sup>。そして、指示の不確定性テーゼを拒否するコストはそれほど高くないように思われる。指示の不確定性テーゼは一つの疑わしい前提に依拠していたからである。人々が何を信じ、何を欲しているのかに関して正解があると考えてる人は、その疑わしい前提、つまり、クワインの行動主義を退ければよい。

#### 4.5 節 本章のまとめ

本章では、クワインが導いた翻訳と指示に関する二つの不確定性テーゼを検討した。これら二つのテーゼはどちらも何らかの三段論法の帰結であり、その前提を根源的翻訳の思考

---

<sup>48</sup> cf. Devitt 1997, pp. 334–335.



実験が補強するという形になっている。二つの不確定性テーゼに関する本稿のスタンスはエヴァンズの影響下にある。すなわち、本稿のスタンスは、翻訳の不確定性に関してはクワインに同意するが（4.2 節）、指示の不確定性に関しては到底受け入れられない、というものである（4.3 節）。指示の不確定性を受け入れられないのは、話者に対して心理学的に不適切な信念を帰属する無数の可能性を許容するからである（4.4 節）。指示の不確定性テーゼの論証は、クワインも認めるように行動主義的な前提に基づいている。したがって、論証の結論を受け入れられない以上は、クワインの行動主義もまた受け入れることもできない。根源的翻訳をめぐるクワインの議論は、意味理論に関する不確定性を導く議論として再構成することで、最終的には行動主義に対する帰謬法となる。

## 5 章 ルイスの解釈理論

### 5.1 節 合理的な解釈

3 章で明らかになった機能主義の限界は次のようなものであった。ある人物の行動を因果的に説明しうる信念・欲求の体系は、原理的には複数存在しうる。その場合、一方の信念・欲求の体系においてある欲求に結び付けられた因果役割と、他方の信念・欲求の体系において別の欲求に結び付けられた因果役割は同一ということになる。別個であるべき心的性質が同一の因果役割と結び付けられるということであるから、心的性質は因果役割によっては個別化できそうにない。より正確に言えば、命題的態度の内容を因果役割によって個別化することはできそうにない。

因果役割機能主義は自然主義者にとって、心的性質とは何かを特徴づけるという課題に対する有望なアプローチであったから、このことは自然主義者には悪い知らせであろう。主体が抱いている命題的態度の内容を決定する物理的事実とはどのようなものか、自然主義者はまだはっきりした答えを手にしていない。ここまでが本稿の第一部で到達した結論であった。

本章では、こうした苦境を抜け出すための一つの手段として、ルイスの解釈理論に注目する<sup>1</sup>。2 章で述べたように、ルイスは同一説の支持者であり、因果役割機能主義を支持しているわけではないのだが、それでも彼は因果役割によって心的性質が個別化されるという考え方を大筋で受け入れている。ルイスは、行動を因果的に説明しうる信念・欲求の体系が原理的には複数存在しうるという問題にどう対処するのだろうか。彼の答えは、一口でいうと、主体が抱いている命題的態度の内容を確定することを可能にする合理性の諸制約があるというものである。つまり、行動を因果的に説明しうる信念・欲求の体系が複数存在しうるというのはいわば幻覚であり、合理性の諸制約を満たす信念・欲求の体系を探すならば、それは一意に定まるであろう、ということである。

この考え方は、それほど反直観的ではない。3.2 節で提示したメアリーに関するシナリオで、アルフレッドが苦痛を感じることの欲求を、アルフレッドに一度も会ったことのないメアリーに帰属するのは倒錯していると述べた。その理由は、一つには、そうした欲求をメアリーに帰属することは、メアリーの行動を解釈する、つまり、メアリーの行動を理解可能なものにするという目的に何の役にも立たないと感じられることにあるだろう。われわれは、メアリーにこうした欲求を帰属するのは不合理だと感じる。こうした不合理な命題的態度帰属を排除するのが合理性の諸制約である。

フォーダーとルポアは、以上で述べたようなルイスのアイデアを、信念や欲求を帰属するという実践は制約充足問題を解くということに他ならない、と表現している。制約充足

---

<sup>1</sup> Lewis 1974; 1986, sec. 1.4, 2.3; 1994.

の問題とは何か。以下のシナリオは具体的で分かりやすい<sup>2</sup>。

例えば、ホームズは次のような制約充足問題を持っていて、モーリアーティがそれに対する最適解だということがあるかもしれない。「犯人は 5 フィート 11 インチより高くはないが、5 フィート 6 インチより低くない。彼は茶色のぶちの入ったダックスフントを飼っていて、その犬の毛はいま盛んに生え変わっている。彼は魚類学に関する実際的知識を持っており、彼の姉はリーズから遠からぬところで生まれた」

ルイスによれば、信念・欲求の帰属にも同じことが当てはまる。ある人の行動を因果的に説明するような信念・欲求の体系は無数にある。しかし、さまざまな合理性の制約を参照することで、その人の行動を最適な仕方では説明する信念・欲求の体系は一つに絞り込まれると言う<sup>3</sup>。

それでは、信念・欲求の体系を絞り込むための諸制約とは具体的に何か。節を改めて検討しよう。

## 5.2 節 制約充足の問題を解くということ

ルイスは命題的態度の帰属が満たすべき多くの制約があることを指摘するが、ここでは三種類の制約にのみ言及する。最初に言及すべきは、合理化の制約である<sup>4</sup>。合理化の制約によれば、信念・欲求の内容は当人の行為を意思決定理論 (decision theory) によって説明を与えられるようなものでなければならない。意思決定理論は、主体が抱いている信念と欲求を、確率関数と効用関数によって与える。ここでいう「確率関数」は、個々の命題についての主体が抱く信念の度合いの表現であり、形式的には命題から  $[0, 1]$  への関数として与えられる。効用関数はさまざまな財 (goods) に主体が見いだす主観的な価値の表現であり、形式的には財の集まりから実数への関数として与えられる。ところで、測定理論の研究者たちが明らかにしたところでは、これら関数が存在するためには、主体の信念や選好の体系が次のような条件を満たしていることが前提になる<sup>5</sup>。例えば、選好関係は推移的で完備 (complete) であるとか、信念の度合いがコルモゴロフの公理を満たす、といった条件である。こうした条件を満たす信念・欲求の体系は斉合的 (coherent) である。斉合的で

<sup>2</sup> フォーダー&ルポア 1997, p.156. ただし、引用元の文脈は、どんな問題もある意味では制約充足の問題とみなせるという指摘をしているところである。

<sup>3</sup> ルイスはここで「命題」の多義性を指摘している。一つの命題観は、命題とは真理値の担い手であるというもののだが、これとは別に、命題とは思考の内容であるという見解もある。ルイスによれば、我々が思考しうる命題は、合理性の制約を通して我々の帰属せらるる信念や欲求の内容に限られるので、我々が思考しうる命題の全体は真理値の担い手としての命題からなる全体の部分となる。ルイスはこの区別を利用して、カプランのパラドクスを解決する。Lewis 1986, sec. 2.3.

<sup>4</sup> Lewis 1974.

<sup>5</sup> Ramsey 1926; 鈴木 2005. 測定理論に関する一般向けの解説は Suppes 1957, pp. 260–274 などを参照。

ない主体には、確率関数や効用関数を割り当てられない。

しかし、斉合的な信念の体系はやはり無数にあるだろう。そして、斉合的な信念体系の中には常識的に考えて不合理なものもある。例えば、目の前に置かれたマグカップを通常の照明条件の下で知覚しているにも関わらず、目の前にマグカップがあるという命題に対する信念の度合いが極めて小さいといった人物を想像しよう。そのような人物の信念体系は外界のあり方に適合していない。特に理由がない限り、信念や欲求を帰属するときに、外界のあり方に適合していない内容をその人が抱いていると考えるべきではない。信念を帰属するやり方が複数あるという場合、その候補の多くは命題的態度の帰属が満たすべき別の制約に違反しているのではないか。

そこで考えられる制約は、人が抱いている信念の大部分は真であるというものである。この制約は「好意の原理 (principle of charity)」と呼ばれる。好意の原理によれば、大きな度合いの信念を抱いている命題の大部分が偽であるような信念帰属は、たとえそれによって行動を説明することができるとしても、不当なものとして排除される<sup>6</sup>。

しかし、真な命題であれば何であれ信念として帰属すべきではない。ミスリーディングな証拠を提示された人には誤信念 (false belief) を帰属することも許される。例えば、ミュラー・リヤーの図形を見たことも聞いたこともない人がはじめてミュラー・リヤーの図形に接したとすれば、彼女に対して一方の矢印が他方の矢印よりも長いという信念を帰属するのは正当であろう。

では、誤信念の帰属はいかなる合理性の制約によって正当化されるのか。「人間性の原理 (principle of humanity)」と呼ばれる制約は、人に帰属すべき信念や欲求の内容は、もし自分がその人の立場だったとすれば抱いたであろう信念や欲求の内容である、という制約を態度帰属に課す。例えば、たとえ自分がすでにミュラー・リヤー錯視について知っていたとしても、自分が無知な人の立場だったとすれば、図形の長さに関する誤信念を抱いたであろう、とすることができる。このようにして誤信念の帰属が可能になる。

1.3.3 節で論じたように、ルイスは心的語彙の意味は心理学理論によって陰伏的に定義されると考えている。そして、心理学理論によって主体にどのような内容の命題的態度が帰属されるのかは一意的に定まらない、というのが第一部で明らかになった機能主義の問題点であった。しかし、機能主義のこうした問題点は次のように考えれば何ら問題でないと判明するかもしれない。すなわち、信念や欲求の帰属を支配する合理性の諸制約を分析的な真理とみなすのである<sup>7</sup>。この提案が認められるなら、分析的な真理は必然的真理でもあるから、合理性の制約は必然的真理ということになる。主体に関するすべての物理的事実

<sup>6</sup> 根源的解釈のような状況における好意の原則の重要性は、デイヴィッドソンによって強調された。ただしフォーダーとルポアによれば、デイヴィッドソンの根源的解釈は、解釈者は主体がどの文をどの条件の下で真とみなすかということしか知らない、と仮定するのに対して、ルイスの場合は、主体に関するすべての物理的事実が与えられ、その上でその主体を志向的用語によって再記述する、という違いがある。フォーダー&ルポア 1997, p. 152.

<sup>7</sup> ただし、この提案は分析的な文と総合的文との間に厳格な区別をつけることなどできないというクワイン以来の懐疑論にさらされうる。分析と総合の区別に関する私自身の立場は、次田 2010 で示した。

が与えられれば、つまり、その主体がある時点においてどのような物的性質を例化しているのかが完全に与えられたならば、彼女がどのような信念や欲求を抱いているのかも決定する。したがって、命題的態度もまた物的性質に付随するのであり、態度帰属という我々が日常的になしている実践は自然主義と両立する。

### 5.3 節 解釈理論への不満

ここまでの議論から浮かび上がってくる態度帰属に関するルイスの描像は、以下の三点にまとめられる。

- ・ 合理的な信念の体系は確率関数で表現される。
- ・ 合理的な欲求の体系は財に対する効用関数で表現される。
- ・ 信念や欲求の帰属は合理化の原則や人間性の原則といった合理性の諸制約を満たす。

態度帰属に関するルイスの描像は、数理的な意思決定理論の研究とも結びついて数多くの成果を挙げてきたという意味で生産的である。実際のところ、私はルイスの描像を決定的に論駁できるとは思っていない。しかし、それでもルイスの描像は、命題的態度に関して自然主義者が採用すべき理論としては、少なくとも三つの点で不満足だと指摘したい。

(1) ルイスは、行動を同じ程度にうまく説明するような、命題的態度を帰属する複数の候補があるという 3.2 節の議論を受け入れており、そうした候補を一つに絞るためには合理性の制約が必要であると考えている。そして、合理性の制約は人々を理解可能な心的態度を抱いている主体として解釈することを可能にする。しかし、この描像は他人に対して心的態度を帰属する当の自分がどのような心的態度を抱いているのかという問題を棚上げにすることで成立しているように見える。この点はとりわけ、人に帰属すべき信念や欲求の内容は、もし自分がその人の立場だったとすれば抱いたであろう信念や欲求の内容である、という人間性の原理において明確に見出される。自分の心的態度をどのように確定すればよいのか、という問題に対してルイスがどのように答えるのかは明らかではない。

(2) ルイスは信念や欲求といった命題的態度を帰属するための必要条件として、主体が合理的であることを前提しており、上の描像のもとでは、不合理な信念や欲求は原理的に排除される。しかし、現実の我々は互いに両立しない二つの命題を信じてしまうこともあれば、証拠に基づいて信念の度合いを適切に改訂しないこともある。そして、こうした不合理性に関わらず、命題的態度の帰属がさまたげられることはないのではないか。

こうした批判に対して、我々はある程度ルイスに対して同情することができる。激しい妄想を抱いている人物は、多くの反証を与えられたとしてもその妄想を手放さない（だか

らこそ「妄想」と呼ばれる)。そのため、もしも妄想が信念の一種であるとすれば、ルイスの描像では帰属することのできない信念があることになる<sup>8</sup>。とはいえ、妄想が信念の一種であるという仮定はそれほど自明ではないし、そもそも精神疾患の患者に対して命題的態度を帰属するという実践が、一般人に対して命題的態度を帰属するという実践と同一視してよいのかどうか大いに議論の余地がある。よって、妄想のケースはルイスの描像にとってそれほど深刻ではないかもしれない。

しかし、精神疾患の患者を命題的態度の帰属という実践の対象から除外しさえすれば、信念や欲求が本質的に合理的なものであるというルイスの描像を維持できるということにはならない。一般人であっても、互いに両立しない二つの命題を信じてしまうことがしばしば起きる。そうしたケースは、命題的態度の帰属を不可能にするほどの不合理性は認めがたい。したがって、ルイスはなぜ些細な不合理性のケースでは命題的態度を帰属できるのかを説明する必要がある。

なお、この問いに答える際には、互いに両立しない命題を信じてしまう原因はさまざまである、という点に注意するのが重要である。実際、ルイスは上の描像を維持するために、互いに両立しない命題を信じてしまうさまざまなケースを場合分けして、個別に対処しようとする<sup>9</sup>。

例えば、現実の我々は諸前提からの論理的帰結を完全に見通すことができないために、前提から矛盾が導出される可能性に気づかず、互いに両立しない二つの命題を信じていることがある。こうしたケースに関して、ルイスは信念体系を分割することで問題を回避する<sup>10</sup>。信念体系を一つの箱のようなものとしてイメージしよう。ただし、信念体系は重複しない複数の領域 (compartment) に分割されている。そして、一方の信念体系ではpと信じているのだが、他方の信念体系ではqと信じていると仮定する。このとき、主体はpとqを信じているにも関わらず、それらの連言p&qを信じているわけではない。こうして各々の信念体系における整合性は当面のところ保たれる。

信念体系の分割は、論理に関する能力の欠陥のみから生じるとは限らない。論理に関して一切の誤りを犯さない理想的な人物であっても、一方でロンドン美しいと信じ、他方でロンドン美しいと信じていることがありうる。例えば、クリプキが論文「信念のパズル」で考察したピエールは、最初フランス語の話者として人々とコミュニケーションするなかでロンドン美しいと信じるようになった。しかし、後にイギリスに渡って英語話者としてコミュニケーションするなかで、彼は“Londre”と“London”の指示対象が同一であることに気づかないせいで、ロンドン美しいとも信じるようになった<sup>11</sup>。

---

<sup>8</sup> cf. Sterelny 1990, p. 94.

<sup>9</sup> Lewis 1986, pp. 34–36.

<sup>10</sup> 信念体系を分裂させるというアイデアは Stalnaker 1984, chap. 5 に由来する。デイヴィッドソンは自己欺瞞の分析において「信念の矛盾」と「矛盾した信念」を区別するときに、同様のアイデアを展開している。Davidson 1986.

<sup>11</sup> もっとも、このケースでピエールの信じている二つの命題が矛盾対立しているということを疑う人もいるかもしれないが。

ルイスにとっては厄介なことに、信念体系の分割というアイデアを採用しても、不整合な信念体系の存在をすべてのケースに関して説明できるわけではない。床屋のパラドクス（「ある村には自分のヒゲを剃らない人のヒゲだけを剃る床屋が存在する」）のように、単一の文から矛盾を導いてしまうような文の内容をうっかり信じてしまうケースがある。このケースのポイントは、単一の前提から矛盾が導いてしまうため、矛盾を導く複数の前提を別個の信念体系に切り分けるという前段落の手法はここでは使えない、ということにある。ルイスはこのケースを未解決問題であると認めている<sup>12</sup>。

また、ルイスが挙げている以外にも不合理性が認められるケースがある。例えば、連言  $p \& q$  が連言肢  $p, q$  よりもステレオタイプに近いために、 $p \& q$  に対する信念の度合いが連言肢に対する信念の度合いよりも高い、といったケースはよく知られている<sup>13</sup>。ルイスの描像を維持するには、心理学で数多く報告されているこの種の確率判断にまつわる錯誤についても何らかの説明を与える必要がある。

(3) ルイスの描像に対する上の二つの問題点が解決されたとしても、別の種類の疑問が残る。それは、ルイスの描像が命題的態度の帰属について語れることのすべてを尽くしているのだろうか、という疑問である。信念や欲求と一口に言っても、さまざまな種類の信念や欲求がありうる。どういうことか。

知識とのアナロジーがここでは参考になる。認識論では、しばしば動物的知識と反省的知識という二種類の知識概念が区別されてきた<sup>14</sup>。反省能力をもたない、つまり自分の心的状態がどのようなものであるかに関する高階の信念を抱くことのできない幼児や動物にも、ある意味で知識を帰属することができると思われる。知識は信念を含意するというテーゼを動物的知識にも適用するならば、動物や幼児にも信念的 (belief-like) な状態を帰属できるのではないか。例えば、カエルのような動物も知覚的表象を形成するが、そうした知覚的表象は信念的な状態といえるかもしれない。また、信念の帰属は欲求の帰属とは切り離して考えることができないので、動物や幼児に信念的な状態を帰属できるならば、欲求的な状態も帰属できるだろう。知覚に基づいて行動するときには、欲求も関わっている。カエルは餌を見つけても満腹ならば食べようとしない。敵を見つけたら逃げることを優先する。こうした能力はほとんど生得的だが、満腹であるとか捕食されたくないといった状態をカエルのような動物に帰属することは、単なる擬人法以上の実質があるはずだ。

以上のような推論が成り立つとすれば、ルイスの描像が妥当かどうかは差し当たり脇に置いておくとして、命題的態度の帰属に関してルイスとは別個のアプローチを探ることに意味があるだろう。次章以降ではこうした作業を試みる。

---

<sup>12</sup> Lewis 1986, p. 36.

<sup>13</sup> この種の誤謬は「連言錯誤 (conjunction fallacy)」と呼ばれる。カーネマン 2014, pp. 276–294.

<sup>14</sup> cf. Dretske 1991.

## 6 章 自然的記号と知覚的信念

“In the beginning there was information. The word came later.”

