

# 形而上学的機能主義と心の哲学

次田 瞬

## 1. 「機能主義」という用法の多様性

「機能主義」という表現は、人類学や社会学などでも特定の学説名として多義的に用いられる。それどころか、心の哲学に話を絞っても「機能主義」という名前はいくつかの異なる意味を持っている。ネッド・ブロックによると、まず「機能主義」は、ある特定のタイプの科学的説明を追究するという方法論的ないし研究戦略上の立場を指すことがある (Block 2007, 27)。ここでいう特定のタイプの説明とは、機能的説明のことである。機能的説明とは、大雑把に言えば、研究対象となるシステムの働きや能力を、当のシステムを構成する諸部分へと分解 (decompose) した上で、そうした構成部分のもつ機能がいかに統合されているのかを述べることで説明すること、と押さえることができるだろう。この意味での機能主義が、複雑な機構の働きについて説明する際には幅広く用いることができる戦略であるのは、言うまでもない。例えば、ある工場がいかにして冷蔵庫を製造するのかは、工場の生産ラインなどの構成要素の機能に言及することで説明されるだろう。人間の認知能力もまた、同様な研究戦略の対象となる。例えば、色を識別する眼の能力を説明するには、眼に対する入出力が何であるかを特定した上で、内的なメカニズム (例えば、色のコーディング) についての数理的モデルを立てる、といった仕方で議論を進めるのが有効である。

しかし、「機能主義」には、方法論に関わる以上のような主張とは別に、より多くの心の哲学者たちの関心を集める用法がある。ブロックが「形而上学的機能主義」と呼ぶこのタイプの機能主義は、「心の本性 (nature of mind)」についての学説であり、「心的状態とは何であるのか」といった問いに解答することを意図している (Block 2007, 28)。形而上学的機能主義は、機能的説明を追究する研究プログラムとは区別されねばならない。というのも、仮に心的状態が何らかの物理状態であると判明した場合に生じるであろう更なる問い、例えば、人間が記憶に情報を蓄える仕方は意味ネットワークでモデル化できるか、といった問いに答えることは、形而上学的機能主義の守備範囲には含まれないからである。

形而上学的機能主義にはいくつかのバリエーションがあるものの、それらは共

通して、「システムがとるある状態を特定の心的状態タイプにするものは、そのシステムの内的組成ではなく、その状態が果たす機能的役割である」といったテーゼにコミットする<sup>1</sup>。心的状態と物理状態との間に抽象的で緩やかな制約関係をおくにとどめる点に、形而上学的機能主義の特徴があるといえる。

形而上学的機能主義は、実体二元論に反対するほぼすべての哲学者によって、心身問題への有力なアプローチとして注目されてきた。形而上学的機能主義の流行は、20世紀中盤における行動主義の退潮と、それに代わって勃興してきたAI研究、そして、それに影響された認知科学革命を背景としている。機能主義に対しては、反自然主義や消去主義を支持する哲学者からの根強い批判もあるが、今日でも心理学の哲学者の多くは、広い意味での機能主義を保持していると思われる。もちろん、こう述べたからといって、機能主義一般に対するいわば一撃論法的な反論があることは否定できない。意識経験の現象学的側面、とりわけクオリアや自己知をめぐる問題圏がそうした反論の代表例とされる。だが、現象学的側面からの批判を重要視するあまり機能主義の諸見解に眼をつぶることにすれば、いくつかの重要な洞察が失われてしまうのではないかと思われる。以下でみるように、一元論者にとって機能主義はやはり魅力的な選択肢であるし、また、機能主義の枠内でクオリアの問題に対処しようとする試みもないわけではない。

そこで本稿では、形而上学的機能主義の中でも人気がある「因果役割機能主義」という立場を主として考察したい。これはデイヴィッド・ルイスや、フランク・ジャクソンなどオーストラリア系の哲学者によって擁護されている立場である<sup>2</sup>。本稿の議論の流れは、以下ようになる。まず、次節では形而上学的機能主義の代表格としてよく知られるパトナムの機能主義に触れ、これを厳しく批判する。3節以降では、因果役割機能主義が立脚している議論の余地のある諸前提を確認した上で、修正案を提案する（もっとも、本稿の機能主義批判は、形而上学的機能主義に対して多少とも好意的という意味で穏当なものである）。

