

(心の哲学 西脇 与作(慶応義塾大学))

はじめに

心の哲学と心の科学は異なるのか。(これは科学と哲学は異なるのかという問いでもある。)物理学の哲学や生物学の哲学が物理学や生物学を無視した形では研究できないように、心の哲学も心の科学を無視しては研究できない。ただ、心の科学は新しい領域であり、まだ未解決の問題がたくさんある。そのことが心の哲学に物理学や生物学と違った印象を与えている。心の科学を無視しては心についての哲学は展開できない。

心に関する省察はギリシャ哲学で登場して以来、多くの哲学者の関心の的となってきた。だが、その歴史は決して恵まれたものではなかった。その不遇はデカルトの心に関する扱い以来一層強まり、心は私たちの知的追求の対象からややもすれば外され勝ちであった。というのも、それは自然科学的な追求に馴染まないもの、適さないものというだけでなく、自然科学的な対象となるべきではないと考えられてきたからである。

今世紀の心に関する追求の目覚ましい進展は19世紀にその出発点を辿ることができる。それはダーウィンの進化論である。そこでは人間も自然科学的な研究対象であると自覚され、その生物学的な特徴は心理学、生態学、行動研究へと進展していくことになる。

今世紀の哲学は言語を中心に哲学の問題を考え、処理する点に特徴があるが、それは言語を操る心の研究に少なからず影響を与えることになった。その一つの成果が認知科学である。論理学、言語学、コンピュータ科学、これらの総合としての認知科学は現在の心の研究の実質的な成果である。このような成果にさらに脳科学の成果が付け加わることによって一層の総合化が目論まれている。

このような展開のなかで、心の謎はどのように説明されるのか。物質とは異なる心独自の特徴は何なのか。心は実在するのか。心と身体、あるいは脳はどのような関係にあるのか。心の内容はどのように扱われるのか。思考と実在はどのような関係にあるのか。信念、欲求、感情、記憶、そして意識といった心の働きはどのように説明されるのか。これらの疑問はいずれも従来明確な解答を与えられなかった。では、現在これらの謎にどのような解答を用意できるのか。

何より私たちはそれぞれ優れた実践的な心理学者である。私たちは自分の心だけでなく、他人の心さえコントロールできるほどに心に関する専門家である。私たちは自分が心をもち自在にそれを使うことができると思っていると同時に、心はしばしば故障して、時には日常生活に重大な支障を与えることも知っている。心は私的で他人には見えないとみなされているが、一方で私たちは他人の気持ちを察することができると思っている。互いに信じ合い、愛し合うことができると考えている。では、私たちは自分自身の心と他人の心のいずれをよりよく知ることができるのか。これもデカルト以来の心に関する哲学の大きな課題である。

0 心の不思議さ

哲学の基本的な謎の一つは私たち人間が自然の中で占める位置である。私たちは他の物体と同じように自然の法則に従う対象なのか、それとも特別の対象で自然法則が適用されない例外者なのか。この問いに対する常識的な答えは、私たちは自然の法則に従うと同時にその法則の適用範囲を超えた特徴や性質ももっており、身体は自然法則に従うが、心はそうではないというものだろう。日常の出来事には、選挙、椅子を見ること、12月3日が僕の誕生日であることを信じること、車の衝突等々がある。これらの出来事を分類してほしいと言われれば、誰もが選挙、車の衝突は公共的な(public)出来事であり、何かを見たり信じたりすることは私的(private)な出来事であると答えるだろう。そして、私的な出来事は心的な出来事(mental events)であり、公共的な出来事は物理的な出来事(physical events)であると分類されることになる。

心の不思議さを表す次の二つの文を考えてみよう。

- (1) 自然を扱うのが自然科学であり、心も自然の産物である。よって、心は自然科学の対象になる。
 (2) 心の対象は表象であり、自然についての表象も含まれる。よって、心は自然科学を対象にする。

いずれの文も正しそうに見える。それぞれの文を別々に考えた場合、私たちは両方とも認めてしまうだろう。しかし、二つの文を同時に考えると心と自然科学のいずれが主人か迷ってしまう。この迷いを起こす関係こそ心の特殊な役割を見事に描き出している。ところで、(1)と(2)のいずれに親近感を覚えるだろうか。私は(1)の立場から以後の話を展開するつもりである。したがって、私の立場では(2)をどのように(1)に還元するかが問題となる。

昔から人々はこの心の謎に挑戦し、様々な答えを出してきた。例えば、アリストテレスは次のように考えた。身体と心は異なる実体ではない。心の性質や特徴は身体の特別な性質や特徴であり、心は身体なしには存在できない。心的性質は直線や平面といった幾何学的対象のようなものではなく、「直線的である」といった幾何学的性質に類似したものである。直線という対象の性質「直線的」がこの特別な性質や特徴に対応している。一方、デカルトの考えでは、身体と心は互いに他の助けなしに存在できる。そして、心的実体(心)の心的性質は物理的実体(もの)の物理的性質と因果的な関係をもっている。ところが、ヒュームはデカルトと違って心的実体の存在を否定する。心的実体に関する印象を私たちはもつことができないからである。私たちは感覚印象、記憶、予期をもつことができるが、自己そのものといった対象に対応する観念をもつことはできない。このように代表的な哲学者の心についての考えでさえ大きく異なっている。心的な性質は物理的な性質の組み合わせに過ぎないと考える哲学者がいたとすれば、物理的な性質は心的性質の組み合わせに過ぎないと考えた哲学者もいる。(例えば、カントは後者の代表。)あるいは、ホッブスは推理や判断は計算の形式(*forms of computation*)であると考えた。身体は計測器として働き、その計測器の状態は外的な世界の出来事を表示するという記号的な役割をもっている。心は特別な仕方で機能するものであり、心的なものはすべて物理的なものに還元できると彼は考えた。人が違えば、そして時代が違えば心についての捉え方も異なる。これを、信頼できる心についての理論などどこにもないと否定的に受け取る人も、難解な心の姿を様々な角度から照らし出してきたと肯定的に受け取る人もいるだろう。

心についての複数の考えが引き起こす問題は多岐にわたっている。その一例が人格の同一性 (*personal identity*)である。椅子や机が同じものか否かには比較的明瞭で、一致した基準があるが、心をもった人格の同一性の基準は曖昧である。10年前の私と現在の私が同一人物とみなされるのはどのようにしてなのか。あるいは、一般的に時間を通じての人格の同一性は何によって決まるのか。この問いに対する答えは椅子や机の場合に比べると簡単ではない。通常、椅子や机の同一性はそれらの物理的な連続性(*physical continuity*)が基準になっている。私たちは心をもっているので、さらに心的な同一性も必要になる。これにはものの場合に対応して心理的な連続性(*psychological continuity*)が基準になる。では、心理的な連続性とは何か。昨日のことを思い出す人は今日のその人と心理的に連続している、つまりは記憶の連続性が心理的な連続性と考えられている。そして、身体的同一性と心的同一性の二本立ての基準によって人格の同一性が捉えられることになる。このような答えは一見すると確実に見えるが、実は多くの疑問をすぐに引き起こす。余りに多くの疑問のため、大抵は人格の同一性に到達する遥か前に追求を放棄してしまうことになる。

(#) 同一性について：次のような状況でどのように同一性が言えるか。思考実験してみよう。

・船の外板：船を造っている板を一枚ずつ取り替えて、すべての板を新しいものにしたとする。修理された船は以前の古い板の船と同じと言えるだろうか。また、古い板を集めて同じ型の船を再現したときの船は以前の船と同一であろうか。

・脳移植：ある人の脳を別の人に移植したとき、移植された人は誰と同一なのか。また、脳を半分移植した場合はどうなるか。

・複製：ある人の複製をつくったとき、いずれの人が以前の人と同一なのか。

私たちは最初に心の科学と心の哲学を区別した。しかし、もし心がものとは異なる実体であり、科学の適用範囲を超えたものであれば、そもそも心の科学はありえないものになってしまう。心の科学が成立することは、例えばデカルトの考えの否定を意味しているし、ホッブスの考えに従うならば心の科学は最終的にも心の科学、つまりは物理学に還元されてしまう。あるいはカントに従うならば心の科学は認識論に還元されてしまう。これら過去の哲学者の心に関する説は心の哲学と心の科学といった一見無害な区別さえ台無しにしてしまう効力をもっている。そこで、最初に述べた心の科学と心の哲学という区別について考えておかなければならなくなる。両者の関係を見定めるためには心を科学的に扱うことの背後に横たわる前提を明らかにする必要がある。その前提が自然主義(naturalism)である。既に私は心の科学を確立するための哲学として心の哲学を位置づけると述べた。この主張の前提になるのが自然主義である。したがって、自然主義とは心に関する哲学的な立場であり、それ自体が哲学的な主張である。当然ながら、自然主義という立場に反対する心の哲学も存在する。そのような哲学の批判ではなく、自然主義的な立場から心の哲学を展開し、それを心の科学に結びつけるという試みが以後の議論である。

1 自然主義(=形而上学的な主張)

自然主義は今日流行しているが、それが正確に何を意味しているかは意見が分かれる。ある哲学者には哲学と経験科学の連続性を主張するものと受け取られている。他の哲学者は二元論の否定こそが自然主義の重要な主張であると考え、あるいは自然主義の本質は認識論の外在主義的な理解にあるとも考えられている。これら三つをすべて含め、二元論と認識論的な内在主義に反対し、哲学と経験科学を連続するものと捉える立場が自然主義と考えている哲学者もいる。自然主義に似た言葉として物理主義も使われてきた。物理主義とはすべての自然現象は物理的な現象であるという主張である。ここでは自然主義を用いるが、それは私の立場が物理学だけではなく、生物学や脳科学を広く含み、「行動」、「主体(agent)」という物理学にはない概念を必須の概念として含むからである。

存在する性質、状態、対象、出来事は自然的なものである。あるものが自然的とは、それが自然科学の基本的な理論によって理解できることである。このような自然理解は決して今世紀特有のものではなく、昔から存在していた。例えば、デモクリトスの原子論は古代の唯物論の典型である。あるいはルクレティオスでもよい。スピノザがアリストテレスの目的因やプラトンの考えを否定し、道徳の自然性や相対性を主張したのも自然主義の例である。ヒュームが実体という概念を否定し、自由、自己、因果性等を心理学的に扱うのも自然主義の典型例である。現代に眼を転じたとき、自然主義的傾向は直接に肌で感じられるほどに強い。その好例がアメリカ哲学と自然主義の関係であろう。アメリカの哲学では自然主義的な傾向が強く、それは既に今世紀の初頭からサンターヤナやデューイに見られる。そのような特徴は「自然化(naturalizing)」という用語に端的に現れている。これらをもとに自然主義の定義を考えれば、次のようになるだろう。私たちがもつ最善の科学理論が存在するもののタイプへの最善の指針となる。状態、性質、対象、出来事は科学理論によって実在的だとみなされる場合に限って実在的である。自然主義の主張は簡単には科学理論が世界の最善の像を与えるというものである。多くの自然主義者は科学理論が実在する世界についての理論であると考え、科学は実在についての科学であるという実在論の主張は自然主義と重なっている。

さて、以上のような自然主義の表面的な理解のもとで、自然主義と心はどのような関係になっているのか。すぐに頭に浮かぶのは「心は自然科学の対象になるか」という問いであろう。この問いはまさに自然主義

と心との関係を探っている。心が自然のものか、それを超えるものかは最初に私たちが問題にした点であり、過去に幾度となく論じられてきたものである。心を自然主義的に理解するというプロジェクトは、したがって、心が自然科学の対象に十分なり得るという前提を認めることであり、これは一つの哲学的な立場、態度である。この態度の是非は実際にこのプロジェクトを実行してみないことには断定的なことは何も言えない。実際、以後の話はこのプロジェクトが遂行されている有様と、遂行上の問題が中心になる。その際の哲学の具体的な作業は一見自然的には見えない心の振舞いや特徴をどのように自然化するかということになる。

自然主義について哲学的に議論する際に気をつけなければならない点は自然主義の内容が科学の進展と共に変化する点である。科学理論が与える自然像は科学理論が変化すればそれに応じて変化する。自然主義の変化は科学の変化に呼応している。科学の内容や形式、そして方法が変化するに応じて自然主義の内容も変化する。このテーゼのもとに自然主義も歴史的に変化することを受け入れるならば、自然主義に関する無用の論争の大半は避けることができる。自然主義の範囲や限界に関して語ることは科学の変化を抜きにしては意味がない。また、科学的でないものも実は科学の変化と同じように変化している。科学的でないものの内容の変化に応じて自然主義の内容は変化する。したがって、自然主義の意味は時代や理論変化に応じて変化する、定まらない。科学と非科学の境界が曖昧であるように、自然主義と非自然主義の境界も曖昧である。

(#) 自然主義の範囲や内容が定まらないことは自然主義が全く流動的ということではない。私たちは手に入る最善の科学知識と神の存在やそこからのいくつかの帰結とが両立しないことを知っている。あるいは、「桃太郎」や「孫悟空」が実在しないことも科学知識に照らして知っている。科学知識と両立しない知識は排除できるが両立する知識については自然主義的に何か言うことはできない。心を因果的に理解できるか、心的な体験は科学的に理解できるかという問いについては自然主義は積極的なことを述べてくれない。端的に、心の自然主義的な理解は可能かといった問いは一見魅力的に見えるが、実は自然主義の流動性ゆえに肯定的にも否定的にも解答できてしまう。心が脳に付随するか、それとも創発的なものかという問いも同じ運命にある。どのような自然主義を採用するかによって自在に解答できてしまう。その意味で自然主義が正しい主張かどうかさえる確な問いとは言えない。自然主義では理解できないものを自然主義的に理解した後でなければ自然主義の善し悪しについては何も言えない。重要なのは、したがって、自然主義ではなく、どのように自然化できるかである。

自然主義、科学主義、物理主義と言った言葉の周辺の事情に触れておこう。非自然主義の練習問題として数学的知識を取り上げてみよう。数学は自然主義的ではないと考えられてきた。というのも、数学的对象は自然の中には存在しないし、その性質も自然の対象のもつ性質とは異なっている。それは概念的な対象と多くの共通点をもっている。しかし、誰も数学的对象が反自然主義的であるとは思わない。それが証拠に、数学的对象は自然的対象を表現するのに実に有効であるし、それをを用いることが自然的対象を損なうということもない。数学は、したがって、昔から自然の言語と考えられてきた。私たちが自然を扱い、利用する典型的な道具の役割を数学は担ってきた。それは決して反自然的ではない。自然主義的に数学的对象を直接に説明し尽くすことは困難であるが、それをを用いると自然の対象を正確に描くことができるのである。このような道具の最も一般的なものが言語である。文法の規則や意味論が自然主義的に扱い難いのは数学の場合と同じであるが、言語を用いることによって私たちは自然を表現し、理解することができる。と同時に、そのような言語がどのように獲得されることになったかの進化論的、心理学的研究は科学的なレベルで遂行できる。したがって、間接的ではあっても言語は立派に自然主義によって扱うことのできる対象である。

自然主義は自然をどのように理解しなければならないかまでは指定しない。実際、自然主義的に描き出される世界像は複数あり、互いに両立しない像さえ沢山ある。その簡単な例が静的自然像(static view)と動的

自然像(dynamic view)である。近代は確かに動的世界観のもとに動いてきた。それは不安と期待の両面をもっている。自由に未来を切り開くことができるという期待と一寸先は闇という不安の同居する世界である。この動的世界観を緩和するのが物理的な知識であり、それは静的世界観を反映している。この二つの世界観を確率概念を通じて考えてみよう。

どのような自然でなければ確率は適用できないか。確率の公理を最低限は満たさなければならない。また、それが偶然という概念に適合することが望ましい。偶然という概念に適合するような確率の適用は動的な自然観を含意する。予測ができないという意味での偶然も、目的をもたない(特定の結果を想定しない)という意味での偶然も動的世界観を含意するなら、自然選択が適用される世界は動的世界であり、力学的世界とは根本的に異なっている。そこからダーウィンの自然とニュートンの自然は世界観において異なっていることがわかる。量子力学の予測、これは物理的な世界での偶然を部分的に表現している。ここでは動的な世界観でも静的な世界観でもいずれであっても成立する。つまり、背後の世界観とは独立に偶然が存在できる。その意味で客観的な偶然である。非線形のシステムの初期条件への感受性もたらす非決定性という偶然は、もし力学的システムであれば、静的世界観を採用した場合は古典的無知による偶然であり、客観的ではない。

動的世界観は常識存在論(folk ontology)の基本原則である。少なくとも、動的世界観は物理学の理論から帰結するものではない。それは半ば直観的な自然観である。心の世界における存在論はどうか。これも動的である。むしろ、心の世界の存在論がそのまま自然に適用されたのが動的世界観である。少なくともそれは知覚的世界の表現としては極めて自然である。認知レベルでの心の世界は力動的で、創造的である。私たちの心が創造的に働くということは動的世界観そのものである。

心は動的、創造的で自由意志をもつ

という文は動的世界観を代表している。知覚を基本にした世界は動的である。それは知覚が動的であるからである。運動や変化は知覚にとって必須のものである。理性的な把握は基本的に静的である。世界そのものが動的か静的かはわからない。しかし、知覚的世界が動的で、理性的世界が静的なのは確かである。この違いは認知の実際の過程と規則の整理された形の違いとして例示できる。論理規則はその使用の細則まで定まっていらない。規則は静的世界しか要求しないが、規則の使用は動的世界を要求する。表象のもつ二面性が静的、動的の区別の基本にある。表現するものとされるものの分離である。変化を表現するにはそれを把握しなければならない。その把握は静的である。静的なのは表現される内容であって、表現自体ではない。

(#) 自然主義批判の例

自然主義への批判は多い。科学が嫌いな人が心までも科学的に理解すべきだ言われたら、その人が嫌悪感をもつのは明らかであろう。このような科学嫌いではないにしても、心の能力、人間の行為の理由、倫理や道徳を考慮したとき、かなりの人はそれらが自然化できるものではない、あるいは自然化されるべきではないと思う。あるいは、科学が実在する世界の記述や説明を与えてくれるのではなく、歴史や文化に相対的なものであり、私たちの世界観に依存したものであるから、心をそのような世界観のもとに考えることは単なる一つの立場に過ぎないと考える哲学者も多い。ほとんどの自然主義批判は次の自然(物理)主義の主張に対する批判である。

The thesis of the explanatory adequacy : It leads directly to the conclusion that mental states are physical states. Pain, for example, is ultimately described and explained in terms of fundamental physics. So pain is physical in the sense that it is completely describable and explainable by physics.

例えば、Rorty は伝統的な哲学が消え、その役割の一部は科学が引き継ぐと考えるが、経験的な結果が哲学の問題にいつも答えてくれるわけではなく、科学以外のものが場合によっては適切に答えてくれると主張する。また、Putnam は科学が真理の唯一の供給者であることを否定し、科学的な合理性は相対的なものに過ぎないと考える。Nagel は科学理論の説明的な十分さを否定する。物理学が十分な説明を与える点に関して、例えば、MacGinn は心

的なものは物理的なものであると考えるが、私たちには心的なものとの物理的なものとの関係を理解することができないと主張する。これらの哲学的な批判の他に、心を自然主義的に理解することへの宗教的、倫理的な批判も多く存在する。このような批判には人間をどのように捉えるかという人間観の違いに依存したものが多い。ここでは人間を特別のものとしてではなく、自然の中の対象の一つであるという立場を保持したい。

まとめ

自然主義は変化する。ものには見られない心の独特な性質や振舞い（と考えられてきたもの）を自然化する、自然主義的に理解するというのが以後の話である。それは心の科学を創っていくことと同じである。そして、それを実行するのが心の哲学の任務である。

2 心身関係あるいは心の哲学の歴史

イリアスでの心とオデッセイアの心、デカルト以前と以後、これらの間には心に対する考えの大きな進展がある。それは心概念の進化と呼ぶに相応しい。特に強調していいのはデカルト以後の心の扱いである。二元論は世界には二つのタイプのもがあると主張する。大抵の場合、一つは物理的で、他は非物理的である。なかでも身体は物理的なものの典型例であり、心は非物理的なものの典型例である。この極めて常識的な二元論は二つの基本型をもっている。実体二元論と性質二元論である。一方、一元論は世界には一つのタイプのものしかないとする。そのものが物理的あるいは非物理的であるのに応じて唯物論と観念論と呼ばれる。

(#) 一元論とか二元論という区別は現在の私たちの常識に合っているだろうか。誰も世界は一種類のものからできていると思っていないし、無限に多くの種類のものからできているとも思っていない。化学元素を文字通り物質の種類と認めるならば、世界は多元論的である。そこで化学元素はみな物質的なものだと考えれば物質一元論となるが、そのようにみならず十分な根拠を私たちはもっているだろうか。十分な根拠がなければ一元論は恣意的な分類の結果に過ぎないことになる。元素をさらに分解して素粒子のレベルまでもっていても、やはり同じように何元論かは今の私たちには断定できない。何元論かという議論は世界の構成に関する真摯な議論というより、私たちの世界に関する考え方を整理するためのものに過ぎないというのが歴史の教訓である。

2.1 デカルト

心と身体はどのような関係にあるのか。この問いに対して、心と身体は世界の二つの異なる実在であり、それらの間には因果的な相互作用がある、というのがデカルトの解答 = 主張である。この心身相互作用の二元論と対照的なのが、今世紀に登場する心身(脳)同一説である。心と脳は同一で、心的な性質は物理的な性質であるというのがこの説の主張である。心的な用語と神経生理学的な用語の違いは、「水」と「 H_2O 」という用語の違いと同じであると考えられる。用語そのものは異なってもその指示対象は同一であり、したがって心と脳は同一であると主張される。同一説以外の今世紀の解答もすべてデカルトの主張を出発点にしてそれを批判する、あるいは継承するという形にまとめることができる。デカルト的な解答は自然主義的な解答ではないこと、今世紀の諸説は自然主義的であること、この二つから自ずと今世紀の諸説の大半は反デカルト的ということになる。

(#) 現代の心身相互作用論には、例えば、Popper-Eccles、Penroseの説がある。しかし、その相互作用の仕方に関しては説ごとに異なっている。相互作用論の最も重要な点は相互作用の特徴づけにある。心がデカルトのものとは異なる実体であると、自ずとその相互作用はもの世界だけで特徴づけることができなくなってしまう。認めやすい相互作用は自然の中の相互作用であり、それは心がこの自然の中であって、そこで相互作用するということを意味している。

まず、デカルトの主張を批判的に吟味してみよう。

デカルトは彼が心をもつことを疑うことができない。
 デカルトは彼が脳を持つことを疑うことができる。
 (ライプニッツの不可識別者同一の原理(後述)を使って)
 それゆえ、心と脳は別のものである。

精神をもつことを疑うことはできないが、身体をもつことは疑うことができるとデカルトは考える。そこから、精神と身体が異なることをライプニッツの法則から導き出す。ここで次のような例を通じて上の議論が誤りを含んでいることを見てみよう。その誤りは意味論的な指示(reference)と意味(sense)の違いに関係している。それは哲学では命題的態度(propositional attitude)と志向性(intentionality)として考えられてきたものである。

伊作は明けの明星を観察したい 伊作は金星を観察したくない

ここにライプニッツの法則を適用するなら、伊作が観察したいという性質は金星にないことから、ライプニッツの法則から明けの明星と金星は異なるものになる。無論、これは誤っている。明けの明星と金星は同じ対象を指している。

伊作は明けの明星を観察する 伊作は金星を観察する

上の文では明けの明星と金星が同じものである。これは、伊作が明けの明星を観察したい、金星を観察したくないという事実と何ら矛盾しない。これは最初の二つの文が示している事実が、明けの明星と金星が異なる性質をもつことを示していない。同様に、

私は脳をもっている 私は心をもっている

という二つの文について、一方を疑うことができ、他方は疑うことができないということから、脳と心が異なる性質をもち、異なる対象であるということは結論できない。

ここで重要なのは「疑う」や「望む」は命題に対する私たちの態度であるということである。たとえある命題が望まれ、他の命題が望まれなくても、最初の命題が述べているものと二番目の命題が述べているものが異なるということは導かれず、というのが上の例から言えることである。「私が脳をもっている」と「私が心をもっている」が異なった命題であって、一方は疑うことができ他方は疑うことができないとしても、そこから脳と心が異なるということが結論できないということである。したがって、デカルトの論証は誤っていたということになる。

(#) 伝統的な心身問題は、デカルトの場合、意識 - 身体問題といってもよい。意識と身体がどのような関係にあるかという問題である。デカルトは心と意識を同じとみなしたので、心的状態は意識の状態のことである。彼の実体二元論は、したがって、意識的なものと延長するものの二元論である。

2.2 行動主義

行動主義は心的出来事に関する話は観察可能な行動に関する話に翻訳されるべきであるという哲学的主張である。したがって、痛みのような心的表現を内観報告に頼って定義することは否定される。