Dretske 1981a, p. vii

### 6.1 節 心のレシピ

われわれは、主体が抱いている命題的態度の内容を決定する物理的事実とはどのようなものか、という問題に取り組んでいる。因果役割機能主義のもとでは、行動を因果的に説明しうる信念・欲求の体系が原理的には複数存在するという難点が生じる。この難点を克服する手段として、前章ではルイスの解釈理論を一瞥した。だが、ルイスの理論は、反省能力をもたない主体に対して信念・欲求を帰属する可能性に対して否定的である。動物や幼児は、解釈されるべき思考を持っていない。

そこで、次に、生命体が抱いている信念の内容を欲求の内容と独立に決定しようとする自然主義的意味論に眼を向けたい。ストルネイカーはこの路線を明確に支持している。ドレツキはそれほど明確ではないが、この路線を支持していると解釈できる。本章ではドレツキの議論を検討のターゲットにする<sup>1</sup>。

信念の内容を欲求と独立に決定する一つの手がかりは、外界についての情報を担う自然的記号 (natural sign) や信号 (signal) という概念である<sup>2</sup>。外界のあり方を表象する信念、とりわけ知覚的信念は、自然的記号と似ている。自然的記号とは何か。例えば、高度計の目盛りは、その高度計が置かれている高度の自然的記号であり、樹木の年輪の数は、その樹木の年齢の自然的記号であると言われる。このように具体的な対象の性質・状態が、通常の条件下では環境の性質・状態と体系的な仕方で共変する (covary) ときに、その具体的な対象は環境についての情報を担う自然的記号 (あるいは、信号) であると言われる。体系的な仕方で性質や状態が共変するとは、ラフに言えば、具体的対象がとりうる性質の集合から環境が通常の条件下でとりうる性質の集合への写像関係があるということである。例えば、高度計の目盛りはその高度計が置かれている高度と共変するので、目盛りが 4,000 の位置にある高度計は、その高度計が高度 4,000 m に位置するという情報を担う。樹木の年輪はその樹木の年齢と共変するので、年輪の数が 78 本の樹木はその樹木が 78 歳であるという情報を担う。他の例も同じように考えられる。

しかし、外界についての情報を担う自然的記号ないし信号と、外界を表象する状態との間には、類似点と相違点がある。順に見ていこう。

ドレツキによれば、情報を担うという関係は、表象がもつ志向性の一つの基準とされる内包性 (intensionality) を持つ。すなわち、「 $x$ は $p$ という情報を担う」という文は、 $p$ に出現

<sup>1</sup> 本章がストルネイカーではなくドレツキをもっぱら参照している理由は、自然的記号に関してドレツキの方がストルネイカーより多くを語っているからである。

<sup>2</sup> 以下では「自然的記号」と「信号」を同義語として用いる。



しているある表現を共外延的な別の表現で置き換えると、文全体の真理値が変わることがある。例えば、コンパスのN極は北極を示すが、たとえ北極とホッキョクグマの生息地が現実一致していたとしても、コンパスのN極はホッキョクグマの生息地を示すわけではないだろう<sup>3</sup>。なぜなら、北極とホッキョクグマの生息地の一致は偶然であり、コンパスのN極がホッキョクグマの生息地と共変しないことも可能だからである。

だが、内包的な文脈を作り出すということは、命題的態度がもつ志向性の一つの特徴でしかない。志向性には存在からの独立性 (existential independence)、つまり、誤りの可能性に開かれているという特徴もある。この点は序章で示唆しておいた通りである。他方で、自然的記号が情報を担うという関係にはこの種の特徴が見られない。例えば、気圧に基づいて高度を示すという原始的な高度計を減圧室のような特殊な環境に持ち込めば、その高度計は高度と共変しない。このとき、たしかに我々は、その高度計はそれが置かれた場所の高度を正しく表象していない、と言えないこともない。しかし、高度計それ自体は、それが置かれた場所の気圧を入力として何らかの振舞いをする（例えば、特定の目盛りの位置に水位が移動する）というだけのことである。高度計に高度を表象させている、あるいは、高度を正しく示すべきものとして利用しているのは、あくまでも我々である<sup>4</sup>。

この違いは歴然として存在するように思える。しかし、それにも関わらず、ドレッキは自然的記号を表象が成立するための重要な構成要素とみなしている。ドレッキは心を工業製品になぞらえる。工業製品は様々な部品から構成されている。そうした部品は完成品が持つべき性質を最初から持っているわけではない。それと似て、自然的記号は外界を表象する心の素材ではあるが、それ自体は表象するわけではない。また、工業製品の部品は何でもよいわけではない。アンプを作るための部品は、何らかの都合のよい性質、例えば、伝導性の高い銅線のようなワイヤである必要がある。それと似て、心を作るためには、適切な性質をもった素材が必要である。内包性という基準をクリアしている自然的記号はその一候補として有力である。自然的記号を素材として、外界を表象する能力を備えた心を作るためのレシピを考案することはできないだろうか<sup>5</sup>。以上が意味論の自然化という課題に対するドレッキのアプローチである。

ドレッキのアプローチは、前章で取り上げたルイスの解釈理論と対照的である。ルイスは、過少決定の問題を克服するために、態度帰属のために必要な合理性の制約を参照することを提案した。このアプローチは、解釈者というすでに信念・欲求を内在的に保持している存在者を前提にしているように思われ、その限りでトップダウン的である。ドレッキのアプローチは、動物の知覚を表象の典型例にすえて、そこから人間が抱くような複雑な

<sup>3</sup> Dretske 1994, p. 211. ドレッキにならって表現を規格化しておくと、「コンパスのN極はあれが北極である (that is the North Pole) という情報を担う」が、北極とホッキョクグマの生息地は同一である (the North Pole is the habitat of polar bears) にも関わらず「コンパスのN極はあれがホッキョクグマの生息地である (that is the habitat of polar bears) という情報を担う」わけではない。

<sup>4</sup> Dretske 1990, pp. 65–66.

<sup>5</sup> Dretske 1994, pp. 209–211.

信念・欲求へと漸近していこうとする点で、ボトムアップ的である<sup>6</sup>。

本章の構成は以下のようなものである。6.2 節では、先に「体系的な共変」と述べた、自然的記号が情報を担う仕方を数学的な道具立ての下で記述する。6.3 節では、ドレッキが自身の情報理論を踏まえて、心のレシピ（表象の目的意味論）をどのように構築したのかを考察する。6.4 節では、ドレッキの目的意味論に対する三種類の批判を検討する。

## 6.2 節 ドレッキの情報理論とその再検討

ドレッキの目的意味論は自然的記号から出発して外界を表象する能力をもった心がいかにして出現しうのかを論じる。そこでは自然的記号の概念は所与にとられている。しかし、ドレッキは目的意味論とは別個に、自然的記号が担う情報内容を特定するための一理論を提案している。本節では、もっとも有力な情報理論の一つだとされているドレッキによる確率的理論を提示し（6.2.1 節）、それに対するいくつかの批判を考察する（6.2.2 節）。結論を先に言えば、ドレッキの情報理論には自然主義的な情報理論とみなす場合には、未解決の問題が残されている。

### 6.2.1 節 情報の確率的理論

1981 年の著作『知識と情報の流れ』において、ドレッキは確率概念を用いて自然的記号が情報を担うという関係を以下のように分析した<sup>7</sup>。

**情報の確率的理論 (PTI) :**  $r$ が $F$ であるという出来事が $s$ は $G$ であるという情報を担うのは、(i)  $s$ が $G$ である確率が 1 未満であり、(ii)  $r$ が $F$ であるときに $s$ が $G$ である条件付き確率が 1 のときそのときに限る<sup>8</sup>。

実際には、ドレッキは (PTI) を天下り式に導入しているわけではなく、上のような定式に落ち着くまでに多くの紙幅を費やしている。だが、ここでは上の定式を所与として受け入れ、それがどのような帰結をもたらすかを追跡する。

まず、(PTI) が何を言っているのかを例解する。例えば、ある高度計が高度 4,000 m に位置する確率が 1 未満であり、さらに、その高度計の目盛りが 4,000 の位置にあるときに高度

---

<sup>6</sup> cf. 戸田山 2014, p. 199.

<sup>7</sup> ここでは「分析」をややルーズに用いており、厳格な意味での概念分析 (conceptual analysis) には対応しない。概念分析の場合には、双条件文の両辺が同義であり右辺がより基礎的な概念のみを含むことが前提となるが、以下の双条件は同義性の条件を満たさない。「情報」はかなり多義的な用語であり、ドレッキはこの双条件が捉えているのは情報概念の一側面に過ぎないことを強調している。情報を分析するやり方としては、確率的理論の他にも反事実的条件法を利用する理論 (Cohen & Meskin 2006. cf. Dretske 1971)、単称因果関係を利用する理論 (Neander 2013) などが提案されている。

<sup>8</sup> Dretske 1981a, pp. 66–67.

が 4,000 m である条件付き確率が 1 であるとしよう。つまり、高度が 5,000 m でないときにその高度計の目盛りは 4,000 を絶対にとらない。このとき、高度計の目盛りが 4,000 の位置にあることはその高度計が高度 4,000 m に位置するという情報を担う。

自然的記号は、特定の対象についての情報を担うことになる。いまの例では、高度計はその高度計がどこに位置するのかという情報を担っている。樹木の年輪であれば、その樹木の年齢がいくつであるのかという情報を担うことになる。一般に、自然的記号は特定の情報源 (source) についての事象的な情報内容 (de re informational content) を担う。(PTI) の「s が G であるという情報を担う」の「s」は “source” の「s」であり、情報内容が特定の対象に関わることを示している<sup>9</sup>。

一般に、測定器具は一定の範囲内でしか働かないが、ドレッツキの定義はこの特徴をうまく捉えている。例えば、ある温度計は気温が 45 度を超えると目盛りが常に 46 度を示すとしよう。その場合、たとえその温度計の目盛りが 46 度であり実際に気温が 46 度だったとしても、その温度計の目盛りが 46 度であることは気温が 46 度であるという情報を担うことではない。目盛りが 46 度であるのに気温が 48 度かもしれないからである。しかし、その温度計は気温が 30 度であるといった情報を担うことはできる。気温が 30 度であるという情報を温度計が担うためには、目盛りが 30 度のときに気温が 30 度であるという条件付き確率が 1 であればよい。目盛りが 30 度でないときの条件付き確率は問題でない。

こうした具体例は直観的には理解しやすい。しかし、ドレッツキはここで問題になっている「出来事」の確率に関して、何らかの客観的解釈（傾向性解釈など）を採用するということを別とすれば、それをどのようにモデル化するのかを説明していない。そこで、この点は掘り下げて考察する価値がある。

確率論では、標本空間 (sample space) の部分集合である出来事に実数値を割り当てる<sup>10</sup>。二つのサイコロを振るという単純なケースでは、標本空間は 1 から 6 の整数の順序対の集合全体である。サイコロの目の和が 7 であるといった出来事は、この標本空間の部分であるような整数の順序対の集合と同一視される。このように、サイコロの和が 7 である確率を問題にするときは、出来事タイプが問題になっているように思われる。

それでは、(PTI) で問題になっている r が F であるといった出来事についてはどうか。先の温度計の例を振り返ってみよう。現実には存在する温度計の中には故障している温度計も含まれるので、もし温度計一般を考慮してしまえば、温度計が 30 度を示すときに気温が 30 度である条件付き確率は 1 ではないだろう。また、たとえ現実には存在する温度計が 30 度を示すときには気温も 30 度であったとしても、それは単なる偶然であって、温度計が 30 度を示すときの気温が 30 度である条件付き確率が 1 であるとは言えない。このため、ドレッツキは、確率の頻度解釈を (PTI) に適用することはできないと述べる<sup>11</sup>。(PTI) が問題にし

<sup>9</sup> Dretske 1981a, p. 65. 厳密に言えば、ドレッツキの定義では情報内容は主体の前提知識によって相対化される。しかし、この点は以下の議論には直接関わらない。

<sup>10</sup> 確率論では “event” の訳として「事象」が好まれる傾向にあるが、ここでは「出来事」と訳す。

<sup>11</sup> Dretske 1981a, p. 245n1.

ているのは、むしろ、ある特定の温度計が特定の時点で 30 度を示すときに気温が 30 度である条件付き確率が 1 であるかどうかである。たしかに、世の中には故障している温度計もあるが、だからといって、いま私が手にしている温度計も信頼できないなどと結論すべきではない。いま私が手にしている温度計が信頼できるかどうかは、私の温度計がいま示している目盛りがいまの気温と法則的に関連しているかどうかにかかっている<sup>12</sup>。(PTI)が問題にしているのは、このように出来事トークン同士が法則的に関連しているかどうかである。してみると、ドレッキは出来事トークンに確率を割り当てていることになる。

出来事トークンへと確率を割り当てるには、出来事トークンを時空領域と同一視するのではなく、出来事トークンを何らかの集合と同一視するような仕掛けがあると便利だと思われる。一つの提案としては、標本空間を法則的に可能な世界の集合とし、 $r$ が $F$ であるという出来事を $r$ が $F$ である可能世界 ( $Fr$ 世界) の集合とみなすことが考えられる。出来事間の法則的な関連性は様相的概念であるから、ここで可能世界を持ち出すのは自然である。この提案の下では、 $r$ が $F$ であるときに $s$ が $G$ である条件付き確率が 1 であるのは、 $Fr$ 世界の集合が  $Gs$ 世界の集合に包摂されるとき、そのときに限られる<sup>13</sup>。実際には、この単純なモデルは 6.2.2 節で修正を迫られることになるが、当面のところはこのように考えておく。

しかし、標本空間について以上のように考えるとしても、気温が 30 度である、といった出来事の確率はどうやって求めればよいのか。ドレッキはこの問いには答えられないことを認める。その代わり、彼は二つの論点を指摘する<sup>14</sup>。まず、個々の出来事の確率を推定することはできないとしても、ある出来事が与えられた場合の別の出来事の条件付き確率については見当をつけることができるかもしれない。喩えるならば、二つの弦の絶対的な長さは分からなくても、どちらの方が長いかは比較によって決定できるようなものである。また、個々の出来事の確率が我々には不可知であったとしても、そのことは (PTI) の下で情報を担うという関係が成立するための障害ではない。情報を担うという関係は、それを認識する主体とは独立に客観的に成立するからである。ドレッキはこの点はどれほど強調してもしすぎることはない述べている。

## 6.2.2 節 情報の確率的理論への批判と応答

ドレッキの情報理論 (PTI) は (i) と (ii) という二つの要素から成り立っている。(PTI) から何らかの不都合な帰結が生じるならば、問題の源泉は (i) と (ii) のどちらかに帰着するだろう。本節では (PTI) に対する批判を (i) と (ii) のどちらが問題を引き起こしているのかという観点から分類し、それぞれに対して可能な応答を探る。

<sup>12</sup> Dretske 1981a, p. 126.

<sup>13</sup> この提案がドレッキの出来事観と整合するかどうかは検討の余地がある。ドレッキの出来事観については Dretske 1977 を参照。

<sup>14</sup> Dretske 1981a, pp. 53–55.

### (1) 確率 1 未満の要求

(PTI) は情報内容が確率 1 未満であることを要求する。この要求の下では、数学的真理を情報として担う信号は存在しない。 $r$ がFであるという信号が $s$ はGであるという情報を担うには、 $s$ がGである確率が 1 未満でなければならない。しかし、 $s$ がGであるということが数学的真理であれば、 $s$ がGであることは任意の可能世界で成り立つので、 $s$ がGであることは確率 1 で成り立つ。よって、数学的真理の情報を担う信号は存在しない<sup>15</sup>。

つまり、(PTI) は偶然的な情報にしか適用できない。しかし、この種の限界をそれほど深刻に受けとる必要はないのではないかと。まず、適用範囲が限定されている理論が常に悪いわけではない。そうした理論は、自然科学では広く見られる。例えば、 $PV = nRT$  という理想気体の状態方程式は、実在の気体には近似的にしか成り立たないが、そのことによって価値が減じるわけではない。(PTI) は偶然的で単称的な情報しか扱えないとしても、自然的記号の情報内容について考えるような場面ではこの制約は問題にならない。

### (2) 条件付き確率 1 の要求

「条件付き確率 1 の要求」とは、情報を担うために信号が情報源のあり方を条件付き確率 1 で伝えなければならないという要求である。(PTI) に対する典型的な批判は、この要求が強すぎるというものである。

6.2.1 節では、信号が情報を担うという関係の一例として、温度計が 30 度の目盛りを指しているときに気温が 30 度である条件付き確率が 1 であるというケースを仮定した。しかし、スプリングが指摘するように<sup>16</sup>、この仮定は文字通りに受け取るならば現実には使用されているどんな測定器具にも誤差があることを忘れていない。19 世紀以降の物理科学では、測定値に含まれる誤差の大きさを見積もることが義務になっている。温度計が 30 度の目盛りを指しているときには条件付き確率 1 で気温がちょうど 30 度であるということとはありえない。

(PTI) は近代科学の実践と不整合をきたしているのではないかと。

スプリングの批判に対するドレッツキの応答は次のようである<sup>17</sup>。たしかに、温度計が 30 度の目盛りを指しているときに気温が条件付き確率 1 でちょうど 30 度であるということとはありえない。しかし、こうした観察が示しているのは、「温度計の目盛りが 30 度であることは気温が 30 度であるという情報を担う」といった言い回しはルーズであるということに過ぎない。我々は一定の誤差範囲の中では気温が何度かを知りうるはずであり、それが可能であるのは、温度計の目盛りが一定の誤差範囲の中では気温が何度であるかを条件付き確率 1 で伝えるからだと考えられる<sup>18</sup>。

<sup>15</sup> Barwise & Seligman 1997, p. 17; Bremer 2003. 自然法則が確率 1 を持つような確率の解釈を採用するならば、数学的真理だけでなく、自然法則の情報を担う信号も存在しない。cf. Dretske 1981a, p. 264n3.

<sup>16</sup> Suppes 1983, p. 81.

<sup>17</sup> Dretske 1983, p. 84.

<sup>18</sup> アームストロングは、条件付き確率 1 の要求は非決定論的な自然法則が支配する世界では満たされないもので、(PTI) に従うと非決定論的な世界では情報が成立する余地がなくなってしまう、とドレッツキを批判した。Armstrong 1983, p. 64. しかし、この批判もスプリングの批判と同様に応答することができる。図式的

しかし、この応答には素朴な疑問が浮かぶ。科学者が物理量を測定する際に一般に利用する誤差関数は本質的には正規分布の確率密度関数と同じであり、誤差範囲をどれほど広げようとも、温度計の目盛りが一定の誤差範囲の中では気温が何度であるかを条件付き確率 1 で伝えるということにはならないのではないかと。ドレッズキが念頭においている誤差関数は、極端な可能性を切り落とすようなものでなければならない。つまり、正常に作動する測定器具であれば、一定の誤差範囲を外れるような測定値は絶対にはじきださないという仮定が措かれているはずである。

実際、極端な可能性は切り落とすというこの戦略をドレッズキは広範に用いている。自然法則や論理法則と両立しうる可能性すべてを考慮するならば、信号が情報源のあり方を条件付き確率 1 で伝えるケースはまずありえないのではないかと思われるからである。ドレッズキは次のように述べている。

いかなる信号も、それが通常的手段によってではなく、何らかの奇妙な宇宙的偶然によって、悪霊によって、あるいは、超自然的な介入によって生成された可能性を排除できない。そうした偶然が真正の可能性として認められるならば、どんな信号も曖昧 (equivocal) である<sup>19</sup>。

この種の懸念から条件付き確率 1 の要求を守ろうとするならば、方法は一つしかない。自然法則や論理法則と両立する可能性のいくつかを、問題となっている出来事がある情報を担うかどうかを考察するときに数えいれないようにすればよい。

極端な可能性を数えいれないという操作を、ドレッズキは「平ら (flat)」や「空っぽ (empty)」といった形容詞の用法との類比によって説明する<sup>20</sup>。現実には幾何学的に厳密な意味で平らなテーブルに出くわすことはありえないが、だからといって、テーブルが平らでないということにはならない。我々が利用しているテーブルに関してそれが平らかどうかを問題にするときには、小さすぎる凹凸は最初から除外されている。小さすぎる凹凸を除外した上で、テーブルの上に凹凸がなければ、そのテーブルは平らと言われる。条件付き確率 1 の要求に関して、ドレッズキは平らの用法と似たようなことが当てはまると論じる。論理法則や自然法則と両立する可能性をすべて考慮するならば、ある出来事が生じたときに別の出来事が生じる条件付き確率が 1 であるということは現実にはありそうにない。しかし、極端な可能性は勘定にいれないという操作が許されるなら事情は違ってくる。ドレッズキは次のような例を提示する<sup>21</sup>。電圧計は回路の電圧を標示するが、電圧計の針の振れは回路の電流に基づいている。オームの法則 ( $E = IR$ ) によれば、電流は抵抗に依存しているので、も

---

にえば、 $r$  が  $F$  であるときに、 $s$  が  $G$  または  $G'$  である条件付き確率が 1 ならば、そこには情報を担うという関係が成立する。cf. Dretske 1981a, pp. 31–32.

<sup>19</sup> Dretske 1981a, p. 130. なお、ここで「曖昧」と訳した “equivocal” は情報理論の専門用語であり、4 章で「曖昧」と訳した哲学用語の “vague” とは意味が異なる。

<sup>20</sup> Dretske 1981a, p. 110.

<sup>21</sup> Dretske 1981a, pp. 111–118.

しも回路のワイヤが抵抗を頻繁に変える可能性があるならば、電圧計の針の振れは回路の電圧を極めて曖昧にしか伝えないだろう。幸運なことに、現実には抵抗はワイヤの物理的組成からしてそう頻繁には変化しないので、回路のワイヤが抵抗を変えるという可能性は勘定に入れなくてよい。回路の抵抗が一定である可能世界に限定した上で、電圧計の針の振れが回路の電圧を条件付き確率 1 で伝えるならば、その電圧計は回路の電圧についての情報を担っている、とすることができる。ドレッキはこのように、ある出来事が生じたときに別の出来事が生じる条件付き確率を、極端な可能性を除外した上で定まるものとみなしている。

しかし、そうすると、何が極端な可能性であるのかを判定する基準は結局のところ我々の興味や関心に依存しているのではないだろうか<sup>22</sup>。何が極端な可能性であるのかを判定する基準が我々次第であるとすれば、たしかに条件付き確率 1 の要求は満たされるかもしれない。だが、その場合、自然的記号が情報を担うということは客観的な事実と言ってよいのかどうかは疑わしい。というのも、適当なやり方で極端な可能性を除外してしまえば、条件付き確率 1 の要求はトリビアルに満たすことができってしまうと思われるからである。したがって、(PTI) がトリビアルでない情報理論であるためには、極端な可能性の範囲を定める基準を与えるメカニズムを用意する必要がある。

ドレッキはこうした疑問がありうることを理解している。だが、極端な可能性の範囲を定める基準をめぐる問題を、ドレッキは語用論の問題として扱おうとする<sup>23</sup>。この態度は、自然的記号が情報を担うという関係を自然主義的意味論の基礎に据える、という企てに照らした場合には適切とは思えない。我々の興味と関心は何らかの信念や欲求の集まりとして特徴づけられると思われるので、何が極端な可能性かどうかを決めるのが我々の興味と関心次第だとすれば、自然的記号が情報を担うという関係は究極的には我々の信念や欲求の集まりによって成立している、ということになってしまうからである。

こうした問題は、(PTI) が特徴づける自然的記号を自然主義的意味論の基礎に据えることはできないということを示唆する。

### 6.3 節 自然的記号から表象へ

本節では、ドレッキ型の目的意味論を定式化する。最初に、自然的記号と表象の相違点がどこにあるのかを論じる。それを踏まえて、自然的情報を材料とする心のレシピをドレッキがどのように描いたのかを明確にする。ドレッキのレシピは、自然的記号の概念に依拠している。前節で論じたように、自然的記号の概念をわれわれの関心と独立に特徴づけることは難しい。しかし、ここでは自然的記号の概念を所与として使えると仮定する。

まず、自然的記号と表象の基本的な違いとして、差し当たり次の二点を指摘できる。

<sup>22</sup> Millikan 2004, pp. 37–38, 下嶋 2009.