なお、以下で言及するタイプの機能主義は、すべて形而上学的機能主義の下位範疇である。そのため、以下では「形而上学的」という限定を省略する。

## 2. パトナムの機能主義とそれに対する批判

機能主義の様々なバリエーションに共通するテーゼとして「システムがとるある状態を特定の心的状態タイプにするものは、そのシステムの内的組成ではなく、その状態が果たす機能的役割である」という定式を前節で提示しておいた。この

テーゼからは、心的状態の「多重実現可能性 (multiple realizability)」という重要なテーゼが導かれる。多重実現可能性は、論理的行動主義や心脳同一説といった立場に対する批判として頻繁に用いられる概念である。心身問題に対するこれらの先行学説は、例えば、痛みという心的状態タイプを、うめき声を上げるといった表出行動であるとか、あるいは C 繊維の発火といった神経生理学的な出来事タイプと同一視するものである。しかし、これらの説は、うめき声を上げることのない (人間以外の) 生物種であるとか、身体の組成構造が人間とは大きく異なる生物種には痛みを帰属させるのに問題が生じるという意味で、排外主義的 (chauvinistic) であるという批判がなされた。他方、機能主義は、痛みとはもっと抽象的な機能的性質 (例えば、組織の損傷を感知するといった機能) であって、生物種に応じて多様な仕方で、痛み状態として実現されるのだと論じる。

ここでのポイントは、機能的性質は物理的な実現からある程度独立しているために、多様な仕方で実現されることを許すということである。機能的性質が多様な仕方で物理的に実現されるという事例は、さまざまな領域に見出される。例えば、心臓などの生物学的器官や、エンジンなどの人工物がその例である。

ところで、生物学的器官と人工物がどちらも、その目的論的な機能について語ることに意味をなすという共通点がある。このことは単なる偶然ではないかもしれない。例えば、ダニエル・デネットは、進化生物学は工学の一種だと主張する (Dennett 1995)。これは必ずしも突飛な主張ではない。実際、進化生物学は、単純な生物種であろうと高等な知能を備えた人間であろうと、すべての生物種は共通の起源を持っており、おのおのの種の個体もつ様々な形質の存在は自然選択によって説明可能だとする、いわゆる適応主義の考えを中核に据えている。そのため、現代の進化心理学者たちは、人間の心を、母なる自然 (Mother Nature) によって形成されたある種の工学的な装置と捉えようとする。

このような主張は、少なくともビジョンとしては強く訴えるところがあると私自身は考える。とはいえ、人間の心的機構の複雑さは尋常ではなく、その機能を容易に特定できる心臓やエンジンと同じようには扱えない。この点を確認するために、まずは試みに、人間の心を機械になぞらえられると仮定して、人間の心はどのような装置と考えられるのかを思案してみよう。以下で論じるのは、心的性質の多重実現性をかなり早い時期から主張していたヒラリー・パトナムがひところ好んで論じた、チューリング機械との類比である。残念ながら、この類比は機能主義の概説にしばしば現れるものの、実際には殆どうまくいかないことが分かっている (以下、パトナム風の機能主義を“PF”と略す)。

さて、かなり大雑把に言えば、チューリング機械は、無限長のテープと、テープの節に記号を読み書きするヘッド、有限個の記号、そして有限状態機械から構成される計算機である。チューリング機械が心のモデルとして選ばれた動機はいくつか考えられる。まず、テープや有限状態機械といった概念は様々な物理的基盤によって実現されるから、多重実現性テーゼの例証として使い勝手がよいということがある。加えて、計算機と志向的システムとの間には、一見すると類比関係が成り立ちそうに見える。計算機は与えられた引数に対して関数の値を計算する。それと類比的に、外界と相互作用する志向的システムは、環境からの刺激を、変換器 (transducer) によって記号表象の形で入力として受け取り、何らかの適切なプロセスを経て、その出力を運動器官に伝えていると想像できよう (Putnam 1975, 409)。

PF のアイデアは、人間以外の生物にも適用しうる。そもそも、単純な生物であれば、振舞いをシミュレートするのにチューリング機械など持ち出すまでもない。例えば、デネットが好んで例に用いるハマグリは、外敵が近づいてくると殻の中に足を引っ込める。しかし、実はこの触発性の反応は、外敵を認知しているのではなく、単に振動や衝突に反応しているに過ぎないのである。こうした生物の振舞いは有限状態機械を用いることで容易にシミュレートできるだろう。また、パトナムが行ったように、こうしたシミュレーションに現実味を持たせるために、状態遷移が非決定的なオートマトンを有機体のモデルとするなどの工夫を施すこともできよう (Putnam 1975, 433)。したがって、ハマグリの内的機構は、有限状態機械を物理的に実現したものということになる。更に言えば、PF が正しければ、人間は、自販機や洗濯機のような機械と比べて、とりうる状態の数などの複雑さにおいてのみ異なる存在者ということが示唆される。