1 論理的行動主義

論理的行動主義 (*Philosophical behaviorism, or Logical behaviorism*)は心理的な用語は意味を損ねることなく行動の用語に翻訳することができるという見解である。翻訳は同じ外延をもつという関係に弱めて考えられる場合もある。また、行動の他に神経的、物理的な用語を含めて翻訳を考える場合もある。論理的行動主義の典型は論理実証主義の意味の検証理論 (=用語の意味はその検証の条件である) に見られる。その代表は ライル(Gilbert Ryle)で、彼の考えは *The Concept of Mind*(1949)で展開され、心身問題に関して大きな影響を与えてきた。彼の行動主義は私たちの日常世界での心的な用語の意味に関する主張で、私たちが個人の思考や信念、意図や感情について語る時に何が意味されているかを明らかにしようとする。その点で科学に関する主張である方法的行動主義(後述)とは異なっている。ライルはデカルトの二元論を批判し、それを「機械の中の幽霊ドグマ (*the ghost in the machine dogma*)」と表現する。観察できる出来事が身体と不可分に結びついているように、内的な出来事は心と不可分に結びついている。これは誤っているというのが彼の考えである。心的用語は人が行動する仕方に言及しているのであり、内的な心的状態に言及してはいない。この心理主義への反対は次の論証を考えればよい。

もし心的状態が行動の内的な原因であるなら、私たちは他人の心的状態についての知識をもてないだろう。ところが、私たちは他人の心的状態についての知識をもつ。それゆえ、心的状態は行動の内的原因ではない。

他人の心的状態についての知識は類推によるというのが伝統的な考え方である。これは「他者の心の問題」としてずっと議論されてきた。ところで、素粒子がものの運動の内的な原因であることは、私たちが素粒子について知ることができないことを意味してはいない。同様に、心的な状態が行動の内的な原因であるという見解は、その内的な原因が存在するかどうかに関する懐疑主義にすぐに結びつくわけではない。ここで次の区別は重要である。

命題 p は観察可能な対象に関するものか、命題 p は観察によってテスト可能か。

これら二つの問いは基本的に異なった問いである。また、同じ解答をもつ必要もない。素粒子や他人の心的状態に関する主張は直接に観察可能な対象についての命題ではない。しかし、それら命題について観察によるテストを実行することができる。

(#) 実際、物理学的な探求のほとんどは私たちが直接に観察できない対象に関するものである。では、そこで直接観察できないような原因や原理を想定することは禁止されているだろうか。答えはノーである。それどころか、積極的に見えない原因や原理が仮定されている。力学的な力を知覚できる人はいないが、その存在と働きは十分テスト可能である。力が働くことによってどうなるかは観察できる。物理学で認められていることは心理学でも認められるというのが妥当な見解であろう。では、先取りになるが、意識はどうか。意識は直接観察できないが、テスト可能であろうか。力や素粒子はそれ自体観察できなくとも、テストが可能なのでその存在を物理的に仮定できるが、意識は、もしデカルトのように空間的な広がりをもたないとすれば、たとえテスト可能であるとしても、どのような仕方で仮定できるのだろうか。

ライルの二番目の反対は自然主義的誤謬(*naturalistic fallacy*)あるいはカテゴリーミステイクである。「デカルトは心をもっている」、「デカルトは身体をもっている」はそれぞれまっとうな命題であるが、「デカルトは心と身体をもっている」という命題はライルによれば許されない。心と身体は別のカテゴリーに属するからである。これはほとんど二元論の否定と言ってもよい。この主張のどこがおかしいのか。

- (i) デカルトは心をもち、脳をもつが、それら二つは異なる。
- (ii) デカルトは心をもち、脳をもつが、それらは同一である。
- (iii) 人が心をもつのは正しく、人が脳をもつのも正しいが、前者が後者の上に存在するものであると考えるのは誤りである。

これら三つの命題はそれぞれ(i)が二元論、(ii)が同一説、(iii)が行動主義の主張を表している。これらは言語の規則に反しているだろうか。反してはいない。またライルは傾向性(disposition)を使って心的状態や性質を分析しようとするが、傾向性で状態を置き換えることはできない。痛みの行動がなくても痛みは存在可能である。(Dennettの例、手術中に体の自由を奪う麻酔と、その記憶を喪失させる薬を併用した場合、それは通常の麻酔と同じ効果をもつだろうか。)

(#) A disposition in general can be illustrated in the *tendency* of glass is to break. Human dispositions are expressed in the form of conditional (if-then) statements. For example, the disposition of Jones to be hungry might be expressed in the following complex conditional: IF Jones has food set before him, and it is not poisoned, and the situation is socially appropriate ..., THEN Jones will eat the food.)

ライルは自分を心理学的行動主義者とは考えないが、痛みと言う例外を除いて、すべての心的状態は行動を通じて分析可能であると考え。また、行動に対応する心的状態の存在も否定する。そして、私たちは意識的な心的出来事をもっていないと結論する。ライルは内的状態は傾向性を説明するのに因果的に関連していない、したがって、適切ではないと考える。心的出来事というのはある仕方で行動する傾向性のもとでの人の行動について予測することである。特定の信念は特定の行動を記述するに過ぎない。

2 方法的行動主義

デカルト的な二元論では心は非物理的で、第三者には観察できないゆえに心理学は科学としては成立不可能である。科学として成立するには、刺激、条件付け、反応といった観察可能なものにだけ言及する行動の理論として心理学を確立しなければならない。方法的行動主義 (*Psychological behaviorism* or *Methodological behaviorism*)は心理学固有の領域は行動の研究であると考え。観察できない内的状態に訴えることは方法論的に不可能である。このような考えは J. B. Watson によって最初に主張された。Watson は人間の活動のうちで公的に観察可能な特徴だけを使って心理学を構成し、その結果、心理学を *hard science* として確立しようとした。人間の心的状態を正確に解釈する唯一の方法は人間の行動を観察することによってであると彼は考えた。彼の弟子である B. F. Skinner は心的状態を記述するのに通常使われる多くの用語と行動を結びつける実験を行うことによって行動主義的な理論を構成した。心的状態の存在に関しては心理学的行動主義者は不可知論を採る場合と、フロジストンやカロリックと同じとみなし否定する場合の二つに分かれる。方法的行動主義は 1913 年から 1957 年までアメリカの心理学の主流であった(1913, when Watson wrote his seminal defense, and 1957, when Chomsky published a devastating review of Skinner's *Verbal Behavior*)。

(#) Noam Chomsky in his critique of Skinner's book *Verbal Behavior*

その時の彼の強烈な反論は次のような内容であった。私たちは無限に多くの異なる文をつくり、理解することができる。行動主義は過去の反応と条件付けの歴史を強調するが、どのようにこの無限の内容を扱うか知らない。最初に経験する新しい内容をどのように習慣的な刺激 - 反応の図式で理解できるというのか。

方法的行動主義は心理学がどのように研究されるべきかについての主張である。この点がまず論理的行動主義と異なる。論理的行動主義は、信念や欲求が行動を引き起こす心的状態であることを否定するが、方

法的行動主義は心理主義的な用語が内的状態を指示することを認める。しかし、それらが内的状態ゆえに信念や欲求について心理学的に語ることは否定する。方法的行動主義の否定的なテーゼは、心理学は信念、欲求による説明を避けるべきであるというものである。この主張は Skinner に代表されるが、心理主義的な説明に対する彼の反対の理由の一つは、信念や欲求は観察できないというものである。二つ目の理由は説明が余りに容易過ぎるというもので、それらを仮定することで観察結果に合う信念や欲求の話を作り上げることができてしまうということである。ここで以下の区別は重要である。

(mentalist thesis) 個人の行動はその人がもつ信念や欲求によって引き起こされる。

(specific belief/desire hypotheses) 特定の信念や欲求は行動のデータに対してテスト可能である。

心理主義のテーゼはテスト可能ではないとしても、特定の信念や欲求をそれぞれの行動に帰することはできる。行動は有機体の内的な状態を記述しなくても説明できるというのが方法的行動主義の肯定的な主張である。方法的行動主義は、(1) 対象のいた過去の環境、過去に行った行動、(2) 現在の環境、この二つから対象の現在の状態を説明できると考える。Skinner の条件刺激の理論はどのようにそのような説明が可能かを示してくれる。

ネズミはベルが鳴るとえさを取る。ベルが鳴る。それゆえ、ネズミはえさを取る。

方法的行動主義の難問は、新しい行動をどのように説明するかである。実際、この点がチョムスキーの批判であった。過去の状態と行動の間の相関関係は現在でもそのまま適用できるのか。過去の相関関係によって条件づけられた行動を説明することは行動主義者には完全に可能であるが、信念や欲求によって行動する場合にはそうではない。つまり、方法的行動主義の環境決定論は誤っている。方法的行動主義は私たちが環境を自主的に変えていくことをうまく説明できない。

(#) 私たちの積極的な行動を際立たせているのは創造性(creativity)である。これを行動主義的に説明することは極めて困難である。せいぜい主張できるのは予測不可能性(unpredictability)である。では、この二つはどのような関係にあるのだろうか。創造性は背後に確実に動的な世界観を必要とする。それがなくては創造性自体を考えることができない。創造性を古典的な無知の結果(単に知識の欠如から知らなかったということ)とみなす消極的な特徴づけは十分ではない。確かにそのような古典的な無知の結果という側面は存在する。「思いつかなかった」が、言われてみれば確かにその通りということがしばしばある。しかし、一方で芸術的な創造性を単に思いつかなかったで済ますには余りに創造的である絵画や音楽が存在している。ところが、予測不可能性は「思いつかなかった」で済ますことのできる部分を沢山もっている。運命論的な世界では創造性も予測不可能性もいずれも古典的な無知の結果でしかない。また、非決定論的世界では創造性と予測不可能性の区別さえつかない。したがって、決定論的世界でこそ二つをどのように扱うかが問題となる。

2.3 同一説

心的性質(状態、過程)は脳の性質(状態、過程)と同じものである。つまり、すべての心的性質はある物理的性質と同じである。気体の温度は気体分子の平均運動エネルギーであるのと同じ意味で、心的性質は脳の性質である。これが同一説の主張である。Smart の 'Sensations and Brain Processes' (1959) は同一説の最初の典型である。痛みと C-fibers の興奮は同義語ではないが、にもかかわらず同じ性質を指示する。痛い状態にあるという性質と C-fibers が興奮している状態にあるという性質は同一である。

(#) (三つの同一説)

1 J.J.C. Smart の主張：しばしば脳過程(に関する)唯物論と呼ばれ、感覚は脳過程と同一であると主張する。

(sensations are identical with brain processes.)

2 David Armstrong の主張：中枢状態唯物論と呼ばれ、心的状態は脳と中枢神経系の状態と同一であると主張する。

(mental states are identical with states of the brain and central nervous system.)

3 Herbert Feigl の主張：神経生理学的な用語は心的用語を指示すると主張する。(certain neuro-physiological terms denote certain mental terms.) Frege の意味と指示の違いに従って、心的用語と神経生理学的用語は意味は異なるが、指示は同一であると考える。

ライプニッツの同一性の法則(Leibniz's law of identity)は、二つのものが同一なら、それらがもつ性質もすべて同じである、つまり、 $(x)(y) [(x=y) \supset (Fx \supset Fy)]$ であることを主張しているが、この法則から、心的出来事と脳の活動は異なる性質をもつので同一ではないという同一説批判が出てくる。例えば、私は様々な記憶をもっているが、それが脳のどこに貯えられているか知らない。だが、同等の知識をもつことは同一性にとって必要ではない。別の同一説批判によると、心的記述と物的記述は次の三つの点で同一ではない。空間的な局所性、客観的観察可能性、志向性がこれら三つの点である。ライプニッツの法則から一つの記述によって述べられる性質は別の記述によっても述べられるが、一つの記述の性質は他の記述には適用できない。

(空間的な局所性を使った議論)

1.脳過程は空間的出来事である。 2.感覚は空間的ではない 3.同一性の法則は一方について正しいものは他方についても正しいと主張する。 4.それゆえ、二つのものは同一ではない。

客観的観察可能性に関しても類似の議論を展開できる。脳過程は観察可能であるが、経験する感覚はそうではない。一方、志向性は心的状態に適用されるが、脳の活動には適用されない。

三番目の批判。同一説は心的な記述を物理的な記述に還元し、心的な記述を消去しようとする。問題はライプニッツの法則からこの消去ができない点にある。椅子と分子の塊が同じ対象を指示しても、そこから一方の記述が消去されるということは出てこない。いずれも有用な役割を演じている。消去には同一であるという以外の事柄が必要である。

四番目の批判。同一説は一辺倒な主張である。心的出来事を生物的なシステムにだけ制限している。心的出来事はコンピュータのような非生物的なシステムにも開かれている。

このように批判を見てくると、別のタイプの同一説の可能性が考えられる。それは翻訳の同一性ではなく、消去の同一性を主張することである。その代表は Rorty や Feyerabend の主張で、disappearance identity theory と呼ばれている。心的なものを含む話は脳の活動だけを含む話に置き換えられ、消去されるべきであると主張される。すべての心的現象についての話は幽霊現象についての話と類似している。これは同一説というより後述の Churchlands の消去的唯物論に近い考えである。したがって、後に詳しく検討してみよう。

(#) 同一説批判の多くはライプニッツの原理に依存している。では、この原理は本当に正しい信頼できる原理なのだろうか。昨日の私と今日の私は異なる性質をもっているが、同一人物である。あるいは、この部屋の酸素分子はみな同一で区別ができないといわれるが、空間的には異なる場所にある。ライプニッツの主張はこのような直観に反するのではないだろうか。彼の原理を正確に述べておくと次のようになる。

不可弁別者同一の原理 Identity of Indiscernibles

二つのものの区別がつかなければそれらは同じである (*Discourse on Metaphysics*)。どんな二つのものも正確に同じ性質をもつことはない。これは集合論の外延性の公理であり、 $(F)(Fx \supset Fy) \supset x = y$ と表現できる。この原理の逆、 $x = y \supset (F)(Fx \supset Fy)$ 、は同一者の不可弁別性を主張している。しばしば両者をまとめてライプニッツの法則と呼ばれている。ライプニッツの法則は、 $x = y \supset (F)(Fx \supset Fy)$ と記号化できる。

(同一説に対する基本的疑問)

同一であるとする際の判定理由はいつも心的な出来事、心的な記述をもとに行われる。これは例えば、DNAと遺伝子の同一が言われる際にも生じることである。心的出来事と脳の活動が同一という判定はいつも心的出来事をもとに行われる。この一方方向性はどのような条件の下で補正されるのか。あるいは、この一方方向の同一性では真の同一性ではないというようなことがあるのだろうか。遺伝子型と表現型の間の複雑な関係は心と脳の間にもそれ以上の複雑な関係が存在することを示しており、したがって、心と脳の状態や出来事の全くの同一という主張は維持が困難で、それに係わらない還元や消去に議論がシフトすることは当然であるのかもしれない。

自然数と実数の対応を考えてみよう。自然数はすべて実数であるから、自然数の計算のシステムは実数の計算のシステムでもある。したがって、自然数の個々の計算はそれぞれ実数の計算と同じである。だからといって、自然数の計算のシステムは実数の計算のシステムと同一ではない。これと同じように心の各状態や出来事が脳の状態や活動であると言っても、脳の状態や活動がすべて心の状態や出来事であると言えない限り、そこから心は脳と同一であるということとは出てこない。

複数実現可能性が問題となるのは「共通外延がない」ことをそれが含意するからである。次の議論(Coextension Argument)を見てみよう。

1. ある心的性質 M があって、それを一つの場合に実現する物理的性質の組 $\{P_1 \vee P_2 \vee \dots \vee P_n\}$ はあるがすべての場合に実現する特定の物理的性質はない。2. もし心的性質 M がある物理的性質と同一であれば、それは $\{P_1 \vee P_2 \vee \dots \vee P_n\}$ のどれかと同一である。3. しかし、 $\{P_1 \vee P_2 \vee \dots \vee P_n\}$ のどれも 1 より M と同一ではない。4. M と共通の外延をもつ物理的性質は存在しないので、 M と同一な物理的性質は存在しない。それゆえ、同一説の主張は誤っている。どのような物理的性質とも同一でない心的性質が存在する。

科学が経験的であるのと同じ意味で、同一説は経験的な主張である。そして、それは明らかに唯物論の立場に立つ。したがって、すぐに同一説の説明構図の方が二元論より単純であることがわかる。その主張は単純明快で、心と脳は同一で、それは経験的に確認されなければならない。その結果として、心は還元・消去され、脳の振る舞いによって置き換えられることになる。脳に関する研究者が心について何か述べようとするとき、脳に関する研究は心に関する研究でもありと考えるのは、背後に同一説があると理解しやすい。しかし、認知科学の内容は必ずしも同一説の立場に立って理解する必要はない。その主張は同一説とは異なり、次の機能主義に基礎を置くものである。

2.4 機能主義 (Functionalism)

機能主義の特徴はタイプ - トークンの区別にある。個々の物理的な対象はトークンであり、性質(種、kinds)はタイプである。概念とそれを表す具体的な対象がそれぞれタイプとトークンである。ある物理的な対象は多くのタイプのトークンである。例えば私のパソコン(トークン)はさまざまな性質(タイプ)をもっている。タイプはいくつかのトークンをもつ、あるいは何もトークンをもたない場合もある。例えば、人間という性質は個々の人間がそのトークンとなるが、ユニコーンはトークンをもっていない。さて、同一説は心理的なトークン、心理的なタイプ両方についての理論であるが、機能主義は同一説が心理的なタイプについて主張することを否定する。つまり、心理的なトークンは物理的なトークンであるが、心理的な性質は物理的な性質ではないと主張する。機能主義は心理的性質が物理的なものに付随する(supervene)、したがって心理的性質は複数の物理的なトークンによって実現される(multiply realizable)と考えるのがその理由である。これは、心をもつものを作り出すのに多くの異なる物理的な方法があることを意味している。したがって、心的な性質は物理的な性質ではない。しかし、心的なトークンは物理的なトークンである。これが機能主義の主張である。

(#) 付随性と複数実現性(*supervenience and multiple realizability*)

この二つの概念は機能主義的な説明をする際に必ず登場する概念である。複数実現性は上に説明があるので、付随性を以下に述べておく。心的状態が脳状態に付随する(Mental states *supervene* on brain states.) とは二人の人(あるいは異なる時間の一人の人)の脳状態に違いがない限り、心的状態にも違いはないということである。二人の人が同じ脳状態にあれば、二人の心的状態も同じであるというのが心的状態が脳状態に付随するということである。心的状態が脳状態に付随するなら、心的内容は脳の中にあることになる。(なぜか?)

心理的な性質は特定の物理的なトークンではないとすると、どのような性質なのか。曖昧ではあるが、機能的な性質であろう。機能的な性質とはどのような性質なのか。次の例で考えてみよう。

ある人がそばにあるのが酒であると信じ、その人が *X* の状態にあるなら、その人はその酒を飲む。

この文は条件文で、特定の信念と状態が満たされれば、その人が何をするかを述べている。状態 *X* は恐らく、酒を飲みたいという欲求であろう。機能主義はこの条件文が特定の欲求をもつとはどのようなことかを記述していると考ええる。適切な因果的役割を演じる状態であればどのようなものでも、それは酒を飲みたいという欲求である。その状態の物理的な構成は問題ではない。心理的な状態は行動や他の心的状態に対する因果的な関係によって理解されることになる。

(#) 代表的な機能主義

- ・ Machine table functionalism (Putnam, Turing machine): Putnam was interested in Turing Machine because the relation of the program governing the operation of the Turing Machine seemed to stand in much the same relation to the physical device as the mind stands to the brain. There seemed to be no reason to be a dualist with respect to a Turing Machine. Also a Universal Turing Machine will be behaviorally equivalent to the mind. Pain is equated with a state in the system that causally produces other states within the system as well as outputs from the system.
- ・ Computational or AI functionalism (Fodor, Block): to identify mental states with computational states inside the system, where the states are defined in terms of kinds of operations performed. Not identify a mental state with a state of a particular machine but with an operation that could be performed in a variety of machines.
- ・ Mind as a syntactic engine that emulates a semantic engine (Dennett)
- ・ Strong AI: computers and the mind perform equivalent procedures in producing their behaviors and these procedures are what this version of functionalism equates with mental processes.

ここでこの章の内容を分類、要約するための問いを考えてみよう。

(問い 1) 心的状態は内在的か。

これを否定するのが論理的行動主義である。

(問い 2) 行動を説明するのに内的状態は必要か。

これを否定するのが方法論的行動主義である。

(問い 3) 心は物質的か。

これを肯定するのが同一説と機能主義である。また、心理的な性質が物理的な性質かどうかに関して肯定するのが同一説、否定するのが機能主義である。

まとめ

デカルトは心はものとは異なり、心の所有者にだけ特権的に理解されるとみなした。今世紀の心に関する見方の大半はこのデカルトの見方に反対する。心はものと原理的に異なる実体ではなく、外から眺めることによって理解できるというのが心についてのその見方である。

付録1 付随性の実践的分析と批判

還元が不可能ならば、付随しない。ここで付随性に対する批判的な考察をしておこう。

(1) 次のような例から考えてみよう。

1 a は b の上にある。 2 a は b のとなりにある。 3 a は b に頼っている。

1 について、 a は b の上にあると同時に、 b が a の上にあることはできないという意味で、「の上にある」という関係は非対称的である。しかし、時点 t で a が b の上にあり、時点 t' では b が a の上にあることは可能である。これに対して2の対称性はいつでも成立する。その理由は「となり」という関係の定義からである。3は1と2に比べると微妙である。 a が b に頼っていることは b が a に頼っていることを排除しない。実際に頼っていないかもしれないが、少なくとも排除はしていない。つまり、対称的か非対称的かの判定が3だけを見たのでは言えないのである。「頼っている」の内容がどのようなものに依存するとはいえ、3が付随性の関係に近いものである。

(2) 基本的な付随性

「状態 a が状態 b にある時点で付随する」

これが最も単純な付随的な関係である。「心の状態はいつの時点でも脳の状態に付随し、脳の状態はいつの時点でも心の状態に依存する」という表現は心脳の付随関係論ということになるが、このような関係は成立していそうもない。あるいは、「心の状態はいつの時点でも脳の状態に付随する」という表現は心の随伴現象論 (epiphenomenalism) の主張であると考えられがちであるが、これは(1)の説明から誤っている。また、付随関係の代わりに相互作用を想定するなら、心身の相互作用論が最初の表現から得られるが、これもありそうにない。状態 a や状態 b の定義される領域は異なっており、したがって、それら領域の扱いは時点に関しても一様にはいかないからである。これは次のような具体的な場面を考えれば明らかであろう。

「ある時点」をある瞬間と考えたとき、瞬間の状態が力学モデルでは確定できるが、確率モデルでは意味をもっていないように、瞬間の心の状態と脳の状態と、それらの対応とはそもそもどのような状態や対応なのか。瞬間的な付随性は物理的にはほとんど無意味の付随性である。逆に、「 $2 + 3 = 5$ 」の計算の心的状態を考えたとき、そこでは時間的な経過や区間は考慮されていない。すると、それが付随する脳の状態はどのように定めたらよいのであろうか。どのような時間的な範囲を想定したらよいのか。実際のところ、私たちには定めようがない。このような考察から明らかのように、付随性の関係は瞬間化しようとしてもうまくいかず、また区間化しようとしてもうまくいかない。したがって、付随性の使用は決定的な結論を導き出さないばかりか、その使用の際には時間に関する文脈に対して十分な配慮が求められる。この結論は付随性概念が階層的に異なる領域の対象を扱う際に用いられる概念としてはそれほど信頼できるものではないことを強く示唆している。

付録2 Bell's Theorem による非付随性

3枚のカードがあり、それらは赤か黒である。これらを使って私とあなたでゲームをすることにしよう。あなたはカードの色を見ることができない。そこで次の三つの文が与えられる。

左端と真ん中のカードは同じ色である。

右端と真ん中のカードが同じ色である。

左端と右端のカードは違う色である。

明らかにこれら三つの文の少なくとも一つは偽である。あなたの課題は偽の文を探すことである。そのためにあなたはこれら3枚のうち2枚のカードを見ることができる。つまり、三つの文の一つを検証できる。もし偽の文を見つけたら、私が200円あなたに払い、そうでないならあなたが私に100円払う。

このゲームは公平だろうか。もしあなたが各文に確率 $1/3$ を与えるなら、偽の文を見つける確率は少なくとも $1/3$ である。「少なくとも」は二つの文が偽になる可能性も残しているからである。したがって、

$$((1/3) \times 200) + ((2/3) \times (-100)) = 0.$$

これはゲームが少なくとも公平、つまり不利にはならないことを示している。

(不正を防ぐために)

このゲームのよいところはあなたの質問を私が知る前に私がゲームを思い通りにできない点である。一つの文は偽でなければならない。残る問題は、質問を聞いた後で、カードをめくる前に不正ができるかどうかである。このような不正ができないようにゲームを少し変えよう。私たちがあなたをだまさないように、あなたは私と私の友人に同じ質問をする。私たちの答えが異なっていたらだましていることがわかってしまう。

最善の方法

私と私の友人は相互に連絡が取れないとすれば、あなたの質問には二人は独立に答えることになる。私たちの答えが一致したら、それをそのカードの「真なる値」としよう。「真なる値」が実在の色であるかどうかは問題ではない。私も私の友人も不正を働かないという仮定のもとでは、あなたは同じ質問に対して確率 $1/3$ を与えることができる。別の形で不正が考えられる。私と私の友人の間に隠れた回路があってこっそり情報が漏れる場合が考えられる。それを防ぐために私と私の友人は光の速度以上の連絡速度が必要であるような遠くに隔離されているとしよう。さらに考えられる不正は私があなたに誘導質問をすることである。あなたがどのような質問をするか推測できるのもまずい。このような誘導質問を一切なくし、質問も推測できないようにしなければならない。誘導質問は単純に禁止すればよいが、質問の推測のほうはどうか。一つの方法はサイコロを振って出た目によって質問をするといった工夫であろう。偶数の目が出たら最初の質問を、奇数の目が出たら二番目の質問をといた具合にすればサイコロ振りが公平になされる限り、どの質問にも偏りはないことになる。したがって、質問の仕方について特定の予測をすることはできないことになる。