<sup>23</sup> Dretske 1981a, pp. 132–134; 1981b.

- (1) **誤りの可能性**：6.1 節で示唆したように、表象は誤りの可能性に開かれている。自然的記号にはこの特徴は見られないと思われる。例えば、気圧に基づいて高度を示すという原始的な高度計を減圧室のような特殊な環境に持ち込めば、その高度計は高度と共変しない。このとき、高度計それ自体は、それが置かれた場所の気圧を入力として何らかの振舞いをする。例えば、特定の見盛りの位置に水位が移動する。ただそれだけである。高度計に高度を表象させている、あるいは、高度を正しく示すべきものとして利用しているのは、あくまでも我々である。また、通常的环境においても、もし高度計が故障しているならば、その高度計は高度を誤って示しているといえる。このように環境との共変から切り離して見盛りの意味について語ることが意味をなすという点で、測定器具はやはり規約的な表象であり、単なる自然的記号ではない。
- (2) **利用者の存在**：表象は利用者をもつ。表象が何を表象しているのかを誰も知らないということはないと思われる。たしかに、無人島の海辺で、押し寄せる波によってチャーチルの顔にそっくりな模様ができるといったことは考えられるが、我々はその模様がチャーチルの絵とはみなさない。また、現在のところ解読不可能な古文書の存在も、昔は理解者が存在していたと考えれば反例ではない。表象には必ず内容の理解者がいるという事実は、表象の内容はそれを描いた人々の意図に強く依存するということを示している。この点は、言語表現のように表象内容を利用者がほぼ完全に恣意的に決められるような表象だけでなく、自然的記号が表象に転用されたケースにも当てはまる。例えば、上で取り上げた原始的な高度計の見盛りは、周囲の気圧と高度の双方と共変しており、気圧と高度のどちらの自然的記号でもありうる。しかし、高度計として利用される限り、この装置の見盛りは高度を表象するのであって、気圧は表象しない。

このように、自然的記号と表象は、一見してかなり明確な対比をなしている。

しかし、ドレツキは自然界には何らかの存在者によって利用されている自然的記号があると指摘する。例えば、生物が知覚のために利用する知覚システムは、測定器具と似ている。外界からの刺激は眼のような感覚器官において特定の電氣的インパルスへと変換される。変換された電氣的インパルスは外界のあり方についての情報を担っているため、知覚システムは神経を伝達してくるこうした電氣的インパルスを利用して行動を制御することができる。感覚器官は自然的記号を作り出し、知覚システムはそれを利用する。例えば、目の前にあるのはエサであることを示す自然的記号が作り出されたなら、満腹でない生物は舌を伸ばすかもしれない。捕食者の接近を示す自然的記号が作りだされたなら、たとえ食料が近くにあっても逃げることを優先するかもしれない。もちろん、知覚システムは適切に作動しないこともあるが、これは機械の故障になぞらえることができる。例えば、感



覚器官が外界の刺激を通常の仕方で電氣的インパルスへと変換しないということがあるかもしれない。もし、知覚システムを測定器具になぞらえることが許されるならば、知覚システムが故障している場合には、高度計が高度を誤って示すのと同様に、知覚システムは外界のあり方を誤って示しているといえる。したがって、もし知覚システムと測定器具の類比がかなり多くの側面に関して成り立つのだとすれば、言語などの規約的表象に関しては別立ての議論がさらに必要だとしても、知覚的表象に関してはそれを非意味論的な概念で説明するという自然主義的意味論の課題を完遂させる見通しが得られるのではないかと期待できる。

もちろん、この路線にもさまざまな障害が立ちはだかっている。最初に生じる疑問は、知覚的表象が誤りうる表象であるのは何故か、というものである。高度計のような測定器具が高度を誤って表象しうるのは、我々が高度計に高度を標示するという機能を意図的に与えたからである。他方で、生物の知覚システムに対して外界のあり方を示すという機能を意図的に付与した存在者はいないだろう<sup>24</sup>。そこで、知覚システムは外界のあり方を示すという機能を自然に有するようになった、と思われる。実際、この見解はかなり直観的にもっともらしい。

しかし、表現型が機能を自然に有するとはどういうことだろうか。機能の概念には目的論的な含意があるのではないかと疑う人もいるかもしれない。それゆえ、表象が誤る可能性をもつことを機能の概念によって説明しようとする目的意味論の支持者は、機能の自然主義的な説明を与える必要がある。

一般に、目的意味論の支持者はここで生物学の哲学に手がかりを求める<sup>25</sup>。生物学の哲学者たちは、生物学において「機能」という語がどのように使われているのかを考察してきた。この問いに対する一つの有力な答えは「起源説 (etiology view)」と呼ばれる。

**機能の起源説：**ある形質cが課題tのための適応であるとき、tの遂行をcの「機能」と呼ぶならば、これは生物学者の「機能」の用法の適切な分析を与える<sup>26</sup>。

適応 (adaptation) とは、何らかの課題を果たすことで自然選択において有利だったが故に現存する形質をいう。例えば、心臓は血液循環させるための適応である。よって、起源説によれば心臓の機能は血液循環である。この答えは心臓の機能を適切に捉えているように思われる。そこで、目的意味論の支持者は、表現型が機能を自然に有するとはどういうことかという問いに対しては、起源説を採用すればよさそうである。

機能の概念を起源説によって理解することによって、目的意味論の支持者は、いまや表象の誤る可能性を自然主義的に説明するための糸口を掴んだ。すなわち、外界のあり方を示すという自然的機能をもつ装置によって作り出された状態であれば、それは (測定器具

<sup>24</sup> 超自然的な創造者が存在するならば話は別だが、自然主義者はその可能性を認めない。

<sup>25</sup> ただし、目的意味論の支持者全員が起源説を支持するわけではない。cf. Nanay 2014.

<sup>26</sup> ソーバー2007, pp. 168, 172.

の目盛りが表象と呼びうるように) 表象と呼べるのではないか。

残念ながら、これではまだ一段落着いていない。外界のあり方を示すという自然的機能をもつとはどういうことか、まだ不明瞭である。実際、ミリカンはそのような機能がありうるのかを疑問視しており、次のような懸念を表明している<sup>27</sup>。生物は、知覚システム以外にも外界のあり方を示すような状態を作り出す装置を持っている。例えば、カメレオンの皮膚はカメレオンが座る環境に合わせて色素配列を調整するという機能をもった装置を含んでいる。この装置は、環境に合った色素配列を作り出すことによって、歴史的に選択されてきたからである。ミリカンはこの装置は表象を作り出す装置ではないと断言する<sup>28</sup>。彼女は理由を丁寧に説明していないが、その直観は正しいと私は思う。ではどうするか。

ここで、表象は利用者から切り離せないという先に確認した論点を思い出そう。カメレオンの色素調整の装置が、外界のあり方を示す状態を作り出しているにも関わらず表象を作り出しているとはいいがたいのは、この装置が作り出した状態に応じて行動を変化させるような存在者がいないからではないか<sup>29</sup>。その点で、知覚システムはカメレオンの色素調整の装置とは異なる。動物の知覚システムは外界からの刺激によって外界のあり方と共変する状態を作り出すが、知覚システムの示す状態に応じて動物は行動を変化させることだろう。目の前に捕食者がいるという情報を担う自然的記号が作られたならば逃げる、エサがあるという情報を担う自然的記号が作られたならば近づいて舌を伸ばすといった具合である。作り出された自然的記号によって行動を変化させる存在者が存在することによって始めて、外界のあり方を示すという機能が生じる<sup>30</sup>。

以上の考察を踏まえて、ドレッツキ型の目的意味論を次のように定式化する。

**ドレッツキ型の目的意味論 (DTS) :** 表象  $r$  が  $p$  を表象するのは次のときそのときに限る。 $r$  は外界のあり方と共変する自然的記号を作り出す機能をもった装置によって作られており、かつ、この装置が正常に  $r$  を作り出したならば  $r$  は  $p$  と共変する自然的記号である。

## 6.4 節 ドレッツキ型の目的意味論に対する批判

(DTS) には様々な批判が寄せられてきた。そうした批判のすべてが強力というわけではな

<sup>27</sup> Millikan 1993, p. 85.

<sup>28</sup> ミリカンは環境と共変するこうした装置のもつ機能を「関係的機能」と呼ぶ。Millikan 1984, p. 39. ただし、以下で示すように、自然的記号を作る機能は関係的機能の一種ではあるが、関係的機能が自然的記号を作る機能であるとは限らない。

<sup>29</sup> ドレッツキは次のように述べている。「ルース・ミリカンは、表象は消費者に依存するという点を強調している。私も賛成である。提供される情報を利用する(消費する)ものがなければ、いかなるものもそれを提供する機能を獲得できない」。Dretske 1995, p. 187n24.

<sup>30</sup> 正確に言えば、ドレッツキは、自然選択は自然的記号を作るシステムが外界のあり方を示すという機能を獲得するための一つのあり方だと考えている。条件付けの学習によって自然的記号を作るシステムが外界のあり方を示すという機能を獲得することもありうる。だが、本稿ではこの点には踏み込まない。

いが、中には無視できないものもある。本節では、ドレッツキ型の目的意味論（DTS）に対する三つの批判を検討する。最初の二つの批判は（DTS）が依拠している機能の考え方に対する批判である。これに対し、三つ目の批判は（DTS）が知覚システムの表象内容を適切に捉えていないという批判である。私の考えでは、この三つ目の批判がもっとも重要である。

#### 6.4.1 節 起源説、および起源説的な機能概念に対する批判

（DTS）は、知覚システムが外界のあり方を示すという機能を自然に持ったという考えを擁護するために、機能の起源説、ないし以下で定義する起源説的な機能概念を利用する。だが、起源説と起源説的な機能概念に対しては、さまざまな批判がなされてきた。本節では三つの批判を考察し、いずれの批判も決定的ではないと結論づける。

##### （1）機能の起源説の妥当性

起源説によれば、形質の機能は、その形質がそれに対する適応であるところの課題である。しかし、適応形質は現在もその課題を遂行しているとは限らない。例えば、ウミガメの前脚は巣を作るのに役立っているが、巣を作るための適応というわけではない<sup>31</sup>。したがって、起源説によれば、巣を作ることは前脚の機能ではないことになる。だが、現に巣を作るために役立っているのならば、前脚の機能は巣を作ることだと考えるのが直観にかなっていないだろうか。また、ウミガメの前脚は現実の例だが、より過激な例を想像することもできる。例えば、自然選択の過程を経ずに突然変異的に発生した形質は、たとえそれが有利な働きををするとしても、起源説のもとでは機能を割り当てることができない。

思うに、目的意味論の支持者は、起源説的な機能概念を利用することを起源説そのものから区別することで、この種の批判に応じることができる。

**起源説的な機能概念：**ある形質  $c$  が課題  $t$  のための適応であるとき、 $t$  の遂行を  $c$  の機能と呼ぶ。

つまり、目的意味論の支持者は、自分が関心を持っているのは適応の概念に訴えることで定義される「機能」らしき概念であって、この概念が現場の生物学者たちが「機能」という語で意味しているものと厳密に一致しているかどうかには拘らない、と論じればよい。目的意味論の支持者は生物学における機能の概念分析に携わる必要はない。起源説的な機能概念を上のように約定しておき、もしその概念が志向性を説明するために有用な理論的概念であることを明らかにできれば、哲学的には大きな収穫であろう。したがって、起源説の機能概念に適用範囲の制限があるとしても、そのことから、起源説的な機能概念まで手放さなければならないという帰結までは出てこない。

---

<sup>31</sup> ソーバー2007, p. 172.

(2) 起源説的な機能概念は自然選択が進化の主要な原動力だと前提している

一般に、集団における遺伝子の頻度が変化することを「進化」と呼ぶが、この意味での進化が生じる原動力は自然選択に限られない<sup>32</sup>。突然変異や遺伝的浮動によっても進化は生じうる。もしある表現型が突然変異や遺伝的浮動によって生じたのだとすれば、その表現型は何らかの課題を遂行したために選択されたのではない。こうした表現型には起源説的な機能概念を適用できない。したがって、起源説的な機能概念を利用する目的意味論の支持者は、生命体のもつ外界の認知能力が自然選択によって生じたと仮定する必要がある。だが、この仮定はトリビアルに真というわけではない。

この診断は正しいと思われる。目的意味論の支持者は外界の認知能力が自然選択によって生じたと仮定しているし、この仮定はトリビアルに真というわけでもないだろう。したがって、起源説的な機能概念を利用する人は、経験的事実に関して一種の賭けを行っていることになる。しかし、この賭けはそれほど分が悪いとは思えない。たしかに、突然変異や遺伝的浮動によって適応的な表現型がたまたま出現することはあるだろう。それでも、その表現型が集団の他の個体にまで広まり、しかも長い期間にわたって維持されるためには、自然選択の力が必要である。適応的な表現型が自然選択なしにたまたま出現しても、自然選択なしに集団中で長い期間にわたって維持されつづける可能性はとても低い<sup>33</sup>。この二つ目の論点は、起源説的な機能概念の利用を不合理にはしないだろう。

(3) 起源説的な機能概念の科学性はみせかけである

起源説的な機能概念は、自然選択に訴えていることから推測されるように、進化生物学の機能概念に一見するとうまく適合している。しかし、形質がある課題を遂行しているだけの状態から、その課題がその形質の真正の機能へと変貌するには、環境とどのくらいの取引をすればよいのだろうか。この問いには正解などないだろう。つまり、有利な形質が適応へと変貌する時点に関する事実問題 (matter of fact) など存在しないと思われる。それに関わらず、起源説的な機能概念は、一つの決定的瞬間をクローズアップして種の最初の個体の誕生をしめす出来事をなんとか特定したいという願いを隠し持っているのではないか。もしそうだとすれば、この種の願いはダーウィンの漸進主義と調和していない<sup>34</sup>。起源説的な機能概念の科学性はみせかけのものであり、むしろ、そこには反ダーウィンの側面が隠れている。

起源説的な機能概念に共感する人々もそうでない人も、この概念が科学を尊重していることを疑ってはいないであろうから、この批判は意外に思われるかもしれない。しかし、起源説的な機能概念が用いている適応の概念は、ダーウィンのアイデアの中核をなす適応主義 (adaptationism) とは独立である。適応主義とは、ラフに言って、表現型レベルの形質

<sup>32</sup> ソーバー2007, pp. 8, 40.

<sup>33</sup> cf. Neander 1997, p. 576.

<sup>34</sup> この問題提起はデネットから大きな示唆を受けている。Dennett 1995, p. 424n10.

に関しては、自然選択が唯一の（あるいは、少なくとも最も重要な）説明要因である、という主張である。この主張は、ある有利な形質が世界とどれだけの取引を行えば適応になるのか、といった問題に厳格な答えがあることを前提していない。適応主義は、動植物が有している複雑な形質が自然選択によって漸進的に形成された、ということしか要求していない。

しかし、たとえ適応主義と適応という二つの概念が独立であり、起源説的な機能概念の科学性が見かけ上のものに過ぎないのだとしても、以上の考察は起源説的な機能概念を採用することに対する論駁になっているとは思えない。形質が適応形質に変貌する瞬間を捉えられないということから、両者の区別が存在しないということを結論するとすれば、それは曖昧な概念に関する滑りやすい坂論法の誤謬と同種の誤りではないか。赤がオレンジに変貌する特定の色合いのペアを見出すことができなくても、赤とオレンジの区別は消滅しない。それと同じように、起源説的な機能概念を採用するために、単なる形質が機能をもった形質に変貌する瞬間が存在する必要はない。

そもそもドレツキが起源説に訴えたとき、彼の力点は、形質がある効果を発揮し始めてからその効果が形質の機能になるまでには一定の時間が必要なので、心が環境のあり方を表象するという能力は一瞬にして作り出せるものではない、という点にあった。ドレツキはパン作りという比喻を提示している。パンを作るには、材料のイーストと小麦粉を混ぜ、その後、発酵してパンがふくらむまでに一定の時間がかかる<sup>35</sup>。ドレツキによる心のレシピはパン作りと少し似ている。外界のあり方と共変する自然的記号を生み出すという形質の課題が機能となるまでには一定の時間をかけなければならない。

#### 6.4.2 節 反応機能は起源説的な機能概念と両立するか

起源説的な機能概念の整合性を受け入れたとしても、自然的記号を作り出す機能などという性質を認めることはできない、という見解もありうる。ミリカンとシェアはこうした見解を示唆している<sup>36</sup>。彼らの見解は次のようにまとめられる。

自然的記号は一般に原因と関わる。例えば、温度計が指す目盛りの位置はその温度計が置かれている環境の気温を原因として定まる。動物の知覚システムが自然的記号を作り出すとすれば、それは外界からの刺激によって何らかの状態に移行することによるだろう。しかし、機能の起源説は、形質 *c* が課題 *t* のための適応であるときに、*t* の遂行を *c* の機能と呼ぶ。そして、ここでいう課題とは、何らかの結果を引き起こすことであるというのが一般的な解釈である。例えば、心臓の機能である血液循環は、心臓の活動によって引き起こされることである。このように、形質の機能は結果に関わっている。それゆえ、自然的

<sup>35</sup> Dretske 1994, p. 223. cf. Dretske 1981a, p. xi. ちなみに、機能の起源説は「タンカー理論」と呼ばれることもある。タンカーはプロペラが相当に回転した後になってようやく進みはじめるからである。MacLaughlin 2001, pp. 101–102.

<sup>36</sup> Millikan 2004, pp. 67–68; Shea 2004, p. 91.

記号を作り出す機能などという観念は、起源説的な機能概念と両立しない。

しかし、機能は原因には関わりえないのだろうか。ネアンダーは、いくつかの生物学的な機能は実際に原因に関わっていると反論している<sup>37</sup>。彼女が列挙している事例の一つを紹介しておこう。松果体は光が弱くなっていくことに反応して、眠気を誘うメラトニンを分泌する。また、膵臓は血中グルコースの上昇に反応してインスリンを分泌する。それでは、これらの器官がもつ機能は何か。機能は結果に関わるべきだという原則からいえば、松果体の機能は眠気を誘うことであり、膵臓の機能は血中グルコースの濃度を下げることかもしれない。しかし、こうした結果だけがこれらの器官の機能だと考える必要はない。松果体は光が弱くなっていくことへと反応することを、膵臓は血中グルコースの濃度が上昇することに反応することをも機能とするのではないか。このように、何らかの出来事に対して反応するという機能も存在するのではないか。

ネアンダーは何らかの出来事に対して反応する機能を「反応機能 (response function)」と呼ぶ。起源説的な機能概念は、反応機能も適切に扱えるような生物学的機能の概念であるべきである。それは次のように考えれば可能である。もし、ランダムにメラトニンを分泌することと、光が弱くなっていくことに反応してメラトニンを分泌することが適応度の違いをもたらさないのであれば、松果体は生存に貢献していないと考えられる。しかし、上に挙げた例はどれも、適応度に関する違いをもたらすはずである。形質の機能は結果にだけ関わるのではなく原因にも関わる。それならば、刺激に対して特定の反応を返すことによって形成された形質はそうした反応を返すことを機能とする、と述べることも可能であろう。

#### 6.4.3 節 識別能力の限界

(DTS) に対する反例として、ゴドフリー＝スミス、ミリカン、ステレルニーらは、捕食者に対する過剰警戒という事例を挙げている<sup>38</sup>。私の考えでは、この種の事例こそが (DTS) に重大な問題を突きつけることになる。

一般に動物の認知メカニズムは、天敵となる捕食者が近くにいるにも関わらずそれに気づいていない、といった事態に陥ることを避けるように進化する。「転ばぬ先の杖」と言うように、近くにいるのが実際には捕食者ではなく無害な存在であったとしても、動物は逃避行動をとる傾向がある。その理由は、捕食者を正確に探知する、つまり、捕食者と正確に対応する自然的記号を作り出す装置を設計するのが難しいということもあるが、より重要な理由は、警戒しすぎても失うものは比較的少ないからである。捕食者を無害な存在と見間違えてしまえばすべてを失ってしまうが、無害な存在を捕食者と見間違えても、コストは逃避する際のエネルギーだけであろう。そうすると、例えば、正常に働いているウ

---

<sup>37</sup> Neander 2013, sec. 2.