しかし、現在では、たとえ人間の心が何らかの計算機だとしても、チューリング機械をモデルとして使うことはできない、あるいは (心の哲学のジャーゴンで言えば)、チューリング機械を人間の心の「アーキテクチャ」とはみなせないという点で論者の間では一般的な合意がある。実際、パトナム自身が 70 年代には以前の自分の立場への批判を行っているので、それを参照するのが好都合である。まず、チューリング機械の状態は、ある時点でちょうど 1 つだけしかとることができないようなものであるのに対して、もしも人間がチューリング機械なら、機械状態は人間全体の瞬間的状态 (total instantaneous state) の表現ということになる。しかし、人の心的状態は一般に、重ね合わせが可能である。ちょうど、鉄の塊が伝導性と伝熱性という別個の傾向性を同時にもつことができるのと同じよう

に、一個人が同時に多様な心的状態（痛み、嫉妬、信念など）をとりうるのは当然のこととして認められるであろう。したがって、日常的な意味での心的状態を、チューリング機械の状態と同一視することはできない。また、こうした問題に加えて、パトナムによれば、記憶や学習は、チューリング機械では新たな状態の獲得として表現することができないという難点もある。もしも人間がチューリング機械なら、心的状態は学習などとは無関係ということになってしまう。

およそ以上の理由で、我々の心的機構をチューリング機械によってモデル化する PF には無理があるとされている。我々もこれ以上 PF の路線を追求しようとは思わない。とはいえ、ここから帰結するのは、せいぜい PF という特定のストーリーが間違っているということであって（Putnam 1975, 298-9）、機能主義一般が反証されたわけではない。だからこそ、先に「試みに」と断っておいたのだった。節を改めて、もう一度機能主義の定式を確認することから出直そう。

### 3. 因果役割機能主義とその可能性

機能主義は心的状態タイプとその機能的役割の同一視を主張する学説であった。ところで、「機能」には二種類の解釈がありうる。前節では、目的論的な解釈を提示したが（心臓やエンジンの例を参照）、因果役割として解釈するやり方もある。既に示唆したように私自身は目的論的な解釈の方に共感するのだが、以下ではオーソドックスな因果役割としての解釈に基づいて議論を進めようと思う<sup>3</sup>。

さて、因果役割とはそもそも何か。ふたたび痛みを例にとると、痛みの因果役割とは、組織の損傷によって引き起こされるとか、集中力を鈍らせるといった出来事間の因果連鎖のパターンにおいて占められる位置づけのことだと言えるかもしれない。機能主義者は、この単純なアイデアを他の心的状態を特徴付けるために拡張するのである。そこで、因果役割に言及する機能主義を「因果役割機能主義（causal role functionalism）」と呼び、“CRF”と略すことにしよう<sup>4</sup>。

CRF は心的状態に因果的効力を認める。だが、このことは、心的状態は他の心的状態とも因果関係に立つ（例えば、痛みによって怒りが引き起こされるといったような）可能性をも認めるということでもある。ここで一つの単純な問題が生じることになる。すなわち、一つの心的語彙（mental state term）を定義しようとすると、他の心的語彙を巻き込まざるをえないという意味で、定義が循環するという問題が生じる。したがって、CRF の支持者たちは、この問題に対処すべく、CRF の定式化を工夫することが求められる。