(ベルの不等式)

確率を使ってあなたがゲームに勝つのは少なくとも $1/3$ である、というのがベルの不等式の内容である。つまり、

(あなたがゲームに勝つ) 確率 $1/3$ 。

(ベルの不等式に違反する)

確率を使った戦略では「少なくとも $1/3$ 」という答えを得ることができたが、これとは異なる戦略を使ってみよう。それが量子戦略である。先に答を言えば、私がゲームに勝つ確率は $2/3$ ではなく、 $3/4$ である。したがって、あなたが勝つ確率は $1/4$ に下がってしまうことになる。平均するなら、

$$((1/4) \times (-200)) + ((3/4) \times 100) = 25$$

となり、私がいつも勝つことになる。

工夫は量子力学に基づいている。既に述べた EPR の設定をうまく使おうというものである。スピン $1/2$ をもつ $P1$ と $P2$ の粒子対を考える。二つの粒子のスピンの合計は 0 である。私が粒子 $P1$ を、私の友人が粒子 $P2$ を取り、測定するとしよう。

あなたがカードについて質問したとき、私は粒子のスピンを測定する。ここで重要なのは、あなたの質問に応じて私が使う偏極器(方解石の角度)の角度が変わる点である。

あなたの質問	私の測定角度	上向き	下向き
左端のカード	60	赤	黒
真ん中のカード	0	赤	黒
右端のカード	-60	赤	黒

私の友人も同じ戦略を使い、しかし反対の答えをする。

あなたの質問	彼の測定角度	下向き	上向き
左端のカード	60	赤	黒
真ん中のカード	0	赤	黒
右端のカード	-60	赤	黒

では、一体どのようにするか。量子力学によれば、あなたが私に質問し、私の測定する粒子 $P1$ が方向 x ($= 60, 0, -60$) でスピン $1/2$ あるいは $-1/2$ の状態にあれば、私の友人の粒子 $P2$ はそれとは逆の状態にある。つまり、スピン $-1/2$ あるいは $1/2$ で方向は同じである。

もし同じカードについて質問されたなら、私と私の友人は上の表に従うなら同じ答えをするだろう。

もし別の方向について聞かれたら、量子力学はその結果を予測してくれる。別の方向でのスピンの測定についての同じ答えを得る確率は垂直方向からの方向のずれに依存する。その値は $\cos^2(x)$ である。したがって、

軸の角度が 60 度ずれていると、 $3/4$ の確率で私が勝つ、

軸の角度が 120 度ずれていると、 $1/4$ の確率で私が勝つ、

ということになる。

(結果)

三つの文はそれらの一つが偽でなければならぬが、各々は $3/4$ の確率で正しいことになる。このようなことは古典的な世界では有り得ないが、にもかかわらず量子的世界では正しい。

この内容を詳しく見てみよう。私も私の友人も同じ角度の場合、つまりあなたに同じ質問をされた場合は下に完全に答えは一致している。

私	赤	黒	黒	黒	赤	黒	赤
私の友人	赤	黒	黒	黒	赤	黒	赤

私と私の友人が異なる質問をされた場合、つまり異なる角度の場合は 60 度だと以下のようになる。

私	赤	赤	黒	黒	黒	赤	黒	赤	赤	黒	赤	赤	黒	黒	黒	赤
私の友人	黒	赤	赤	黒	黒	赤	赤	赤	黒	赤	赤	黒	黒	赤	赤	赤

赤と黒の不一致は $1/4$ ということになる。したがって、私と私の友人の答えが一致するのは $3/4$ ということになる。ゲームを拡張して、 n 枚のカードとすれば更に一般的な議論が展開できる。いずれにしろ、以上のことが示しているのは Bell の不等式が成立するような古典的な世界は量子論的な世界に付随しないことを明確に示している。

3 心の科学の歴史

心の哲学の歴史と心の科学の歴史を区別することは難しい。その第一の理由は心の科学はせいぜい 19 世紀の後半からであり、それまでは哲学の一分野と考えられていたからである。第二に、近年の認知科学の展開は哲学と科学の共同作業であり、哲学と科学の境界など問題でなくなってきたからである。それは意識についての最近の研究に端的に表れている。したがって、以下の話は文字どおりの心の科学の歴史ではなく、心理学、認知科学、その他の周辺領域についての話と考えてほしい。

(#) 方法的行動主義の歴史

行動主義は心的状態が観察可能な行動と同一である、あるいは私たちの信念や心的状態は私たちの行為を通じて説明されると考える哲学的な立場である。したがって、心的状態の存在やその適合性は否定される。現在では多くの人々によってその正当性が否定されている。

Skinner の *Science and Human Behavior* での攻撃的は彼が心理主義 (mentalism) と呼んだもので、それは人間の行動の説明に内的な心的状態を使う考えである。私たちは日常生活でこの心理主義的な説明や理解に慣れ親しんでいる。彼はこのタイプの説明の何が誤っていると考えたのだろうか。彼が心理主義の誤りとして挙げたのは次のような点である。

1. 心的な現象を私たちは直接に観察できない。
2. 心理主義的な説明は本当の説明ではない。
3. 心理主義的な説明は不必要である。
4. 入力と出力の媒介項としての内的状態は無視できる。
5. 心理主義的な説明は人間中心主義の反映

である。

3.1 認知科学(COGNITIVE SCIENCE)とはどのような科学か

認知科学者は情報や知識を扱う自然あるいは人工の認知過程を研究する。神経科学は人間の神経装置、つまりは脳を研究する。そこでは脳の物理的な性質や機能だけでなく意識のような高次の現象も研究されている。認知心理学は知覚、学習、記憶といった(脳ではなく)人間の心の性質を研究する。では、脳と心は異なるのか。心理学者が人間の記憶について語る時、何かを思い出している誰か、別の人間の心を想定している。心理学者が理論をつくる場合、この想定された心を使って考えている。心理学においては人間主体が何かをするということが前提になっている。

人工知能(AI)は人工的な認知過程の研究であり、心を機械的につくりだすことを目標にしている。どのようにしたら心を持つ機械をつくることができるか。これが目標である。よく聞く例はチェスを指すコンピュータである。現在のコンピュータはチューリングマシンによって概念的に基礎付けられている。

言語学は人間の言語を研究する。心理言語学(psycholinguistics)は言語獲得とその神経的な基礎を研究する。認知言語学(cognitive linguistics)は言語とそれ以外の認知過程、例えば、空間認知との関係を研究する。言語教育は応用言語学の対象である。計算言語学(computational linguistics)は言語過程のモデルを扱うが、言語に基づく人間とコンピュータの結びつきを考える際に重要になる。

哲学では心の哲学と呼ばれる領域で、人間の心とは何かを考える。私たちは機械に過ぎないのか。コンピュータは考えることができるのか。心と脳の関係はどのようなものか。心的状態はどのような状態か。知覚とは何か。哲学では心が登場する科学において心について語り、考える概念的な枠組みを追求する。

以上の異なる領域の研究が統合され心の科学がいつの日か成立するのだろうか。心の科学の最終的な姿はどのようなものか。

「歴史的に眺めると、認知科学は心の科学に対する極めて有力な候補であった。それは人間の心を記号処理装置、ある種のコンピュータとみなす。したがって、まず認知科学は人工知能と認知心理学の融合であった。この試みは1950年頃から始まる。今日、これに言語学と神経科学が更に融合されようとしている。心理学のほかの領域もこれに加えられている。この融合において出てくる問題は哲学的な議論として活発に論じられており、これには物理学さえ巻き込まれている。私たちはまだ心の科学の真の姿を見ていないが、人間の心の最終的な理論が認知科学と呼ばれるだろう。」

このような考えは正しいだろうか。何か誤りが含まれていないか。認知科学は最終的な心の科学として定義されない。認知科学という言葉はそれ自身の歴史をもっている。認知科学は心に関する自らの研究方法や手法をもつべきである。様々な領域が勝手に融合されて心の科学が出来上がるわけではない。それら領域が還元されてなくなるとも限らない。それでは、認知科学も残らない。

認知科学は人間の心を情報を受け取り、蓄え、変形し、伝達するシステムとみなした。これは情報処理システムとして心を考えることである。これら情報過程は様々なレベルで分析できる。このレベルの存在が認知科学の中に多くの領域を取り込む鍵になる。一つのレベルが具体的な実現(implementation)であり、神経科学の対象である。他は実現される計算的な仕組み(computation)で、人工知能や論理が対象にしている。これら二つのレベルが融合し、神経ネットワークを生み出し、別のレベルをつくることになる。心理学的なレベルは更に高次のレベルで、主体全体が対象である。哲学的レベルはメタレベルである。情報過程は表象的(representational)である。表象は認知科学の概念のなかでも極めて重要である。私たちの神経装置は外部世界を表象しなければならない。それら表象は有意味で、志向的である。表象は何かについての表象である。

人間の認知はまず形式的な記号処理とみなされた。これらの古典的見解は認知主義あるいは強い人工知能 (cognitivism or strong artificial intelligence) と呼ばれている。それは、もし正しいソフトウェアを実行するハードウェアを手に入れるなら、それが実際に人間の心であることを意味している。今日、認知科学はこの強い哲学的なテーゼと結びついてはいない。人間の情報処理は形式的に記述できると述べるだけである。

3.2 認知科学の仮定

心の研究は Wundt が実験心理学を 19 世紀に打ち立てるまでは哲学の領域に止まっていた。この実験心理学は行動主義によってその研究対象を観察可能な刺激と反応の間の関係に狭められ、その結果、ここには意識や心的表象に関する話は登場ほとんどなくなる。だが、1950 年代の Miller の思考に関する研究に端を発し、McCarthy, Minsky, Newell, そして Simon らによって人工知能の研究へと急転回を見せる。Chomsky の言語に関する非行動主義的な考えも心の研究に拍車をかけることになった。そして、機能主義 (functionalism) として哲学的にまとめられる。このような認知科学はそれを一つの科学的理論としてまとめるのにどのような前提を置いたのか。その前提は表象 (= 表示、Representation) と計算 (Computation) である。この認知科学の中心仮説は、思考は心の表象構造とそれら構造上で働く計算的な手続きによって理解される、というものである。

現在、心の哲学の研究の多くは自然主義的である。自然主義的な研究では認知科学と心の哲学に基本的な考えの違いはない。認識に関する基礎付け主義 (foundationalism) は否定され、心の本性はアприオリな考察からではなく、認知科学の経験的な研究から導き出されると考えられている。人間の表象と計算は認知科学の大半の研究ではコンピュータのデータ構造と計算アルゴリズムに類似したものと想定されている。一方、コネクショニズムの考えではニューロンとその結合がデータ構造に、ニューロンの発火と活性化の広がりアルゴリズムにたとえられて、新しい表象と計算の考えが提案される。ここには心、脳、コンピュータという 3 組の間で互いに相手を研究し合うという認知科学の特徴が現れている。そのような複数の心のモデルの中でもっとも適切な心のモデルは何か哲学にとっては重要な問題となってくる。心のモデルの候補は実に多い。重要なのはどれがもっとも適切に心の振舞いと本性を実現してくれるかである。

このような認知科学の基本枠組みはすべての人に受け入れられているわけではない。当然、批判者はいる。私たちの心が表象と計算によって働くという前提は経験的な仮定である。確かにこの仮定は問題解決、学習、言語使用といった分野で成功を収めているが、Dreyfus や Searle はこの仮定が基本的に誤っていると批判する。それら批判の概略は以下のようなものである。

- 1 認知科学は感情や意識を十分に扱うことができない。
- 2 認知科学は外部環境のもつ重要性を無視している。
- 3 認知科学は社会的な人間思考の特徴を扱えない。
- 4 心はダイナミックなシステムであって、計算システムではない。

ここではこれら批判を直接に扱うことはせず、批判の背後にあって重要な役割を演じている常識心理学について以下で考えてみよう。感情、意識、環境としての自然や社会、心の本性は私たちが日常生活で常識的に慣れ親しんでいるものであり、常識心理学が重要な役割を演じているからである。

3.3 常識心理学の捉え方

近年常識心理学 (commonsense psychology or folk psychology) は哲学者の間だけでなく、(社会) 心理学者や動物行動学者の間でも論争の主題になってきている。「常識心理学」は二つの異なる意味をもっている。(1) 信念、欲求、意図、希望、恐れ等によって人間の行動を説明する、常識 (commonsense) 的な心理

学、(2)常識的理論の一部としてそのような説明に解釈を与えること、の二つの異なる意味がある。常識心理学は theory-theory と呼ばれることになるが、これは(1)の常識心理学の哲学的説明である。(1)の常識心理学は人間の行動を説明する概念的な枠組を定め、行動の説明として信念や欲求を使うものである。この説明をどのように解釈するかが(2)の常識心理学の問題である。行動の説明として使われた信念や欲求が仮説となつて説明理論が構成されることが問題にされる。次の例が両者の違いを明らかにしてくれる。

「花子は赤ん坊を寝かしたい」

もし(1)の意味での常識心理学が正しいなら、「花子は赤ん坊を寝かしたい」ので電気を消した」というのは「電気を消した」の適切な説明になっている。(2)の意味での常識心理学はこの説明がどのように解釈できるか問題にする。(2)の常識心理学が正しいなら、「花子は赤ん坊を寝かしたい」は、花子は赤ん坊を寝かしたいという内的な状態にあることとその状態が花子が電気を消すことの原因になったという仮説を私たちがもっていることになる。

常識心理学には二つの異なる論争がある。

(1) Theory or Simulation?: 信念や欲求を使って行動を説明する際何をしているのであろうか。ある哲学者たちは常識心理学はシミュレーションであり、それはある技能を実行することだと考える。ある信念を帰属させることは自らをその信念をもつ者の状況に置いて考えることである。これに対して支配的な考えは、信念や欲求をもつ者は(2)の意味での常識心理学を使っているというものである。信念の帰属はその人の内的な状態についてある仮説を想定することであると考える。

(#) 常識的な立場からみれば、一元論も二元論も現実から遊離した机上の空論という感じが否めない。私たちは異なる状況で異なる立場、視点をとることに躊躇しない。朝は一元論的に考え、夜に二元論的に考えることに何の不都合も感じていない。私たちは心に関して、一貫性のないモザイク的な考えを使って生活している。では、哲学の授業で一元論や二元論が議論され、いずれが正しいのかといったことが話されるのは単なる頭の体操に過ぎないのか。過去の哲学者たちの心に関する諸説は心を理屈の通った一貫した考えで、統一的に理解しようとする際にどのような立場が一番筋の通った立場かを問題にしている。常識心理学はこのような論理的整合性や一貫性を求めている。与えられた状況に最善の考え方で対処するというのが常識心理学の立場である。常識心理学は理論であっても、継ぎ接ぎだらけの一貫性のない、時には矛盾した理論である。とはいえ、全くのデタラメでは決してない。私たちは一貫して家族や友人が自分と同じように問題に直面し、解決できたり、できなかったりしていると考えている。一つの主題について議論している場合には互いにそこの約束を了解して合理的に議論している。このことは常識心理学が理論かどうかを考える際に気をつけなければならない点である。

(2) True or False?: どの程度常識的な信念 - 欲求の枠組が正しいのであろうか。それは科学的に明らかにできるのであろうか。科学的な証明は常識心理学を(2)の意味で考えたときに可能になる問題である。常識心理学の枠組みが科学によって証明されるとするのが Fodor のような内在的実在論者である。一方、常識心理学は誤りで科学的な心理学によって置換えられるべきであると考えるのが Churchlands に代表される消去的唯物論者である。しかし、科学的証明の問題だけで常識心理学の身分が判明するわけではない。次の論証を考えてみればよい。

- a. 常識心理学は物理主義的な科学的心理学によっては証明されない。 b. 常識心理学が正しいのはそれが物理主義的な科学的心理学によって証明される場合であり、その場合に限る。
- c. したがって、常識心理学は正しくない。

ここで重要なのは前提 b である。もし前提 b が(2)の意味での常識心理学であるなら、この前提は妥当に見

える。しかし、結論は(2)の意味での常識心理学にしか適用できない。一方、b を(1)の意味での常識心理学と考えるなら、b は正しくなさそうである。(1)の意味での常識心理学は科学理論ではない。そうなら、科学理論がそれが正しいかどうか判定すると考える理由はなくなる。したがって、消去的な唯物論の主張は前提 b が(2)の意味での常識心理学の場合にしか適応できない。常識心理学は理論ではなく、シミュレーションであるという主張は消去的な唯物論の対象ではなく、それとは別の形での、(2)の意味での常識心理学の一層根本的な否定である。

常識心理学の身分に関してより穏健な立場は、それが科学的心理学によって部分的に証明され、部分的には証明されないというものである。常識心理学はその物理的基盤を科学的心理学に頼らなくとも見出すことができる。常識心理学は因果的であるから、行動の説明は厳密な法則によって裏付けることができる。とにかく、常識心理学は日常生活において有用である。

ところで(1)の意味での常識心理学は実際にどのようなものなのか。その正体をはっきりしないところに理論かシミュレーションかという問題が出てくる。それを理論と考える理由を述べるには十分な余裕がないが、次のような比較からその一端を述べておこう。文法は Chomsky に依れば暗黙的な理論として身につけていると考えられている。無論、学習によってそれは身につけられるもので、最初から内蔵されているものではない。私たちは文法を自覚的に知っているわけではないが、正しい文を作り、使うことができる。文法についての理論は言語学者がそのような言語実践から抽出するもので、これは自覚的な、暗黙的ではない理論である。文法はしたがって、人が暗黙的にもっているものと自覚的に取り出された文法理論の二つがあることになり、それは二つの意味での常識心理学と同じである。私たちは誤った文を多く作るが、その誤りはたいていは文法的な誤りではなく、事実に関する誤りである。人が文法を誤って使う場合、大抵はうっかりミスである。それらミスは単純なミスとして文法習得上の誤りとして処理される。十分に文法を習得すれば簡単に克服できるものと考えられている。したがって、私たちが問題にするのは文法上のミスではなく、文が事実を正しく述べているかどうかである。

このような文法と常識心理学を比較してみると、常識心理学が正しいかどうかは常識心理学の規則ではなく、その規則を使って表現される内容ということになる。しかし、消去的唯物論が問題にするのは規則それ自体が正しいかどうかである。そして、それは常識心理学(2)が誤りということになる。さらに問題は、常識心理学が誤りということから、その常識心理学で使われている概念が存在しないということが直接には出てこない場合があることである。確かに、魔女や熱素は存在しなかったが、星や太陽は誤った天文学理論のもとで使われても、存在しなかったわけではない。それらは無事に現在の天文学理論でも通用している。また、魔女も星も明らかに規則ではない。規則が正しいかどうかはそれを使っている人には無意味である。規則には真偽ではなく、採用不採用があるだけである。この採用不採用は歴史的にしかわからない。

3.4 内在的な理論

思考の言語仮説(The Language of Thought Hypothesis)は、信念、欲求、恐れといった認知状態は表象システム上の計算である、と主張する。それはまたより一般的な二つの仮説、つまり、志向の実在論(intentional realism)と心の表象主義(representationalism)の結びついたものである。志向の実在論は常識心理学によって態度を帰属させられた状態は行動を決定する実在的な状態であるという見解であり、心の表象主義はそれら実在的な状態は自然言語の文に類似した性質をもっている、つまり、それら状態は内容をもっており、その内容は構成的に決定される、と主張する。この二つの仮説は認知科学の仮説そのものであり、認知科学を心の理解に転用するということの表明でもある。Fodor による思考の言語仮説は次の二つの構成要素からなっている。

- (a) 信念、欲求、その他の志向的状态は脳内に存在する物理的、心理的に実在的な表象であって、行動の原因となるものである。
- (b) これら志向的状态はそれに結びついた志向的な対象の構造に類似した構成的構造をもっている。

「私が左手をあげたい」と思っているとしよう。(a) は私が左手をあげるというという意味論的な内容の心的表象をつくり、その表象は私が左手をあげるということを真にすることを私に引き起こすであろうという事実によって志向的である、ということ述べている。(b) は表象が構成要素からなり、その各々はそれ自体で表象である。ある部分は私の左手を、他のものは手を挙げるという行為を指示している。これら部分は他の表象においても同じように起こり得る。

思考の言語仮説は心的な過程をモデル化するのにどのようなタイプの計算的構築が最も適切かに関する認知科学者の違いを反映してくれる。コネクショニストは心的表象がこの仮説が主張するような意味での構成要素ではないと考える。彼らは表象がその因果的な役割によってのみ個別化されると主張する。だから、彼らは認識の神経ネットワークモデルで満足する。これに対して思考の言語仮説を採用する者は、心的状態は transportable な部分をもつ構成要素であると考え。だから、Turing/Von Neumann 型の計算的構築を好むのである。

私たちが認知のシステムを内蔵化しているのは事実であるが、それがまったく自給自足的に内部にその動きの原理をもっているわけではない。どれだけシステムの内部情報をもとにシステムを説明できるかは疑問である。むしろ、そのようなことが可能であるに越したことはない。それが不可能であることが明らか以上、どのようにすれば合理的なシステムを維持することができるのであろうか。この対処の方法は以下に述べていくが、そこには認知科学に限らず、どこにでも見られる一般的な特徴が圧縮されて現れている。その先例は生物学である。細胞学と進化論はいずれも近代的な生物学の基盤となった考えである。細胞学は現代の分子生物学にいたるまでの物理化学的な研究の出発点であった。一方、進化論は生物の目的論的、あるいは生気論的な理解を因果的な理解に変え、集団としての生物を大域的に研究する集団生物学の支柱となっている。

個体(内在)主義：合理的システムによる生物の因果的理解は近い原因に基づく

集団(外在)主義：歴史による生物の因果的理解は遠い原因に基づく

このような哲学的な表現は生物学では個体発生と系統発生という用語によって象徴的に表現されている。それはまたニュートン的な世界理解とダーウィンの世界理解の違いでもある。それらが整合的であるかどうかは論争が絶えないが、少なくともダーウィンは矛盾しないと考えていた。思考の言語は力学的でない規則から成立している。しかし、その実行は力学的になさなければならないというところに問題が出てくる。外在的な理解は人間の認知を含む自然が歴史的であるという前提の下で一層自然になる。この前提は実在論的な理解を促進する。あるいは、意味の世界を前面に出して認知を理解しようとすることでもある。二つの研究方法は「機能」の理解に端的に現れる。機能を力学的に理解する場合と歴史的に理解する場合の違いとなって現れる。

3.5 外在的な理論

外在的に心の振舞いを理解する方法や結果は哲学よりは認知に関する動物行動学や進化的行動学といった分野に求めたほうがよいのかもしれない。しかし、その基本は人間中心な理解であり、それが可能かどうか哲学では問題にされてきた。そのような例として、次の試みを考えてみよう。

Millikan にとって重要な概念は機能、表象、真理である。そのようにデザインされたものの機能とはそれ

が進化してきた環境で正しく機能するならば、実行されるような性能である。彼女はそれを proper function と呼んでいる。表象はそれが世界のある状態に対応するときのみその機能を発揮することができるものである。表象を理解するには表象の作成者、表象の消費者、表象そのものを考えなければならない。表象内容は有機体の表象消費部分が表象しようとするものに依存する。表象内容を決定するためには私たちは消費者がそれをうまく使う条件、つまり、normal conditions に注目しなければならない。表象それ自体についての彼女の考えは Wittgenstein の像理論に近い。適切なレベルで表象はそれが指し示す像を描かなければならない。表象は表象されるものの変異に応じて変換される。それは適切に文節化されている。それらは構成的であり、したがって、世界の状態変化に対応している。

真理に対する彼女の立場は対応説であるが、それは彼女の実在論と表裏一体となっている。私たちの言語能力は私たちをルールに従わせるという機能をもった生物学的な機能の集合である。彼女は二つの規則に従うことを区別する。proximal rules と distal rules である。proximal rule に従うとは入ってくる感覚情報と対応する生理学に従うことである。distal rule に従うとはさらに世界の実際の状態にも従うことである。これは近い規則と遠い規則と置き換えてもよい。

Dummett や Putnam による検証主義的な真理規則 (Verificationist truth rules) は proximal rules であり、それらは感覚的な情報にだけ依存している。対応的な真理規則 (Correspondence truth rules) は世界の状態に依存している点で distal rules である。近い規則に従うことはしばしば遠い規則に違反するように見えることがある。このずれが真理の規範性を説明してくれる。この規範性とは生物がその進化の中で正しい機能を遂行するのに外部世界を参照していることから出てくる。検証主義的な規則は外部世界とのつながりがなく、真理規則としては不十分である。

行動は有機体の行為の機能的な形態であるので、環境と生物学的な機能なしに記述することはできない。これは生物学的な事実が物理的な事実が付随していないことを示している。なぜなら、行動は生物学的な機能に言及することによってのみ定義でき、それら機能は歴史的に決定されるからである。物理的に同じ対象が異なる機能をもつことは可能なのである。