<sup>38</sup> Godfrey-Smith 1989, sec. 8; Millikan 1993, p. 85; Sterelny 1990, pp. 123–124.

サギの感覚器官は、近くにいるのはキツネではなくウォンバットであるにも関わらず、逃避行動を引き起こすことにつながる信号を脳に送るかもしれない。直観的には、この信号は近くにキツネがいるということを誤って表象していると言いたい。だが、ウサギの正常な感覚器官が作り出した信号は近くにキツネがいるということとは共変しない。よって、これは（DTS）の左から右に対する反例となる。さらに次のことも成り立つ。正常に働いているウサギの感覚器官が作り出した信号は、近くにキツネもしくはウォンバットがいる、といった選言的な内容の自然的記号である。しかし、直観的には、この信号は近くにキツネもしくはウォンバットがいるといった選言的な表象内容をもつとは言いがたい。これは（DTS）の右から左に対する反例となる。したがって、表象は正常な条件下で作りに出されたとしても外界のあり方に関する自然的記号とは限らないのは明らかである。

ミリカンは、こうした議論がドレツキ型の目的意味論を決定的に論駁すると考えていたように思われる。しかし、過剰警戒の事例がドレツキ型の目的意味論に対して本当に致命的だと結論するには、もう少し議論が必要だと私は思う。というのは、ドレツキ型の目的意味論の支持者は次のように応答するかもしれないからである。たしかに、直観的には上のウサギの感覚器官が作り出した信号は、感覚器官が正常な場合ですらキツネと不正確にしか共変していないとしても、キツネを正しく表象しそなっていると思えるかもしれない。しかし、この直観に従う必要はない。ウサギは結局のところキツネとウォンバットを識別できないのだから、キツネまたはウォンバットがいるといった選言的な表象内容を帰属させることに何の問題もない、と。

この応答は神経生理学の研究を参照することで動機付けられるかもしれない<sup>39</sup>。神経生理学者は、例えば、舌を伸ばしてハエなどの小動物を捕食するカエルが、実際にはどの程度の識別能力を持っているのかに関心をもっている。ネアンダーによると、カエルの視覚システムは目の前に小動物がいるという自然的記号を作り出しているわけではない。カエルの視覚システムを司るとされるのは中脳に位置するT(5)2という細胞なのだが、この細胞は視野を横切る虫状の運動（worm-like motion）に反応するという<sup>40</sup>。虫状の運動の担い手は栄養価のあるハエやクモといった小動物だけでなく、BB弾の運動のように実験室で人為的に作り出された刺激にも反応してしまう。こうした研究を背景として、ネアンダーは次のように論じる<sup>41</sup>。たしかに、カエルが生息している野生の環境では、虫状の運動は栄養価のある小動物の運動と外延が一致しているのかもしれない。しかし、実験室で明らかになったように、カエルの視覚システムは虫状の運動を識別する能力を有してはいるが、栄養価のある小動物を識別する能力は有していない。したがって、カエルは野生の環境においても、実際には栄養価のある小動物の運動を表象するというよりは、むしろ虫状の運動を表

<sup>39</sup> カエルの視覚システムについては Neander 2006a が有益なサーヴェイである。

<sup>40</sup> 1950年代になされたレトヴィンらの古典的研究は網膜神経節細胞（retinal ganglion cells）の働きに注目していた。しかし、ネアンダーによれば、彼らはカエルの視覚システムをかなり単純化しており、対象の種類を識別するのは実際には中脳の働きであることが最近になって分かってきた。

<sup>41</sup> Neander 2013, p. 31.

象しているのは明らかである、と。

カエルの知覚システムのあり方は、捕食者に対する過剰警戒の事例ではないが、過剰警戒の事例に対して応答する一つの自然な方針を示唆している。すなわち、ドレツキ型の目的意味論は知覚的表象の内容がもつ肌理の細かさを、感覚器官がもつ識別能力の程度に対応するように設定すればよいのではないかと。たしかに、ウサギやカエルの視覚システムはキツネがいることやハエがいることを表象できない。しかし、それは彼らの視覚システムがいわば単純であるからだ。感覚器官のメカニズムが複雑になって識別能力が向上した場合には、より肌理が細かい表象内容を獲得することになる。

しかし、この方針を一貫させようとする、ドレツキ型の目的意味論は厄介な問題に直面する。ウサギやカエルよりも高度な識別能力をもつと思しき感覚器官、例えば、人間の視覚システムに目を向けよう。ウサギやカエルほどでなくても、人間の視覚システムもまた、正常に働いているにも関わらず錯覚に陥る可能性を免れていない。図式的に述べるなら、視覚システムが故障しているわけではないのに、その設計上どうしても  $F$  と  $F'$  を識別できないということは十分に考えられる ( $F$  の候補としては、自然種や特定の対象などが考えられる)。そのため、直観的には目の前に  $F$  があるという表象内容を帰属したいのに、(DTS) に従うと、目の前に  $F$  or  $F'$  がある、といった表象内容しか帰属できなくなる。

この帰結は重要である。どうすればドレツキ型がこの問題に対処できるのかははっきりしない。ミリカン、ドレツキとは違うやり方で自然的記号を特徴づけることでこうした問題を回避する可能性を示唆しているが、ミリカンによる自然的記号の理論はドレツキの理論と比べるとスケッチの域を出ていないし、また、ミリカン自身も最終的にこの修正案を別の理由で拒否している<sup>42</sup>。ネアンダーは知覚過程とそれ以後の概念的な過程とを区別し、概念的表象に関しては識別能力によって許容されるよりも肌理の細かな内容をもちうるとし、知覚的表象に関してはそれを認めないことを提案している<sup>43</sup>。この提案は理解できる。知覚的表象と概念的表象の内容が何らかの点で異なることは、ほとんどの論者が認めよう。とはいえ、概念的表象がいかんにして識別能力より肌理の細かい内容をもちうるのかがいずれ説明されねばならないことには変わりない。本節の議論はドレツキ型の目的意味論を退ける決定的な議論ではないかもしれないが、ドレツキ型の目的意味論の問題点は明確になったと言える。

## 6.5 節 本章のまとめ

自然的記号と表象はそれぞれに特有の仕方で外界のあり方を表している。ドレツキの情報理論は自然的記号に関する確率論的に分析する。目的意味論は自然的記号の概念を基礎として外界についての表象能力を備えた心がいかにして出現しうるのかを論じている。これ

---

<sup>42</sup> Millikan 2004, pp. 75–77.

<sup>43</sup> Neander 2013, pp. 32–33.



らの組み合わせによって、ドレッキは外界のあり方を表す存在者についての包括的な理論を提示しようとしたのであった。

しかし、本章の議論が正しければ、ドレッキの包括的な理論にはいくつかの重大な問題に直面するだろう。まず、6.2.2 節の末尾で示したように、ドレッキは自然的記号が情報を担うかどうかを決める基準が話者の興味や関心に依存すると考えていたが、話者の興味や関心に依拠している自然的記号を自然主義的意味論の基礎に据えることは不適切ではないかという疑念が生じた。次に、ドレッキの目的意味論には、6.4.3 節で特定した問題、すなわち、捕食者に対する過剰警戒という諸事例を適切に扱うことができそうにないという問題がある。ドレッキ型の目的意味論がどのようにしてこの種の事例を扱うことができるのかははっきりしない。

これらの難点は目的意味論一般に共通しているわけではない。例えば、ミリカン型の目的意味論は、一見したところ、自然的記号の概念を利用しておらず、自然的記号の概念を特徴づけるという課題を免除されているように見える。また、そもそもミリカンの目的意味論は、過剰警戒の事例を適切に扱えないドレッキ型の理論に対する不満に（少なくとも部分的には）動機付けられており、実際に、過剰警戒の事例を適切に扱えるよう設計されている。次章では、このミリカン型の目的意味論を検討しよう。

## 7 章 生得的な表象内容

“You can’t always get what you want  
But if you try sometime  
You just might find  
That you get what you need”  
Sterelny 2001, p. 260

前章では、有機体が抱いている信念の内容を欲求の内容と独立に決定しようとするアプローチの一例としてドレッツキの議論を取り上げ、その問題点を指摘した。本章と次章では、欲求の内容を信念の内容と独立に決定しようとするアプローチを検討する。

### 7.1 節 欲求ベースの自然主義的意味論

自然主義的意味論にはさまざまなアプローチがある。目的意味論はその一種だが、目的意味論の中にもさまざまなバージョンが存在する。本章で取り上げるミリカン型はドレッツキ型とは別である。二つの目的意味論の共通点は起源説的な機能概念を利用するという点にのみ求められる。ドレッツキ型は信念ベースのアプローチを採用するのに対して、ミリカン型は欲求ベースのアプローチを採用する。

第一次近似としては、ミリカン型の目的意味論による欲求内容の定義は以下のようだ。

欲求Dの内容がpであるのはDの生物学的機能がpをもたらすことであるとき、そのときに限る<sup>1</sup>。

起源説的な機能概念がここでも使用されていると考えてよい。

しかし、この双条件はすべての欲求が何らかの生物学的な機能を持つと前提している。欲求の内容にはさまざまであり、どの欲求内容にも進化論的な起源があるとは考えにくいのではないか。例えば、ローリングストーンズの音楽を聴きたいといった欲求は、ごく最近になって出現したはずであり、遠い昔に選択圧のふるいにかけて生き残った欲求ではない。また、ゴールデンゲートブリッジから飛び降りて有名になりたいなど、欲求の中には、その欲求を抱いた有機体の生存とはほとんど関係ないものもあるだろう<sup>2</sup>。どうすれば上の双条件がこの種の事例を適切に扱えるのかは、まったく明らかではない。

これらの難点は真正のものだと思われる<sup>3</sup>。しかし、こうした難点にしばらく眼をつぶる

<sup>1</sup> Millikan 1986; Papineau 1993, p. 58.

<sup>2</sup> Crane 2003, pp. 190–191. cf. Sterelny 1990, p. 133.

<sup>3</sup> ミリカンとパピノーは上述の批判がありうることに気づいているので、このように言い切るのは本当はフェアではない。本章は彼らの著作全体がもつポテンシャルを一部切り出して検討しているだけである。

ならば、上の双条件にも使い道が見出せるかもしれない。例えば、比較的単純な生物が生得的に用いている信号の内容を決定するという目的に利用してはどうだろうか<sup>4</sup>。これはミリカンが実際に試みたことである。

動物行動学は、動物たちがさまざまな信号によってコミュニケーションしていることに注目してきた。しかし、そこにはコミュニケーションの内容が恣意的に与えられる危険性がある。例えば、蜜蜂は 8 の字を描くダンスをすることが知られているが、そのダンスは何を表しているのだろうか。長くつづける蜜蜂は「長旅で俺は疲れたよ」と言っているのだ、と解釈する人もいるかもしれない。こうした解釈が不適切である理由は、動物行動学の中では、擬人法に頼っているから、という以上の理由は見出しがたい。これに対して、以下で導入するミリカンの理論は、原則に基づいた仕方では蜜蜂のダンスが表す内容を特定するよう意図されている。動物行動学は、蜜蜂のダンスが餌のある方向と距離を表していることを明らかにした、などと言われる。その際に「表している」という意味論的概念が使われるわけだが、ミリカンの理論はこうした用法の内実を明らかにしようとするものだ。ミリカンの試みは、もし成功しているとすれば、興味深い結果と言えよう。

ミリカンの理論は複雑であり、彼女の立場が十分明確に定式化されることは、二次文献まで含めても稀である。考えてみれば奇妙なことだが、ミリカンに対する批判は、彼女の立場を十分明確に定式化しないまま様々な反例を持ち出すものさえある<sup>5</sup>。本章では、ミリカンの立場を明確に定式化し、利点と弱点を正確に見積もる。

本章の構成は以下のようなものである。7.2 節ではミリカン型の目的意味論の大枠を説明する。7.3 節では、ミリカン型は表象概念をあまりに単純な生命活動にまで拡張して適用しているという疑念に応答する。7.4 節では、ドレッキ型の目的意味論において深刻な問題だった過剰警戒の事例を、ミリカン型の目的意味論はより適切に扱える見込みがあると論じる。7.5 節では、ミリカン型の目的意味論が表象の内容を特定するために用いる手続きについて因果論的な観点からの批判を行い、その批判に対する二つの応答戦略を検討する。

## 7.2 節 ミリカン型の目的意味論

本節では、ミリカン型の目的意味論の大枠を説明する。あらかじめ断っておくと、本節で素描するミリカンの立場は 1980 年代に提案されたものであり、後年のミリカンが採用している立場と単純に同一視できない。しかし、ミリカンの基本的なアイデアを説明するには 80 年代の立場の方が分かりやすいので、テキスト解釈上の厳密さを犠牲にしても、ここ

---

<sup>4</sup> このような路線変更は退行的だと言われるかもしれないが、ドレッキ型の目的意味論もそれほど事情は変わらない。実際、先に述べたような批判は信念ベースの目的意味論にも当てはまる。例えば、すべての信念に進化論的な起源があるとは考えられない、バラク・オバマが第 44 代の合衆国大統領であるという信念は最近になって出現したものであり、遠い昔の選択圧のふりを生き延びたものではない、といった批判を向けることができるだろう。前章では、知覚的信念に話を限定することで、暗黙のうちにこの種の批判を回避していた。

<sup>5</sup> cf. Luntley 1999, sec. 7.5.

ではあえて 80 年代の立場に注目する。

ミリカンは、何かを表象するという性質は、ある対象がそれ単独で持てるようなものではない、と言う。彼女は、表象はその生産者 (producer) とそれを利用する消費者 (consumer) が協働することで成立するというモデルを立てる。表象の生産者は事実に対応している表象をもたらすことで消費者に対して情報伝達する、といった協働的なコミュニケーションの成立が、表象にとって本質的だとする。

ミリカンのアイデアは具体例を見た方が分かりやすい。彼女は蜜蜂のダンスを表象の例として好んでいるので、これを利用しよう。8 の字を描く蜜蜂のダンスは、巣から餌の位置への方向と距離を数量的に伝える信号になっていると言われる。ここでいう「餌の位置への方向」とは、太陽の位置を基準とした角度のことである。ダンサーはこの角度を測り、巣の中の垂直面でその角度だけ傾いたダンスをする。また、巣から餌への距離はダンスにおいて尻振りする部分の長さによって示される。こうしたダンスに接した蜜蜂の個体は、ダンスを解釈して採餌場所へと飛行する。蜜蜂のダンスによるコミュニケーションをミリカンの図式に当てはめると

表象：ダンス

生産者：ダンサー

消費者：ダンスを解釈して採餌場所へと飛行する蜜蜂

となる。生産者は、自身をもたらした表象を消費者が適切に利用することで利益を得る。表象は時間と空間に位置をもつ出来事トークンであり、これが消費者の行動を引きおこし、消費者は利益を得る。

しかし、消費者がダンスを「解釈する」とはどういうことなのか。たとえダンスが表象だとしても、その表象内容を一体どうやって特徴づければよいのだろうか。

ミリカンは二通りの答えを用意している。生産者がもたらす表象は、消費者の何らかの行動を引きおこす。消費者の行動は何らかの機能を担っている。そこで、生産者がもたらす表象は消費者に対して機能を果たすように指令していると考えれば、その指令が表象の内容である (指令的表象)。しかし、表象は指令的であると同時に記述的でもあるかもしれない。消費者の行動が担う機能が果たされるためには、正常な条件 (Normal condition) が成り立っていなければならない。この正常な条件こそが、生産者がもたらした表象の内容であると考えられる (記述的表象)<sup>6</sup>。

ここで重要な概念は「消費者の行動が担う機能」と「機能が果たされる正常な条件」の二つである。消費者の行動が担う機能はともかく、機能が果たされる正常な条件の方は解説を要する。ミリカンは、機能の多くは限られた条件下でしか果たされない、と述べてい

---

<sup>6</sup> 指令的表象と記述的表象はそれぞれ欲求と信念に相当する。単純な動物の場合には欲求と信念が未分化の状態にあり、同一の表象が指令的かつ記述的になっている、とミリカンは考える。

る。例えば、嘔吐反射は、胃の中に毒物があるときに限って、毒物を吐き出すという機能を果たすことができる。嘔吐反射は胃の中に毒物があるときにそれを吐き出すことによって歴史的に選択されてきた。この例では、胃の中に毒物があるという条件が、毒物を吐き出すという嘔吐反射の機能の正常な条件である。ミリカンが口を酸っぱくして強調するように、正常な条件とは必ずしも統計的にありふれた条件ではない。実際、胃の中に毒物があるのは稀である。正常な条件とは、むしろ、その機能をもつ形質が歴史的に選択されたことを説明するために言及せざるをえないような条件のことである<sup>7</sup>。

以上を踏まえて、蜜蜂のダンスの例を検討してみよう。ダンスが引きおこす消費者の行動は、ダンスの角度と尻振りの長さに対応する方向と距離を飛行することだが、こうした飛行の機能は蜜や花粉などの食料を巣へと持ち帰ることであろう。食料を採餌場所から持ち帰るのに成功することによって、飛行が選択されてきたからである。それでは、この機能の正常な条件とは何か。もっともらしい答えは、ダンスに対応する方向と距離に採餌場所があるということだろう<sup>8</sup>。ミリカンは正にそう考えている。そして、これこそが蜜蜂のダンスが表象する内容である。

しかし、ここまでの議論は、蜜蜂のダンスを安直にミリカンの図式に当てはめただけに見えるかもしれない。そもそも、意味論的な概念を用いずに、どうやって表象と生産者と消費者を切り分ければいいのか。この疑問に対する答えは次のようなものだろう<sup>9</sup>。最初のステップは、多様な入力を受け取って、その入力に応じて振舞いを変化させているメカニズムを見つけることである。そうしたメカニズムは消費者の候補であり、入力は表象の候補である。次に、この消費者候補となったメカニズムの機能とそれが働く条件を調べる。もし、入力に応じて振舞いを変化させることが適応であるならば、消費者候補は表象候補を、自身の機能が果たされる正常な条件を表していると考えられる。更に、その表象候補を生産しているメカニズムに目を向けよう。もし、消費者候補の機能が果たされる正常な条件が成立しているときに限って、表象候補を生産する生産者が利益を得るのであれば、生産者はその表象候補を生産することを機能としていると言える。これらのステップすべてがクリアされれば、表象と生産者と消費者を切り分けられたことになる。

ミリカンは、単純な動物が用いる信号は指令的表象と記述的表象のどちらともいえることを認めているが<sup>10</sup>、ここではドレッキ型の目的意味論との違いを強調するため、記述的表象に話を限定して、ミリカンの目的意味論を次のように定式化しておこう。

**ミリカン型の目的意味論 (MTS) :** 表象  $r$  が  $p$  を表象するのは次のときそのときに限る。消費者と協働している生産者が  $r$  をもたらし、かつ、 $p$  が成り立つことが、

<sup>7</sup> Millikan 1989, pp. 86–87.

<sup>8</sup> この点については7.5節でより丁寧に検討する。

<sup>9</sup> Shea 2005, pp. 34–37.

<sup>10</sup> ただし、7.5節では、別の理由で蜜蜂ダンスの表象内容は不確定だと論じることになる。

rによって引き起こされるrの消費者の行動が担う機能が果たされる正常な条件である。

(MTS)は表象が世界のあり方を誤って表象することを許容する。例えば、ダンスの角度と長さに対応する位置に食料がない場合、表象の消費者が機能を果たするための正常な条件が成立していないので、そのダンスのトークンは偽になる。

蜜蜂のダンスは、生物界における表象を用いたコミュニケーションの一例に過ぎない。本節の冒頭で述べたように、表象するとはそれ以上分割できない単一のシステムの働きではなく、表象と生産者と消費者という三つ組からなる複合的な構造の働きであるというのが、ミリカンの基本的な発想であった。蜜蜂のダンスという例の利点は、ダンサーとダンサーに接する個体は別であるために、この基本構図を適用しやすいことにある。しかし、表象の生産者と消費者は同じ個体の二つの部分であってもよい、とミリカンは考える。ミリカンはカエルの知覚を例に挙げる。カエルの視覚システム（生産者）は、網膜神経節細胞の発火という表象を作り出す。この表象は舌の運動系（消費者）の反応を引き起こす。舌を伸ばすという行動の機能は獲物を捕らえることであり、この機能が果たされる正常な条件は、目の前に黒くて運動する小粒（BB弾やコンピュータのスクリーン上のドット）があることではなく、獲物がいることだろう。なぜなら、目の前に黒くて運動する小粒があることは、獲物がいることと違って、舌の運動系の反応が歴史的に選択されてきたことを説明するためには言及しなくてもよい条件だからである。したがって、網膜神経節細胞の発火は獲物が存在することを表象する。こうした説明は、多くの生得的な認知メカニズムに適用できそうである。

### 7.3 節 単純な生命活動と表象

ミリカン型の目的意味論は単純すぎる生物の活動にも表象の概念を適用する羽目に陥るのではないかと懸念する人もいる<sup>11</sup>。たしかに、前節のようなやり方で表象とその生産者／消費者を切り出した場合、そうして切り出された表象は「表象」と呼べるのかどうかいささか疑わしい場合がある。例えば、ある種の嫌気性バクテリアは内部に磁石を備える。地球の磁界によって向きを変える磁石によってバクテリアは進行方向を決める。北半球では磁力線は下向きなので、酸素が少ない海中へ進むことになる。(MTS)によれば、このバクテリアの磁石の向きは、酸素の少ない領域がその先にあることを表象している。こうした原始的な生物の活動にまで表象が関わっているのだろうか。ミリカンは「表象」を不必要に拡大解釈しているのではないかと。

ミリカンによれば、この疑問は0を「自然数」とみなすのかという問題と似て、「表象」

---

<sup>11</sup> cf. Sterelny 1990, p. 49.