ここで、ルイスのような CRF 論者は、科学における理論語は理論全体によって陰伏的に定義されている、という論理実証主義以来のアイデアを応用する。すなわち、心的語彙は、心理学的理論において用いられている一種の理論語だとみなすのである。まず、心的語彙を含み、かつ真であるすべての文の集合を“Th”と呼ぼう。Th はいわば心理学的理論であり、痛みや怒りといった心的状態と、感覚入力、そして行動による出力との関係を特定していると仮定する。次に、理論 Th に含まれる全ての文を連言でつなぐことで、ある単一の文  $s$  が得られる ( $s$  は途方もなく長い文になるだろう)。ところで、Th で用いられている心的語彙を  $M_1, \dots, M_n$  とすると、文  $s$  は“ $s(M_1, \dots, M_n)$ ”という風に表記できるだろう。ここで、 $s$  に出現しているすべての心的語彙を変項で置き換えて（二階の）存在量化を施すと、“ $\exists F_1 \dots \exists F_n s(F_1, \dots, F_n)$ ”が得られる。このような仕方から科学理論から理論語を除去して得られるものを「ラムジー文」という。もちろん、ラムジー文それ自体は、理論から理論語（ここでは心的語彙）を消去しているだけであって、定義するわけではないし、そもそも存在量化が現れていることから示唆されるように、ラムジー文は現実の世界に関する実質的な主張を含んでいる。しかし、ラムジー文は心的語彙を定義するために転用することができる。すなわち、任意の心的語彙  $M_i$  は“ $M_i(x) \equiv_{df} \exists F_1 \dots \exists F_n [s(F_1, \dots, F_n) \& F_i(x)]$ ”という式によって定義できる。およそ以上のようにすれば、CRF のアイデアに基づきながら、心的語彙を定義することができる。

しかし、ただちに気づかれるように、CRF によるラムジー文の使用は、あまりに非現実的と言ってよいほど強い前提を措いているように思われる。例えば、あらゆる心的状態タイプの意味を定める理論が、単一の文として書き下せる保証はどこにあるのか、と問われるかもしれない。

この疑問は極めてもつともだと思われるのだが、さしあたり次のように答えてみたい。まず、ラムジー文を用いて個々の心的語彙を定義するという議論は、定義が循環するという困難に対処するところに眼目があったことに注意しなければならない。また、機能主義はそもそも、心的状態とは何であるのかという形而上学的問題に対して回答することを意図した学説であった。この問いは個々の心的語彙の意味は何かという問いとは区別しうる。もしこの区別が成り立つならば、文  $s$  が無限連言になるかどうかは、形而上学的問題とは関連しないかもしれない。

他にも多くの疑問が提出されるであろうが、その内のいくつかは次節で議論することにしたい。その前に、ここでは、ラムジー文を経由した心的語彙の定義がどのように使われるのかを確認しておきたい。すなわち、上のような CRF の定

式は、機能主義の要件ともいうべき心的状態の多重実現可能性をどのように確保しているのかを確認する。

ルイスによれば、痛み状態が多重実現可能であるという場合には、2つの論理的可能性がある。その2つを彼は、「狂人の痛み (mad pain)」と「火星人の痛み (martian pain)」と呼んでいる。これらのうち、どちらかといえば理解しやすいのは火星人の痛みの方である。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、など。しかし、火星人の身体組成は人間のそれとは大きく異なっている。CRF はこうした火星人对して痛み状態を帰属することができる。痛み概念に相当するのが  $M_i$  であるとする、人間の場合  $F_i$  の値となるのは例えば C 繊維の発火という出来事タイプである。これに対し、火星人の場合には  $F_i$  の値となるのは別個の神経生理学的状態タイプである、と考えればよい。

実は、この議論では、本稿がここまでの論述において曖昧にしてきた区別が用いられている。ルイスは痛み概念と痛み状態とを区別している。ルイスによれば、痛み概念とは、特定の因果役割のことであり、その因果役割をみえす物理状態はなんでも痛み状態なのである。この精妙な区別は、「勝者は勝者でなかったかもしれない」という文を事象様相で解釈するときのことを考えると理解しやすい。「勝者」は非固定的 (non-rigid) な記述であるから、可能世界に応じて指示対象は異なる。したがって、現実世界で勝者であった人が、他の可能世界でも勝者であるとは限らない。これと同じように、「痛み」の概念は非固定的であり、可能世界ならぬ集団 (population) に応じて、外延が異なるというわけである<sup>5</sup>。

さて、CRF にとってより問題となるのは、狂人の痛みである。ルイスのいささか奇抜な設定によれば、狂人の痛み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。痛みはうめき声ではなく、指を鳴らす反応を引き起こす。痛みを避けようとも、ジタバタもしない。このように、痛みに関わる狂人の振舞い方は常識的な人間とは全く異なるのだが、狂人の身体組成は我々とほとんど変わるところがないと仮定する。問題は、狂人の振舞い方は他の健全な人間の振舞い方と大きく異なるため、 $F_i$  の値が存在しないという点である。