彼女は意味の合理主義 (meaning rationalism) を批判する。意味が同じとか違うとかは認識的に思考する者に与えられているという個体主義の基本にある意味の合理主義は正しいだろうか。意味は認識上与えられるものではない。意味のような合理性は頭の中にはない。したがって、論理的な可能性はアプリアリに与えられるのではなく、心と環境との統合の性質である。Millikan は意味の合理性批判に基づいて志向的な心理学の再構成を考える。心理学は今以上に生態学的な科学でなければならないというのがその主張である。人間が信念や欲求を形成するのに用いるメカニズムは生物学的なメカニズムであり、したがって、それらは生物学的な機能をもっている。それら生物学的機能は進化論と適切な環境に言及しなければ決定できないので、実験室での入力、出力の観察からは理解できない。志向的な心理学は行動の法則を見つけるのではなく、認知システムが正常な環境のもとで正常に機能する場合を説明することに専念すべきである。それは生理学が心臓や肝臓の正常な機能を説明するのと同じである。

3.6 コネクショニズムとその拡張

コネクショニズムの歴史はイギリス経験論の連合主義にまで遡ることができるが、Minsky と Papert は Rosenblatt のモデルが連合を学習できないパターンが存在することを示し、その結果、コネクショニズムは認知モデルとしては不適切とみなされ、冬の時代に入ることになる。しかし、80年代には多層パーセプトロンが Rosenblatt の単純なモデルの不備を克服することがわかり、Rummelhart と McClelland の大著 *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* によってまとめられた。そこで重要な役割を果たすのが多層のパーセプトロンの学習ルール、back-propagation である。

コネクショニズムは人工的な神経回路を使って人間の知的な能力を説明しようとする。神経回路は多くの

単位からなる単純化されたモデルである。各単位は単位間の結合の強さを示す重みをもっている。哲学が関心を寄せるのは、この主義が心の古典的な考え、つまり、心は記号言語を処理するコンピュータに似たものであるという考えとは異なる新しい考えを約束するよう見えるからである。人間の認識はデジタルコンピュータの記号計算に類似したものというのが古典的な見解である。そこでは、情報は私たちが紙に文字を書くのと同じように記号列によって表示される。コネクショニズムは情報が記号ではなく神経網の単位間の結合の重さによって貯蔵されると主張する。心的な過程は力学的で、神経網の活動の時間的な発展である。

このコネクショニズムの考えを更に発展させた一例が力学系(dynamical systems)として心の振舞いを捉えようとするものである。認識はデジタル形式の計算ではなく、力学的な現象であり、従って力学的な用語で理解されるべきであると主張される。現代の力学的アプローチは90年代に開花する。力学派と計算派の違いの一つは時間の役割にある。計算派は認識は時間を超えて存在するものと考え、力学派は認識を時間の中で捉えようとする。認識は本質的に時間的な現象であるとする。認識は感覚 - 思考 - 行為という系列として捉えるべきではなく、連続的な展開として捉えるべきである。それは物理的な力学系の時間発展と同じように考えなければならない。認識の状況性も強調される。認知過程は状況の中での脳の遂行過程を記述することで明らかにされる。表象と計算によって認識を考える代わりに、認識の力学モデルは表象を力学的な存在(系の状態、軌跡等)として扱う。表象は静的な存在ではなく、変化の直中での一時的な安定性と捉えられる。

Fodor と Pylyshyn は人間の知性の特徴を systematicity に見て、コネクショニズム、したがって、その拡張である力学系もそれを説明できないと考えた。ある文をつくったり、理解したりする能力は関連する構造の別の文をつくったり、理解したりする能力に内在的に結びついているというのが言語の systematicity である。「太郎は花子が好きだ」を理解する者はだれでも「花子は次郎が好きだ」を理解できる。これは「太郎」、「花子」、「好きだ」という構成要素の意味を使って「太郎は花子が好きだ」の意味を計算しているからである。彼らはこのような計算がコネクショニズムでは説明できないと主張する。確かにそのような計算のできるようなモデルをコネクショニズムの範囲でつくることはできるが、それは偶然的にしか過ぎなく、内在的に人間の認知に備わったものであることを説明できない。この批判は初期のコネクショニズムについては有効であった。しかし、近年この批判に対する再批判が出ている。古典的な記号処理のモデルのなかにも systematicity を否定するようなモデルが存在するというのがその骨子である。

4 哲学的諸問題

これまでの話は私の心の哲学への態度と心に関する先人の試みを垣間見るだけであった。したがって、話は比較的理解しやすかった筈である。そこには哲学的な議論がほとんどなかったからである。哲学の議論や論証は人を哲学嫌いにする犯人であるが、それを避けたのでは哲学ではなくなってしまう。ここではその議論や論証を中心にして心の科学の哲学的な基礎の有力な候補である機能主義に関する問題を考えてみよう。実際、機能主義は現在の認知科学の理論的な支柱である。心の科学が認知科学であるとするれば、機能主義はまさに心の科学を支える哲学ということになる。

認知科学の哲学的な基盤は機能主義であると述べてきた。そこでこの章では機能主義を巡る問題のいくつかを考えてみよう。いわば機能主義の哲学的な分析である。機能主義は行動主義と同一説の欠点を克服しようとして、心的過程を神経生理的なレベルよりは抽象的なレベルにある内的な過程として理解しようとする。この抽象的なレベルの特徴づけが認知科学の哲学的な基礎となる。心的状態はその具体的な実現が果たす形式的な役割と同じものとみなすのが機能主義であった。ある人が痛い状態にあるのは皮膚組織の損傷によって引き起こされ、その人に叫び声をあげさせ、逃げるといった一連の行動を引き起こすような

内的な状態がその人に存在することである。より複雑な信念や欲求もその機能的な役割と同じであるとみなされる。つまり、あるタイプの心的状態は内的状態があるパターンの行動を引き起こすのに果たす機能的な役割をもつものであり、その内的状態そのものではない。この見解の重要な帰結は二つのシステムが物理的に全く異なっているにもかかわらず、同一の心的状態をもつことができるということである。これが既に複数実現可能性と呼んだものである。火星人と私たちはたとえ異なる生理学的な構造をもっているにもかかわらず、同じ痛みをもつことができることになる。このような機能主義の考えで果たして人間の心を十分に理解することができるのだろうか。

4.1 Ned Block の機能主義の問題

最初の考察は複数実現可能性についてである。既に述べたように機能主義の特徴の一つは複数実現可能性によって心の性質を捉えたところにあった。しかし、この概念は以下のような議論で逆に機能主義の誤り、不十分さを指摘するのに使うことができる。機能主義はコンピュータの仕組みや働きを使って説明される場合が多い。これを Ned Block は巧みに次のように説明する。その内部状態が F_i で、入力 I_i ならば、出力が O_i であることを表す Turing machine T の完全な記述を想像しよう。

(#) Turing Machine

A Turing machine is an *abstract representation* of a computing device. It consists of a read/write head that scans a (possibly infinite) two-dimensional tape divided into squares, each of which is inscribed with a 0 or 1. Computation begins with the machine, in a given "state", scanning a square. It erases what it finds there, prints a 0 or 1, moves to an adjacent square, and goes into a new state. This behavior is completely determined by three parameters: (1) the state the machine is in, (2) the number on the square it is scanning, and (3) a table of instructions. The table of instructions specifies, for each state and binary input, what the machine should write, which direction it should move in, and which state it should go into. (E.g., "If in State 1 scanning a 0: print 1, move left, and go into State 3".) The table can list only finitely many states, each of which becomes implicitly defined by the role it plays in the table of instructions. These states are often referred to as the "functional states" of the machine.

A Turing machine, therefore, is more like a computer program (software) than a computer (hardware). Any given Turing machine can be realized or implemented on an infinite number of different physical computing devices. Computer scientists and logicians have shown that Turing machines can compute any function that any conventional digital computers can compute. Also, a 'probabilistic automaton' can be defined as a Turing machine in which the transition from input and state to output and state takes place with a certain probability (E.g. "If in State 1 scanning a 0: (a) there is a 60% probability that the machine will print 1, move left, and go into State 3, and (b) there is a 40% probability that the machine will print 0, move left, and go into State 2".)

(#) Turing machine から機能主義へ

Turing machine とコンピュータ科学、論理学との関係は特に密接であるが、機能主義の心の概念にも重要な役割を果たしている。Turing Machine の状態は入力、出力、他の状態によって完全に定められる。つまり、ある状態である入力を受け取ると何を出力し、どのような状態に移行するかだけで Turing machine の一つの状態が定まる。それが何からできているか、何を意味しているかといったことは不要である。機能主義は人間の心についてこの定め方を適用する。心的状態は基本的に Turing Machine の状態と同じであるとみなす。神経生理学的な脳の細部は不要である。こうして、心的状態は知覚（入力）、行動（出力）、他の機能的状態によって定義できる。行動主義と機能主義の間には共通点が多くある。違いは心的状態にある。Turing machine の状態は入力と出力だけでは定まらなかった。同じように、人間の心についてその定義には入力と出力（＝刺激と反応）だけでなく、心的状態が不可欠であるというのが機能主義の立場である。この点で行動主義と異なっている。

すると、このシステム T は $M = T(F_1, \dots; I_1, \dots; O_1, \dots)$ の形の文で記述できる。この文は各内部状態 F_i 、入力 I_i 、それに対応する出力 O_i の指令の集合を表現している。 M を機械記述と呼んでおこう。すると、次のように機械記述を一般化することによって、可能な入力、出力に関するシステム内部の機能的体制の抽象的な記述が得られる。

$T(F_1, \dots; I_1, \dots; O_1, \dots)$ であるような内部状態 F_1, \dots の集合が存在する。

これは元の機械記述 M の Ramsey sentence と呼ばれるものである。それは抽象的な意味で T の状態の内部の機能的体制が入力と出力の間の中にあることを示している。これを使うと、機能主義の主張は、私たちの心の仕組みに関して正しい機能的な記述 M が存在し、その記述 M に対して M の Ramsey sentence が内部状態 F_1, \dots のある系列に対して真であるどんなシステムも私たちと同じ心をもつということになる。さらに、 F_{17} が私たちの痛みの機能的役割を演じる状態であるとする、「痛み」は私たちの機能的な体制から次のように定義されることになる。

システム S は痛い状態にある iff $T(F_1, \dots; I_1, \dots; O_1, \dots)$ が S の機械記述で、 S は状態 F_{17} にあるような内部状態 F_1, \dots の集合が存在する。

Block によれば、上のように特徴づけられた機能主義は余りに広すぎるか、あるいは余りに狭すぎるかのいずれかになってしまう。彼の議論は上の機械記述の入力と出力の曖昧な定め方に基づいている。もし入力と出力が観察可能な周囲の環境と行動としてできるだけ詳しく記述されるなら、機能主義は余りに広すぎる主張になる。例えば、あるロボットがあなたの心的生活の機能的体制をすべて模倣するようデザインされるとする。ロボットはあなたと同じように見るし、行動もする。しかし、あなたの脳が内部の神経細胞によって作られる機能的な体制であるのとは違って、仮にロボットの機能的体制が中国人から作られ、各中国人があなたの内部状態 F_i の一つの役割を演じるようにすることができるとしてみよう。そして、人々は機械記述 M で定義された内部構造を模倣するように相互に行動し、中国人の状態に機能的に同型な状態にあなたがある場合に生み出されるのと同じ出力 O を生み出すともしてみよう。機能主義によれば、中国人からなるシステムとロボットはあなたと同じ心的生活(mental life)をもつことになるだろう。これはどう考えても奇妙である。機能主義はこの奇妙な結論をうまく説明できない。

Block によれば、機能主義者が上の結論から逃れる一つの方法は私たちの実際の神経的な入力と出力によって入力と出力を定めることである。しかし、このように定めると今度は機能主義は狭くなり過ぎてしまう。他の動物やロボットは私たちと異なるつくられ方をしているので、入力や出力の物質的な記述は当然異なってくる。しかし、私たちの心的概念は人間だけがもつ特別な生理学的特徴によってだけ決まっているのではない。こうして機能主義者はジレンマに直面することになる。私たちはどの程度の詳しさで入力と出力について記述すればよいかの基準をもっていないのである。

Block はまた感覚質(qualia)の問題も論じている。感覚質とは私たちの感覚・知覚の質的な内容、つまり、痛み、色、触感といったものである。中国人からなるようなロボットは私たちと同じような経験 (= 心的生活) をもつことができない。赤を見る、痛みを感じる、信念を自覚するといった経験をそのようなシステムはもつことができない。これが示しているのは機能主義には感覚質が不在であることである。さらに、機能主義では感覚質の逆転も可能である。あなたと同型の存在があなたが緑色を見たときにする経験と同じ経験を赤を見たときにするといったことが可能になる。もしこのようなことが可能であれば、機能主義が正しいということは考え難いことになる。

既に繰り返し述べてきたように、機能主義は認知科学が抽象的な計算過程として思考の本性を解明するこ

とを支える哲学的支柱であった。したがって、機能主義への攻撃は心の科学としての認知科学への攻撃である。私たちの思考には抽象的な計算的側面がある。だが、それが私たちの心の働きのすべてではない。ひょっとすると、科学的な探求や説明の範囲を超えるものが心にはあるかもしれない。

4.2 サールの「Chinese-room argument」

Chinese-room argument の議論で前提されているのが Turing Test である。このテストに合格するなら、合格したコンピュータはテストされた事柄に関して人間と区別がつかないというのがこのテストの趣旨である。

(#) The TURING TEST for Machine Intelligence

1950年に Turing は人工的な知能に関する Turing Test を提案した。このテストは二人の人間とテストされるコンピュータからなっている。一人がコンピュータの端末のある部屋にいる。この端末を使って部屋にいる人は二つの会話を行う。一つはコンピュータと、他はもう一人の人の会話である。但し、その人は会話の相手がコンピュータか人間のいずれであるか教えられていないとしよう。まず二つの会話がなされ、誰と話したのか尋ねられる。話した相手がコンピュータなのか、それとも人間なのか答えなくてはならない。この質問は何度も繰り返される。繰り返しの後には何回の中したかの結果が出る。50%的中率なら単なる当てずっぽうということになるが、適切な会話時間の長さのもとで、コンピュータが人間と間違われることが十分な頻度で起こったならば、そのコンピュータには知能があると言ってもよいだろう。

Turing Test が知能のテストとして十分である理由：テストにパスするには世界や人間に関する広範囲の知識をもつことが必要であるばかりではなく、新しい環境に的確に反応できる相当な柔軟性が求められる。この二つの条件を満たすには知能無くしては不可能である。これに対して異議を唱えたのが John Searle だった。

(#) Searle's article "Is the brain's mind a computer program?" was published in *Scientific American* in January 1990, along with a response by Paul and Patricia Churchland, entitled "Could a machine think?" The main argument in Searle's article, now known generally as the "**Chinese-room argument**," was first published in *Behavioral and Brain Sciences* in 1980 in an article entitled "Minds, Brains, and Programs."

さて、サールの論証の意図は人間の思考過程に形式的に同型である計算アルゴリズムを実行しても思考を生み出すには十分では有り得ないことを示すことにあった。もしこの論証が正しければ、これは様々な機能主義に対して同じように言えるので、機能主義は人間の心の解明の原理としては誤っていることを意味している。さらに、サールが強いAI(strong AI)と呼ぶ考えは根本的に誤っていることになる。というのも、強いAIは「思考 = 形式的規則に従う記号操作」を仮定する機能主義であり、思考するコンピュータをこの仮定に基づいて作りだすことができ、かつ人間の認知過程をそのコンピュータを使ったシミュレーションによって研究できると主張するからである。

サールは次のような思考実験を考える。漢字を一切知らない人に漢字を操作する全く形式的な規則の組を与えたとしてみよう。だから、その人は記号を識別できても、それが何を意味しているのかわからない。規則は漢字が何を意味しているかは述べていない。規則は単にある形の記号が部屋に入ってきたら、紙に別の形の記号を書かなければならない、また、あるグループの記号は別のグループの記号と一緒に、そしてその順序はしかじかである、といったことを述べているだけである。その人は部屋に座り、規則を次々に適用して、紙の上に規則によって得られた漢字の別の集合を書き下す。そして、その紙を部屋の外で待っている別の人に渡す。その人には何のことが一切わからないが、彼が適用した規則は文法的に正しい中国語の文を作り出す。しかし、その部屋の人には自分が何をしたのかさえわかっていない。Searleによれば、その人はコンピュータが行っているのと正確に同じことを行っている。したがって、形式的な規則を使っての漢字の操作は、その人が中国語を理解するには十分でないならば、コンピュータにとっても

中国語を理解するには十分ではない。そして、実際その人には中国語が理解できない。この議論をまとめれば次のようになる。

公理1：コンピュータのプログラムは形式的（統語論的）である。

公理2：人間の心は心的内容（意味論）をもっている。

公理3：統語論は意味論を構成しないし、またそれに対して十分でもない。

結論1：プログラムは心を構成しないし、またそれに対して十分ではない。

言語を定めるには少なくともその言語の統語論(syntax)と意味論(semantic)を定めなければならない。言語の統語論は語彙とその他の補助的な記号を挙げ、カテゴリーに分類し、さらには文法を定めなければならない。ここには意味は一切入っていない。これに対して、意味論は記号の解釈である。ある記号列が何を指示し、何を意味しているか定めることが意味論の役割である。真理、指示、意味はいずれも意味論的な概念である。言語を知るとは、したがって、少なくともその言語の統語論と意味論の両方を知る必要がある。

サールが論じていないのは何だろうか。まず、彼はコンピュータが思考できないとはしていない。彼は二元論者ではない。人間は単に考える（生物的な）機械であると言い切っている。第二に、彼は考える有機体は生物的な材料からできていなければならないとも言っていない。生物でない材料から考えるコンピュータをつくることは可能である。最後に、彼は思考は記号操作を含まないとも言っていない。では、サールの積極的な主張は何か。人間の脳で実行される思考に形式的に同型の計算を備える他に「考える」ものをつくりだすには何が要求されているのか。サールによれば、何かが考えているかどうかはその計算アルゴリズム（ソフトウェア）だけではなく、そのアルゴリズムを実行するものの本性（ハードウェア）にも依存している。機能主義が問題なのは、彼によれば、思考に含まれる計算過程の物理的な実現から余りに抽象的なレベルにあることである。彼はどのようなものが思考を生み出す因果的な能力をもっているかの基準は提供していない。それは将来の研究対象である。現在の私たちが置かれているのは、私たちは思考の幾つかの具体例と思考しないものの幾つかの具体例を知っているという状況である。

（サールの論証への反応）

サールの論証への反論や批判は実に多い。彼の論証は認知科学の哲学的な支柱である機能主義が十分に私たちの心を捉えることができないという主張であるから、機能主義的に心を捉えることができると考える哲学者、認知科学者の関心を特に強くかった。したがって、彼の論証への反応は実に多い。次の例は認知科学的な反論の代表である。

（#）Robot Reply

Searle is correct to say that a computer sitting on a desk could not be said to understand English even if it could pass the Turing Test. This is because a stationary computer without sensory apparatus would lack a necessary condition of understanding a language: It would lack the necessary causal connections between the WORDS of one's language and the OBJECTS IN THE WORLD to which those words refer. However, the robot reply continues, if you were to put the computer inside a ROBOT which had sensory apparatus, that would fundamentally change the situation. Unlike a simple computer that could only manipulate the word "pig", give definitions, etc. the right kind of robot could go to a farm, identify the pigs, separate the pigs from the goats and even slop the pigs. In that case you would have a computer that does understand a language.

According to this theory of meaning and understanding, the robot's sensory apparatus would establish a causal connection between the robot's words and the world which would be sufficient to give the English words like, "pig" (as used by the robot), genuine meaning. (Using Searle's categories, the robot reply claims that the robot's words

would have genuine semantic properties and not merely syntactic ones.) Which would be to say (glibly) that if a robot can recognize a duck and can say that it's a duck then it knows what the word "duck" means.

William Robinson in his helpful book, *Computers, Minds & Robots*, p.37 makes the point this way: Searle has not shown that a robot could not have understanding. Searle has not shown--indeed, he does not try to show--that we could not add sensing devices and motor devices to the input channels and output channels of a computer in such a way as to make a robot that makes word-world connections. If we imagine such devices added in appropriate ways, we can imagine a robot that both makes the right word-word connections and the right word-world connections. Such a robot would possess exactly the feature whose absence from a mere computer is what enables Searle to show that the computer lacks understanding. Thus, it cannot follow from Searle's argument that such a robot would not understand the symbols to which it responds appropriately in both verbal and nonverbal ways.

このような反論と並んでより哲学的な反応の代表は Churchlands であろう。彼らは消去的な唯物論という立場とそれを実証するコネクショニズムに基づいてサルへの反論を試みる。Paul and Patricia Churchland はサールの論証に満足していない。サールの公理 3 を信じる根拠がないというのが彼らの考えである。

公理 3：統語論は意味論を構成しないし、またそれに対して十分でもない。

Churchlands は部屋の中の人には中国語を理解していないという点ではサルに賛成する。その人、部屋、規則集、何枚かの紙からなるシステムが中国語を理解していると言うのもばかげている。しかし、このことから記号を操作・処理するシステムが意味論を一切もち得ないと考える根拠はないと彼らは主張する。私たちは人間の認知の働き、そして記号操作システム一般について十分知っていない。したがって、ただ一つの思考実験から一般的な結論を得ることはできない。

サールの思考実験では公理 3 を証明するのではなく、それを仮定しているだけである。公理 3 は強い AI の基本的な仮定の否定に過ぎない。それゆえ、彼の論証は論点先取である。この点を明らかにするために Churchlands は次のような類比的な論証を考案し、電磁力が光を構成できないことを示す。

公理 1*：電気と磁気は力である。

公理 2*：光の本質的な性質は輝きである。

公理 3*：力それ自体は輝きを構成しないし、またそれに対して十分でもない。

結論 1*：電気と磁気は光にとって必要でも十分でもない。

誰かが手で磁石を振動させても光はつくりだすことはできない。私たちに見える光を生み出すには振動が余りにゆっくり過ぎるからである。いずれにしろ、これは電磁現象のただ一つの例でしかなく、そこから一般化することはできない。

サールの論証が論点先取であると言う一方で、彼らは伝統的な AI 研究が問題を孕んでいることを認める。従来の系列的なデジタルコンピュータを使って知覚やパターン認識のような複雑な心的過程のモデルをつくるのは誤っている。思考をつくりだすには脳のような並列的な機械(connectionist machines)が必要である。それらコンピュータは記号を操作・処理することによって作動していない。

サルはこれに対して「中国人のジム」の話で対抗する。一人ではなく、多くの人がいる部屋で計算の仕事は分担して並列的になされる。それでも中国語の理解はできない。さらに、並行的な機械で計算できるものはまた系列的な機械でもできる。それでも Churchlands は納得できない。彼らにとっては理解できる、できないはあくまで経験的なことである。

4.3 心的内容に関する意味論

今まで私たちが考えてきたのは機能主義を巡る問題であった。今度は機能主義的に捉えた心の仕組みについて考えてみよう。機能主義は私たちの心を情報処理のシステムとみなす。では、どのように心において情報が処理されていると考えるのか。入力や出力はどのようなものとして扱われているのか。そこで心的表象、志向性、心的内容といった概念を巡って、私たちの心の中身をどのように捉えるかが議論されることになる。この議論は実は今世紀の論理学、言語の哲学を背景にもっている。

(意味の素朴な指示理論)

言語の統語論と意味論は独立に特徴づけられる。文法を理解するのに意味は無関係であると考えられている。実際、意味は解釈(interpretation)を通じて与えられる。単語の意味はその単語の指示対象であるというのが意味の素朴指示理論である。「Clinton」という語の意味は多くの女性と関係したアメリカの大統領である。また、「鳥」は特定の対象ではなくある集合を指示している。さらに、動詞「飛ぶ」は飛ぶものすべての集合を指している。このようにそれぞれの品詞、そして文に関して構成的にその解釈を与えることによって、最終的に文すべての真偽を構成的に与えることができる。論理学を知っている者は「充足(satisfaction)」概念を定義する際の解釈を考えればよいだろう。無論、そこでの言語は通常第一階の形式言語であるが、このような仕方を自然言語にそのまま流用して考えたのが意味の素朴指示理論である。

しかし、この素朴な理論にはいくつかの問題がある。この世界の対象を指示しない場合にはどうすればよいのか。「現在のフランス王」、「サンタクロース」はこの世界には指示対象が存在しないが、明らかに互いに異なる意味をもっている。また、「腎臓をもつもの」、「心臓をもつもの」は意味は異なるが、この世界の同じ集合を指示する。この二つのことからだけでも素朴な指示理論は意味の理論としては不適切である。指示が同じであるからといって意味が同じであるとは言えないからである。「Samuel Clemens」と「Mark Twain」は同一人物を指す。一方に他方を代入することによって真理値が変わらない文はたくさんあるとはいえ、次のような二つの文では異なる真理値をもってしまう。

John は Samuel Clemens が Samuel Clemens であると信じている。

John は Mark Twain が Samuel Clemens であると信じている。

(上の二つの文が同じ真理値をもつ場合と異なる真理値をもつ場合を具体的に考えてみよう。)

(意味(Sense)と指示(Reference))

意味は指示と同じではないことになる、意味と指示はそれぞれどのようなものなのか。これを最初に組織的に明らかにしたのが Frege であり、彼は意味と指示を明確に区別した。Frege は心理主義的な意味の理論に反対する。心理主義的な意味の理論では、語の意味は人によって変わる。この変わり方が全く個人に依存するのであれば、「Clinton」についての正確なコミュニケーションができなくなってしまう。たとえ、人それぞれに彼についての異なる心的表象、信念をもっている、みな同一の個人について話している。