という言葉の用法の問題に回収できる<sup>12</sup>。お望みなら、ある種の出来事が表象という身分を獲得するためには、それはある程度の複雑さをもったメカニズムによって生産される必要がある、と考えるてもよい。ここでいう「複雑さ」というのは、外的環境についての多様な自然的情報をどのくらい統合しているのかどうかに関わる。磁石の向きのみに頼るバクテリアは、この意味ではあまりに単純である。カエルの眼球は外界から受ける光刺激に含まれる多くの情報を利用している。さらに高等な視覚システムは、網膜に対する光刺激のみを符号化するわけではなく、運動視差とか両眼視差、それに自分の身体がどういう姿勢にあるかという固有受容器からの情報なども利用することで恒常性のメカニズムを発達させている。こうしたスケールのどこかで外的環境に関する表象がはじめて成立する、と考えるてもよい。こうした修正はミリカンの立場にとって本質的ではない。

逆に、人間のように複雑な生命活動を行う生命体にも、(MTS)を適用しうる心的状態があるかもしれない。一つの候補は感情(emotion)である<sup>13</sup>。信念・欲求が志向性をもつという合意があるが、感情に関してはそのような合意はない。しかし、仮に感情にも志向性があるとしよう。例えば、目の前のヘビに対する恐怖を取り上げる。ヘビ恐怖が生得的であることは、起源説的な機能概念が利用できる可能性を開いている。それでは、ヘビに対する恐怖感情はどのような内容をもつのか。一つの可能性は、この動物は危険である、といった内容であろう。この動物が危険であるとき、この感情によって引き起こされる逃避行動は機能を正常に果たす。

## 7.4 節 比較研究

### 7.4.1 節 ドレッキ型との比較：過剰警戒の事例をめぐって

ドレッキ型とミリカン型の目的意味論の主要な違いは何だろうか。この疑問に対する本稿の答えは、ドレッキ型は信念ベースだがミリカン型は欲求ベースだ、というものである。しかし、文献では、ドレッキ型は自然的記号を作り出す装置から出発して表象内容を考察するがゆえに、表象の生産を重視しているのに対し、ミリカン型は表象の消費を重視している、という仕方でドレッキ型とミリカン型を対比させることがある<sup>14</sup>。

しかし、「表象の消費を重視する」というフレーズは多義的であり、この回答はミスリーディングだと思う。6.3 節で論じたように、表象は自然的記号のように宇宙に遍在していないわけだが、この違いを説明するため、我々は表象には利用者(消費者)が存在しなければならないという点を強調した。この意味では、ドレッキ型ですら表象の消費を重視している<sup>15</sup>。ミリカン型に固有な意味での「消費の重視」は、表象の内容を決定する要因は表

<sup>12</sup> cf. Millikan 2004, p. 158.

<sup>13</sup> この可能性は宮園健吾博士から示唆された。

<sup>14</sup> 戸田山 2014, pp. 219–221.

<sup>15</sup> Dretske 1995, p. 187n24.

象が発生する仕方ではなく利用される仕方である、という考え方である。ミリカン型は、表象の利用者がどのような機能を担っているのかをまず確定し、その機能が果たされる正常な条件を表象内容とみなす、という順序になっている。

ドレッキ型とミリカン型の違いを以上のように押さえたところで、次に、二つの理論では表象の扱い方にどのような差があるのかを検討しよう。ここで取り上げるべき事例は、捕食者に対する過剰警戒である。6.4.3 節で論じたように、過剰警戒の事例はドレッキ型の目的意味論（DTS）では扱いが難しい。しかし、（MTS）の下では過剰警戒の事例をうまく扱える見込みがある。

捕食者に対する過剰警戒の一例として、ミリカンはビーバーの水面叩き（water slap）を挙げる。ビーバーは他の個体に危険を伝えるためにしっぽで水面を強く叩いて水しぶきを上げる。動物がこのように危険を知らせるために特定の信号を用いる場合、コストの非対称性により、結果的には危険がないとしてもその種の信号を発しておく方が生存にとって有利に働きうる。危険があるときよりも危険がないときの方が信号がより頻繁に発せられることすらありえる。実際、ビーバーは警戒心が強いので水面叩きを実際には危険ではないものに対しても頻繁に行う。よって、ビーバーの水面叩きは危険の自然的記号とはいいいがたい。しかし、ウサギの視覚システムの場合と似て、こうした信号もまた、直観的には危険や捕食者の接近を意味するように思われる<sup>16</sup>。

7.2 節で論じたように、表象は生産者と消費者が存在することによって成立する。ビーバーの水面叩きの場合、表象の生産者はビーバーであり、表象の消費者は水面叩きを解釈して水中にもぐる他のビーバーだと考えられる。水面叩きが引き起こす消費者の行動（水中にもぐる）がもつ機能は、捕食者から逃避することであろう。捕食者から無事に逃れることによって、水面叩きという出来事によって引き起こされるこうした行動が選択されてきたからである。それでは、この機能の正常な条件とは何か。もっともらしい答えは、水面叩きがなされた付近に危険がある、といったことだろう。これこそが水面叩きの表象内容である。もし危険がないにも関わらずビーバーが水面を叩いたならば、表象の消費者が機能を果たすための正常な条件が成立していないので、その水面叩きの出来事トークンは、その付近に危険があると誤って表象していることになる。

重要なのは、ビーバーの水面叩きが捕食者の存在を表象するために、それが正確に捕食者の存在と共変している必要はないことである。表象内容は生産者が正常に働いている場合にそれと共変するような内容をもつわけではない。むしろ、表象の消費者から出発して、消費者の機能が果たされる正常な条件をもって表象の内容とする。ミリカン型の目的意味論をドレッキ型から分けるのはこの点である。ミリカン型にとって、表象の消費を重視するとは、表象が成立するためには消費者の存在が不可欠であるというだけでなく、表象の内容を決定するのは消費者の機能とそれに結び付けられた正常な条件である。

---

<sup>16</sup> Millikan 1989, p. 85.



#### 7.4.2 節 動物行動学のコミュニケーション理論との比較

7.2 節ではミリカンの目的意味論をスケッチしたが、その際に、蜜蜂のダンスを表象の具体例に用いた。このことから、ミリカンの目的意味論は動物行動学におけるコミュニケーション理論と関連があると考えるのは自然だと思われる。本節では、二つの理論の共通点と相違点を指摘する。

ミリカンの目的意味論とコミュニケーション理論の間には、コミュニケーションを意味論的語彙を用いずに特徴付けようとしているという共通点がある。動物行動学における「コミュニケーション」の定義（以下、(E) と呼ぶ）は、例えば以下のようである。

Aの行動がBの感覚器官を操作し、その結果としてBの行動が変わるとき、AとBのあいだにコミュニケーションが成立したという<sup>17</sup>。

しかし、(E) と (MTS) には隔たりもある。(E) はコミュニケーションの成立を定義しているだけであって、コミュニケーションされた内容を特定する方法には触れていない。したがって、(E) それ自体は、観察者が擬人法によって自分の好きなように動物のコミュニケーションに内容を与えることを許容する。例えば、尻振りを長くつづける蜜蜂は「長旅で俺は疲れたよ」と言っているのだ、と解釈する人もいるかもしれないが、こうした解釈は排除されない。これに対し、(MTS) は原則に基づいた仕方で表象内容を特定するべく意図されている。(MTS) はコミュニケーションの内容が恣意的に特定される危険性を回避する方法を提供すると期待される。

#### 7.5 節 正常な条件について

「正常な条件」という概念は、表象内容を特定する上で重要な役割を担っている。ミリカンの目的意味論にとってまさに要となる概念であり、慎重な検討を必要とする。本節の前半では、表象の消費者の機能が果たされる正常な条件に関するカミンズの提案を批判的に検討する。後半では、前半の問題提起に対する二通りの応答戦略を提示する。

##### 7.5.1 節 カミンズの提案とその問題点

7.2 節では、機能が果たされるための正常な条件とは、その機能をもつ形質が歴史的に選択されたことを説明するために言及せざるをえないような条件だと述べた。ミリカンはここでいう「説明」を、自然法則と正常な条件とからその機能の発現を演繹的に導出すること

---

<sup>17</sup> マクファーランド 1993, p. 291. なお、ここで引用した文章の執筆者はリチャード・ドーキンズである。

だと考える<sup>18</sup>。逆に言えば、機能が果たされる正常な条件とは、その機能の発現を自然法則と組み合わせることで導出するために最低限必要とされる条件ということである。

しかし、ここで一つの問題が生じる。蜜蜂のダンスの例に立ち戻ってみよう。7.3節では、ダンスに接した別の個体がなす行動の機能が果たされるための正常な条件、すなわち、ダンスの角度と尻振りの長さに対応する特定の方向と距離に餌があるということである、と述べた。さて、一つの問題というのは、食料を巣に持って帰ってくるという機能が果たされるための正常な条件が、特定の位置に餌があるということだけだという保証はないのではないかということである。餌を持ち帰ることができるために環境が満たすべき条件は他にもある。例えば、強風が吹いていないこと、天敵がいないこと、周辺の地域に DDT が散布されていないこと、などなど。消費者である蜜蜂の機能が果たされるための正常な条件が、このように様々な条件を含むのだとすると、(MTS) は蜜蜂のダンスがこれらの条件すべてを表象するという帰結をもたらす。しかし、これでは肌理が細かすぎるようにみえる。むしろ、ダンスは餌の位置だけを表象しているべきなのではないか？

この問題に関連して、カミンズが興味深いアナロジーを提示しているので、それを参照してみよう<sup>19</sup>。カミンズは、まず振り子の法則に言及する。 $l$ を糸の長さ、 $g$ を重力定数、 $T$ を周期とすると、これらの間の関係は

$$2\pi\sqrt{l/g} = T$$

という等式で表現できる。この法則は振り子の運動を説明する上で極めて重要である。ただし、この等式は、空気抵抗や摩擦を無視しているという意味で、理想化を含んでいる。実際には、空気抵抗や摩擦もまた、周期に因果的な影響を与えるからである。しかし、空気抵抗や摩擦のない状況に振り子は存在しうるのに対し、重力や糸のない状況では振り子という装置は存在しえないので、重力と糸の長さは、周期を決定する上での基本要因 (basic factor) になっている。さて、ここで蜜蜂のダンスに戻ろう。我々が直面していた問題は、特定の位置に餌があるという条件を、強風が吹いていないといった他の条件から区別する方法を探すことであった。この問題に関して、特定の位置に餌があるという条件は、餌を巣に持ち帰るという機能が果たされるための正常な条件における基本要因なのだ、とカミンズは示唆する。もしも、この示唆が正しいとすれば、表象内容の定義 (MTS) における「正常な条件」への言及を「正常な条件における基本要因」へと置き換えれば、蜜蜂のダンスは正に餌の位置だけを表象していることになる。

しかし、カミンズの議論には疑わしいところがある<sup>20</sup>。彼は実際には基本要因の概念を定

<sup>18</sup> cf. Millikan, 1984, pp. 33–34.

<sup>19</sup> Cummins 1989, pp. 78–80. カミンズは表象の歴史的説明は不適切であるとして、目的意味論を最終的には拒否するのだが、そのことは以下のアナロジーの興味深さを損なうものではない。カミンズによれば、ミリカン自身もこのアナロジーを認めているようである。op cit., p. 163n6.

<sup>20</sup> cf. Neander 2006b, pp. 387–388.

義しておらず、上のようなアナロジーしか提示していない。そのため、このアナロジーがどうして成立しているのか、特定の位置に餌があるという条件がどうして基本要因なのかについての正確な説明は与えられていない。特定の位置に餌があることだけでなく、強風が吹いていないことやDDTが散布されていないことなどの条件はすべて、蜜蜂が繁殖する上で満たされていなければならない必要条件になっている。これらの条件の間に非対称性を持ち込むのは容易ではないように思われる。

この状況は、マッチを擦ってマッチに火をつける、という因果論において持ち出される典型例と類比的なところがある。マッチを擦ったことは、マッチに火がつくことの原因である。しかし、周囲に酸素があることやマッチが湿っていないことも、マッチに火がついたことに因果的な影響を及ぼす。ここで、マッチを擦ったことはマッチに火がついたことの真の原因であって、他はマッチに火がついたことを可能にする条件 (enabler) に過ぎないという誘惑に駆られる。しかし、真の原因と可能にする条件を区別するのは容易ではない。両者の区別は、観察者の関心と相対的にしか成り立たないかもしれない。因果性の研究者の中には、そもそも因果論にとって何よりも重要な区別は、因果関係を単なる相関から区別することであって、真の原因と可能にする条件を区別することの重要性は低いと論じる人もいる<sup>21</sup>。

消費者の行動の機能が果たされる正常な条件という概念に関する以上の疑念が適切であるならば、特定の位置に餌があることだけを特別視するための根拠を我々はまだ手にしていない。もしも、ミリカンの意図が、蜜蜂のダンスが餌の位置（方向と距離）だけを意味・表象しているということを保証するような理論を作ることにあるならば、その課題は果たされていないということになりそうである。

## 7.5.2 節 二通りの応答戦略

では、こうした帰結が出てくることは、ミリカン型の目的意味論に欠陥があることを示しているのだろうか。そう結論するのは勇み足であるように思われる。ミリカン型の目的意味論の立場から、少なくとも二通りの応答戦略が考えられる。

### (1) 穏健な応答戦略：(MTS) は適用範囲を制限すれば正しい

そもそも、本章で展開してきたミリカン型の目的意味論は、単純な動物の生得的な認知メカニズムに適用範囲が限定されている。人間の思考は、ゴールデンゲートブリッジから飛び降りて有名になりたいといった欲求のように、進化の歴史とおおよそ関連のない内容をもちうる。この種の表象内容を扱うことは本節の枠組ではできないし、それは最初から意図していない。

従来、表象や志向性といった概念は人間の心的現象に適用されてきた。これに対し、

---

<sup>21</sup> 例えば、Sloman 2005, pp. 26–27 を参照。

(MTS)の適用範囲は、単純な生物の生得的な認知メカニズムに限定されており、表象や志向性といった哲学的概念を蜜蜂やゾウリムシといった生物にまで真剣に適用した哲学者はミリカン以前にはほとんどいなかった。そのため、蜜蜂のダンスが表象内容をもつとミリカンがいうとき、どの程度きめの細かい内容をダンスに帰属すべきかについては、それほど明確な基準があるわけではない。

そうであるならば、蜜蜂のダンスは消費者の行動の機能が果たされるための正常な条件すべてを多重に表象していると考えても、すぐさま不合理だと結論づけるべきでもないのではない。たしかに、これは蜜蜂のダンスに対してかなり肌理の細かい表象内容を帰属することである。しかし、この考え自体はそこまで奇妙ではないと論じる余地がある。絵画的表象と比較してみよう。例えば、目の前のコーヒーカップを一枚の写真に撮るとき、この写真はそのコーヒーカップが特定の色や大きさといった諸性質をもつことを多重に表象していると言っていることができる。これと類比的に、蜜蜂のダンスは特定の位置に餌があること、強風が吹いていないこと、DDT が散布されていないこと、などなどを多重に表象している、と考えることもできる。

## (2) ラディカルな応答戦略：(MTS)は適用範囲を制限しても間違っている

ミリカンのもともとの目論見は、単純な動物にも人間が信念や欲求を抱くのと似たような表象内容を帰属できると考えることで単純な動物と複雑な認知メカニズムを備えた人間との間の連続性を見出すことにあったと思われる。ところが、前節の議論が正しければ、動物の生得的な認知的メカニズムの表象内容は人間の抱く信念内容とはまったく異質であることになってしまう。この結果は単純な動物と複雑な認知メカニズムを備えた人間との間の連続性を見出すという当初の目論見を裏切っている。したがって、(MTS)の定式化は不適切であり、さらに改良すべきである。

(MTS)は、表象の内容を決定する上でもっぱら表象の消費者に注目している。ここに問題の元凶があるのではない。例えば、蜜蜂のダンサーは、特定の位置に餌があるという事実に対応しているだけでなく、強風が吹いていないといった事実にも対応したダンスを生産しなくては、その機能を果たしたことになる。それゆえに、蜜蜂のダンスは強風が吹いていないという内容をもつことになってしまうというのが7.4.1節のポイントであった。そこで、この問題事例に対処する一つの方針として、表象の生産者にも表象内容を決定する上で何らかの役割を負わせるということが考えられる。

それでは、どのようにして表象の生産者にも表象内容を決定する上で役割を負わせればよいのか。後年のミリカンは次のような考え方を示唆しているように読める<sup>22</sup>。まず、表象の生産者の機能は事実に対応する表象を消費者に伝達することだが、しかし、表象が正常に生産された場合には、その表象は自然的記号とみなすことができる。そして、自然的記号とみなした場合にはそれらは不自然な内容、蜜蜂のダンスの例でいえば、強風が吹いて

<sup>22</sup> Millikan 2004, pp. 85–86.

いないとかDDTが散布されていないといった不自然な内容をもつわけではない。したがって、先の穏健な応答戦略のように蜜蜂のダンスが連言的内容をもつと考える必要はない。

この応答戦略が本当に上手くいくのかどうかは私にはそれほど明らかではない。ここでは二つの点を指摘したい。まず、この応答戦略は7.2節で素描した基本的なアイデアからの逸脱を含んでいる。というのも、表象の生産者の機能は事実に対応する表象を消費者に伝達することだが、どの事実と対応することが重要であるかはもっぱら消費者が決めるというのが、ドレツキ型とミリカン型の目的意味論を分ける一つのポイントだったからである<sup>23</sup>。また、この応答戦略は自然的記号の概念を使用しているわけだが、なぜ蜜蜂のダンスが強風が吹いていないとかDDTが散布されていないといった内容の自然的記号ではないとされるのか不明瞭である。これらの間には、例えば、強風が吹いていたとすれば蜜蜂のダンスは〜でなかっただろう、といった反事実的な依存関係があるからである。自然的記号が情報を担うという関係をミリカンが本当のところどう分析しているのか、もっと明確にする必要がある<sup>24</sup>。こうした理由で、私はこちらの応答戦略には満足していない。

## 7.6 節 本章のまとめ

本章では、ミリカン型の目的意味論について考察した。7.2節では、ミリカン型の目的意味論のテーゼを(MTS)として定式化した。7.3節では、(MTS)は表象概念を拡大解釈しているという疑念に対して、その疑問は0を自然数とみなすかどうかで争うようなものだと言った。7.4節では(MTS)が(DTS)の鬼門である過剰警戒の事例をうまく扱える見込みがあると論じた。7.5節では、表象の消費者が機能を果たす正常な条件とは何かを問題視した。この点が明確にならない限り、(MTS)に従うと表象が過剰に肌理が細かい内容をもつことは避けられそうにないだろう。基本要因の概念に訴えるカミンズの議論は、基本要因の概念が定義されていないために不十分である。この困難には二通りの応答戦略がありうる。私自身は穏健な戦略を好む。すなわち、蜜蜂のダンスなどの生得的な表象は、実際に、過剰に肌理が細かい内容をもっている。そのように考えることを否定する積極的な理由は、特に見出せない。

<sup>23</sup> Millikan 1989, pp. 88–89.

<sup>24</sup> 自然的記号についてのミリカンの集中的な考察は Millikan 2004, chap. 3 を参照。ただし、そこでは本稿 6.2 節の (PTI) のような定式化はなされていない。私自身はミリカンの自然的記号についての見解を二つの点でうまく理解できないでいる。(1) ミリカンは (PTI) を批判する際に関連する参照クラスの不明瞭さを指摘しているが、関連する参照クラスをミリカン自身がどのようにして決定しているのかが分からない。(2) ミリカンはドレツキとは違って条件付き確率 1 の要求を自然的記号に課さないが、他方で、情報を担うという関係には推移性が成り立つという考え (ゼロックス原理) を擁護している。しかし、ドレツキは条件付き確率 1 の要求とゼロックス原理は一蓮托生の関係にあると考えている。両者の結びつきをミリカンがどのように考えているのかはよく分からない。Dretske 1981a, chap. 3; Millikan 2004, p. 53.