たしかに、もはや「痛み」と呼ぶにふさわしい因果役割を果たしていない以上、狂人に痛みを帰属できるかどうかは疑わしいという直観はありうる。しかし他方で、C 繊維の発火という物的出来事は痛みの因果役割をもはや果たさなくなったということを、ちょうど機械が故障したのと同じように考えた場合にはどうか。

故障した電卓はそれでも「電卓」と呼ばれ続ける。同じように、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ならば、通常の人間の場合に痛み状態であるところの C 繊維の発火という出来事タイプを、狂人の場合にも「痛み」と呼んではいけないだろうか。必ずしもそうではあるまい。

こうした直観を取り込むためにルイスが訴えるのは類似性の観念である。火星人と通常の間は、身体組成が異なっても痛みの振舞いがよく似ている。狂人と通常の間は、もはや痛みの振舞い方を共有しないが、身体組成はよく似ている。だからこそ、狂人にも火星人も共に、我々は痛み状態を帰属することをそれほど躊躇わないかもしれない。しかし、類似性関係が推移性を満たさないことはよく知られる。例えば、4 番目のカテゴリーとして狂った火星人というものを考えると、彼は振舞い方も身体組成も通常の間とはあまりに似ていないので、そうした個体に対して痛み状態を帰属するのは躊躇われることになる。

心的状態の帰属には普通の間との類似性が重要な役割を果たすというルイスの見解には聞くべきところがあるように思われる。私見によれば、同じような見解はウィトゲンシュタインの『哲学探究』にも見られる。

人は、生きている人間、およびそれに似たもの（似た振舞いをするもの）についてのみ、感覚を持つ、見る、盲目である、聞く、難聴である、意識がある、意識がない、とすることができる。（Wittgenstein 1953, §281）

マルコムはこの節を次のように解釈している（Armstrong & Malcolm 1984）。我々は、人間について「見る」とか「意識がある」といった言葉を適用できるが、動物については（残酷にも？）そういう言葉を使えない人というのを想像できる。では逆に、動物に対してしかそうした語を使えない人を想像できるだろうか。マルコムによれば、無理である。そういう人は単に「感覚」とか「意識」といった言葉の用法を端的に理解していないのである。ここでのポイントは、心的語彙の適用には非対称性があるということである。犬や猫のような動物に「恐れている」とか「喜んでいる」といった言葉を適用できるのは彼らが人間に似ている限りでのことである。実際、ミミズや蠅に対して同じ表現を使えるかどうかは疑わしくなる。心的語彙の帰属先のパラダイムはあくまでも人間なのである。

とはいえ、類似性関係が我々の関心から独立に、客観的に成り立つ関係であるかどうかは大いに議論の余地のある問題である。心的状態の帰属が結局のところ我々の関心から独立でないのだとすれば、志向的態度も含めてあらゆる心的状態



に関して事実問題は存在しないということになるのだろうか。それとも、ルイスやウィトゲンシュタインの考えは完全に誤っており、むしろ人間をパラダイムとする先入観を捨てて、志向的システムについての理論は唯1つではありえないと考えるべきだろうか<sup>6</sup>。これは興味深い問題ではあるが、機能主義の学説それ自体を検討するという本稿の目的からすると、いささか脇道に逸れる。いずれにせよ、CRF が心的状態の多重実現可能性について正面から取り組む場合に、ルイスのような立場に落ち着くのは自然であるように思われる。

#### 4. 因果役割機能主義の限界と修正案

本節では、ここまで曖昧にしてきた CRF の定式化をより厳密にするため、以下 2 つの問題を順に検討していく。まず、(1) 心的状態の因果役割に関する理論とは何か。前節では「心的語彙を含み、かつ真であるすべての文の集合 Th を心理学的理論と呼ぶ」と述べた。だが、理論 Th はどのように特定されるのだろうか。そもそも心理学には理論があるのだろうか。(2) 前節では理論 Th は単一の文として書き下せるという前提を措いた。この仮定には本当に問題がないのだろうか。とりわけ後者の問題は機能主義にとって大きな問題となるはずである。

##### 4. 1 心理学的理論として何を採用すべきか

CRT の支持者の多くは、心的語彙の意味を定義するための理論として、常識心理学 (folk psychology) を念頭においている。それでは、常識心理学という理論は、私たちの常識をどのように反映するのだろうか。