指示の性格を保存しながら意味の違いを捉える意味論の代表例として可能世界意味論(possible worlds semantics)がある。多くの可能な世界があるというのがこの意味論の基本的な考えである。Clinton はある世界では現在のアメリカ大統領であり、別の世界ではそうでない。このように考えることによって、「現在のアメリカ大統領」の外延(extension)と内包(intension)の区別ができるようになる。例えば、「現在のアメリカ大統領」はこの現実の世界で Clinton を指示するが、他の可能な世界では Bob Dole を指示する。しかし、「Clinton」はすべての可能な世界で同じ人物を指示する。すると、「Clinton」と「現在のアメリカ大統領」は異なる指示対象をもつことができる。それらは指示の仕方が異なっている。したがって、意味を内包と考えるなら、「現在のアメリカ大統領」と「Clinton」は異なる意味をもつことになる。これは素朴な指示理論より入念につくられており、心理主義的な要素も少ない。

可能世界意味論では次のような関係が成立している。固有名詞の意味はその語が指示する個体である。つまり、あらゆる可能世界に対して同一の個体を指定する定数関数である。述語の意味は可能世界からその世界で対応する性質をもつものの集合への関数である。

可能世界意味論とその変形は現在も活発に研究されている。そしてこれ以外の意味論も多くの候補が出され、心的内容の形式的な表現として実際に使われ出している。無論、そこには未解決な問題が数多く存在している。このような意味論は本来心的内容そのものの研究というより、既述のように知識の形式的な表現装置として考えられてきた。しかし、現在では明らかに認知科学への応用が意識され、それによって研究が進められている場合が多い。

5 心的内容について

心的表象はどのようなものによってなされるか。今までの諸説は心的表象についてそれぞれ次のような答えを出すだろう。結果だけ書いておくので、どうしてそうなるかしっかり確認してほしい。

- (1) デカルト的な二元論ではある種の心的実体が心的表象を担っている。
- (2) 同一説ではタイプとしての脳状態が心的表象の役割を果たす。
- (3) 機能主義ではトークンとしての脳状態、つまりは個々の状態そのものである。
- (4) 行動主義では表象の役割を演じているものはない。というのも、(論理的行動主義では) 心的表象についての話は表象的でない行動の傾向性についての話に過ぎないからである。
- (5) 消去的唯物論でも心的表象は何の役割も演じない。心的表象が存在するという常識心理学は端的に誤っているからである。

このように心的表象はそれぞれの説によってめまぐるしく異なった解釈を受けている。ここでは何が表象であるかという超哲学的問題ではなく、表象するとはどのようなことか、つまり、何が何かを表象するとはどのようなことかを考えよう。心的実体、脳状態のタイプ、個々の脳状態が表象しているというのが非消去的な説の答えであったが、では、それらがどのように、そして何によって表象することができるのか説明されなければならない。つまり、何が心的実体や脳状態にその表象内容 (= 志向的内容) を与えているのか説明されなければならない。

志向性や表象は心的なものの特徴であると多くの哲学者は考えてきた。物理世界の対象は志向性や表象をもたない。これに対して、思考、信念、欲求は何かについて(about)のものである。物理的なものは何かについての表示をもたないが、思考は何かについてのものである。だが、厳密に言えば、多くの物理的对象も何かを表象 (= 表示) する。実際、この文は物理的な記号で書かれた物理的对象だが、ある情報を表示している。絵画も同じようにある情景や画家の意図を表示している。温度計は温度を表示している。しかし、温度計は私たち人間が温度を表示するためにつくり、使うものである。温度計が最初から表示をもっているのではなく、それは温度表示のためにつくられた私たちの工夫に過ぎない。文や絵画もその作成者が表示の手段として使うものであり、本来的に表象作用をもっているわけではない。いわば、それらの表象は派生的なものである。私たちが以後考えるのは派生的ではない、本来的な表象である。そこで、自然主義的にこの本来的な表象を考えてみよう。

5.1 Intentional Stance

Daniel Dennett は心的表象についての問いより広い内容をもった問いを考える。あるシステムが一般に表象的な状態をもつとみなされるのはどのようにしてか、という問いである。その答えは、あるシステムの

行動を「合理的」と解釈しない限り、そのシステムを志向的システム(intentional system = one which has intentional states such as belief and desire)、つまり表象状態をもつシステムとして一貫して考えることはできないというものである。

(#) physical strategy(stance), design strategy(stance), intentional strategy(stance)

Dennett はシステムの運動や行動を予測するために私たちが用いる戦略や態度を上のに分類し、それぞれの長所、短所を述べた上で、心的表象を問題にする。物理的な態度やデザインを使つての見方は自然現象の説明や機械の説明を考えれば容易に想像できるであろう。志向的態度も、私たちが仲間と接する際に取り合う態度である。あなたは決して自分の恋人や友人を単なる石ころや自動車とはみなしていない。同じ人間で同じように振る舞うとみなしている。だが、しばしば人間は奇妙で非合理的な振舞いをする。そのため、行動の予測や解釈は通常極めて困難になる。

彼の考えで重要なのは志向的態度という考えではない。重要なのは、システムの行動が志向的態度を通じて予測できるなら、そのシステムを真に志向的状態をもつものとみなす点にある。この見方は分析的な哲学のなかで、Quine の根源的解釈(radical interpretation) に触発された多くの哲学者によって論じられてきた。Word and Object (1960)において、Quine は未知の言語を母国語に翻訳するには寛容の原理(principle of charity = a dictum that we translate the utterances of a person speaking that language so that they are generally reasonable given that person's environment.)が必要であると論じた。Donald Davidson はさらにこの原理を単なる方法的な格言から思考と信念の存在のための構成的で規範的な原理(constitutive normative principle)へと格上げした。この傾向は Dennett にも見られる。彼は、有機体を信念や欲求のような命題的態度をもつものとして解釈するならば、その有機体を合理的なものと考えなければならないと論じる。したがって、命題的態度を実際に帰属させることは不可避免的に規範的なものをそこに含めることになる。ある信念や欲求を有機体に帰属させる言明はそのシステムの行動の解釈全体がその有機体は信念や欲求をもつことであるときに正確に正しいというのがこの見方の基本的な考えである。この種の解釈は有機体の行動の合理化を引き起こすような解釈である。つまり、この種の解釈によって、新しい状況で有機体がどのように行動するかを予測し、そのように行動するのがその有機体の観点から意味があるのかどうかを説明することができるようになる。こうして、私たちは信念や欲求をもつものは一般に論理的な能力をもつとみなさなければならないことになる。

「もし私が市場に行くなら、新鮮な野菜を手に入れることができるだろう。私は市場に出かける。私はそこで新鮮な野菜を手に入れることができないだろう。」

誰かが上のように言ったなら、あなたは困惑して、その解釈に困ってしまうだろう。Dennett が正しいなら、上の三つの信念をその人が持っているとは思いたくないから、単に冗談に過ぎないと考えるだろう。私たちは人間に矛盾なく志向的態度、つまりは信念と欲求を帰属させることができる限り、その人間を論理的な能力をもった思考者とみなさなければならない。これが Dennett の主張である。

では、彼の主張は説得力があるだろうか。まず、彼の主張は自然主義的ではない。それは志向的概念(信念、欲求等)を志向的でない概念を使って自然化するような試みではない。志向的システムとは、そのシステムを志向的とみなすことによってその行動が予測できるようになるシステムである。また、当然ながら(非自然的な概念である)合理性とはどのようなものかも説明されていない。

5.2 類似性(Resemblance)

私たちが実際に知覚しているのは世界そのものではなく、外部の対象が私たちの心に因果的に引き起こし

た心像あるいは観念であるという考えが今でもある。これは17世紀以来の名残である。例えば、夢を考えると、実際に見ていないにもかかわらず、夢の中で見ているかのように思われる。これは知覚だけではなく、思考のレベルにも適用される。 x を知覚する際に x のイメージをもつように、 x を考える際にも x のイメージを考えてしまう。もしこれが思考の姿であるなら、表象の問題への解答は次のようになる。心的対象を x の表象とするものは x のイメージである。 x の表象は x に類似している。

(#) 知覚の類似説と表象の類似説は同じ構造をもっている。しかし、思考レベルでの類似は知覚レベルの類似とは多くの点で類似していない。それが表象の類似説の欠点となる。次のような問題を考えてみよう。本物のネコとネコのイメージは一体何が類似しているのだろうか、また何が異なっているのだろうか。

類似しているという考えの問題は物理主義と類似説が両立しない点である。ある人の脳とその状態がもっとも似ているのは他人の脳とその状態である。もし思考が類似性を含むなら、脳は他人の脳しか考えることができなくなってしまう。しかし、私たちは脳以外のものも考えているので、類似説が誤っているか、あるいは脳に全く結びつかない心的イメージをもたなければならないことになる。今日の評価は前者を選択する。さらに、イメージと命題の内容は随分と異なる。抽象的、一般的な命題をどのようにイメージしたらよいか。無限や多次元空間のイメージをもつことはそもそもできるのだろうか。

5.3 共変性(Covariance)

「牛」という語は牛を意味している。というのも、牛の刺激のコントロールのもとに「牛」という言語的な反応がなされるからである。これは既述の Skinner が意味について行った説明である。Fodorはこの考えをもとに心的内容の共変説(covariance theory)を展開した。ある表象が牛を意味するとは、それが牛のコントロールのもとにある場合である。つまり、牛の現存と、頭の中のその表象のトークンの間に規則的な因果連関があるならば、その表象は牛を意味している。そして、牛の表象は牛とともに変化する。

この説を展開する前に二つのことが解決されなければならない。まず、「牛」があらゆる牛を表象しなければならない。さらに、因果連関の鎖の中で「牛」が牛の表面、網膜の一部あるいは神経細胞ではなく、牛そのものを意味しなければならない。つまり、あらゆる牛を含み、正確に牛を含むような因果連関を選んで、解釈しなければならない。最初の候補とすれば、

表象「牛」が牛を意味する iff すべての牛がそしてそれだけが「牛」をトークンとする、

となる。しかし、すべての牛が実際に「牛」をトークンとすることを要求はできないし、ユニコーンのような架空の対象をどうして私たちが語れるかもこれでは説明できない。存在していないユニコーンとの因果連関は不可能である。さらに、牛だけが「牛」をトークンとすると要求することもできない。というのも、私たちは誤りの可能性を確保しておきたいからである。遠くの馬が「牛」をトークンとすることを許したいのである。誤表象(misrepresentation)は私たちにつきもので、それも説明されなければならない。

最初の二つの問題はそれほど難しくない。Fodorは個々の牛と「牛」の間の因果連関ではなく、「牛」に牛を意味させる肝心の部分は二つのものを関係させる自然法則の存在であると考えた。この自然法則は「牛は「牛」の原因である」といった形式をもっている。法則の重要な性質は反事実的な文を支持する点である。(反事実的な文とは if there were a P , then it would cause a Q という形式の条件法の文である。)法則が反事実的な文を認めることはユニコーンのような対象についても法則が適用できることを意味している。(なぜだろうか?) さらに、もし牛が適切な条件のもとにあり、それは「牛」を意味するという反事実的な文が成立するなら、たとえ適切でない条件の下でほとんどの牛が「牛」を意味しなくとも構わない。つまり、すべての牛が実際に「牛」の意味でなくとも構わない。これが最初の問題への返答だが、誤表象の問題は

どうか。

しばしば牛以外のものが「牛」のトークンの原因になる。もし牛同様に馬も「牛」のトークンの原因になるなら、「牛」は牛であるという性質と共に変化するのではなく、牛あるいは馬であるという性質と共に変化することになる。もしそうなら、「牛」は牛を意味するのではなく、牛あるいは馬を意味することになる。この後者の性質は選言的性質(disjunctive property)と呼ばれることから、この問題は「選言問題」と呼ばれている。

(#) なぜ「牛」のトークンの原因が馬でもあるときに、「牛」は牛かつ馬を意味しないのだろうか。そしてこの連言的性質(conjunctive property)を「牛」と共に変化する性質としないのだろうか。

この問題への最初の解答は Dretske によって与えられた。その基本的な発想は、概念を学んだ時期と、その後単にそれを使用する時期を区別することにある。学習時には牛と「牛」の相関は厳密である。正確な相関によって学習がなされる。それが単に使用する段階になるとルーズになり、相関は弱まる。この弱い相関が誤表象を説明してくれる。確かに、この説明は語の学習についてもっともらしいが、心的表象の意味についても彼の考えを適用できるだろうか。語のないものでも表象はできる。その際の選言問題は彼の考えで解けるだろうか。(後に、Dretske は自分の発想がうまく行かないことから共変説をあきらめ、内容に関する目的論的な立場に立つことになる。)

「牛」がしばしば馬によって引き起こされても、「牛」は牛を意味することを主張するために、Fodor は牛が正しく「牛」を引き起こす状況と馬が誤って「牛」を引き起こす場合の間にある対称性を壊す必要があると考えた。「牛」が牛を意味することが既に決まっていな限り、馬を見て「牛」とは言わない。馬を牛と見誤るのは牛によって引き起こされる記号をトークンとすることである。この逆は成立しない。牛を見て「牛」と言うのは馬もまた「牛」の原因であると言う事実には依存しない。これが彼の考えである。馬が原因の「牛」は牛が原因の「牛」に非対称的に依存する(asymmetrically dependent)。共変説 + 非対称的依存性、これが彼の解答である。

(#) Fodor の説は上述のことから一応の解答を与えることができるが、別の問題に直面する。それは検証主義の問題である。と言うように、議論は現在にいたるまで延々と継続している。それを丹念に追求して、問題を解決する、すると次の問題が現れる、またその解決に努力する、という風に遂行されていくのが現場の哲学者の行っていることである。

5.4 目的論

Dretske、Papineau、Godfrey-Smith、そして中でも(既述の)Millikan がこの立場の代表である。この立場は人によって微妙に異なるので、もっとも基本的な特徴だけを述べてみよう。

意味の決定を表象を引き起こすものだけではなく、表象の因果的な役割全体(つまりは、その典型的な原因と結果)をも考えることによって行うというのが目的論の基本的な発想である。特に、表象の生物学的な機能や目的が注目される。(これが「目的論的」、つまりは、「目標」や「ゴール」という呼び方につながっている。)自然選択と適応を表象に適用すると次のようになる。

有機体Cにおいて、「トラ」のような心的記号はトラを意味する。なぜなら、その記号はトラを表象する機能を持ち、「トラ」がトラに対して適切な行動を引き起こしたCの祖先がそうでなかった祖先より進化的に有利であった限りでその機能をもっている。

このように考えると目的論的な説は誤りの問題をうまく扱うことができる。眼や心臓はそれぞれの機能を果たしているが、それがうまく働かない場合がしばしばある。病気や老衰によってうまく働かなくなった

器官や組織に私たちは日常接している。したがって、「トラ」の機能がトラを表象することであるなら、「トラ」が時にはライオンによって引き起こされても、やはりトラを意味している。心臓が止まる場合があっても、「心臓」は正常に機能している心臓を意味している。

この説にも多くの批判がある。Fodorはこの説が非決定論的であると批判するし、Cumminsは目的論的な説が歴史的な性格をもち、それが歴史的でない物理主義と衝突することを指摘している。このような批判とそれへの反論という往復作業はやはり現在進行形の形で行われている。(後述参照)

5.5 概念役割意味論(Conceptual Role Semantics)

記号の内容を考えるのに記号の原因だけでなく、システム全体における結果も見べきであるというのがこの意味論の主張である。この点では目的論に似ている。概念役割意味論は機能主義と提携している。多くの哲学者、認知科学者がさまざまに異なる意味論を展開している。ここでは Ned Block の意味論を考えてみよう。

意味は二つの因子によって基礎付けられている。一つは頭の中にある概念役割因子で、他は内的な記号とその世界の中の指示対象の間の関係に關する外的因子である。興味深いのは記号の概念的役割における内的因子で、それは推論、熟慮において記号がどのように働くか、他の記号とどのように結合・分離するか、感覚入力と行動出力をどのように媒介するかという機能をもっている。このような概念的役割は抽象的であって、記号の因果的役割とは異なっている。

Block の具体例を使って、この二因子意味論が信念を具体的に扱えるかどうかを見てみよう。西脇が次の二つの文によって表現している信念を比べてみよう。

- 1 私は破産の危機に瀕している
- 2 西脇与作は破産の危機に瀕している

西脇が西脇であることを信じない(で、安藤だと信じているかもしれないので、その)限り、彼が2を信じる場合は彼が1を信じる場合(彼が彼自身であると信じないことはできない)とは異なった破産の避け方をするかもしれない。異なる破産の避け方があるということは、1と2が同じ場合があるとはいえ、少なくともそれらが異なる信念を表している場合があることを示している。二人の異なる人が「私は喉が渴いている」と言ったとき、ある場合にはそれは同じ信念を表した文であり、別の場合には(違う人が話しているのであるから)異なる信念を表した文になっている。

この二つの違いを概念役割意味論は説明できる。1と2は概念役割が異なっているが外的関係は共有する、つまり、意味の外的因子は(「私」と「西脇与作」の)同一の指示対象をもつが、意味の内的因子は異なる。つまり、推論や行動における役割が異なる。二番目の場合、二つの(あるいは二人の人が言った)「私は喉が渴いている」という文の内的因子は同一であるが、外的因子は異なっている。つまり、その文の「私」は異なる個人を指している。これを別の言い方で表現すれば、内的(概念的役割)因子は文(あるいは信念)の狭い(narrow)意味を決定するが、外的因子はその広い(broad)意味を決定する。すると、1と2は同一の広い意味をもつが、狭い意味において異なっている。それに対して、二番目の文は二人の異なる人が発話した場合、同じ狭い意味をもつが、広い意味において異なるということになる。

この広い、狭いを使い分けると、共変説、目的論的な説は広い意味に、一方概念役割の意味論は狭い意味により焦点を当てていることになる。あるいは前の二説は x についてのあなたの概念よりは x とあなたの因果関係を重視しており、その逆が概念役割意味論だと言ってもよい。西脇と安藤が「トラ」について異なる信念をもつなら、「トラ」はこの二人の間で異なる狭い意味をもつことになる。

5.6 内在主義と外在主義

既に説明した内在主義と外在主義を上述のことと合わせて考えてみよう。まず、デカルト的な二元論を思い出してみよう。そこでは心と身体は全く異なる二つの実体で、私のことは私が直接に内観によって理解できた。この考えは未だに私たちに巣くっている。その一つが私たち自身の考えや信念から出発して世界の事柄を考えようという姿勢である。私の思考や信念の内容には私だけが特権的に近づくことができるので、この私の内的経験を基礎にして外的な世界の像をつくる。これがデカルト的な二元論の目標になる。その出発点を言い換えれば次のように表現できる。私たちの思考や経験は外界から独立している。全く心のうちにある純粋に内的な事柄である思考の表象内容が何にも増してすべての出発点である。

心と脳の二つを認めたとすると、心的状態が外界から独立している、あるいは遮断されているということは、私たちの思考がすべて「頭の中」にあって、脳の状態以外のものに依存していないということになる。心が脳にどのように依存しているかは既にいくつか述べてきた。なかでも付随性と複数実現性は広く使用されている概念であった。心的状態が脳状態に付随するとは二人の人(あるいは異なる時間の一人の人)の脳状態に違いがない限り、心的状態にも違いはないということであった。言い換えれば、二人の人が同じ脳状態にあれば、二人の心的状態も同じであるというのが心的状態が脳状態に付随するということである。したがって、「思考は全く頭の中にある」とは、ある人の思考の内容はその人の脳状態に付随するということになる。これが内在主義(internalism)と呼ばれている考えである。これに対して、思考内容は頭の外にある、あるいは外部世界に関連しているというのが外在主義(externalism)である。以下に外在主義を支持する駆動力の一つとなった Putnam の考えを紹介しよう。

多くの哲学者が外在主義に関する思考実験を試みてきた。彼は桶の中の脳(brain in a vat)という身体から分離された脳が人工的に生かされて、活動している例を考え、そのような脳の思考は外部世界のものごとについてのものではなく、脳自身のメカニズムのついてのものであることを論じている。彼の考えでは、志向性のもつ「...について」という特徴(例えば、牛についての表象や思考)は、思考や表象が外部世界の適切な対象への正しい因果関係を必要とする。彼の主張は、外的な因果関係が志向性には必要である、ということになる。桶のなかの脳は世界からは遮断されている。したがって、そのような脳にとって外部世界についての思考はないことになる。

(意味は頭の中にあるのか。「意味」の意味)

私たちは二つの異なる意味論、素朴な意味の指示理論と可能世界意味論について触れた。いずれも心理的でない意味についての理論であった。語の意味は心的表象といった心理的なものではなかった。今一度この問題を考えてみよう。「意味は頭の中にあるのか」という問いに対して Hilary Putnam はノーと答える。以下に彼の議論を考えてみよう。

(双子地球(Twin Earth))

1750年の地球とほとんどすべての点で同じ惑星があったとしてみよう。観察によっては二つの惑星は区別できない。これを「双子地球」と呼ぶことにする。ただ一点だけ地球と双子地球に違いがあった。それは地球でも双子地球でも「水」と呼ばれるものが地球上では H_2O だが、双子地球では異なる化学物質であった。それをXYZと呼んでおこう。化学的には異なるが、双子地球での「水」は地球と全く同じ役割をもち、同じように使われていた。

さて、「水」は1750年に地球と双子地球で何を意味していただろうか。その時代の誰もが知らなくとも、その時の地球での「水」は私たちが現在 H_2O として知っているものを指示していた。地球での心理学も双子地球での心理学も同じである。「水」の心的表象は、したがって、地球でも双子地球でも同じである。しかし、一方の「水」は H_2O であり、他方の「水」はXYZである。したがって、頭の中にある心的表象は「水」の意味を決定しないし、思考の内容も決定しない。それゆえ、意味は頭の中には存在しない。これ

が Putnam の結論である。Putnam はこの議論を水だけではなく、他の自然種にも適用する。

(内包の本性、形而上学的必然性と認識的必然性)

もし Putnam が正しければ、「水は H_2O である」は(話者の言語共同体に依存して)必然的に真あるいは偽を表現していることになる。あるものが H_2O でなければ、それは水ではなく、したがって、 H_2O でない水が存在するような可能世界はない。1750 年には誰も「水は H_2O である」と知らず、したがって、その文を述べていなくとも、彼らが水について語っているときには彼らは H_2O について語っているのである。「水は H_2O である」はいつでも真である「水は水である」とは違うので、これは考え難いことに見える。「水は水である」が分析的に真であるのに対し、「水は H_2O である」はその真偽の決定に科学的な知識を必要とする。何の科学的知識ももっていない人にとってさえ「水は水である」はその人に理解できる限り、真である。それゆえ、「水」と「 H_2O 」は同義語ではないように見える。というのも、もし同じ意味なら、「水」と「 H_2O 」の意味を学ぶだけで「水は H_2O である」の真偽を決定できることになってしまうからである。実際はこの文の真偽は科学的な追求によってしか決定できない。では、どのような意味で「水」と「 H_2O 」は同じ意味をもっているのか。Putnam によれば、双子地球の思考実験が示しているのは次の二つの言明が共に真では有り得ないということである。

(1) 語の意味(内包)はある心理的な実体、例えば心的表象と同一である。

(2) 内包は概念を決定する。

Putnam は(1)を否定し、(2)を採用する。しかし、語がどのように意味を獲得するかは心理学や話者の意図が役割をもっているのは明らかである。ある言語共同体の人々は「水」を彼らの環境で呑むことができ、味がなく、色もなく、においもないものを指示するのに使う。これらの性質はみなその共同体における水のステレオタイプの一部である。それらは共同体のメンバーが水に結びつけ、水とみなすのに使う性質である。Putnam によれば、語の意味とステレオタイプを同一視してはならないが、ステレオタイプは語の意味の一部である。同じように、意味と内包の概念はステレオタイプと同じではない。ステレオタイプが行うのは与えられた語の意味を与えるのではなく、その指示を固定することである。Putnam の主張をまとめると、次のようになる。「水」の意味はその外延からなる。その外延はそのものを同定するのに使うステレオタイプの性質を使って選び出されるものによって決定される。「水は H_2O である」が必然的なのは、 H_2O でないものは水では有り得ないという意味においてである。「水」は、「Clinton」や「私」と同じように、厳格作用子(rigid designator)である。それはあらゆる可能世界で同一のものを指示する。したがって、「水」のような自然種を表す語は指標詞(「私」、「ここ」、「今」のような語)に似ている。しかし、 H_2O という性質が水のステレオタイプの一部ではないため、水が H_2O かどうかは科学的な研究が必要となる。それゆえ、「水は H_2O である」は認識上は偶然的である。

(#) この章の説明の多くは Goldberg, S. and A. Pessin (1997) *Gray Matters: An Introduction to the Philosophy of Mind*, Armonk: M. E. Shape. に従っている。また、本文の内容についての文献は以下の通りである。

(参考文献)

Block, N. (1986) "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy*, Volume X, edited by P.A. French, et. al., 615-678.

Block, N. (1987) "Functional Role and Truth Conditions," *Proceedings of the Aristotelian Society* LXI, 1987, 157-181.