## 8 章 基本的な信念・欲求

“it is, for instance, possible to say that a chicken believes a certain sort of caterpillar to be poisonous, and mean by that merely that it abstains from eating such caterpillars on account of unpleasant experiences connected with them.”

Ramsey 1927, p. 40

### 8.1 節 導入

欲求ベースの自然主義的意味論は、欲求内容を信念内容と独立に決定し、しかる後に、明らかになった欲求内容を所与として信念内容を決定しようとする。このアプローチを機能の概念を持ち出すことなく、できるだけ首尾一貫した形で展開しようとした論者として、ジェイミー・ホワイトを挙げることができる。

ホワイトはまず、本稿ではこれまで疑うことなく済ませていた仮定に検討を加える。これまで信念と欲求のどちらか一方の内容を所与として他方の内容を決定する何らかの適切な原理があると仮定してきた。しかし、本当にそんな原理があるのかどうかは、きちんと検討されるべき事柄であった。ホワイトは、欲求内容を所与として信念内容を決定するための原理を提出し、ありうる批判に応答している。その上で、ホワイトは条件付けの概念を利用して欲求の内容を決定することを試みる。これら二つの部分をうまく組み合わせることができれば信念と欲求は自然化されるはずなのだが、はたしてそう上手くいくのかは慎重に検討する必要がある。

本章の構成は以下のようなものである。8.2 節では、欲求内容を所与として信念内容を決定するための原理としてホワイトが提出した原理 (R) を紹介する。8.3 節では、(R) に対する反例に対処できるよう (R) を修正する。8.4 節では、条件付けの概念を用いて欲求内容を決定するホワイトの方法を紹介し、8.5 節では、ホワイトの方法が抱えている限界を指摘する。8.6 節で、そこまでの議論を手短に要約した後、8.7 節では、ホワイトの失敗を踏まえて、新たな提案を行う。

### 8.2 節 原理 (R)

ホワイトは信念の真理条件を以下のように定式化する<sup>1</sup>。

(R) 信念の真理条件は、その信念が何らかの欲求と結びついて引きおこすである

---

<sup>1</sup> Whyte 1990, p. 150. ゴドフリー＝スミスは、ホワイトは後にこの定式化を撤回したと報告している。Godfrey-Smith 1996, p. 200n14. しかし、どのような理由で撤回したのかよく分からない。Whyte 1997 は依然として原理 (R) を擁護しようとしている。

う行動がその欲求を充実するのを保証する、そういう条件である<sup>2</sup>。

原理 (R) はラムジーに因んでそう呼ばれる。(R) のルーツはラムジーの論文「事実と命題」にさかのぼる。この原理がどういうアイデアを表現しているのか、そのままではやや分かりにくい、おおよそ次のように理解すればよい。

ある行動が、その行動を引き起こした欲求を充実するとき、その行動は成功したといえる。ところで、行動が成功するかどうかは、その行動を欲求とともに引き起こした信念が真であるかどうかにかかっている。例えば、体重が減って欲しいという欲求が運動を引き起こしたとする。ダイエットに成功するかどうかは、運動を引き起こした信念、つまり、運動すれば体重が減るとい信念が真であるかどうかにかかっている。運動すれば体重が減るのであれば、それは行動の成功を保証するだろう。つまり、運動すれば体重が減るということは、行動が成功するための十分条件である。このことを踏まえると、行動の成功を保証する条件（行動の成功条件）を信念の真理条件的内容とみなせるのではないか。おおよそ以上が (R) の表現しているアイデアである<sup>3</sup>。信念内容を行動の成功条件と同一視することから、(R) は「成功意味論 (success semantics)」とも呼ばれる。

それでは、原理 (R) は真だろうか。ブラックバーンによれば、信念に関する「おなじみの全体論 (familiar holism)」が (R) にとって問題になる<sup>4</sup>。ある行動の原因とみなすべき信念は一つだけではなく、関連する諸信念も原因とみなすべきである。そして、行動の原因となる諸信念の中に偽な信念が紛れ込んでいる場合には、その他の信念が真だったとしても、行動は成功しない可能性がある。

次のような例を考えよう。私はクッキーを食べたいという欲求を持っている。そして、この欲求と何らかの信念が、私によるなんらかの行動を引き起こしたとする。もし私の行動を引き起こした信念の真理条件的内容が戸棚の中にクッキーが入っていることであるならば、(R) によると、実際に戸棚の中にクッキーが入っているということは、私の行動がクッキーを食べたいという問題の欲求を充実することを保証するだろう。しかし、このような含意は成り立たない。なぜなら、クッキーを食べたいという欲求をある行動が充実することを保証する条件が、戸棚の中にクッキーがあるということだけであるはずがないからである。戸棚の中にクッキーが入っていると正しく信じていても、戸棚が置かれている場所について思い違いしていたとすれば、私は当の戸棚のところ（台所だとしよう）までたどり着くことができない。それゆえ、クッキーを食べたいという欲求を充実することはできないだろう。

---

<sup>2</sup> 原文も記しておく。"A belief's truth condition is that which guarantees the fulfilment of any desire by the action which that belief and desire would combine to cause."

<sup>3</sup> 一ノ瀬正樹教授から、こうした考え方はパースに通じるものがあるのではないか、という指摘をいただいた。私自身の力不足で、パースについて調べるどころまで手が回らなかったが、ラムジーの論文「事実と命題」はプラグマティズムに何度か言及しており、ラムジーがラッセル経由でパースから影響を受けた可能性はあると思う。

<sup>4</sup> Blackburn 2005, p. 23. cf. Mellor 2012, pp. 71–73.

ホワイトはこうした事例の存在に気づいており、以下のようにコメントしている<sup>5</sup>。たしかに、クッキーを食べたいという欲求の充実を保証するのに、戸棚の中にクッキーがあると正しく信じているだけでは不十分である。当の戸棚が台所にあることも正しく信じていなければ、クッキーを食べたいという欲求を充実することを保証できない。しかし、戸棚の中にクッキーがあるという信念は、クッキーを食べたいという欲求とは別の欲求と結びついて行動を引き起こすこともありうる。例えば、戸棚の中にはないお菓子を食べたいという欲求が、戸棚の中にクッキーがあるという信念と結びついて何らかの行動を引き起こしたとする。この欲求の充実を保証する条件の中には、何が含まれるだろうか。戸棚の中にクッキーがあるという条件は含まれるが、戸棚が台所にあるという条件は含まれないだろう。こうした観察を踏まえて、ホワイトは、さまざまな欲求を考慮に入れば、戸棚の中にクッキーがあること以外の余計な条件をはじき出すことができるはずだ、と主張する。そのための方法は、二つのステップに分けられる。まず、信念の集まりに対して、真理条件を割り当てる。次に、問題となっている信念の内容を決定するために、その信念が要素として含まれる任意の信念集合の真理条件に共通して含まれる条件を、その信念の内容とみなす。

以上の議論は、実質的に (R) を次のような原理に改訂することを提案していると考えてよいだろう<sup>6</sup>。

(R') 信念 B の真理条件は、B を要素として含む任意の信念集合  $\Gamma$  の真理条件に共通する条件である。ただし、信念集合  $\Gamma = \{B_1, B_2, \dots, B_n\}$  の真理条件とは  $B_1, B_2, \dots, B_n$  が何らかの欲求 D と結びついて引き起こすであろう行動が D を充実するのを保証する、そういう条件である。

(R') はブラックバーンが問題にするような事例を扱うことができる。

### 8.3 節 一切の障害はないという信念？

(R') に対する反例はないだろうか。次のような事例が考えられる<sup>7</sup>。私はクッキーを食べたいという欲求を持っており、戸棚の中にクッキーがあると正しく信じている。また、戸棚は台所にあるといった、戸棚やクッキーに関連する諸々の信念はすべて正しかったとす

<sup>5</sup> 実際、先の段落で示した例はホワイトが挙げているものである。ちなみに、ブラックバーンの例は、ケンブリッジに行きたい人が、ケンブリッジはロンドンの北北東にあると正しく信じているのだが、ロンドンから北北東に行くにはパディントンで電車に乗ればよいと間違えて信じているならば、ブリストルに着いてしまうだろう、というものである。ホワイトの例と趣旨は同じだが、イギリスの地名は日本人には馴染み薄いので、本文ではホワイトの例を借用した。

<sup>6</sup> この定式化は私が再構成したものであり、ホワイトは (R) をこのように修正してなどいない。しかし、この修正はホワイトの意図を適切に反映していると思う。

<sup>7</sup> この事例は Whyte 1997 による。



る。しかし、それにも関わらず、これらの信念や欲求によって引き起こされた行動は成功しなかった。その日は湿度があまりにも高かったため、湿気で戸棚の引き出しが膨らんで開かず、クッキーを食べられなかった。さて、このとき私がとった行動の成功条件の中には、湿気で戸棚が開かないということはない、という条件が含まれるはずである。しかし、私は湿気に関しては何も考えていなかったのだから、私の行動を引き起こした諸信念の集合の真理条件が、湿気で戸棚が開かないということはないという条件を含むというのは奇妙である。

この種の事例は、行動の成功条件の中には主体にとって思いもよらないような可能性の否定が含まれているということを示唆する。湿気によって戸棚が開かないといった可能性の他にも、例えば、クッキーが手に取った瞬間に爆発するとか、さまざまな奇抜な可能性を想像することができる。こうした奇抜な可能性の否定をどうやって信念内容から排除すればよいのか<sup>8</sup>。

ホワイトはこの問題に対して、上のケースにおいて私は、実質的に湿気で戸棚が開かないということはないということに相当する事柄を信じているのだ、と論じる。すなわち、私は、私が行動する上で一切の障害はないと信じて、クッキーを食べに戸棚へと向かっている、という。ホワイトは、行動の成功を保証するために設けたこの種の信念を「一切の障害はないという信念 (No impediment belief)」と呼ぶ。

ネネイはこの応答を批判して、一切の障害がないという信念が抱かれているとは限らない、と指摘する<sup>9</sup>。一切の障害がないという信念は行動を動機付けるかもしれないが、何らかの障害があるとは考えていない場合でも、その他の信念が行動を動機付けるには十分かもしれない<sup>10</sup>。実際、何らかの障害があるとは考えていなくても、私は台所の戸棚に向かうかもしれない。台所の戸棚の中にクッキーがあるという信念だけが私の行動を動機づけている、という可能性はあつてよいと思われる。

しかし、ネネイの指摘が正しいとしても、次のように考えれば (R) を適切に修正できると私は思う。前節の最後に提示した (R') の修正案は二つのステップから構成されていた。湿気の事例は信念集合に対して真理条件を割り当てるという一つ目のステップが反直観的な帰結をもたらすということを示唆している。だが、そもそもこのステップは必要なのだろうか。例えば、(R') を次のように修正するのはどうか。

(R'') 信念 B の真理条件は、B が何らかの欲求 D と結びついて引き起こすであろう行動が D を充実することを保証するのにつねに必要な条件である。

(R'') は、任意の欲求と結びついて引き起こされる行動が、その欲求の充実を保証する条

<sup>8</sup> この問題は 7.5.1 節でミリカン型の目的意味論において生じた問題とよく似ている。

<sup>9</sup> Nanay 2013, p. 154.

<sup>10</sup> 信念論理では  $Bp \leftrightarrow \text{not-}B(\text{not-}p)$  という置換が成り立つが、この原理が成り立つのは理想的な思考者だけだろう。現実には、すべての命題 p に関してこの原理が成り立つとは考えにくい。

件すべてに含まれ、その意味で、その欲求の充実を保証するのに必要不可欠な条件を、信念の内容とみなす、と主張する。湿気で戸棚が開かないということはないという条件は、戸棚のクッキーを食べることを保証する条件には含まれるが、戸棚の中にクッキーがあるという信念が引き起こす任意の行動の成功条件に含まれるわけではない。例えば、戸棚の中にはないお菓子を手に入れるためには、湿気で戸棚が開かないということは必要不可欠の条件でない。こうして、(R'') は「一切の障害はないという信念」という問題含みの存在に訴えることなく反例に対処できる<sup>11</sup>。

ここまでは順調だが、原理 (R'') を最終的に支持できるかどうかは実は微妙な問題をはらんでいる。原理 (R'') によって信念内容を決定しようとするなら、あらゆる欲求内容を信念とは独立に決定できる必要があると思われる (8.5 節)。しかし、まもなく明らかになるように、欲求内容を決定するそのような方法があるのかは疑わしい。

#### 8.4 節 条件付けの概念を用いた欲求内容の決定方法

本節では、起源説的な機能概念に依拠せずに欲求内容（欲求の充実条件）を決定するホワイトの方法を手短に紹介する<sup>12</sup>。その批判的な検討は次節で行う。

まず、ホワイトは次のような提案から出発する<sup>13</sup>。

(F<sub>0</sub>) 欲求 D の充実条件とは、それが成り立つときそのときに限って、欲求 D が消え去るであろう (would go away) 条件である。

「消え去るであろう条件」という回りくどい表現は、欲求内容の肌理の細かさを区別するのに役立つ。例えば、甘いものを食べたいという欲求は、クッキーを食べることで消え去るが、甘いものを食べたいという欲求はクッキーを食べたいという欲求と同じではない。なぜなら、甘いものを食べたいという欲求は、任意の甘いものを食べることによって消え去るであろうが、クッキーを食べたいという欲求は、クッキー以外の甘いものを食べることによって消え去らないだろうからである。甘いものを食べたいという欲求とクッキーを食べたいという欲求はこうして区別することができる。以上の議論を一般化すれば、肌理の細かさにおいて内容の異なる欲求をある程度区別できる見込みがある。

しかし、この提案はただちに多くの反論に出くわす。重要な反論は、欲求が充実するた

<sup>11</sup> 戸棚の中にクッキーがある、ということ以外にも何か余計な条件がすべての成功条件に含まれてしまう可能性はある。しかし、そういう条件は、戸棚の中にクッキーがあるという信念以外のどんな信念に関しても、それが引き起こす行動の成功条件に含まれると考えられる。もしそうだとすれば、その種の条件は端的に除外してしまえばよい。なお、原理 (R'') は  $2+2=4$  のような数学的信念は取り扱うことを意図していないことにも注意。

<sup>12</sup> Whyte 1991.

<sup>13</sup> この (F<sub>0</sub>) という名称は私が勝手につけたもので、ホワイトがこういう名称を使っているわけではない。以下同様である。

めの十分条件を適切に捉えていないというものである。たしかに、欲求が充実するときにはその欲求は消え去るだろうが、単に欲求が消え去るだけでは充実したとは言えない場合がある。

次のようなケースを考えよう。私はオムレツを食べたかったのだが、あなたが私の腹に強烈な一撃を加えたために、もう何も食べたくなくなった。もっと極端な例も考えられる。私はオムレツを食べたかったのだが、天井が崩れ落ちて私を押しつぶしたため、私の食欲は消え去った。もし欲求の充実条件が欲求が消え去るような条件だとすれば、これらのケースにおいて、私は腹に強烈な一撃を食らうこと、あるいは、天井が崩れ落ちることを欲していたということになってしまう。これは不合理である。したがって、欲求は単に消え去るだけでは充実したとは言えない。

この反例に対処するため、ホワイトは行動主義的な強化の概念に注目する<sup>14</sup>。一般に、欲求の充実をもたらす行動はその欲求の充実によって強化される。クッキーを食べることがクッキーを食べたいという欲求の内容であるのは、クッキーを食べることでクッキーを食べたいという欲求が消え去るだけでなく、クッキーを食べることでそこに至るまでの行動を強化するからである。他方で、腹に一撃を食らってクッキーを食べたいという欲求が消え去っても、いかなる行動も強化されない。それゆえ、クッキーを食べたいという欲求の充実条件は腹に一撃を食らうことではない。そこで、欲求の充実条件を次のように述べなおしてみる。

(F<sub>1</sub>) 欲求 D の充実条件とは、それが成り立つときそのときに限って、欲求 D が消え去るであろう条件であり、かつ、D を消し去る行動が強化されるような条件である。

しかし、この定式化でもまだ不十分だと思われる。我々は自分には識別できない二つの対象の一方だけを欲求することがある。例えば、私は洋梨を食べたいのだが、味音痴であるために和梨から洋梨を識別できないとする。このとき、洋梨を食べたいという私の欲求は、和梨を食べることによって消え去るであろうし、さらに、和梨を食べるに至るまでの一連の行動は強化されるであろう。だが、私が食べたかったのは洋梨であって和梨ではない。和梨を食べることで洋梨を食べたと思い込み、和梨を食べたいという欲求は消え去るかもしれないが、それは洋梨を食べたいという欲求が充実することではないだろう。

この反例に対して、ホワイトは次のような修正案を提示する<sup>15</sup>。

(F<sub>2</sub>) 欲求 D の充実条件とは、どれほど識別能力が向上したとしても、それが成り立つときそのときに限って、欲求 D が消え去るであろう条件であり、かつ、D

<sup>14</sup> Whyte 1991, pp. 67–68. cf. Chisholm 1957, p. 182. 強化については本稿 1.2.1 節を参照せよ。

<sup>15</sup> Whyte 1991, p. 72.

を消し去る行動が強化されるような条件である。

先に述べたように、現実の私は味音痴であり、洋梨を和梨から識別できない。しかし、私の味覚がより鋭敏で識別能力が向上していたとすれば、和梨を食べることで当の欲求が消え去ることはなかったであろう。したがって、和梨を食べることは私の欲するところではない。かくして、この修正案は問題となっていた反例にも対処できる。

## 8.5 節 前節の方法では扱えないさまざまな欲求

8.3 節では、欲求内容を所与として信念内容を決定する原理 (R'') を提案した。つづく 8.4 節では、欲求内容を信念内容とは独立に決定する原理 (F<sub>2</sub>) を提案した。それでは、これら二つのアイデアを結びつけば、信念・欲求の内容を決定する物理的事実は特定されたことになるだろうか。

そう結論づけるのは性急である。前節で紹介した原理 (F<sub>2</sub>) は、すべての欲求に対してその内容を決定するための方法ではありえない。なぜなら、欲求の中には、充実することが自分自身の生存と両立しないような欲求もあるからである。例えば、火葬して欲しいという欲求について考えよう。火葬するときには、すでに死んでいるので、そのような欲求は存在せず、消え去ることもできない。よって、この種の欲求の内容は、原理 (F<sub>2</sub>) では決定できないだろう。

他にも、次のようなケースが考えられる。私は友人からエスカルゴが美味であるという話を聞いて、エスカルゴを食べたくなったのだが、実際にレストランに行ってエスカルゴを注文して食べたところ、美味とはまったく思えず、二度とエスカルゴを食べるまい、と決意したとする。この場合、私はエスカルゴを食べたいという欲求をもち、エスカルゴを食べたことでその欲求は充実した。しかし、エスカルゴを食べて欲求は消え去ったものの、エスカルゴを食べるに至るまでの行動は強化されないだろう。よって、欲求が消え去るに至る行動が強化されることは、欲求が充実するための必要条件ではない<sup>16</sup>。さらに想像を膨らませれば、原理 (F<sub>2</sub>) では扱えないような欲求の事例は膨大な数にのぼるのではないかと思われる。

ホワイトはこの種の異論がありうることに気づいており、原理 (F<sub>2</sub>) は基本的欲求 (basic desire) の内容だけを決定するのだと述べる。彼によれば、火葬して欲しいといった欲求は基本的欲求ではない欲求である。したがって、原理 (F<sub>2</sub>) がこの欲求を扱えないことは驚くことではない、とホワイトは述べる。

<sup>16</sup> この事例は反例としてあまり説得力がないかもしれない。エスカルゴを食べたが不味くて二度と食べるまいと決意した場合、エスカルゴを食べたいという欲求が生じなくなるだけで、この欲求が生じたときにレストランに行って注文しない、ということには必ずしもならない。なんらかの理由で決意を翻してエスカルゴを食べたくなったなら、同じ行動をとるのではないか。このように論じることができるとすれば、行動は強化されたと言えるかもしれない。この点を指摘された Richard Dietz 博士に感謝する。

こうした開き直りは、いくつもの問題を含んでいると思われる。私がもっとも深刻だと考える問題は、信念内容と独立に内容を決定できる欲求が、欲求全体の一部にすぎない場合でも、原理（R'）と原理（F<sub>2</sub>）を結びつけることはできるのか、という問題である。原理（R'）は以下のように定式化されていた。

（R'）信念 B の真理条件は、B が何らかの欲求 D と結びついて引きおこすであろう行動が D を充実することを保証するのにつねに必要となる条件である。

問題は、この原理（R'）を経由して信念の内容を決定するには、基本的ではない欲求の内容もあらかじめ決定されていなければならないだろう、ということである。すでに確認したように、信念の内容を決定するには複数の欲求を参照しなければならない。例えば、戸棚の中にクッキーがあるという信念をその他の信念から区別して切り出してくるためには、クッキーが欲しいという欲求だけでなく、戸棚の中にお菓子を手に入れたい、という別の欲求も参照する必要があった。だが、戸棚の中にお菓子を手に入れたい欲求は基本的欲求とみなしてよいのかは、かなり微妙ではないだろうか。あるお菓子を単純に提示されただけでは、それが戸棚の中にある種類のお菓子なのかどうかを識別することはできないだろうからである。

十分な識別能力を持っている人ならば、その程度の識別は可能だと言われるかもしれない。そういう人は、戸棚の中にどの種類のお菓子が入っているのかも知っているであろう、と。この提案に応答するには、そもそも「識別（discrimination）」とはどういう意味かを反省する必要があると思う。われわれが求めているのは、心的現象に特有とされる志向性の解明なので、ここでいう「識別」には、志向的態度に言及しない解釈が与えられるべきである。さて、一般に「識別」には二つの解釈がある。一つの解釈は、非同一性の知識を獲得することである。つまり、aとbを識別するとは $a \neq b$ という知識を獲得することだと解釈する。しかし、知識は信念を含意する以上、この意味での「識別」は志向的とみなしうる。「識別」にはこれとは別の解釈もある。二種類の刺激に対して単なる偶然以上の確率で異なる反応を返すことのできるシステムは、それら二種類の刺激を「識別している」と言われうる<sup>17</sup>。この意味での「識別」は信念を前提していない。したがって、「識別能力が向上する」とは、別の世界では同じ反応を返す二種類の刺激に対して、その世界では異なる反応を返せる、ということではない。この非志向的な意味での識別に関する限り、戸棚の中に入っているお菓子とそうでないお菓子を識別することは無理ではないかと思う。よって、原理（R'）を原理（F<sub>2</sub>）に接続することはできない。

---

<sup>17</sup> cf. Williamson 2013, p. 5.