ルイスはかつて、心的状態と感覚刺激と運動反応 (motor response) の間の因果的關係に関して集団の中で共有知となっているような言い回し (platitude) をとにかく取り集めてくればよい、と主張した (Lewis 1999, 257-8)。言い回しには様々な形式の表現がありうる。前節では心的語彙は基本的に述語の形をしているとみなしていたが、「歯痛は痛み的一种である」といった形式の言い回しもあるだろう。ともあれ、ルイスによれば、ある集団における共有知、すなわち、その集団のどんな愚か者でも知っている常識を取り集めてくれば常識心理学の理論になる、といういささか楽観的ともいべき見通しをもっていた<sup>7</sup>。このように、常識心理学によって心的状態の因果役割が特定されるとする機能主義は、しばしば「常識的機能主義 (common sense functionalism)」と呼ばれる。

いささか寄り道になるが、実は、概念分析のために、常識として我々が所有し

ている言い回しを収集するという発想は、今日では「キャンベラ計画（Canberra Plan）」と呼ばれる哲学の方法論に発展している。キャンベラ計画は、分析される概念についての言い回しを収集する第一段階と、言い回しを充分に実現（realize）するものを採る第二段階という 2 つのステップによって遂行される。第一段階とは、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言い回しを、我々の日常的な信念から選りわける作業であり、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を明らかにする。第二段階の作業では、アポステリオリな科学的知識を持ち込むことが許されている。例えば、痛み概念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、という風にある。

キャンベラ派の哲学者たちは、哲学とはアプリオリな概念分析に携わるもの、というメタ哲学的な直観を擁護している。この点で彼らの考えは、哲学は自然科学と本質的に同じ目的を共有し、同じ手法によって研究を進めることができるとする方法論的自然主義（methodological naturalism）と対立することになる<sup>8</sup>。

さて、本稿はメタ哲学について特定の論陣を張るものではないが、分析と総合の区別にコミットしているという意味で、キャンベラ計画には保守的な側面があることは指摘しておきたい。それは以下のような理由による。

キャンベラ計画の第一段階では、哲学的諸概念をあらゆる語彙の意味を定義するために、言い回しを収集して理論を手に入れることが目指される。だが、どんな理論であっても、そこからラムジー文を経由して理論語の意味を定義できると考えるのは早計であるように思われる。例えば、バナナに関して真である文を集めてくると想定しよう。そこには「バナナは亜熱帯でよく育つ」とか「バナナは 10 kg より軽い」といった文が含まれることだろう。だが、こうした文からなる常識バナナ理論によって、語「バナナ」の意味が定められるとするのは奇妙である。これに対し、常識心理学は常識バナナ理論とはまったく性格が異なる理論であり、心的語彙の意味は、常識心理学の理論によって与えられると主張するならば、心的状態の因果役割について述べる常識心理学の言い回しは分析的ということになる<sup>9</sup>。つまり、痛みや信念といったものがしかじかの因果役割をもつことは分析的に真だということである。

キャンベラ計画の第一段階が、分析的な文を収集するという作業に携わるもの

だとすれば、それがアプリアリな概念分析であるという主張も正当化されることになるだろう。なぜなら、分析性はアプリアリ性と必然性を含意すると一般に認められているからである (Williamson 2008, 51)。

しかし、分析と総合の区別が存在するという主張は擁護できないとするか、あるいは、少なくともこの主張には慎重でなければならないと考える論者は、クワインの論文の登場から半世紀が経った現在でも、数多い<sup>10</sup>。分析的な文と総合的な文との区別がつけられないとすれば、キャンベラ派の哲学者は、言い回しを収集してくる際に、いったいどの文が哲学的概念をあらわす語の意味を決定するような文であるのかを判定する原則なしに収集しなくてはならなくなる。

とはいえ、すべての機能主義者は分析と総合の区別を擁護しなければならないのかというと、実際にはそうではないように思われる。分析と総合の区別に懐疑的な機能主義者は、常識心理学の理論が偽な帰結をもたらす可能性を認めることができる。それならば、心的状態の因果役割を特定する理論として、常識心理学ではなく認知科学の理論を採用するという選択肢もあることになる。心理学的理論からは分析性の身分をはぎとるだけで、常識的機能主義の利点はすべて引き継ぐことができる<sup>11</sup>。よって、ここでは、分析と総合の区別へのコミットは機能主義者にとって必須ではないのだと結論し、もう一つの問題を検討しよう。