- Cummins, R. (1989) *Meaning and Mental Representation*. Cambridge, Massachusetts: The MIT Press.
- Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press.
- Dennett, D. (1987) *The Intentional Stance*, Cambridge, MA: The MIT Press.
- Dretske, F. (1982) *Knowledge and the Flow of Information*, Cambridge, MA: The MIT Press.
- Fodor, J. A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA: The MIT Press.
- Godfrey-Smith, P. (1992) "Indication and Adaptation," *Synthese*, 92, 283-312.
- Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: The MIT Press.
- Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Basil Blackwell.
- Putnam, H. (1975). *Mind, language, and reality*. Cambridge: Cambridge University Press.
- Putnam, H. (1981) *Reason, Truth, and History*, Cambridge: Cambridge University Press.

6 心的因果(Mental Causation)

(#) 二元論をとらない限り心的因果という主題が意味をもっていないのではないかという疑問をもつ人がいるだろう。あるいは、二元論をとっても心と身体の間相互作用を認めないのであればやはりこの主題は意味がないと言われるかもしれない。確かにこの問いはもっともらしく見える。心的なものを物理的なものに還元する試みのもとでは、還元されるものと還元するもの間に相互作用を考えることはしないのが普通である。このような見方は私たち自身が対象でない場合にはそのまま言い放っておいても大きな支障は起こらない。しかし、実際に心の振舞いを経験している私たちの場合には多に支障が起こる。心が還元されても、心の性質や現象の経験が消えてしまうわけではない。還元の過程が明確にされない限り、心の性質や現象をどのように考えたらよいのか途方に暮れるだけである。経験される心の性質や現象を説明するには心と脳の間の一元的な関係を一元論、二元論の主張とともに考えておかなければならない。以下を読めば、その必要性が実感できるであろうし、伝統的な二元論が常識に合わないこともわかるだろう。類似の問題は意識に関する消去的唯物論に対しても適用できる。「意識は消去されるのであるから、意識について議論するのは無意味である」という主張に対して消去的唯物論者はどのような態度を取るべきだろうか。納得のいく消去の仕方でない限り、意識についての問題はやはり残るだろう。

心的状態は物理的状态と因果的に相互作用する。 (例えば、痛みは身体の傷が原因で起こり、その痛みが原因となつてうまく歩けない、という自然な表現を考えればよい。) この極めて常識的な考え(これを以後 MC と呼ぶことにする。)は多くの人が人間と世界に関するもっとも基本的なものとして受け入れてきた。だが、不思議なことにどのように MC が真であるかの説明には納得できるものは一つもない。その歴史的な証拠はデカルトの心身相互作用論である。どのような相互作用が心と身体の間にあるのか説明できないことが、デカルト以後の哲学者の研究を励まし、悩ませてきた。その試みも空しく、伝統的な心の諸説はどれも大きな障害に突き当たったままである。ここでは、この相互作用の実態を考えてみよう。私たちはまず MC に係わる具体的な例から考えてみよう。

6.1 MC の妥当性

・ 心的因果の例 1 : 心から物理的なものへ : 私たちの意志的な行為がその代表例である。

常識的な心的因果の例のタイプは私たち自身や他人を行為する主体として考え、話す場合である。人が適切な心的状態(信念、欲求)をもち、それが身体的な運動や音を引き起こすという出来事がそこでの行為である。行為という出来事が心的因果の例になるのはなぜなのか。日常の言葉遣いでは、行為は人が云々の信念と欲求をもつゆえに起こる行動であると言われる。「ゆえに」という表現は信念や欲求をもつことがあの行動の原因であるということを示している。身体の運動や音には行為とみなされないものが多くあ

る。心臓の規則的な運動や咀嚼の音は行為ではない。では、行為とみなされない常識的な理由は何か。それらはその人自身が自分の意志によって始めたものではない。それらは心的原因をもっていない。心臓の動きや咀嚼の音の原因を尋ねられた場合、誰も誰かの信念や欲求をもち出さない。したがって、行為と単なる身体的な運動は心的原因をもつ行動と心的原因をもたない行動の違いとして考えることができる。

・心的因果の例2：物理的なものから心へ：感覚や知覚がその代表例。

行為が物理的な出来事を引き起こす心的出来事を含むものであるとすれば、逆に心的出来事を引き起こす物理的な出来事もある。感覚や知覚を考えてみよう。その過程は、例えば、次のように述べられる。視覚野に入っている対象から反射した光が網膜にあたり、そして最後にはその人の対象についての視覚像を生み出す。ここでは（網膜に光があたるという）物理的な出来事が（ある視覚像をもつという）心的出来事を引き起こしている。痛みの場合もよく似ている。傷のような身体の損傷という状態が原因で痛いという心的状態が起こる。痛みの現象は（傷のある）身体的状態のあるタイプによって通常引き起こされる（痛いという）心的状態のあるタイプからなっている。

・心的因果の例3：心から心へ：新しい信念がその一例。

私たちの心的状態（ある信念をもつこと）は別の心的状態（別の信念をもつこと）の獲得の原因となる。私の信念があなたの信念を変えさせ、その結果あなたの運命が一変するといったことはしばしば起こる。物理的なものから物理的なものへの因果関係があるように、心的状態が別の心的状態を引き起こすのも不思議ではない。

このような諸例はMCが常識的な考えとしては極めて確からしいことを雄弁に語っている。MCの確からしさは私たちの世界観の核心にある次のような二つの考えからもさらに強まる。

（責任）

通常、人があることに責任があり、別のことに責任がないというのは、その人（の心的状態）があることの原因になっているかどうかで判断される。責任と心的因果はいつも対になっている。人は、自らが原因である行為には責任があり、自らが原因でない行為には責任がない。つまり、行為と責任の外延は一致している。これは次のように表現してもよい。

Aは行為Bをする iff AはBに責任がある

次の例を考えてみよう。

- (1) 史門は医者を意図的にぶって、怪我を負わせた。
- (2) 史門の手が反射的に医者にあたり、その結果、医者は怪我をした。

私たちの自然な反応は、(1)の場合は史門に責任があり、(2)の場合は史門に責任がないというものだろう。しかし、MCが成立していないなら、この反応はどうして正しいといえるだろうか。MC以外の基準で私たちの責任に対する自然な反応を正当化してくれるような基準は見当たりそうにもない。そこで、結論として、標準的な責任に関する議論では心的因果が仮定されているということになる。

（#）責任はその人が原因になっていることと同じであるという考えは、その原因が自由に引き起こされるものでなければならない。その人の自由意志によってその人自らが原因でなければならない。すると、自由意志と因果的な決定論の関係はどのようになっているのかという伝統的な問題が思い起こされることになる。図式的にその内容を以下にまとめておこう。詳しくは後述の行為の因果連関を参照。

（決定論的論証）

1.すべての出来事には原因がある。 2. 人間の行為は出来事である。 3. それゆえ、人間の行為には原因がある。 4. したがって、決定論は正しい。

(決定論の証拠)

1. 科学は最終的にすべてのものの原因を見出すだろう。 2. 日常生活では、すべてが原因をもつと認めている。どんな出来事にも原因があると考えている。

(客観性)

世界は客観的なものであるという信念が正しいと思う際、そこではMCが前提になっている。世界の客観性についての信念は常識的なものであるが、それは何に基づいているのか。多分、多くの異なる人々の知覚的な判断が類似しているという事実に基づいているのであろう。人の知覚的な信念はその信念内容の世界の適切な特徴によって引き起こされる。この知覚の信念の類似性は共通原因の結果として説明される。今の場合の共通原因は人々が共有する、客観的な性質をもつ世界である。世界は客観的な性質をもち、それを原因にする知覚の結果は、同じ原因に対して同じ結果を生むことから、みな類似の結果をもつということになる。一つの、共通の世界が原因であるから、その結果である信念も類似している。人の知覚の信念は世界の適切な特徴によって引き起こされるという考えがここで使われている。これはMCの過程の一部である。つまり個人の間での知覚の信念の類似性が世界が客観的であるという説の基礎であるなら、その時にはMCが仮定されているのである。

ここまで如何にMCが常識的な世界で確からしい仮説であるか述べてきた。この僅かな例だけでもMCは正しいと結論したくなる。経験的な証拠は山ほどあるからである。たとえMCは正しいと認めたとしても、どのような意味でMCが正しいのかと聞き返されると、その返答は簡単ではない。今まで述べてきた心の理論を思い出して、どれが適切にMCを説明してくれるものがあるだろうか。以下、MCが正しいかどうかという問いをMCの説明を与えてくれる心の理論があるかどうかという形で考えてみよう。

6.2 心的因果の問題

MCは心的出来事と物理的出来事の関係性を述べていた。では、そもそも心的出来事とは何か。

(二元論と懐疑論的テーゼ)

二元論は懐疑論に陥ることを示すのがここでの目的である。そのために二元論の主張を復習しておこう。

心的性質と物理的性質はその種類からして基本的に異なっている。特に、物理的性質は空間的な性質であるのに対し、心的性質は空間的性質ではない。

これは性質二元論の主張である。デカルトの二元論は心は物理的でない実体であるという実体二元論であった。ここでは私たちは性質二元論だけを考えることにする。性質二元論のほうが実体二元論よりもっともらしく、信奉者も多いからである。(信奉者の絶対数は少ないが。)さらに、性質二元論の問題はそのまま実体二元論に一般化できるからである。性質二元論の最低限の主張は、心的状態は非物理的状態である、あるいは心的性質は非物理的性質である、という主張である。この主張に、物理的なものだけがこのような状態を取ることができる、という主張を加えると、心的状態は物理的なものの非物理的状態であるという主張が得られる。この主張が通常性質二元論と呼ばれているものである。では、このような見解がMCの説明に成功するだろうか。ここで別の原理を一つ仮定しておこう。それは因果メカニズムの原理で次の主張である。

すべての因果性は機械的な因果性である。ある出来事が別の出来事を起こす場合はいつでも、二つの出来

事を結びつけるあるメカニズムが存在する。

さて、論証は次のように進行する。まず、MC は正しい；心的出来事と物理的出来事は因果的に相互作用する、と仮定しよう。すると、MC は因果性に関する主張であるから、それは因果メカニズムの原理を満たさなければならない。特に、心的性質と物理的性質の間を媒介するメカニズムがなければならない。しかし、すべての因果メカニズムは物理的なメカニズムであるという強い経験的根拠がある。しかし、性質二元論者の MC の説明ではこれを否定しなければならない。というのも、心的出来事は非物理的であり、その非物理的なものと物理的なものとの間にあるメカニズムは非物理的になってしまうからである。従って、彼らの説明は正しくない。

この論証が正しいかどうかは因果メカニズムの原理と論証中の物理的性質と非物理的性質の間には仲介メカニズムがないという仮定である。まず、因果メカニズムの原理から考えてみよう。因果メカニズムの原理は通常の科学研究において科学者がその仕事を実行する際の常識的な仮定である。因果的な解明に携わる科学者の作業仮説そのものである。この仮定に疑問を差し挟むということは原因や結果に関する形而上学的な考察が必要であることを意味している。Hume の因果性の議論に入る前に次の点を確認しておこう。それは、時空的に隣接する二つの出来事があっても一方が他方の原因あるいは結果とは言えない、つまり、時空間的な隣接性は因果連関の十分な証拠にはならないという点である。では、因果性は隣接性以外に何を含んでいるのか。それが因果メカニズムの原理であった。

私たちが知っているメカニズムのすべては物理的なメカニズムである。心的状態は非物理的状态であり、すべての因果性が仲介的なメカニズムを含むという仮定のもとでは、MC が成立するなら、非物理的状态と物理的状态を結びつけるメカニズムが存在しなければならず、これを証明することは困難である。非物理的状态と物理的状态の間にメカニズムが存在することを「二元論的メカニズム仮定」と呼んだとすると、これが誤っているという主張は多くの根拠、証拠をもっている。しかし、これが誤っているという証明はできない。ただ、次のことは経験的に確かなことである。今日までに知られている因果的メカニズムのすべては物理的な性質と物理的な状態しか含んでいない。

(物理主義と懐疑論的テーゼ)

前節の結果は次のようであった。もし二元論が正しい心の理論であるならば、MC は誤りであると結論できる可能性がある。結局、MC がどのように正しいかを二元論は述べることができない。そこで過去に出されたのが並行論であった。並行論の意図は相互作用抜きに心身関係を捉えようということにあった。

だが、現在はこの逆で、MC を否定する代わりに、それを正しいとし、二元論を否定する傾向が圧倒的に強い。私たちもこの傾向を受け入れることにする。

心的状態が物理的状态と因果的に相互作用する場合がある。すべての因果作用は仲介的なメカニズムを含んでいる。そのようなメカニズムの唯一の候補は物理状態と性質である。したがって、心的状態は少なくともある物理的性質をもたなければならない。この結論を一步前進させるのが物理主義である。前進させた主張は心的状態は物理状態そのものである、となる。心的因果に関する物理主義的な説明は、したがって、心的状態はそれがもつ物理的性質によって物理状態と因果的に相互作用する、ということになる。

しかし、本当に物理主義は心的因果を説明するのだろうか。そこで、次の批判を考えてみよう。二元論は心的なもの¹と物理的なものを余りに異なるものと考え過ぎた。一方、物理主義は心的なもの²と物理的なものを余りに似たもの³と考え過ぎていないだろうか。確かに、心的状態は物理状態と因果的に相互作用するとは言えるが、それは心的状態として相互作用しているのではなく、むしろ物理状態として因果的に相互作用しているに過ぎないのではないか。これでは心的なものが何の因果的な役割も果たさない、単なる随伴現象に過ぎないことになる。

やはり問題は因果性をどのように考えるかに行き着くことになる。Hume の問題は、時空的な隣接性のほかに因果性には何が必要かであった。そして、その解答は因果法則の存在であった。因果性は隣接性の他に因果法則の存在を要求する。

- 1 史門がマッチを擦ると、その瞬間に火が点く。
- 2 史門が転ぶと、その瞬間に伊作がテレビの電源を入れる。

明らかに 1 は因果関係を述べている。それに対して 2 は二つの出来事間に因果的な結びつきは（常識の範囲で）認められない。1 の二つの出来事は法則的に結びついているのに対し、2 のそれらは単に偶然に結びついているに過ぎない。

（#）因果性は実に様々に考察されてきた。私たちも後でじっくり考えることにするが、ここではその複雑さを理解してもらうために本文の因果性とは別のタイプの因果性の一例を挙げておこう。私たちが因果関係と呼ぶものが多岐にわたっていることを以下のものを参考に考え直してほしい。

Downward Causation

Donald T. Campbell formulated the principle of downward causation as: all processes at the lower level of a hierarchy are restrained by and act in conformity to the laws of the higher level.

Reductionism can be defined as the belief that the behavior of a whole or system is completely determined by the behavior of the parts, elements or subsystems. In other words, if you know the laws governing the behavior of the parts, you should be able to deduce the laws governing the behavior of the whole.

Systems theory has always taken an anti-reductionist stance, noting that the whole is more than the sum of the parts. In other words, the whole has "emergent properties" which cannot be reduced to properties of the parts. Since emergence is a rather slippery concept, which has been defined in many different ways, most of which are highly ambiguous or fuzzy, I prefer to express this idea with the more precise concept of downward causation.

Downward causation can be defined as a converse of the reductionist principle above: the behavior of the parts (down) is determined by the behavior of the whole (up), so determination moves downward instead of upward. The difference is that determination is not 'complete'. This makes it possible to formulate a clear systemic stance, without lapsing into either the extremes of reductionism or of holism: the whole is to some degree constrained by the parts (upward causation), but at the same time the parts are to some degree constrained by the whole (downward causation).

このような Hume の考えと物理主義を結びつけると、心的因果に対してどのようなことが言えるだろうか。その前に二つの物理主義を区別しておこう。

（タイプ物理主義、因果法則、懐疑論的テーゼ）

タイプ物理主義の主張は次のようになっていた。

どのようなタイプの心的状態もあるタイプの物理的状態である。

さて、この賛同者が MC の正しいことを主張したとしてみよう。Hume の考えに従えば、心的状態と物理的状態が因果的に相互作用するような場合を支配する法則があることになる。どのような法則が存在するかを探るために神経生理学の教科書や論文を漁ってみても、そこに見出すのは脳の状態や神経の性質の間に成立する法則だけであろう。その限りでは MC が正しいかどうかは不明のままである。したがって、タイプ物理主義者が心的因果を説明しようとするれば、心的状態のタイプと物理的状態のタイプを結びつける法則が存在しなければならないことになる。そのような法則は、

ある人が痛いと感じているなら、その人の脳のある領野に云々のニューロンがある

といった形をしているだろう。この形の法則は心理 - 物理法則(psycho-physical laws)と呼ばれてきた。このような法則が存在するかどうかはいくら哲学的に議論しても埒があかない。これは経験的な研究によって決着する事柄である。もしそのような研究が成功し、法則が確証されれば、タイプ物理主義者の主張は認められることになる。むしろ、その場合には「痛いと感じる」がどのように科学的に扱われるかがわかっていなければならない。そして、その際には MC の説明も認められる。しかし、その逆の場合も同じように可能である。

こうして得られる暫定的な結論は、MC のタイプ物理主義的説明の妥当性は神経科学の発展待ちということになる。しかし、痛みのような感覚ではなく、信念や欲求を考えた場合、心理 - 物理法則の将来の獲得は期待薄である。少なくとも、現在のところそのような法則の兆候さえ見られない。

(#) 心理 - 物理法則を追求する歴史はヘルムホルツにまで遡れるが、近年の代表例は Roger Penrose であろう。彼の考えは以下のようなものである。(詳しくは後述を参照)

One form of dualism involves the mysteries of quantum mechanics. Roger Penrose argues that consciousness is the link between the quantum world, in which a single object can exist in two places at the same time, and the so-called classical world of familiar objects where this cannot happen. Speculation that quantum mechanics and consciousness are linked is based on the principle that the act of measurement--which ultimately involves a conscious observer--has an effect on quantum events.

Moreover, with Stuart Hameroff, he has proposed a theory that the switch from quantum to classical states occurs inside certain proteins call microtubules. The brain's microtubules, they argue, are ideally situated to perform this transformation, producing "occasions of experience" that with the flow of time give rise to stream of consciousness thought.

(トークン物理主義、因果法則、懐疑論的テーゼ)

タイプ物理主義では心的因果がうまく説明できないという心配から出てきたのがトークン物理主義であり、その典型は Davidson に見られる。彼の有名な論証を試みよう。

1 ある心的出来事は物理的出来事と因果的に相互作用する。(MC)

2 因果作用が存在するところには決定論的な法則が存在しなければならない。(Hume の考え)

3 しかし、心的出来事を物理的出来事に結びつける決定論的法則はない。(心的なものの非法則性)

したがって、タイプの同一性はなくとも、特定の心的状態のトークンは特定の物理状態のトークンと同じである。

もし彼が正しいならば、MC と両立する物理主義があるだけでなく、MC が正しいことがトークン物理主義の正しさを含意することになる。Davidson によれば、心的状態が物理的状態を引き起こしたと仮定してみると、そこには厳格な法則が存在する。この法則は物理的であり、その心的状態は物理的な記述をもつことになる。これは、その心的状態が物理的でもあることを示している。そして、物理的な出来事に因果的に関係するすべての心的出来事は物理的な出来事であると一般化できる。ここでは二つの点が基礎になっている。まず、心理 - 物理法則が存在しないということはタイプ物理主義者の MC の説明を成立させないが、物理主義的な MC の説明そのものを排除するわけではない。第二に、心理 - 物理法則の存在しないことは Hume の考えと MC を使うことによって、トークン物理主義を成立させる議論として利用できる。結果として、心的因果の説と両立するだけでなく、それによって実際につくられる物理主義の一つがあ

ることを示唆している。私たちは MC と両立可能な物理主義、つまりはトークン物理主義があるということと満足できるだろうか。満足してはいけないという反論が二つある。

(因果的関連の自然主義的原理からの論証)

対象のもつすべての性質が因果的に関連しているのではない。これが最初の議論の核心である。黄色い本が床に落ちたときに「黄色」は本の落下に因果的に関連しているだろうか。明らかに因果的に関連していない。では、因果的に関連しているとはどのようなことなのか。因果的な関連性に対する答えは現在の最善の科学によって与えられるというのが自然主義である。その原理 (NCR) は以下になるだろう。

様々な科学によって前提される基本法則に現れる性質だけが因果的に関連する性質である。

この原理によれば、基本的な物理法則に登場する性質こそが因果的に関連する性質であるということになる。この原理を以前の黄色の本に適用するとどうなるだろうか。黄色の本が床に落ちるという出来事に適用でき、かつ黄色に言及する法則がないとすれば、上の原理は黄色それ自身は本が落ちることに因果的に関連している性質ではない。さて、この原理を使って Davidson のトークン物理主義に反対する議論を考えてみよう。

NCR は与えられた結果に対してある性質が因果的に関連するかどうかを決定する途を与える。手が上に動くような出来事と手を挙げたいという性質を仮定し、さらにこの性質が手が上に動くことに因果的に関連しているかどうか決定したいとしてみよう。NCR が私たちに教えてくれるのは、任意の出来事 e とその因果的な関連が問題になっている性質 p が与えられた場合、 p が e を含む科学法則において言及されたなら、 p は e に因果的に連関しているということである。だから、手を挙げたいという性質は手の上への動きに因果的に関連しているなら、手の動きに適用され、手を挙げたいという性質に言及する科学法則が存在することになる。しかし、そのような法則の存在は Davidson が否定しているものである。心的なものの非法則性は、心理 - 物理的法則が存在しないことと同じである。従って、心的性質は因果的に関連していないことになる。

この反証が何を言い、何を言っていないかは重要である。まず、Davidson のトークン物理主義は MC と両立しないとは言っていない。反証は両者が共に真でありうることを許している。それが言っているのは、トークン物理主義が真であれば、たとえ心的状態が物理状態と因果的に相互作用する場合があったとしても、心的性質は因果的に関連していないということである。前の例で言えば、本が黄色であることとそれが床に落ちることとは両立可能である。だが、同時に本が黄色いことは床への落下に因果的に関連していない。たとえ本が赤色であっても、その本は床に落ちるかもしれない。トークン物理主義の難点は、人間の心的性質は本が黄色であることに似ていることにある。したがって、身体はたとえ異なる心的性質をもっている(あるいは心的性質を全然もっていなくとも)、同じように行動するかもしれない。これは、トークン物理主義が心的性質に因果的な役割を与えないことを示している。そうであれば、MC のトークン物理主義的な説明は誤っていることになる。

Davidson はこの反対にさまざまに答えている。二つの人が物理的に区別できないならば、彼らは心的にも区別できないという付随性の考えを使って彼は反論を始める。この対偶から、二人の心的状態が異なれば、二人の間に何らかの物理的な違いがなければならない。物理的な性質の違いは因果的な違いを引き起こすゆえ、因果的な効力をもっている。したがって、心的性質も因果的に関連している。これが彼の答えである。しかし、この最後の結論は出てこない。出てくるのは、心的性質は因果的に関連した性質と連動して変化するというだけである。因果的に連動して変化するということと因果的に関連した性質の間には重要な違いがある。

付随性を使った議論は心的性質が因果的に関連する性質と共に変化するというを示すだけであり、心的性質が因果的に関連する性質ではない、というのが再批判である。したがって、トークン物理主義とNCRのもとでは、心的状態は物理状態と因果的に相互作用するとしても、心的性質は何の因果的な役割も演じない、という結論になる。

(因果的説明排除の原理からの論証)

MCのトークン物理主義的説明に対する二番目の反論は、Davidsonの非法則的一元論から得られる。トークン物理主義は心的性質の因果的無関連性を含意するという点では最初の反論に似ているが、NCRに直接訴えることはしない。使われる原理をKimにしたがってCEEと呼んでおく。Davidsonのトークン物理主義とCEEは心的性質の因果的無関連性を含意するというのがその主張である。CEEは次のことを述べている。

ある出来事の原因あるいは因果的説明は、それが十分な原因や説明とみなされるなら、その出来事の別の原因や説明を排除する。

CEEは人々がある出来事の説明を複数喜んで受け入れるような場合はないということを言っているのではなく、そのような場合があっても、十全な説明はせいぜい一つであると述べている。KimはCEEにNCRを加えることによって、「もし与えられた出来事の十分に受け入れることのできる説明があれば、その出来事に因果的に関連する性質だけがその説明で言及される性質である。」という結論を導き出す。これをDavidsonのMCのトークン物理主義的説明に適用してみよう。それは次のようになろう。人間の行為はある種の物理法則によって支配されている。もしそうなら、それらは十分な物理的説明を原理上はもっている。しかし、CEEのもとで、もしそれらが十分な物理的説明をもつなら、関連する性質はその説明のなかで言及されているものだけである。心的性質はその説明のなかには言及されていない。したがって、心的性質は因果的に無関係である。

(最後に)

MCについての妥当な説明を与えることができるものはなかった。これでは出発点と何ら変わりはないことになる。MC自体、あるいは私たちの因果作用についての考えが誤っているのではないかという疑問が出てくる。しかし、ここでそのような疑問から科学的な心的因果なる理論をつくっても、出てくるのは今まで考えてきたのと同じもの、つまり、そのような理論によってMCは説明できるのかという疑問である。ここにはさらなる哲学が求められている。