## 8.6 節 小休止

ホワイトの議論の検討はここまでに留め、本章の道のりを振り返っておこう。われわれは欲求ベースの自然主義的意味論を検討している。欲求内容を信念内容と独立に決定する方法には、生物学的機能の概念に訴えるものと条件付けの概念に訴えるものがある。本章では、後者の路線をホワイトの議論に沿って検討してきた。

欲求内容を決定する方法を提案するのに先立って、ホワイトは、欲求内容を所与として信念内容を決定するための原理 (R) を提出した。(R) はおなじみの全体論や、予期せぬ障害などいくつかの反例に出くわす。そこで、原理 (R) を原理 (R'') に置き換えることで、これらの反例に対処することを試みた。

では、欲求内容の方はどうやって決定するのか。欲求内容とは欲求の充実条件である。欲求の充実条件とは何か。ホワイトは、欲求が消え去る条件の中でも欲求の消去をもたらした行動を強化するような条件を、欲求内容とみなすことを提案する。原理 (F<sub>2</sub>) には、古典的な反例を退けられるという魅力がある。しかし、欲求が充実することは当の欲求が消え去ることを含意しない。火葬して欲しいといった欲求が反例となる。したがって、原理 (F<sub>2</sub>) は、どれほど上手くいっているとしても、ごく限られた範囲の欲求の充実条件しか決定できていない。そして、欲求の一部しか内容が決定できなければ、原理 (R'') によって信念内容を決定することもできない。こうして、ホワイトの欲求ベースの理論は暗礁に乗り上げる。ホワイトの議論に対する本稿の診断はおおよそ以上である。

しかし、こうした問題点を指摘しただけで、ホワイトの議論すべてをが台無しになるわけではないと思う。次節では、ホワイトの議論に多少の修正を加えることで、彼の目標を部分的にでも実現できないか検討する。

## 8.7 節 二つの提案

私見によれば、欲求ベースの自然主義的意味論には、信念ベースの理論と比べて、二つの点で優れている。一つの利点は発達論的な観点から、もう一つの利点は概念的な観点から見出すことができる。

信念内容を特定するのに先立って欲求内容を特定しようと試みることは、発達心理学の観点からは妥当な選択に思われる。例えば、マインドリーディングの研究者たちは、幼児は信念よりも欲求を先に学習しはじめるという点で、大まかな合意をみている。人によって欲求するものが違うことを、幼児は人によって信じていることが異なることより先に学ぶ。前者はおおよそ 18 ヶ月程度で、後者はおおよそ 4 年程度で学ぶようである<sup>18</sup>。こうした観察は、信念内容より欲求内容を優先する理由になるかもしれない。

また、6 章の議論が正しければ、信念ベースの理論には生命体の識別能力を超えた内容を

---

<sup>18</sup> Gopnik, Melzoff & Kuhl 1999.

帰属することに関する困難がある。信念ベースの理論では、知覚システムが正常に機能していてもFとF'を識別できないならば、Fがいるといった信念を帰属できなくなってしまう。しかし、検証主義的な立場に後退することを避けようとするならば、識別能力を超えた内容を抱くことが可能でなければならない<sup>19</sup>。欲求ベースの理論にはこの種の困難を克服できる見込みがある（7.4.1, 8.4 節）。

ホワイトの原理（F<sub>2</sub>）は、欲求全体のごく一部ではあっても、生得的ではない欲求の内容を決定する方法を提供している。条件付け・強化の概念を参照していることから示唆されるように、原理（F<sub>2</sub>）で取り扱うことのできる基本的欲求は、生物の生存と密接に関わっている。本節の前半（8.7.1 節）では、基本的欲求の範囲を策定することを試みる。

しかる後、私は基本的欲求の内容に対応づけられるような内容をもつ信念について論じる（8.7.2 節）。こうした信念を私は「基本的信念」と呼ぶ。基本的信念の内容は、基本的欲求の充実条件のみを所与として決定できるかもしれない。ホワイトは基本的欲求を所与として任意の信念に関して内容を特定しようと試みていたが、少なくとも、私の提案はずっと穏健である。

### 8.7.1 節 基本的欲求

8.4 節では、欲求内容を信念内容と独立に決定するための原理（F<sub>2</sub>）を紹介した。もっとも（F<sub>2</sub>）は基本的欲求という欲求全体の一部の内容だけを決定する方法だということが判明したので、以下では次のように言い換えることにする。

（F<sub>3</sub>）基本的欲求 D の充実条件とは、どれほど識別能力が向上したとしても、それが成り立つときそのときに限って、D が消え去るであろう条件であり、かつ、D を消し去る行動が強化されるような条件である。

しかし、次のような疑問がまだ残っている：（1）基本的欲求とはどのような種類の欲求なのか？（2）原理（F<sub>2</sub>）では扱うことのできなかつた欲求は、基本的でない欲求とみなしてよいのか？以下では、これら二つの疑問に簡単な回答を与える。

#### （1）基本的な欲求とはどのような欲求か

そもそも基本的欲求とはどのような欲求なのだろうか。これがある程度明確にならなければ、上の提案が適切かどうかを判定することはできない、と言われるかもしれない。私の考えは次のようである。

欲求と一口に言っても、ピアノ算術は矛盾していて欲しいといった抽象的なレベルの欲求から、食欲のように生物の生存と密接に関わるレベルの欲求まで、さまざまな欲求があ

---

<sup>19</sup> cf. Macdonald & Papineau 2006, p. 7.

る。この点に関して言えば、「基本的欲求」によって、ホワイトが生物の生存と密接に関わるレベルの欲求を意図しているのは明らかである。典型的には、水が欲しいという欲求、クッキーが欲しいという欲求、チョコレートが欲しいという欲求、などといった食欲が挙げられる<sup>20</sup>。この種の欲求は、形式的には、性質Fの一例を手に入れるという内容をもった欲求として特徴づけられると思われる。

基本的欲求の別の特徴として、信念を経由せずに形成されるという特徴も考えられる。信念を経由することで派生的に形成された欲求とはどのような欲求だろうか。単純な例としては（食欲とは関係ないが）、部長と話をしたい人は、部屋の向こう側にいる身なりの良い女性が部長だと信じているので、その身なりの良い女性と話したいという欲求を形成したとする<sup>21</sup>。身なりの良い女性と話したいという欲求は信念を経由して形成された欲求である。もともと、この場合、部長と話をしたいという欲求も、何らかの信念を経由して形成されていると思われる。他の信念をいっさい経由せずに形成されるのが基本的欲求である。

基本的欲求には身体的な変化が不可欠であるように思われる。ボブとメアリーが仲良くしてほしいと思っていて、自分の知らないところで二人が結婚していたとすると、私の欲求は充足されているが、何の身体的な変化も生じる必要がない。ホワイトは、識別能力の改善に訴えることでこうしたケースも射程に収められると言うかもしれないが、この応答は「識別能力」を拡大解釈するという代償を負っているように思われる。

基本的欲求には身体的な変化が伴うべきであるものの、基本的欲求を単なる生理的現象と同一視する必要はないと思う。空腹のような生理的現象は因果役割によって適切に特徴づけられる。例えば、空腹とは、カロリーのある食料を摂取することや腹を強打されることによって軽減・消滅するような性質であり、地球上の動物の場合は生理学的に解明可能なホメオスタシスのメカニズムによって実現されていると考えられる<sup>22</sup>。これに対して、8.4節の方法で扱える基本的欲求は、水とかクッキーといった特定の種類の食料に対する欲求である。われわれが何らかの基本的欲求を抱くときには空腹も生じるだろうが、そのことは基本的欲求と空腹の同一視を帰結しない。基本的欲求の内容は、因果役割機能主義では捉えることができない。

## (2) 基本的／基本的でないという区別

8.5 節では、原理（F<sub>2</sub>）に対する反例をいくつか挙げた。しかし、原理（F<sub>2</sub>）を基本的欲求に限定した（F<sub>3</sub>）には、それらの事例はもはや反例ではない、と言うことは許されるだろうか。

おそらく、許されるだろう。まず、火葬にして欲しいという欲求だが、この欲求は生存とは明らかに関係がなく、Fの一例を手に入れたいという形もしていない。よって、基本的欲求ではないと判断できる。この判断は直観的に十分理解可能だと思う。次に、エスカル

<sup>20</sup> Bermúdez 2003, p. 68.

<sup>21</sup> cf. Dretske 1988, pp. 127–128.

<sup>22</sup> cf. ジョンソン＝レアード 1989, p. 399.



ゴを食べてみたいという欲求だが、そもそもこうした欲求を人が抱くいかにしてか。例えば、テレビなどでエスカルゴの料理なるものの存在を知ること、おいしそうだから食べてみたい、といった経路でこの欲求が形成されたとすれば、それは基本的欲求ではない。このように考えれば、エスカルゴを食べてみたかったのだが、実際に食べたら二度と食べるものかと思った、という事例は、原理 (F<sub>2</sub>) を基本的欲求に限定するようにしておけば反例ではないと言える。

### 8.7.2 節 基本的信念

基本的欲求は、性質 F の一例を手に入れたいたいという内容をもった欲求だった。次に、基本的欲求に対応づけられるような信念を切り出すことを試みる。そうした信念を仮に「基本的信念」と呼ぶこととする。

基本的信念は、目の前にある何らかの対象にFという性質を帰属する指標的な信念である。したがって、基本的信念の内容は言語的には「これはFである (it is F)」などと表現される<sup>23</sup>。この種の信念は、Fを手に入れたいたいという欲求を充実する行動を引き起こす原因として不可欠であるように思われる<sup>24</sup>。

例えば、クッキーを食べたいといった欲求が行動を引き起こすときには、戸棚の中にクッキーがある、戸棚は台所にある、などといったさまざまな信念が因果的に媒介しており、この欲求が引き起こす一連の行動は欲求が充足されることで終結する。ところで、こうした一連の行動は、どんな場合であれ、目の前の対象 F として判断するという心的作用によって媒介されるのではないか。もしそうだとすれば、この心的作用を「これは F である」と言語的に表現される内容をもった信念の獲得とみなせるのではないか。

こうした仮説に基づいて、私は基本的信念の内容を決定するために、以下のような原理を提案する。

ある信念が「これは F である」と表現される内容をもつ基本的信念であるのは、F を手に入れたいたいという基本的欲求が充実するまでの間にこの欲求が引き起こすであろう一連の行動の原因の中には必ずその信念が含まれるとき、そのときに限る。

以下では、この仮説の内容を明確にすべく補助的な説明を加える。

一般に「これは F である」という発話は、発話の状況によって真理条件の内容が異なるとされる。発話の状況によって、「これ」の指示対象が異なるからである。それにも関わらず、「これは F である」という発話はそれぞれ、同一の文タイプの発話とみなすことができ

<sup>23</sup> 言語哲学では、指標詞を直示 (demonstration) を必要とするかどうかに応じて、直示詞と純粹指標詞に分ける場合があるが、ここでは「指標的」を、発話の文脈によって内容が変化するという程度の緩い意味で用いる。

<sup>24</sup> cf. Dretske 1981, pp. 202–203.

るだろう。同じようなことが基本的信念にも当てはまる。まず、基本的信念は、主体がそれを抱く時点や場所によって内容が異なる。それにも関わらず、それぞれの基本的信念は同種の実現者をもつだろう。

基本的信念が指標的である、という点に奇妙なところはない。実のところ、こうした指標的な側面は蜜蜂のダンスにすら見られる。7.2 節で取り上げたように、8 の字を描く蜜蜂のダンスは、巣から餌の位置へ方向と距離を数量的に伝えると言われる。ここで注目したいことは、8 の字の角度や長さがまったく同じ二つのダンスであっても、それらのダンスがなされた場所や時点によって、これらのダンスは異なる表象内容をもつ、ということである。一方のダンスは、巣 a から餌の位置へ方向と距離を伝えるのに対して、他方のダンスは、巣 b から餌の位置へ方向と距離を伝える。こうした観察は動物のコミュニケーションにおいて伝達される内容には指標的な側面があることを示しており、したがって、基本的信念が指標的であることは奇妙ではない。

ところで、基本的信念の内容は、ある対象にしかじかの性質が帰属するという趣旨の内容であるから、基本的信念を抱く主体には、個々の対象を切り出すための認知的なリソースが備わっていなければならない。ミリカンが取り上げている架空の毒ヘビの例を使ってこの点を説明しよう<sup>25</sup>。この毒ヘビは視覚によってラットを捉えて襲いかかり、嗅覚によって死にかけているラットの足跡を追い、触覚で頭を捉えて飲み込む。それぞれの感覚様相が別々の役割を割り当てられているため、自分が襲い、追いかけ、飲み込んでいるものを同じ対象として把握していない。この種の動物には、基本的信念を帰属することはできない。

それでは、「対象を切り出すための認知的なリソース」とは何か。私の考えでは、感覚器官に対して与えられた刺激をそのまま行動へと短絡させないような行動形成メカニズムが備えていること、がそれである。ミリカンが挙げる架空の毒ヘビは、各々の感覚器官に与えられた刺激を特定の行動へと短絡させているため、当然この条件を満たさない。それどころか、この毒ヘビには信念や欲求を帰属するということがそもそも意味をなさないようにすら思える。

私の考えでは、信念や欲求を帰属しうる動物はすべて、感覚器官に対して与えられた刺激をそのまま行動へと短絡させないという条件を満たしている。第2部（特に5章や8章）では自然主義的意味論の課題を、内容が何であるのかは分からないがとにかく信念や欲求をもつ主体が与えられたと仮定して、その主体が抱く信念や欲求の内容はどうやって特定されるのか、という形で考えてきた。信念や欲求を持つということは、信念や欲求の因果役割の担い手をもつということである。したがって、この仮定は、信念や欲求を帰属しうる主体が感覚器官に対する刺激を特定の行動へと短絡させることはないことを正当化するだろう。

---

<sup>25</sup> Millikan 2004, p. 173.

## 8.8 節 結論

われわれは、主体が抱いている信念・欲求の内容を決定する物理的事実とはどのようなものか、という問題に取り組んできた。この取り組みの発端は、因果役割機能主義のもとでは、行動を因果的に説明しうる信念・欲求の体系が原理的には複数存在するという過少決定の問題が生じたためであった。第 2 部では、この問題を克服するため、これまでに提案されてきた命題的態度に関する自然主義者の諸見解を比較検討してきた。比較のための分類枠として、自然主義者の諸見解を

- ・ 消去主義
- ・ 解釈理論
- ・ 信念ベースの理論
- ・ 欲求ベースの理論

の四種類に分類した。これら四種類の代表的な見解を検討し終えたいま、私の心は欲求ベースの理論に傾いている。

まず、欲求ベースの理論以外の三つの選択肢を、どのように批判してきたのかを振り返っておこう。4章で取り上げたクワインは、言語習熟に関する行動主義を前提として指示の不確定性や態度帰属の不確定性を導いた。クワインは、指示が不確定であること、態度帰属が不確定であることは、それらに関する事実問題が存在しないということだと考えており、それゆえに彼の立場は消去主義的である。しかし、多くの論者とともに私はこの結論を受け入れがたいと考える。クワインは行動主義的な前提に対して説得的な正当化を与えてはいない。それゆえに、指示の不確定性を帰結するクワインの議論は行動主義的な前提に対する帰謬法として理解することができる。

5章では、ルイスの解釈理論を一瞥したが、この立場は態度帰属を行うにあたって、すでに命題的態度を抱いている解釈者の存在を前提しているように思われる。それに加えて、高度な合理性を仮定しなければ命題的態度を帰属できず、動物にも備わるような内在的志向性を取り扱えないといった弱点を指摘できる。ルイスの解釈理論は、反省能力をもたない主体に信念・欲求を帰属する可能性に対して否定的であり、動物や幼児は解釈されるべき思考を持っていないことになるだろう。この批判は解釈理論にとって致命傷にはならないかもしれないが、これが命題的態度に関して自然主義者が採用すべき理論なのかどうかは疑問符がつく。6章以降で取り上げた、信念・欲求に関するボトムアップの研究を自然主義者は追及すべきである。

6章では、信念ベースの理論の一例として、ドレッキ型の目的意味論を検討した。ドレッキ型の目的意味論は、三種類の批判に直面する。最初の二つの批判は致命的ではないが、三番目の批判は重要である。過剰警戒の事例は、生命体がしばしば外界の認知に対して大

きなコストを負担しようとしないうことを教えるが、ドレツキ型の目的意味論では、生命体に帰属する信念内容はその生命体の知覚システムによって識別できる範囲に限られる。この帰結は、自然選択による進化が最適にデザインされた外界の認知能力をもった生物を生み出すという非現実的な前提を抜きにしては成り立たない。現実には、人間も含めてどんな生命体であれ、知覚システムが正常に機能していても誤りを犯すことがありうる。それゆえ、ドレツキ型の目的意味論は支持できない。

こうして自然主義者が採用する三つの選択肢を退けた後、7章と8章では残された最後の選択肢である欲求ベースの理論の検討に向かった。7章で検討したミリカン型の目的意味論には、ドレツキ型の目的意味論に向けられた重要な反例である過剰警戒の事例にも適切に対処できる見込みがある、という長所がある。ミリカン型の目的意味論は、動物のコミュニケーション行動や生得的な認知メカニズムに帰属される内容を決定するという目的には利用できる可能性が高い。

7章は動物のコミュニケーション行動を検討するのに留まることは断っておかねばならない。たしかに、動物のコミュニケーション行動において伝達される内容の分析と、信念・欲求帰属の分析はまったく無関係というわけではない。信念・欲求の帰属は行動を説明する文脈で価値をもつが、それと同様に、動物のコミュニケーション行動において何らかの内容が伝達されているとみなすことは、複雑な行動を説明する上で有用でありうる。ミリカン型の目的意味論が信念・欲求に関しては何が言えるのかをほとんど検討していないという点で、7章の議論はミリカン型の目的意味論に対して不十分ではある。この点について少し弁明すると、7章の議論が不十分であることは、ミリカン型の目的意味論が起源説的な機能概念に依拠していることに起因している。起源説的な機能概念を採用した場合、自然選択を経て形成ないし維持されていない形質がもたらす結果を、その形質の機能として割り当てることはできない。この特徴は自然主義的意味論という企てを完遂する上では弱点となりうる。というのも、人間は新奇で奇怪な信念や欲求を抱くものだからである。7.1節では、ゴールデンゲートブリッジから飛び降りて有名になりたい欲求、などという極端な例を挙げたが、ホワイトチョコレートが欲しいといった欲求に対してすら自然選択が働いたと考える理由はない。こうした観察は、目的意味論を諦める理由としては弱いだろうが、それでも、少なからぬ哲学者たちに、信念・欲求を自然化するという目的のために起源説的な機能概念に依拠するという選択肢を棄てさせる理由にもなってきた<sup>26</sup>。

そこで、8章では、ホワイトによる欲求ベースの理論を批判的に検討した。形質の生物学的機能に注目するミリカン型の目的意味論と違い、ホワイトは条件付け強化の概念に注目して欲求の内容を決定することを提案した。この方法により、基本的欲求という一部の欲求に関しては内容が決定する。問題は、それだけでは基本的でない欲求の内容や、信念の内容までは決定しないということである(8.5節)。しかし、こうした問題点があることは、ホワイトが欲求に関して述べたことがすべて無駄だったことを意味しない。8.7節では、基