#### 4. 2 命題的態度を個別化できるか

2つ目の問題は、心的状態タイプの因果役割を特定する理論 Th が単一の文として書き下せるという想定に問題はないかというものだった。この想定は痛み状態のように、内容 (content) を持たない心的状態だけを考察している分にはもつともらしい。だが、信念や欲求などの命題的態度について考え始めると、心的語彙の数が有限であると想定することはできないと思われる。それは次のような理由による。まず、命題的態度の内容の数には原理的に上限がない、ということは広く受け入れられている。また、信念状態や欲求状態は内容が異なれば別個の状態タイプとみなさざるをえないだろう。というのも、異なる内容をもつ2つの信念状態は、異なる因果役割を持つと考えるのが自然だからである。しかし、そうすると、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ことになる。かくして、心的語彙のリストは当然ながら有限では済まない、というわけである (Block 2007, 21)。

こうした問題は、そもそも信念をそれ以上分節化できない主体の性質とみなす、

副詞説的な立場からの帰結だとするのは自然である。むしろ、信念を「P と信じる」と「Q と信じる」の共通項とみなし、信念を主体と命題との間に成り立つ二項関係として分析するべきではないか。そこで、例えばルイスは、信念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みなすことを提案する (Lewis 1999, 258n13)。理論 Th の役割は信念や欲求といった態度 (attitude) を特徴付けることであって、一度それらが定義されれば、個別の命題とのペアによって更に肌理細かな心的状態の分類を手にすることができると期待される。もっとも、命題という存在者の身分について疑問が残るかもしれないが、それは形而上学の問題であって心の哲学の問題ではない (Lewis 1983, 110)。

私自身は、理論 Th は態度を特徴付けるに留まる、というルイスのアイデアは機能主義を擁護するという目的にとっては最も現実的ではないかと考えている。実際、このような立場は科学的心理学とストレートに接続することが可能であるように思われる。例えば、心語 (mentalese) の仮説について考えてみよう<sup>12</sup>。心語とはさしあたり、言語表現がもつ合成性 (compositionality) を思考の合成性へと直ちに結びつける仕掛けと理解しておこう。すなわち、心語の仮説は、きわめて大雑把な言い方をすれば、我々が何らかの思考をはたらかせるとき、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。この仮説にしたがうと、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 (いわゆる) 信念ボックスに p を表現する心語の文が含まれることだ、という風に説明されるだろう。

もちろん、心語の仮説には様々な批判があることは公平のために付け加えねばならない。しかし、二点ほど注意しておくことがある。第一に、心語は重要な意味で経験的な仮説であり、形而上学的機能主義の議論とは独立である。第二に、我々のあずかり知らぬ言語が我々の脳内で作動しているという考えは直感に反するかもしれないが、この仮説をまともに批判するのはそれほど容易ではない<sup>13</sup>。

さて、以上の議論が正しければ、我々はルイスの機能主義を心語の仮説は整合的であることを示したことになる。既に述べたように、私自身はこれが機能主義を擁護する最も現実的なやり方だと考える。しかし、すべての機能主義者がこの結論に同意するわけではないということは注目に値する。例えば、ジャクソンらは、機能主義と心語仮説とを明確に分離して、あくまで機能的役割によって心的状態の命題内容が定まるとまで主張できなくては機能主義の名に値しないと主張している (Braddon-Mitchell & Jackson 2006, 175)。

## 5. まとめと今後の研究課題

本稿では、主としてルイスらによる因果役割機能主義を検討した。我々は因果役割機能主義を提示する過程で「機能主義」に関する幾つかの用語法を区別したが、4 節での議論を通じて因果役割機能主義にも更なる下位区分が必要であることが明確になった。また、4. 2 節では思考の内容を個別化する一つの考え方として心語の仮説を導入することを提案したが、ジャクソンらはあくまで機能的役割によって心的状態の命題内容が定まると主張して、機能主義を貫徹する可能性を追究している (Braddon-Mitchell & Jackson 2006)。彼らの見解については、別の機会に更なる検討を加えたい。

---

\*本稿は科学研究費補助金（特別研究員奨励費）による研究成果の一部である。

<sup>1</sup> J. Levin, “Functionalism” 参照。

<sup>2</sup> Lewis (1983), Lewis (1999), Braddon-Mitchell & Jackson (2006) 参照。また「人気がある」という点については (Kim 2005, 124) を参照。