6.3 行為の因果連関

一元論によって心を捉えるという立場を取ることと、心の性質や現象を追求することは異なることである。遺伝子の仕組みがわかれば、その遺伝子をもつ生物個体の行動がわかるものではないように、心の仕組みがわかったからといって、その性質や現象が説明できるわけではない。特に、遺伝子の仕組みの解明に使われた考え方や手法がそのまま生物個体の性質や現象に適用できない(実際、遺伝学と生態学は考え方も手法も異なる)ように、心の場合もそれが働いて行為に結びつく場面では、心が物理的ではあっても物理学の考えや手法ではその行為を扱うことはできない。

ところで、心的因果についての議論は私たちをどこに導くのだろうか。多くの哲学者が参加して心的因果の姿を捉えようとしても議論のための議論が横行しているようにさえ映る。この理由の一つは心的因果が心の内だけでなく、外とも結びついているためである。外との結びつきとは信念や欲求が原因になって行為が実行されることである。ただ、問題を行為を含めた場面に拡張することは一層解決を困難にするという危険性も持っている。とはいえ、以下に行為の因果連関の基本的な考え方と問題点を挙げておくことは

有意義なことと思われる。

私たちの行為は世界の中で実行される。その世界は因果的な連関をもつ出来事や状態の集まりである。行為もまた出来事の一つであるから、世界の因果連関の中に組み込まれていることになる。このような因果の鎖の一部を取り出してみるなら、次のような一連の関係が見えてくるだろう。

環境、遺伝子 心 (= 信念 + 欲求) 行為

心の関与は私たちの常識的な世界では重要な役割を果たすとみなされている。しかし、既に議論を重ねてきたように心の働きと物理世界との関係は決して明らかなものではない。この明らかでない関係が引き起こす典型的な問題が自由と決定に関するパズルである。それは次のように表現できる。人間の信念、欲求、そして行為がその人自身のコントロール外のものによって引き起こされるなら、そこに私たちの自由な裁量が入ってこない。環境や遺伝子は人間の心のあり方を決定している。環境や遺伝子は私たちのコントロール外のものである。そうであるなら、どのようにして行為が自由選択の結果と言えるのか。これが伝統的なパズルである。

自由な行為に見えないような行為は私たちの周りに溢れている。そして、そのような行為は社会にさまざまな問題を生じさせている。例えば、病的な盗癖をもつ人の窃盗は自由になされた行為ではないことから、法的な責任を免れるのだろうか。このような問題の背後にあって、私たちを悩ませている概念が因果性(因果作用)である。

因果性とは何か、ある出来事が別の出来事を引き起こすとは何を意味するのか。原因と結果が因果作用を構成しているが、原因はその結果の十分条件である必要はない。また、必要条件である必要もない。つまり、原因、結果と前提、帰結の関係は類似していても基本的に異なったものである。私たちは前提と帰結の論理的関係についてはある程度知っているが、原因と結果の因果関係については思っているほどは知っていない。

(#) 「ならば」の二義性について

「A ならば、B である」という表現は単純であるが、原因 - 結果と前提 - 帰結の二つの(根本的に異なる)関係を二重に意味している。それを次の例で実感してほしい。

- (1) $x + y = z$ ならば、 $2z = x + y + z$ である。
 (2) 伊作が怒るならば、史門が泣く。

(1)の「ならば」は論理的な「ならば」であり、前提 $x + y = z$ と帰結 $2z = x + y + z$ の含意関係を主張している。(実際、変数が自然数や実数であれば(1)は正しい文であり、含意関係は成立している。) それに対して、(2)の「ならば」は因果的な「ならば」で、伊作のある状態変化と史門のある状態変化の間に因果的な関係があることを主張している。二つの「ならば」の違いは極めて重要である。ここでは違うことだけを確認しておこう。例えば、(1)の前提と帰結はそれらがいつ成立するかは考慮されないが、(2)の二つの状態は時間的な制約を受けている。伊作が先に怒り、次に史門が泣くのでなければ、因果関係は成立していない。日常的な表現である「ならば」が論理的、因果的の二つの意味を併せもつことは日本語だけの偶然的な特徴ではない。英語でも「if then」は二義的に使われている。

(問い) 物理学は数学を使って物理現象を記述するが、数式の間論理的な「ならば」はどのように物理現象の因果的な「ならば」に関係しているのだろうか。

因果的な連関に関して古来議論されてきた考えは次の二つである。

Determinism (決定論) : 因果的な事実を完全に記述すれば、何が将来生じるか決定できる。

Indeterminism (非決定論) : 現在の完全な記述が与えられても、将来に二つ以上の可能性を残す。

因果的決定論はあらゆる因果的に関連する事実が与えられれば、将来はただ一つだけ決まると主張する。もしすべての物質が決定論的で、心も物質であれば、人間の行為は物理的に決定されていることになる。実際、古典力学は物質の決定論を主張してきた。この世界観は今世紀まで信じられてきたが、量子力学の登場と共に非決定論的な世界観が浸透し始めている。(この非決定論の心の領域への浸透は後述参照。)物理学以外の領域では人間の行為の自由選択、意志の自由が古くから認められてきたため、19世紀には社会科学で既にその自由の入った出来事や状態を確率・統計概念を用いて取り扱ってきた。この二つの流れと心の特徴づけは上述のパズルとなって私たちに突きつけられている。心が物理的なものかどうかを未定のままにしておいても、次のような二つの選択肢の間で決断を迫られることになる。決定論が正しいなら、確率は私たちの知識や情報の欠如であり、主観的なものになる。決定論が誤っているなら、確率は世界についての客観的な事実を述べていることになる。では、非決定論は私たちを自由にするのか。それは私たちに自由を保証する理論的な根拠になるのだろうか。確率が世界の客観的な事実であるなら、上述の行為の因果連関は次のように修正しなければならない。

環境、遺伝子、偶然(chance) 心 (= 信念 + 欲求) 行動

そこで、あなたの信念と欲求が環境、遺伝子、そして偶然によっているとしてみよう。もし決定論があなたの自由を奪うなら、偶然もやはりあなたの自由を奪うことになる。あなたが自ら自由に決定するのではなく、偶然に左右されるままになるというのは、あなたにとっては厳格に決定され自由の入る余地がない場合と大同小異である。とにかく、あなたの自由な決断は入る余地がない。

因果作用は決定論が成立しない世界でも可能である。非決定論的な世界でも因果作用は存在する。原因と結果の間に確率的なつながりがある場合を考えてみればよい。決定論的でない因果連関が想像できるだろう。つまり、因果性と決定論、因果性と非決定論はそれぞれ両立する。したがって、真の問題は決定論と自由が和解できるかどうかではなく、因果性と自由が和解できるかどうかである。量子力学が決定論は誤りであると主張しても、因果性そのものを否定はしていない。このような意味で、因果性こそが自由を考える上での鍵である。

ここで、決定論と運命論 (fatalism) の区別も重要である。決定論は、もし過去が異なっていたとすれば、現在も異なっていたらという考えを排除しない。決定論はまた、現在私がある仕方ではなく別の仕方を選ぶならば、私は未来に起こることに影響を与えることができるという考えも排除しない。しかし、運命論はこれを否定する。現在あなたが何をしようとする未来はそれとは無関係であるというのが運命論の主張である。つまり、決定論と運命論はほとんど正反対のことを主張している。運命論は私たちの信念や欲求が無力であることを主張するが、決定論では信念や欲求は因果的に私たちの行動をコントロールできることが主張されている。

6.4 自由意志

私たちは既に両立可能性 (compatibility) という論理的な概念を何回か使ってきた。ここでその意味を確認しておこう。

二つの命題は両立可能である iff 命題の一方の真が他方の真を排除しない

したがって、二つの命題が両立しないなら、一方の命題が真のとき、他方の命題は偽でなければならない。この概念を用いると自由意志についての私たちの考えは次の二つに分けることができる。

非両立主義：決定論が正しいならば、私たちは自由ではない。

両立主義：決定論の正しさと自由の存在は両立する。

(Daniel C. Dennett on Determinism and Free Will)

1983年にDennettはオックスフォード大学のJohn Locke Lectureで自由意志 (Free Will)を論じ、その内容が翌年出版の*Elbow Room: The Varieties of Free Will Worth Wanting*に述べられている。人間が自由意志をもつとは何を意味するかを「私たちは行為の自由を全くもたない決定論的な機械なのか、それとも私たちは行動を真に選択できる余地、余裕をもっているのか」という問いによって彼はこの問題を追求している。

生物学的な主体としての人間はどのようなもので、自由意志がなぜ人間には重要なのかを彼は考える。ハチの自動的な行動と人間の行動の比較：ハチはある行動を中断されるとまたそれを繰り返す、それを何度も繰り返す。人間はそこで選択し、新しい行動を選ぶ。ハチと人間の生物学的な違いはどこにあるのか。人は環境との相互作用を通じて、状況の内的、心的モデルをつくり、うまい行動戦略を考え出す。ハチは環境から学ぶことをせず、遺伝的に決められたプログラムを忠実に実行するだけである。この違いこそ私たちが自由意志をもつことであり、それはモデルを使って可能な選択肢の中から選び出すように見えるということである。

自由意志の哲学的な議論はパラドックスとして語られる。私たちは自分が自由意志をもっていると思っている。一方、生物学は人間をハチと同じような決定論的ルールに従うメカニズムとして記述する。この二つをどのように和解させたらよいのか。

決定論とは何か。可能な行動の中から一つを選択すると言うとき、真に選択というものが存在するのか、それとも選択があるように見えるに過ぎないのか。Dennettの決定論の定義は、次の通りである。すべての物理的な出来事はそれ以前のすべての出来事の総計によって引き起こされる、あるいは決定される。もし宇宙全体を同じある時点から再出発させたなら、いつも同じ未来に到達するだろうか。これは実験できないため、古くから哲学者や物理学者の関心を集めてきた。間接的な証拠の一例は、複数の宇宙が存在し、量子論的な非決定性が生じるたびに新しい宇宙が分岐するという説である。量子レベルの非決定性が自由意志を説明するという考え（後述 Penrose 参照）をDennettは退ける。ランダムな量子レベルの出来事と行動を人間がコントロールすることは関係ないと考えるからである。

彼は多くのタイプの自由意志について議論している。決定論と自由意志の両立可能性は、例えば、脳が量子レベルの非決定性を使って、複数の可能性の中から一つを選択するという仕方考えられてきた。しかし、彼は別のタイプの両立可能性を主張する。彼が考える自由意志は、理性的な行為によって環境に対応する生物としての主体の能力である。もし私たちの脳が私たちの行為をコントロールできるなら、行為の選択とは脳の不完全さでしかない。行為の選択は単なる錯覚ではないのか。人間はただ一つの仕方振る舞う機械でしかないなら、行為の責任はどうなるのか。それは人々が社会的に受け入れられる仕方振る舞うための効果的な方法である。さらに、もし行動上の選択が錯覚であれば、運命論が正しいことにならないのか。行動の選択はないがコントロールはできる。これが私たちの自由意志なのか。

(#) 「決める」世界観から「決まる」世界観への移行は自然世界で始まり、人間の心の世界にまで及んでいる。そして、「決める」を「決まる」に還元することが科学では目指されてきた。その最後の砦が自由意志の存在である。「決まる」世界で「決める」ことが可能となるのはどのような条件によってか。これは自由と決定の問題である。それは、主観的主体と客観的対象と区別されてきたものでもある。この二元論をどうすべきか。それを解決する方法の一つの方法が、状態と出来事の違いを導き出す相互作用と視点の枠組みにあるように思われる。

7 自己知識と他者の心

7.1 デカルト的な自己知識

自己についての知識と他人についての知識についての問いは哲学のなかで長い伝統をもつ問題であるばかりでなく、日常生活の場面でも毎日のように登場する問題である。それらは日常生活のほとんどの場面

で重要な役割を果たし、場合によってはその人の一生を左右するような力さえもっている。デカルトが自己知識を強調して以来、それは私的で内的な世界で進行する意識的な知識であり、私だけが特権的に知ることのできるのが私の自己知識(self knowledge)である、と一般に受け取られてきた。現在でも私たちの常識として通用している。このような直接に知ることのできる、証拠の要らない、自分自身についての知識と対照的なのが行動の観察を通じて得られる他人の知識である。心の科学は当然ながらこのような観察による知識に基づいている。日常生活でも他人の観察とその利用は実に重要な役割を演じている。以下では、まず自己知識から考えていくことにしよう。

(自己知識)

自己知識に関する最初の明確な表現はデカルトのものである。各個人が自らの心について唯一の、特権的な立場を取ることができるという主張はデカルトの深い哲学的な直観であった。人が自分自身の(信念、欲求、思考等の)心的状態についてもつ考え、意見を「一人称の意見(first-person opinions)」と呼ぶことにすると、自らの心に対して特権的であるとは、自らの心的状態について信頼できる意見をもつということである。一人称の意見が表明される言明を「一人称の報告(first-person reports)」と呼ぶと、それらは「私は考える」、「私は信じる」、「私は疑う」といった表現を含んでいる。それら報告は、まず心的状態の内容(何かについての心的状態であるので、そのような何かが内容である。具体的には信念の内容、疑いの内容と呼ばれているものである。)、次に、話者のその内容に対する態度を含んでいる。つまり、一人称の報告はある人の心的状態の内容とその内容に対するその人の態度の両方に関する一人称の意見を表明している。(デカルト批判ではこれが既に使われていたことを思い出してほしい。私たちは命題に対する態度を使ってデカルトの誤りを指摘した。)

自己知識に関する仮説を再度書き記しておこう。

一人称の報告は、その人が何を信じているか(疑っているか、考えているか等)を確定しようとする時には特権的である。

これを一人称特権説と呼んだとすると、これは私たちの心の概念の中心に位置している。では、どのような理由でこの説が認められているのか。心をもつということのなかには自らの心の状態について知る立場にいたることが含まれている。これを可知仮説と呼んだとすれば、一人称特権説から可知仮説が導き出され、そこから一人称特権説は心の概念の一部をなしていることがわかる。可知仮説は心的状態と物理的状態の違いを捉えている。「雨が降るだろうと考えている」ような心的状態と体重が65キロだという物理的状態は何が異なっているのか。「雨が降るだろうと考えている」状態にある人は自分が「雨が降るだろうと考えている」状態にあることを知っている。心的状態にあることとその心的状態にあることを知っていることの間には強い結びつきがあるように思われる。一方、自分の体重が65キロであること(65キロの状態)を知らない人はたくさんいる。体重65キロとそれを知っていることの間には何の結びつきもなさそうである。それは偶然的なものである。可知仮説がないなら、一体どのように心的状態について話ができるのか。一人称の報告をそもそもどのようにつくることができるのか。このような点から可知仮説は正しいようにみえる。

次に可知仮説が想定している知識はどのようなものか。私たちは自分自身の考えを知る時、その証拠を求めない。その知識の獲得には証拠を集める必要がない。その意味で一人称の意見は無根拠(groundless)である。この無根拠という性質を使って自己知識の仮説を書き直すと、

一人称の報告は、その人自身の心的状態を確定する時にはその本性上特権的である、

となる。哲学が議論の出発点とする第一の原理、それ自身は何かによって根拠付けられることがない原理という特徴を自己知識の仮説はもっていることになる。

では、心をもつことが自分自身の心的状態を知る立場に在ることであるという可知仮説は疑う余地はないのだろうか。日常経験は疑うことができる理由を与えてくれるように見える。史門は哲学の試験が簡単だと考えていると自分で言ったとしてみよう。彼の一人称の報告は「哲学の試験は簡単だと彼が信じている」という内容である。だが、実際には史門は毎日試験勉強に励み、彼が信じていると言ったことを本当は信じていないような行動をとっていたとしてみよう。私たちは彼の行動からその一人称の意見は嘘であると言うだろう。史門は哲学の試験が簡単だと自分で誤って信じている。実際は彼は簡単だと信じていない。このような自己欺瞞があるなら、私たちは自己知識が完璧で完全であることを否定しなければならない。ところで、可知仮説には知識が完璧で完全であるという条件は入っているだろうか。そのような条件は可知仮説には含まれていない。

(他人の心の知識についての仮説)

他人の心的状態の知識を手に入れることができる場合があるという仮説について考えてみよう。これは仮説というより、日常当たり前のことではないのか。他人の気持ちを知るということは日常茶飯事のことであり、他人の気持ちの報告が正しいかどうかはその報告内容に十分な証拠があるかどうかによって決定される。そして、私たちの社会生活はこの仮説が正しいというもとで成立している。実際、社会科学はこの仮説無くしては研究不可能である。社会心理学が扱うのは他人の心であって、社会心理学者自身の自己知識ではない。

(「心的状態を知る」という概念)

ここで第3の仮説として、既述の二つの仮説をまとめ、自己知識と他人の心の知識は共に心的状態の知識を含むものとして正しい表現になっていると仮定してみよう。これを認知仮説と呼ぶことにする。

認知仮説のもとでは自己知識も他人の心の知識もある型の知識、つまり心的状態の知識を含んでいる。では心的状態を知ることにはどのようなものが含まれているのだろうか。心的状態を知ることの分析結果が二つの仮説を真にするものを含んでいるなら、心の相互作用に関する理論の足がかりになるだろうが、残念ながら両方の仮説を真にするものは見当たらない。「心的状態を知る」はどのように特徴づけても、両方の仮説に共に合致するようにはできない。この結果は二つの仮説の再吟味を引き起こすことになる。まず、「心的状態を知る」とはどのようなことが考察してみよう。

(#) Hume's "Bundle" Theory of Self

(心に関する Descartes の伝統は克服されるべきであるが、Hume の懐疑論は現在でも生きた問題である。) Hume は心と身体を分けるが、Descartes と違って心を実体と見なすことに反対する。それは彼の二つの原理に基づいている。(1) 心の知覚すべては印象と観念に分けられる。(2) すべての観念は印象のコピーである。(知覚は意識を含み、印象は観念より生き生きとしていて、強い。)(2)は、次のことを意味している。すべての観念に対して、それを生じさせた印象(の組)があり、観念はそれを生じさせた印象に似ている。したがって、以前の印象から作られない、印象から独立した観念は存在しない。意識は経験の主体が存在することを示しているという考えに対して、Hume は次のように反対する。印象が観念をつくるのであるから、存続する自己の印象をもたなければ、自己の観念ももてない。印象はいつも移り変わっている。存続する自己の印象はどこにもない。それゆえ、存続する自己の観念はない。この Hume の批判は、時間を通じて存続する、経験の統合としての自己という哲学的な解釈に向けられている。

"I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement."

Treatise of Human Nature, Book I, Part 4, section 6.) (国家はある関係で結ばれた個人の集まり以外の何ものでもない。)

では、このような誤りはどこからくるのか。Humeによれば観念の連合に原因がある。ある印象と別の印象が恒常的に起こるなら、原因と結果のような恒常的な関係があると考えてしまう。自己についても同様である。

Humeの試みは次の点で問題を残す。(1)同一人物のものとして考えられる印象の束は因果的に関係していない。(2)他人の経験を直接に知ることはできない。

StrawsonのDescartes二元論批判

Strawsonは日常言語に反映されるPerson概念を明らかにしようとする。そして、なぜこの概念がDescartesやHumeのものより優れているか示そうとする。そこで彼は次の問いを考える。「なぜ意識の状態は物理的な性質をもつ、正にその同じ対象に帰属させられるのか。」つまり、なぜ「I」は意識の状態を指すと同時に、物理的な身体をも指すのか。これはHumeの考えに反するよう見える。そもそも意識を何かに帰属させるということがなぜ必要なのか。

Strawsonは述語を物理的なものと心的なものに分け、Descartesの心、つまりは意識状態の実体が物理的な述語をも満たさなければならないとし、意識だけの実体を批判する。

StrawsonのHume束説批判

なぜ意識状態は何かに帰属しなければならないのか。Humeでは人は知覚の束であり、それが原因と結果の鎖をなしていた。ここでは知覚は区別できる、離散的な出来事である。この離散的な出来事は通常は確定記述によって与えられる。この確定記述で指される個別的なものは個人的な経験によって確定されるものでもある。個別経験はこの人やあの人の経験としてのみ確定できる。では、あるPersonをつくるために個々の経験はどのように関係していなければならないのか。これをHumeは説明できない。

他人の心についての懐疑論

緩やかな懐疑論：他人の経験は私の経験に似ているか。

強い懐疑論：私の経験以外に経験は存在するのか。

- (1) 他人の心を直接に知ることはできない。
- (2) 他人の心を推論によって間接的に知ることはできない。
- (3) したがって、直接にも間接にも他人の心を知ることはできない。

帰納法あるいは類推による推論が可能なら懐疑論は克服できる。しかし、これについてはHumeの有名な論証がある。Humeは因果性についての考察とともに、経験的な知識の正当化についても有名な論証を与えている。その理由をHumeは「帰納法は合理的に正当化できない」という形で示した。彼は経験的な主張を結論として導き出すために新たな前提が必要であると考えた。それが次の原理で、Principle of the Uniformity of Nature (*PUN*) (自然の一樣性原理)と呼ばれている。その主張は、未来は過去に似ている、という単純な内容である。経験的に繰り返し生じる現象は帰納法を使って、未知のものや未来のものに一般化されるのが普通である。そこで、Humeは帰納法を使った論証はみなこの原理を仮定しなければならないと考えた。

(Hume's skeptical argument)

- (1) Every inductive argument requires *PUN* as a premise.
- (2) If the conclusion of an inductive argument is rationally justified by the premises, then those premises must themselves be rationally justifiable.
- (3) So, if the conclusion of an inductive argument is justified, there must be a rational justification for *PUN*.
- (4) If *PUN* is rationally justifiable, then there must be a good inductive argument or a good deductive argument for *PUN*.
- (5) There is no good inductive argument for *PUN*, since any such argument will be circular.
- (6) There cannot be a good deductive argument for *PUN*, since *PUN* is not a priori true, nor does *PUN*

deductively follow from the observations we have made to date.

(7) So, *PUN* is not rationally justifiable.

Hence, there is no rational justification for the beliefs we have that take the form of predictions or generalizations.

* エメラルドの色や明日の太陽についての私たちの信念は正当化できない。その理由は正当化できない自然の一樣性原理という仮定に基づいているからである。(4) - (6)を詳しく見てみよう。自然の一樣性とは未来は過去に似ているという主張である。これは帰納法に基づいて真であると知ることができるものであろうか。もしそうなら、帰納的な論証は次のようになるだろう。

Nature has been uniform in my past observations.

Nature in general is uniform.

この論証は帰納的であり、Hume によればすべての帰納的な論証は *PUN* を前提としてもつから、この論証は循環することになる。

(It assumes as a premise the very proposition it tries to establish as a conclusion.)

では、*PUN* の演繹的な正当化はできるだろうか。Hume の答えは no である。上の論証は演繹的に妥当ではない。

また、それは定義の問題でもない。では、Hume の懐疑論は論駁できるか？ 帰納的な論証には *PUN* が用いられなければならないが、*PUN* は正確に何を意味しているのか。

未来はあらゆる点で過去に似ている。

未来はある点で過去に似ている。

上のいずれが意味されているのか不明であることからしても、*PUN* の正確な特徴づけは困難である。これは Hume の論証を *PUN* に依存しない形で再定式化すべきであることを示唆している。そこで、新しい定式化の一例を以下に挙げてみよう。それは信頼可能性(reliability)という概念を使ったものである。ここで信頼可能性の度合いを次の例で考えてみよう。

some degree of reliability

(1) All emeralds are green.

(2) All emeralds are green until the year 2050; thereafter, they are blue.

どちらの方がより妥当性が高いか。(1)の方が(2)より妥当性が高いことをどのように正当化するかが哲学の問題となる。その正当化は別にして、帰納法に適用すると次のようになる。帰納法は過去において極めて信頼性が高かった。したがって、帰納法は現在も未来も極めて信頼性が高いだろう。これを使って論証を構成すると次のようになる。

(reformulation of Hume's argument)

(1) To rationally justify induction, you must show that induction will be reliable.

(2) To show that induction will be reliable, you must construct an inductive argument or a deductively valid argument.

(3) You can't show that induction will be reliable by an inductive argument; that would be question-begging.

(4) You can't validly deduce that induction will be reliable from premises describing the past reliability of induction (or from definitions).

Hence, induction cannot be rationally justified.