<sup>26</sup> cf. Blackburn 2005; Fodor 1987, chap. 4; Whyte 1993

本的信念という一部の信念に対してではあるが、内容を決定する方法を提案した。

7章と8章の議論が成功しているとするれば、動物のコミュニケーション行動、そして信念と欲求のうち基本的な部分に関しては内容を決定できたことになる。つまり、少なくとも志向性をもつ心的状態の一部は、それがどのような物理的事実によって成立しているのかが解明されたことになり、穏健な自然主義は擁護された。

われわれは、主体が抱いている信念・欲求の内容を決定する物理的事実とはどのようなものか、という問題に取り組んできた。どうすればこのプロジェクトを完遂できるのか、その見通しは現時点では明らかではない。しかし、ごく一部とはいえ、命題的態度のもつ志向性に関して、自然主義者がそれを良心の呵責なく取り扱うための道筋がつけられたことは収穫である。このことを確認したところで、議論を終えましょう。

## 文献表

本文中で文献を指示するときには、著者名と初出の出版年を用いる。著者名の後のカッコ内は初出の出版年である。ただし、論文の場合には再録された論文集の出版年を別個に記載していることがある。その場合、本文中で文献を引用するときのページ数は再録版に依拠する。

## 英語文献

- Armstrong, David (1968). *A Materialist Theory of the Mind*, Routledge.
- (1973). *Belief, Truth and Knowledge*, Cambridge University Press.
- (1983). “Indeterminism, proximal stimuli, and perception,” *The Behavioral and Brain Sciences* 6, pp. 64-65.
- Barwise, Jon & Jerry Seligman (1997). *Information Flow*, Cambridge University Press.
- Bermúdez, José (2003). *Thinking without Words*, Oxford University Press.
- Bird, Alexander (2007). *Nature’s Metaphysics*, Oxford University Press.
- Blackburn, Simon (1984). *Spreading the Word*, Oxford University Press.
- (1991). “Losing your mind: physics, identity and folk burglar prevention,” in his *Essays in Quasi-Realism*, Oxford University Press, (1993), pp. 229–254.
- (2005). “Success semantics,” in Hallvard Lillehammer & David Mellor eds., *Ramsey’s Legacy*, Oxford University Press, pp. 22-36.
- Block, Ned (1997). “Functionalism,” in his *Consciousness, Function, and Representation: Collected Papers*, MIT Press, (2007), pp. 15–26.
- Block, Ned & Jerry Fodor (1972). “What is wrong with functionalism” in Ned Block, *Consciousness, Function, and Representation: Collected Papers*, MIT Press, (2007), pp. 45–61.
- Braddon-Mitchell, David & Frank Jackson (2007). *Philosophy of Mind and Cognition: An Introduction*, 2nd edition, Blackwell.
- Braddon-Mitchell, David & Nola, Robert eds. (2009). *Conceptual Analysis and Philosophical Naturalism*, MIT Press.
- Bremer, Manuel (2003). “Do logical truths carry information?,” *Minds and Machines* 13, pp. 567–575.
- Byrne, Alex (2007). “Soames on Quine and Davidson,” *Philosophical Studies* 135, pp. 439–449.
- Carruthers, Peter (2002). “The cognitive functions of language,” *Behavioral and Brain Sciences* 25, pp. 657–726.
- Chisholm, Roderick (1957). *Perceiving: a Philosophical Study*, Cornell University Press.
- Churchland, Paul (1988). *Matter and Consciousness*, Revised Edition, MIT Press.

- (1992). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press.
- Churchland, Patricia (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press.
- Chomsky, Noam (1959). “A Review of B. F. Skinner's *Verbal Behavior*,” *Language* 35, pp. 26–58.
- Cowie, Fiona (1999). *What's within?*, Oxford University Press.
- (2008). “Innateness and language,” in *Stanford Encyclopedia of Philosophy*, available online at: <http://plato.stanford.edu/entries/innateness-language/>.
- Crane, Tim (2001). *Elements of Mind*, Oxford University Press.
- (2003). *The Mechanical Mind*, 2nd edition, Routledge.
- Cummins, Robert (1983). *The Nature of Psychological Explanation*, MIT Press.
- (1989). *Meaning and Mental Representation*, MIT Press.
- Davidson, Donald (1963). “Action, cause and reason,” in his *Essays on Actions and Events*, Oxford University Press, (1980), pp. 3–20.
- (1969). “The individuation of events,” in his *Essays on Actions and Events*, Oxford University Press, (1980), pp. 163–180.
- (1973). “Radical interpretation,” in his *Inquiries into Truth and Interpretation*, Oxford University Press, (1984), pp. 125–139.
- (1986). “Deception and division,” in his *Problems of Rationality*, Oxford University Press, (2004), pp. 199–212.
- (1995). “Could there be a science of rationality,” in his *Problems of Rationality*, Oxford University Press, (2004), pp. 117–134.
- Davidson, Donald & Patrick Suppes (1956). “A finistic axiomatization of subjective probability and utility,” *Econometrica* 24, pp. 264–275.
- Dennett, Daniel (1978). “Skinner skinned,” in his *Brainstorms*, MIT Press, pp. 53–70.
- (1984). “Foreword,” in Ruth Millikan, *Language, Thought and Other Biological Categories*, MIT Press (1984), pp. ix–x.
- (1995). *Darwin's Dangerous Idea: Evolution and Meaning of Life*, Simon & Schuster.
- Devitt, Michael (1997). *Realism and Truth*, Paperback edition, Princeton University Press.
- Devitt, Michael & Kim Sterelny (1999). *Language and Reality*, 2nd edition, MIT Press.
- Dretske, Fred (1971). “Conclusive reasons,” in his *Perception, Knowledge and Belief*, Cambridge University Press, (2000), pp. 3–29.
- (1977). “Referring to events,” in *Midwest Studies in Philosophy*, pp. 90–99.
- (1981a). *Knowledge and the Flow of Information*, MIT Press.
- (1981b). “The pragmatic dimension of knowledge,” in his *Perception, Knowledge and Belief*, Cambridge University Press, (2000), pp. 48–63.

- (1983). “Author’s response,” *The Behavioral and Brain Sciences* 6, pp. 82–90.
- (1988). *Explaining Behavior: Reasons in a World of Causes*, MIT Press.
- (1990). “The epistemology of belief,” in his *Perception, Knowledge and Belief*, Cambridge University Press, (2000), pp. 64–79.
- (1991). “Two conceptions of knowledge: Rational vs. reliable belief,” in his *Perception, Knowledge and Belief*, Cambridge University Press, (2000), pp. 80–93.
- (1994). “If you can’t make one, you don’t know how it works,” in his *Perception, Knowledge and Belief*, Cambridge University Press, (2000), pp. 208–226.
- (2008). “Epistemology and information,” in Pieter Adriaans & Johan van Benthem eds., *Philosophy of Information*, Elsevier, pp. 29–47.
- Enç, Berent (1983). “In defense of identity theory,” *The Journal of Philosophy* 80, pp. 279–298.
- Evans, Gareth (1975). “Identity and predication,” in his *Collected Papers*, Oxford University Press, (1985), pp. 25–48.
- (1981). “Semantic theory and tacit knowledge,” in his *Collected Papers*, Oxford University Press, (1985), pp. 322–342.
- Field, Hartry (1973). “Theory change and the indeterminacy of reference,” *The Journal of Philosophy* 70, pp. 462–481.
- (1978). “Mental representation,” *Erkenntnis* 13, pp. 9–61.
- (1986). “Critical notice: Robert Stalnaker, *Inquiry*,” *Philosophy of Science* 53, pp. 425–448.
- Fish, William (2009). *Perception, Hallucination and Illusion*, Oxford University Press.
- Fodor, Jerry (1975). *Language of Thought*, MIT Press.
- (1987). *Psychosemantics*, MIT Press.
- Friedman, Michael (1975). “Physicalism and the indeterminacy of translation,” *Noûs* 9, pp. 353–374.
- Gibson jr., Roger (2004). “Quine’s behaviorism cum empiricism,” in Roger Gibson jr. ed. *The Cambridge Companion to Quine*, Cambridge University Press, pp. 181–199.
- Godfrey-Smith, Peter (1989). “Misinformation,” *Canadian Journal of Philosophy* 19, pp. 533–550.
- (1992). “Indication and adaptation,” in *Synthese* 92, pp. 283–312.
- (1996). *Complexity and the Function of Mind in Nature*, Cambridge University Press.
- Gopnik, Alison, Andrew Meltzoff, and Patricia Kuhl (1999), *The Scientist in the Crib*, William Morrow.
- Gozzano, Simone & Christopher Hill (2012), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge University Press.
- Graham, George (2015). “Behaviorism,” in *Stanford Encyclopedia of Philosophy*.  
<http://plato.stanford.edu/entries/behaviorism/>
- Hacker, Peter (2006). “Soames’ history of analytic philosophy,” *The Philosophical Quarterly* 56, pp. 121–131.



- Hall, Ned (2001). "Review of Helen Steward, *The Ontology of Mind*," *Mind* 110, pp. 1123–1127.
- Hawthorne, John & David Manley (2005). "Review of Stephen Mumford's *Dispositions*," *Noûs* 39, pp.179–195.
- Hookway, Christopher (1988). *Quine: Language, Experience and Reality*, Polity Press.
- Jackson, Frank, Pargetter, Robert & Prior, Elizabeth (1982). "Functionalism and type-type identity theories," *Philosophical Studies* 42, pp. 209–225.
- Kim, Jaegwon (1976). "Events as property exemplification," in his *Supervenience and Mind*, Cambridge University Press, (1993), pp. 33–52.
- (1998). *Mind in a Physical World*, MIT Press.
- (2006). *Philosophy of Mind*, 2nd Edition, Westview Press.
- Kripke, Saul (1980). *Naming and Necessity*, Harvard University Press..
- Laurence, Stephen and Eric Margolis (2001). "The poverty of the stimulus argument," *British Journal for Philosophy of Science* 52, pp. 217–276.
- Lewis, David (1966). "An argument for identity theory," in his *Philosophical Papers: Vol. 1*, Oxford University Press, (1983), pp. 97–107.
- (1970). "How to define theoretical terms," in his *Philosophical Papers: Vol. 1*, Oxford University Press, (1983), pp. 78–95.
- (1972). "Psychophysical and theoretical identification" in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, (1999), pp. 248–261.
- (1973). *Counterfactuals*, Blackwell.
- (1974). "Radical interpretation," in his *Philosophical Papers: Vol. 1*, Oxford University Press, (1983), pp. 108–121.
- (1980). "Mad pains and martian pains," in his *Philosophical Papers: Vol. 1*, Oxford University Press, (1983), pp. 122–132.
- (1983). "New work for a theory of universals," in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, (1999), pp. 8–55.
- (1986). *On the Plurality of Worlds*, Blackwell.
- (1994). "Reduction of mind," in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, (1999), pp. 291–324.
- (1997). "Finkish dispositions," in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, (1999), pp 133–151.
- Loewer, Barry (1997). "A guide to naturalizing semantics," in Bob Hale & Crispin Wright eds., *A Companion to the Philosophy of Language*, Blackwell, pp. 108–126.
- Luntley, Michael (1999). *Contemporary Philosophy of Thought*, Wiley-Blackwell.
- Lycan, William (1987). *Consciousness*, MIT Press.
- Machery, Edouard (2009). *Doing without Concepts*, Oxford University Press.

- MacCorquodale, Kenneth (1970). "On Chomsky's review of Skinner's *Verbal Behavior*," *Journal of Experimental Analysis of Behavior* 13, pp. 83–99.
- McCulloch, George (1995). *The Mind and Its World*, Routledge.
- Macdonald, Graham & David Papineau (2006). "Introduction: Prospects and problems for teleosemantics," in Graham Macdonald & David Papineau eds., *Teleosemantics*, Oxford University Press, pp. 1–22.
- McLaughlin, Peter (2001). *What Functions Explain*, Cambridge University Press.
- Mellor, David (1974). "In defence of dispositions," *The Philosophical Review* 83, pp. 157–181.
- (2012). "Successful semantics," in his *Mind, Meaning and Reality*, Oxford University Press, pp. 60–77.
- Miller, Alexander (2007). *Philosophy of Language*, 2nd edition, Routledge.
- Millikan, Ruth (1984). *Language, Thought and Other Biological Categories*, MIT Press.
- (1986). "Thoughts without laws," in her *White Queen Psychology and Other Essays for Alice*, MIT Press (1993), pp. 51–82.
- (1989). "Biosemantics," in her *White Queen Psychology and Other Essays for Alice*, MIT Press (1993), pp. 83–101.
- (2004). *Varieties of Meaning*, MIT Press.
- Morton, Peter (1997). *A Historical Introduction to the Philosophy of Mind*, Broadview Press.
- Mumford, Stephen (1998). *Dispositions*, Oxford University Press.
- Nanay, Bence (2013). "Success semantics: the sequel," *Philosophical Studies*, pp. 151–165.
- (2014). "Teleosemantics without etiology" *Philosophy of Science* 81, pp. 798–810.
- Neale, Stephen (1995). "The philosophical significance of Gödel's slingshot," *Mind* 104, pp. 761–825.
- Neander, Karen (1997). "The function of cognition: Godfrey-Smith's environmental complexity thesis," *Biology and Philosophy* 12, pp. 567–580.
- (2006a). "Content for cognitive science," in Graham Macdonald & David Papineau eds., *Teleosemantics*, Oxford University Press, pp. 167–194.
- (2006b). "Naturalistic theories of reference," in Michael Devitt & Richard Hanley eds., *The Blackwell Guide to Philosophy of Language*, pp. 374–391.
- (2013). "Toward an informational teleosemantics," in D. Ryder et al ed., *Millikan and Her Critics*, Wiley-Blackwell, pp. 21–36.
- Papineau, David (1993). *Philosophical Naturalism*, Blackwell.
- (2015). "Naturalism," in *Stanford Encyclopedia of Philosophy*.  
<http://plato.stanford.edu/entries/naturalism/>
- Paul, L. A. & Ned Hall (2013). *Causation: A User's Guide*, Oxford University press.
- Pinker, Steven (1994). *The Language Instinct*, Harper Collins.

- (2005). “So how does the mind work?” *Mind & Language* 20, pp.1–24.
- Prior, Elizabeth, Robert Pargetter & Frank Jackson (1982). “Three theses of dispositons,” *American Philosophical Quarterly* 19, pp. 251–257.
- Putnam, Hilary (1970). “On properties” in his *Mathematics, Matter and Method*, Cambridge University Press, (1975), pp. 305–322.
- (1983). *Realism and Reason*, Cambridge University Press.
- Quine, Willard (1960). *Word and Object*, MIT Press.
- (1969a). *Ontological Relativity and Other Essays*, Columbia University Press.
- (1969b). “Linguistics and philosophy,” in his *The Ways of Paradox and Other Essays*, MIT Press, (1976), pp. 56–58.
- (1970). “On the reasons for indeterminacy of translation,” *The Journal of Philosophy*, pp. 178–183.
- (1987). “Indeterminacy of translation again” *The Journal of Philosophy*, pp. 5–10.
- (1992). *Pursuit of Truth*, 2nd edition, Harvard University Press.
- Ramsey, Frank (1926). “Truth and probability” in his *Philosophical Papers*, Cambridge University Press, (1990), pp. 52–94.
- (1927). “Facts and propositions” in his *Philosophical Papers*, Cambridge University Press, (1990), pp. 34–51.
- (1929). “Theories” in his *Philosophical Papers*, Cambridge University Press, (1990), pp. 112–136.
- Rey, Georges (1997). *Contemporary Philosophy of Mind*, Blackwell.
- Ryle, Gilbert (2009). *The Concept of Mind*, 60th Anniversary Edition, Routledge.
- Schaffer, Jonathan (2003). “Overdetermining causes,” *Philosophical Studies* 114, pp. 23–45.
- Shea, Nicholas (2004). *On Millikan*, Wadsworth.
- Simon, Herbert (1996). *The Sciences of the Artificial*, 3rd edition, MIT Press.
- Sloman, Steven (2005). *Causal Models*, Oxford University Press.
- Soames, Scott (2003). *Philosophical Analysis in the Twentieth Century, Volume 2: The Age of Meaning*, Princeton University Press.
- (2006). “Hacker’s complaint,” *The Philosophical Quarterly* 56, pp. 426–435.
- (2015). “David Lewis’s place in analytic philosophy,” in Barry Loewer & Jonathan Schaffer eds., *A Companion to David Lewis*, John Wiley & Sons, pp. 80–98.
- Stalnaker, Robert (1976). “Propositions,” in Alfred MacKay & Daniel Merrill eds, *Issues in the Philosophy of Language*, Yale University Press, pp. 79–91.
- (1984). *Inquiry*, MIT Press.
- (1997). “Reference and necessity,” in Bob Hale & Crispin Wright eds., *A Companion to the Philosophy of Language*, Blackwell, pp.534–554.

- (2014). *Context*, Oxford University Press.
- Sterelny, Kim (1990). *The Representational Theory of Mind*, Blackwell.
- . (1995). “Basic minds,” in his *The Evolution of Agency and Other Essays*, Cambridge University Press (2001) pp. 198–220.
- . (2001). “The evolution of agency,” in his *The Evolution of Agency and Other Essays*, Cambridge University Press (2001) pp. 260–287.
- (2003). *Thought in a Hostile World*, Blackwell.
- Steward, Helen (1997). *The Ontology of Mind*, Oxford University Press.
- Suppes, Patrick (1957). *Introduction to Logic*, Dover.
- (1983). “Probability and information,” *The Behavioral and Brain Sciences* 6, pp. 81–82.
- Tanney, Julia (2009). “Rethinking Ryle,” in Ryle (2009), pp. ix–lvii.
- de Vignemont, Frédérique. & Massin, Oliver (2015). “Touch,” in Mohan Matthen ed., *The Oxford Handbook of Philosophy of Perception*, Oxford University Press, pp. 294–313.
- Wetzel, Linda (2009). *Types and Tokens*, MIT Press.
- Whyte, Jamie (1990). “Success semantics,” *Analysis* 50, pp. 149–157.
- (1991). “The normal rewards of success,” *Analysis* 51, pp. 65–74.
- (1993). “Purpose and content,” *British Journal for Philosophy of Science* 44, pp. 45–60.
- (1997). “Success again: replies to Brandom and Godfrey-Smith,” *Analysis* 57, pp. 84–88.
- Williamson, Timothy (1994). *Vagueness*, Routledge.
- (2007). *The Philosophy of Philosophy*, Wiley-Blackwell.
- (2013). *Identity and Discrimination*, 2nd edition, Wiley-Blackwell.

## 邦語文献

- アイゼンク、ハンス (1988) 『精神分析に別れを告げよう』 批評社.
- 飯田隆 (1989) 『言語哲学大全 II』, 勁草書房.
- 飯田隆 (2002) 『言語哲学大全 IV』, 勁草書房.
- エヴニン、サイモン (1996) 『デイヴィッドソン』, 勁草書房.
- 大山正 (1994) 『色彩心理学入門』, 中央公論社.
- 金杉武司 (2002) 「哲学的行動主義」, 渡辺恒夫ほか編『心理学の哲学』, 北大路書房, pp. 92–106.
- カーネマン、ダニエル (2014) 『ファスト&スロー (上)』, 早川書房.
- 下嶋篤 (2009) 「情報意味論における真理概念」, 加賀裕郎ほか編 『現代哲学の真理論』, 世界思想社, pp. 147–159.
- 下條信輔 (1994) 『サブリミナル・マインド』, 中央公論社.
- ジョンソン＝レアー、フィリップ (1989) 『心のシミュレーション』, 海保博之ほか訳, 新曜社.

- 杉山尚子, 島宗理, 佐藤方哉, リチャード・マロット, マリア・マロット (1998) 『行動分析学入門』, 産業図書.
- 鈴木聡 (2005) 「信念のスタティクス, 信念のキネティクス」, 博士論文.
- 鈴木貴之 (2000) 「翻訳の不確定性と意味の懐疑論」, 『哲学・科学史論叢』, pp. 105–128.
- ソーバー、エリオット (2009) 『進化論の射程』, 松本俊吉ほか訳, 春秋社.
- 高橋正子 (1991) 『計算論』, 近代科学社.
- 丹治信春 (2009) 『クワイン』, 平凡社.
- 戸田山和久 (2014) 『哲学入門』, ちくま書房.
- 次田瞬 (2010) 「クワインの規約主義批判」, 『論集』 29, pp. 188–202.
- 次田瞬 (2015) 「目的意味論について」, 『科学哲学』 48-1, pp. 17–33.
- 朴嵩哲 (2011) 「理論説 vs. シミュレーション説: 両説は結局どこが違うのか?」, 『哲学・科学史論叢』 13, pp. 123–167.
- ファインマン, リチャード (1968) 『ファインマン物理学 II 光 熱 波動』, 富山小太郎訳, 岩波書店.
- フォーダー, ジェリー, アーネスト・ルポア (1997) 『意味の全体論』 産業図書.
- マクファーランド, デイヴィッド編 (1993) 『オックスフォード動物行動学事典』, 木村武二監訳, どうぶつ社.
- 山口尚 (2012) 『クオリアの哲学と知識論証』, 春秋社.