<sup>3</sup> 実際には、どちらの解釈をとっても以下の議論にはそれほど大きな違いは生じない。

<sup>4</sup> 理論の対象を因果役割によって特定するという戦略は、自然科学で一般的に見られるという意味では、正当な手続きといえる。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、集団遺伝学では遺伝子を表現型に与える影響によって同定していた。

<sup>5</sup> 実際には、単称名の場合と違って、一般名の固定性をどう定義すべきかは色々な問題がある。Devitt & Sterelny (1997, chap. 5) 参照。

<sup>6</sup> Sterelny (1990, sec. 1-3) 参照。

<sup>7</sup> 共有知の形式的な定義は van Ditmarsch *et al.* (2007, sec. 2-3) を参照。

<sup>8</sup> キャンベラ計画についての以上のまとめは、Papineau, “Naturalism”を参考にした。

<sup>9</sup> Nolan (2005, 126) 参照。ただし、意味の全体論をとるルイスがこのような区別を本質的とみなすのかどうかは、議論の余地がある。

<sup>10</sup> Churchland (2007) 参照。分析と総合の区別に関する私自身の見解は次田 (2010) で示した。

<sup>11</sup> Botterill & Carruthers (1999, 11) 参照。

<sup>12</sup> 命題的態度の説明に心語の仮説を導入することは「心の表象主義」とも呼ばれる。

<sup>13</sup> しばしば見られる誤解や疑問にここで回答するのも意味があるだろう。まず、「言語的对象」と言っても、脳中に例えば日本語の文が書き込まれているというのではない。言語表現

---

は媒体に関して中立なので、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可能性はアプリアリには退けられない。次に、ふだん我々が思考をするときには、たとえ表立って声にださなくとも内語という仕方で、日常言語の諸表現を用いて思考している。それで十分ではないのか。なぜ余計に言語の数を増やす必要があるのか。この疑問には、ここでは思考一般の本性は何かが問題になっている、と応えることができる。日常言語の表現を用いて黙って考えるという一人称的な経験についての現象学は、思考がいかに関現されるのかを説明する心理学的理論とは水準が異なる。

[参考文献]

- Armstrong, David & Malcolm, Norman. 1984. *Consciousness and Causality*, Blackwell Publisher.
- Block, Ned. 2007. *Consciousness, Function, and Representation: Collected Papers*, MIT Press.
- Braddon-Mitchell, David & Jackson, Frank. 2006. *Philosophy of Mind and Cognition: An Introduction*, 2nd edition, Blackwell.
- Botterill, George. & Carruthers, Peter. 1999. *The Philosophy of Psychology*, Cambridge University Press.
- Churchland, Paul. 2007. "The evolving fortunes of eliminative materialism," in *Contemporary Debates in Philosophy of Mind*, Brian P. & Cohen, Jonathan (ed.), McLaughlin, 160-81.
- Dennett, Daniel. 1995. *Darwin's Dangerous Idea: Evolution and Meaning of Life*, Simon & Schuster.
- Devitt, Michael. & Sterelny, Kim, 1997. *Language and Reality*, 2nd edition, MIT Press.
- van Ditmarsch, Hans et al. 2007. *Dynamic Epistemic Logic*, Springer-Verlag.
- Kim, Jaegwon. 2005. *Philosophy of Mind*, 2nd edition, Westview Press.
- Levin, Janet. "Functionalism," in *Stanford Encyclopedia of Philosophy*.
- Lewis, David. 1983. *Philosophical Papers Vol. 1*, Oxford University Press.
- . 1999. *Papers in Metaphysics and Epistemology*, Cambridge University Press.
- Nolan, Daniel. 2005. *David Lewis*, McGill-Queen's University Press.
- Papineau, David. "Naturalism," in *Stanford Encyclopedia of Philosophy*.
- Putnam, Hilary. 1975. *Philosophical Papers: Volume 2, Mind, Language and Reality*, Cambridge University Press.
- Sterelny, Kim. 1990. *The Representational Theory of Mind*, Blackwell.
- 次田瞬. 2010. 「クワインの規約主義批判」, 『論集』 29, 188-202.
- Williamson, Timothy. 2008. *The Philosophy of Philosophy*, Wiley-Blackwell.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, Basil Blackwell.