7.2 心的状態の知識

ここでは伝統的な心的状態の知識についての概念を考えよう。伝統的な立場はデカルト的な見解である。自分自身の心的状態の知識が説明上出発点となる。まず、自己知識から解明し、次に他人の心的状態を知ることにその結果を拡張するというのが伝統的な追究の仕方である。これはデカルトに端を発している。一人称の特権性は私たち自身が自身の観念について疑うことができないという点にその起源がある。自己知識のモデルを他人の心の知識へと拡張しようとする、多くの問題に遭遇し、他人の心の存在に関する懐疑論が出てくる。これに対する対応は他人の心をもとにするという戦略の転換である。これが行動主義の方針である。しかし、この方針でも自己知識の十分な概念が得られない限り懐疑論に落ち込むしかない。

では、第3の途はあるのか。それは Wittgenstein を出発点に Wright、Davidson、McDowell 等によって展開された試みで、私たちが他人や私たち自身に心的状態があるとする時に使ういくつかの特徴や手掛かりによって、自己知識と他人の心の知識の両方が説明されるような戦略を用いるものである。

(1) デカルト的方法：自己知識をモデルにした理論

自分自身の心についての知識を説明の出発点にするという姿勢はデカルトに顕著に見られる。彼の「第二省察」の最後に知識の本性に関する考察がある。ある人が対象の時間的な変化にもかかわらず、その対象の知識ををもち続ける際にその人は何を知っているのか。これが彼の問いであった。彼の例は蠟であった。蠟についての知識はその感覚的な性質から得られる知識からなっている。しかし、デカルトによれば、蠟はあらゆる変化を被りながら感覚的な性質を変えていくが、それが蠟であるという私たちの知識は変わらない。ここから彼が得る否定的な結論は、蠟が変化するにもかかわらず明確な蠟の知識が得られるのは感覚的な観察によるのではないということであった。そして、デカルトは蠟の知識に関する別の説明を提案する。蠟の知識は変化する感覚的知識ではないので、蠟の変化を通じて保存されるようなものの知識である。蠟の物理的な変化を通じて不変なものは蠟自体の中にはない。それは私たちが蠟について創る観念である。これがデカルトの答えである。つまり、蠟自体の知識は特定の蠟についてのある人自身の観念の知識である。デカルトにとって、蠟の知識の直接の対象は蠟自体ではなく、蠟の観念である。したがって、このことがこの観念や思考の知識が基本になってそれ以外の知識が説明されるというデカルト的な計画となるのである。この計画では、私たちは世界を直接に知るのではなく、世界についての知識は直接に知られるもの（世界についての私たちの観念）から推論によって導き出されるものである。

(自己知識のデカルト的モデル)

自己のもつ観念には二つの特徴があるとデカルトは考えた。まず、それは私的な領域の知識である。その人自身の意識の領域の知識である。したがって、その人自身にしか近づくことができない。二番目に、この知識は誤ることができない知識である。私的な意識のなかで起こるものに直接に結びついているので、途中で誤りを引き起こすものの入り込む余地はない。

心は本質的に私的であるというのは法律上のものではなく、形而上学的なものである。原理的に、ある人自身の心で起こっていることに近づくことは他人にはできない。このように考えると、心はその所有者以外には入ることのできない完全にその人の私的な劇場のようなものである。この比喻では他人の心を知る可能性が排除されているように見えるが、そうではない。劇場に直接入らなくとも中で何が上演されているかは知ることができる。但し、直接知ることはできない。ある人が自分自身の意識的な思考についても一人称の意見は常に正しいというのがデカルトの二番目の説である。このようなデカルト的枠組みは次の三つに整理できる。

- 1 自己知識は今まさに起こっている観念の知識である。
- 2 これら観念はその人自身にしか近づくことのできないものの中にある。
- 3 その中のものの存在と内容についての知識は誤ることのないものである。

その人自身の知識という考えは観念自体の本性に向けられることになる。デカルト的な哲学は観念についての考察を通じて、観念についての説を心的表象(mental representation)に適用する。また、志向性の類似説もここから出てくる。

観念が心的表象であれば、それらが何を表象するかを決めるものがなければならない。牛の思考についての何かはその思考を牛についてのもので、馬についてのものではないことを決めなければならない。観念

が表象するものを決めるのはその観念の何なのか。心像説は、観念が心像であるゆえに観念の表象するものを決めると主張する。では、どのように心像は表象に成功するのか。それを説明するのが志向性の類似説である。像はそれが表象するものに類似することによって表象するというのがこの説である。この説が心的表象に適用できるとするならば、心的表象自体は絵画的な特徴をもっていなければならない。心像説はこのことを主張しており、観念は心の眼の前の絵画のような像である。

こうして、その人自身の観念を知るとは、その人の心の劇場にある心像をその人の心の眼で見ることである。この比喩的な表現を避けようとするならば、内観(introspection)という概念を用いなければならない。心の眼で見るとは内観することである。

(他人の心の知識についてのデカルト的方法)

デカルトの自己知識は私的で他人には近づくことができないということを考えれば、他人の心を直接知ることはできない。では、他人の心が存在することをどのように知ることができるのか。この問いに対する答えは Mill によれば次のようになる。まず、その人自身の心とその人自身の身体の間因果関係があることはその人の自己観察からわかる。第二に、他人を観察することからその人の一連の身体の運動についての知識を獲得する。第三に、他人の動きで観察された系列と自分自身での観察による系列を比較することができる。第四に、他人の観察が増えるに連れ、自分の観察と比較しながら、自分の動きの場合と同じ型の原因が他人の動きにもあるということを確認することができる。最後に、自分の身体的な動きの原因は心的な原因であることを既に知っているのだから、他人の場合も同じであると結論できる。そして、このようなアナロジーをもとにして、世界には自分自身と同じような他人も存在すると結論できる。この推論の系列は他人の心の存在だけでなく、それがどのようなものかを決める方法も示している。こうして、他人の心の存在とその働き方がわかることになる。

(デカルト的方法批判)

デカルト的な説明の意義は次の二点にあった。まず、自分自身の心の知識と他人の心の知識は異なるという私たちの直観にその説明を与えた。さらに、一人称の特権性に対する基礎を与えることができた。このような意義をもつデカルト的な説に対して三つの重要な批判を取り上げてみよう。

Mill の他人の心が存在するという論証がアナロジーに基づくものである点をまず考えよう。アナロジーに基づく推論では結論の確からしさはせいぜいそれを支える証拠の確からしさと同じである。実際、Mill の結論は「他人の心の仮説」であり、いくつかの証拠の最善の説明(the best explanation) のための仮説である。このような仮定は科学においては珍しいことではない。原子の存在はその証拠が提出される遥か以前から仮定されていた。しかし、デカルト的な枠組みにおいてさえ、他人の心の存在を仮定して他人の行動を説明する以外の説明方法は考えることができる。心的でない、機械的な原理の結果として他人の行動の観察結果を説明することが別の説明方法としてあるならば、他人が単なるオートマトンではなく心を持っているということがどのようにして言えるのだろうか。他人の心の存在を認めることのほうがオートマトンを認めることより単純である、そして仮定は単純なほうがいい、これが Mill の答えである。

次に二番目の批判に移ろう。Strawson と Malcom の批判は次のように表現できる。デカルト的な見解はある人自身が心的状態をもつとはどのようなことかということから、他の人自身が心的状態をもつとはどうということかを導き出している。しかし、注意深く調べるならば、他人が心的状態をもつことをこのように導き出すことには矛盾がある。デカルト的な概念ではその人自身以外の人が心的状態をもつという考え自体意味をもたない。したがって、他人の心の存在は懐疑論に陥ってしまう。これが概略である。

この論証は言語的な論証である。与えられた文を理解するのに何が含まれているかについてのテーゼは次のものである。

任意の話者が与えられた文を理解することになれば、話者はその文の正しい使用についての基準をもっていなければならない。

ところで、「正しい使用の基準をもっている」と「正しい使用の基準を明確に述べることができる」とことは異なる。したがって、話者はその基準を明瞭に述べることができなくとも、文を理解することができる場合があるだろう。さらに、「正しい使用の基準をもっている」とは「その基準が満たされる時にはそれが満たされることを知る」とことも異なっている。この二つの区別がなされるならば、このテーゼは確からしく見える。では、この文理解の基準を他人に心的状態を帰属させる文に適用してみよう。具体例で考えるのがよいだろう。

伊作は哲学がつまらないと信じている 史門は痛みを感じている 花子はパチンコをしたい

いずれの文もある心的状態（信念、痛み、欲求）をある人（伊作、史門、花子）に帰属させている。常識的に私たちはこれらの文を理解できる。「哲学がつまらないと信じる」、「痛みを感じている」、「パチンコをしたい」という（心的）述語だけを取り出しても、私たちはみな理解できる。さて、もしデカルト的な見解を受け入れるなら、他人の心で起こっていることは論理的に私的なことで、その人以外の誰にも近づくことはできない。すると、ある心的述語の正しい適用の基準はその人以外にはもつことができない。デカルト的な見解が正しければ、私たちは心的状態を他人に帰属させる文を理解できないことになる。せいぜい自分自身に対してだけしかできない。そして、心的状態を他人に帰属させる文を理解できないのであれば、他人が心をもつとはどのようなことかを実際には知らないということになる。この論証の核心は、

もしデカルト主義が正しければ、その人以外の誰も心的述語の正しい適用基準をもつことができない

という前提にある。この前提はデカルト主義と心的状態をもつことが私的であることからの必然的な結果である。

最後の批判はさらに厳しい。デカルト的な見解では、ある人自身の心像の一人称報告は心像を正しく記述する概念を適用するということである。心像を正しく記述する概念は心像の表象を記述するようなものである。概念の適用について語ることに意味があるということを厳しく批判したのが Wittgenstein である。*Philosophical Investigations* での批判は概念の適用はその本性上規範的な事柄（正しく使う、誤って使う）であるという考えに基づいている。この規範性は、今の場合、与えられた概念が与えられた場合に正しく適用されるか、誤って使用されるかの決定の仕方を想定している。だが、この想定は私的な心像への概念適用に対しては満たされない。彼は有名な private language argument でこれを示す。真の規範の存在は誤りをする可能性を想定する。概念適用における誤りの可能性は、その概念が正しく適用される条件と自らの概念の適用を比較でき、その正しい適用の基準を見出せる可能性を必要とする。しかし、デカルト的な見解では、概念を適用するのは論理的に私的なことである。すると、概念の正しい適用の唯一の基準はその人がその適用の時に正しいと考えるものだけとなる。これは、概念の正しい適用に見えるものと概念の正しい適用であるものの間に区別がないことを意味している。もしそうなら、概念適用に関して語るべきではないことになる。

（＃）Wittgenstein の私的言語に関する主張はどのようなものか。懐疑論によれば、他人の経験が同じ状況での私の経験と同じかどうか私は知ることができない。「私は頭が痛い」とあなたが言う時、あなただけが頭が痛いということが何を意味しているか知っており、「私は頭が痛い」と私が言う時には、その意味を知っているのは私だけである。この懐疑論は私が「痛み」のような語の意味を痛みに関する私の私的な経験から学んだという仮定に基づい

ている。Wittgenstein は *Philosophical Investigations* において語が伝達不可能な意味をもつことができないことを論じた。その論証は私的言語という概念が論理的に矛盾していることを示すことにあった。感覚と公的に観察可能なもの間にあるすべての結合を取り去って、名前と感覚を結びつけることによってできあがるのが私的言語である。名前と感覚の結びつけは、私が感覚 S を経験する時に「S」という記号を書き、以後もそれを守るという仕方で行われる。Wittgenstein によれば言語は規則に支配されている。、それゆえ、語が意味をもつためには、それを使用する規則がなければならない。だが、もし「S」が私的な感覚を指示するのに使われるとすれば、私に正しくみえることは正しいことになる。これは「正しい」について私たちが語ることができないことを意味している。ある状況でのある語の使用が正しいか誤っているかを示すためには、何か独立のものが必要である。私は今の感覚を「S」と命名した感覚像と比べることができる。しかし、想定された感覚像はその正しさを独立にテストしてやることができない。それゆえ、わたしは「S」を正しく使っているかどうか言うことができない。したがって、「S」は有意義な記号ではない。

以上がデカルト的な見解に対する批判である。これらの批判を受け入れるなら、心を私的なものではないものとして考えなければならなくなる。それが次の私たちの課題である。

(2) 反デカルトの見解：他人の心の知識

ここで考える反デカルト的な見解は行動主義である。心の知識に関する行動主義の見解はデカルト的な仕組みを逆にしたものである。心の私的性格を公的な性格で置き換える。それは他人の心についての公共的知識を基礎にして、それを自己知識に拡大するという説明上の戦略を採用する。そこでは他人の心の知識は他人の行動、行動的な傾向性についての知識と解釈されている。

(他人の心の知識の行動主義的モデル)

他人の心の存在と他人の心的状態の存在の両方の知識をもつための基礎は何か。それを調べる時に利用できるのは他人の行動である。この点では行動主義者もデカルト主義者も変わりはない。行動と心との関係をどのように考えるかで二つは根本的に異なっている。行動主義は行動と心のギャップを消去しようとする。心を行動に還元することを主張する行動主義のテーゼは、心的用語を含む文を行動的な用語しか含まない文に翻訳できるという主張と考えることができる。この主張を哲学的な行動主義のテーゼとして次のように表現できる。

心の状態を述べる文は行動を述べる文と意味において同じである。

これを他人の心を知る場合に適用すると、他人の心を知ることは他人の行動を知ることと意味において同じであるという主張になる。デカルト主義と行動主義が如何に異なるかを他人の心に関する次の文で比較してみよう。次の三つの文をエステラが述べたとしてみる。

「伊作は自民党の経済政策では日本の将来はよくなると信じている。」

「史門は激しい痛みを感じている。」

「私の心以外に他人の心が存在する。」

デカルト主義で考えるなら、エステラは他人の意識のなかで進行している事柄について主張している。したがって、彼女の言うことが正しいためには、彼女の主張することが実際にそれぞれの意識のなかで起こっていなければならない。文の真理条件(truth condition)を使ってデカルト主義者のこの主張を述べ直すと次のようになる。

最初の主張が真である iff もしエステラは自民党の経済政策では日本の将来はよくなると信じていたなら、そのエステラの意識の中にあることに完全に類似したことが伊作の意識の中にもある

二番目の主張が真である iff エステラが彼女自身の意識の中で直接にわかる、彼女が「激しい痛み」と呼ぶものに類似したものが史門の意識の中にある

三番目の主張が真である iff 何かが起こっている（エステラの意識の他の）意識が存在する

一方、行動主義的な見解ではエステラは他人の行動について上の主張をしている。したがって、行動主義的にそれら主張の真理条件を考えると次のようになるだろう。

最初の主張が真である iff 伊作の過去と未来の行動は自民党の経済政策に否定的に反応する行動性向として解釈できるような行動である

二番目の主張が真である iff 史門が激しい痛みの典型的な行動を示す

三番目の主張が真である iff 心をもった人に典型的に認められるのと少なくとも同じ程度の複雑さをもった行動がある

以上のことから、他人の心の知識についてのデカルト主義と行動主義の違いは明らかであろう。実際の私たちはいずれかの見解の基づいて首尾一貫した見方をしてはいない。両方の見方を適当に使い分けている。

（自己知識への行動主義的見解）

他人の心の知識は他人の行動性向の知識であるというのが行動主義である。これを自分自身の場合に適用すると、自己知識は自分自身の行動性向についての知識ということになる。このように考えることによって、自己知識のさまざまな特徴は説明できるのだろうか。特に、一人称の特権性を説明できるのだろうか。

デカルト主義での自己知識の特権性は、自らの意識の中で演じられることを直接に経験できるということに基づいて考えられていた。もし自己知識が自身の行動性向の知識に過ぎないのであれば、他人のほうがその人自身の行動性向をその人よりはうまく知ることができる場合がしばしばあることになる。これが示唆しているのは、一人称の特権性も、その人自身の心を知ることと他人の心を知ることの間にある違いも行動主義的用語ではうまく説明できないのではないかということである。自分自身の知識と他人の心の知識の違いは次の一般化が大抵の人の大抵の場合に成立することから行動主義的にも説明できる。

他人の行動の証人になる場合よりその人自身が自分の行動の証人になる場合が多い。

むろん、この命題が誤っている場合はある。しかし、大抵の場合、自分の行動はいやおうなしに眼に入ってくる。この命題には心の形而上学的性質はないが、それを使うことによって、一人称の特権性は説明できる。そこではデカルト主義で考えられた自己知識の直接性や特権性という特徴をあえて想定する必要はない。実際、このことは行動主義の利点と考えることができる。

（行動主義の自己知識への批判）

行動主義は自己自身の心的状態についての知識をその行動性向の知識であると主張するのであるが、行動に現れないような意識状態はどのように扱うことになるのか。痛みの場合がこの例となる。痛みの意識はどのような行動性向なのか。痛みと信念の場合は随分事情が異なっている。これは Ryle への批判と同じ批判である。

(3) 行動主義的でない理論の可能性

今まで見てきた二つの理論は心についての知識の典型を自己知識に見るか、他人の心の知識に見るかの違いにあった。そのため、基本とする心の知識のタイプとは異なるタイプの説明には困難がつきまわっていた。その結果は対照的である。いずれもその観点が狭すぎたのである。

デカルト的観点は自己自身の意識状態の知識は説明できるが、他人の心の知識は十分に説明できない。行動主義的観点は他人の心の知識は説明できるが、自己自身の意識状態の知識は十分に説明できない。

このような結果から、行動主義の長所を生かしながら、心の知識への行動主義的理解に伴う難点に陥らずに心の理論をつくりたくなる。このような動機で生まれたのが公言理論と呼ばれるものである。それは心の公共性という点では行動主義と同じであるが、心の知識を行動性向とは考えない。Wittgenstein に刺激を受け、Wright, Davidson, McDowell らによって展開されたこの説の特徴は、「自己知識(self-knowledge)」や「自分自身の心を知る」という表現はどのようなタイプの知識にも言及していないという点にある。私たちは他人の心について知識をもつことはできるだろうが、自分の自分自身の心への関係は表現されるほど認知的(cognitive)ではなく、「自己知識」と呼ぶものは各人がその心的状態を一人称の状態報告において表現しなければならない能力として理解した方がよい。これが基本主張である。

(公言理論 (Avowal theory) : 一人称の特権性の新しい説

行動主義の欠点である自己知識をうまく扱うことを目指すこの説は当然ながら、一人称の特権性についてのより優れた扱いを目標としている。一人称の特権性についての行動主義的欠陥は次の点にあった。ある人の一人称の意見がその行動とどのように合っているか観察する前に、その人の一人称の意見が真であることが正当化できるのはなぜかを行動主義は説明できなかった。説明できなかった正確な理由は、行動主義者はある人の心的状態の報告をその人自身の行動性向についての言明と解釈し、その報告のもつ真や偽を説明することができなかったからである。

この問題を解決する方法は、心的状態を主体に帰属させる言明はみなその主体の行動に関する言明に意味において同じであるという行動主義のテーゼを放棄することである。しかし、行動主義すべてを放棄したのではその長所も一緒に放棄しなければならない。そこで考えられるのは、次のようなことが保持できる仕方での放棄である。

命題的態度の帰属は人の行動にかなうが、命題的態度を帰属させる文は主体の行動を記述する文とは意味において異なっている。

命題的態度の主体への帰属が行動にかなうとはどのような意味なのか。もしある人に心的状態を帰属させる文が真とみなされることになるなら、その心的状態の帰属はその主体の行動(過去に言ったこと、過去に行ったこと、未来に言ったり行ったりすること)と矛盾することなく一致していなければならない。例えば、「伊作は今日は晴れると信じている」が真とみなされるなら、その信念を伊作がもっていると考えることは伊作の行動にかなっていなければならない。もし彼が外を見て傘をもって外出したとすれば、彼が晴れるという信念をもっていると考えることはその行動にかなっておらず、したがって、「伊作は今日は晴れると信じている」という文は偽とされなければならない。伊作の傘をもっての外出という行動はその文を真とみなすことができる帰属の目録を制約している。行動をそのまま真理条件に使うのではなく、真や偽のための制約として使うのである。これは行動主義とは一線を画した考えである。

では、一人称の報告がどのように真理値を取ることができるかを見てみよう。Crispin Wright は次のように言う。心的状態は公然たるものである。それらは基礎のない、特権的な自己帰属に服従する。この服従は心的状態の本質である。言い換えれば、自分自身の状態についてのその人の意見が特権的であるように、

信念、欲求、意図といった状態はそれらの本性から特権的である。すると、他人の特権を仮定することは、その特権が心的状態それ自身に植え込まれているゆえにアプリアリに正当化される。要するに、心的状態が公然たるものであるとは、自分自身の心的状態についてのその人の意見がその本性上特権的であると言うことである。

この主張に論点先取はないだろうか。この答えは求めたいものを既に仮定していないだろうか。なぜこのような公言仮説を受け入れなければならないのか知りたくなるだろう。これに対する Davidson の答えは、それは他人を解釈する際に必要となる仮定であるというものである。Wright は自己知識と他人の心に関するあらゆる性質を有意味にする唯一の仕方が公言仮説であると答える。

まず、Davidson の擁護を考えてみよう。史門の信念や欲求を明らかにしたいとする。彼の信念や欲求が何であるかの証拠は彼の言語的でない行動の中にある。しかし、言語的でない行動はいつも複数の解釈を可能にする。史門が黙って食器を取りに台所に行くとしてみよう。その行動に対してさまざまな信念や欲求が想像できる。さらに彼のその後の行動を観察することによっていくつかの信念や欲求は取り除かれるだろう。しかし、一般に一つだけを行動から選び出すことはできない。これを人の心的状態はその人の行動によっては決定されないということができる。これは、私たちが人の信念や欲求が何であるか決めようとするれば言語的でない行動以外のものが必要であることを示唆している。私たちに必要なものは、その人が何を信じ、欲するか言うことを取り上げることである。心的状態の行動による非決定性は、他人の信念を決定する際、信念について述べられたことに特別の地位を与えなければならないことを保証しているように見えることである。私たちは信念についての一人称の報告を天気予報を扱うようには扱うことができない。私たちは人の一人称の報告を推定的に真であると扱わなければならない。その本性上、人の心的状態はその人の言語的ではない行動から推論できなく、その人の一人称の報告に密接に関連したものとみなされなければならない。これは公言仮説と同じである。したがって、人の心が他人にも解釈できるものであるならば、心的状態は公然たるものでなければならない。

事実、私たちはデカルト主義と行動主義の誤りの結果として、この公言仮説を擁護する二番目の仕方を考えることができる。デカルト主義も行動主義も心に関する知識のある側面は説明することができた。しかし、すべての側面の説明はできなかった。この片手落ちを救う方法として公言仮説を考えることができる。

このように考えると公言仮説は自己知識と他人の心の知識の理論の基礎になるように思われるが、その答えは否定的である。公言仮説によれば、ある人のその人自身への関係は認知的ではない。つまり、知るとか知ることによって失敗することによって理解されるのではない。むしろ、表現的である。つまり、その人自身の心的状態を適切な種類の発言で表現するその人の能力によって理解される。ある人自身の心の事実に対するその人の関係は認知的ではなく、表現的である。すると自己知識のほうが他人の心の知識より適切に記述されることになる。というのも、犬が痛がって出す鳴声と私たちのそれに対応する表現は私たちのほうがはるかに念が入っている。表現能力は私たちのほうが数段上である。

(最後に)

この章でも前章に引き続いて否定的な結論しか得られなかった。しかし、悲観しないでほしい。日常生活での自己知識、他人の心についての知識を考えれば、その大半は不完全で、曖昧なものである。もし私たちが完全性や整合性をむやみに追求しないのであれば、上に述べられてきた考えはそれぞれ相当に通用する考えである。多分、私たちは哲学の理論よりはるかに巧みに上の考えを使い分けているのであろう。誰も自分や他人が一貫した態度で行動し続けるとは思っていない。

1980年代の後半から意識は熱い関心の的になり、心の哲学の議論の中心に位置するようになっていく。と同時に、神経・認知科学の先端研究の注目の対象となっている。William James 以来哲学や心理学から忘れられていた意識はどうして復活したのだろうか。鍵となる意識の問題はそれがどのように物理世界で生じることができるのかである。どのように意識経験が物理的な基盤から生じることが可能なのか。意識の謎は、宇宙の起源と並んで、最大の理論的謎である。この謎が経験的な研究によって解けるということは科学革命が起こることと同じことを意味している。三つの理由が考えられる。意識の謎とは実際に何なのか、そして何をもって解決とするのが明らかでないというのが最初の理由である。第二は、私たちが理解したいのは私たち自身の意識であるから、私たちは極めて強い意味で私たち自身に係わっている。それゆえ、哲学や科学の領域だけでなく、すべての人に関わる自己知識の問題でもある。最後は、上の二つのことから、もし革命がなされるなら、それは計り知れない変化を私たち自身、社会、文化に及ぼすことになる。

(意識の問題)

意識の研究者の多くはどのように主観性が生じたのかの説明と理解を見出そうとしている。そのために、知覚、記憶、言語、認識を理解しようとしている。しかし、その前に、あるものが意識的であるかどうかはどのように決められるのか。Robert Kirk は有機体が意識的であるためにもたなければならないものを考え、意識的かどうかの判定基準として 次のものを挙げている。

- (1) 意図的に使うために情報を集めることができる。
- (2) その情報に基づいて、活動を始め、それをコントロールできる。
- (3) 情報を評価・解釈し、それを使って何をするか決定することができる。

では、主観性はなぜ謎なのか。私たちが主観的に知る感覚や知覚がどのように生じるか生理学はまだ十分に語ってはくれない。私たちが知る現象的なものは物理的な世界には関係していないように見える。それら現象は私たちが自らの内につくりだすもので、物理的には存在していないように見える。私たちは紛れもなく物理的なものである。と同時に、自己意識はこの物理的なものから独立しているという考えも歴史的に引き継いできた。これがデカルトの伝統であった。そして、この二元論的な見解のもたらす問題に悩んできた。その悩みの中で、哲学者とは違って実証的な脳科学者は、意識は複雑な脳のシステムの不可避的な結果であることを実証しようとしてきた

(#) 意識に関する「容易な問題(easy problems)」と「難解な問題(hard problem)」

The “easy” problems of consciousness include the problems of explaining the following phenomena, among others: (i) the ability to discriminate, categorize, and react to environmental stimuli; (ii) the ability to produce verbal reports of mental states; (iii) the ability of a system to scan its own internal states; (iv) the difference between wakefulness and sleep; (v) the ability of a cognitive system to integrate information and exploit it in the control of behavior. All of these problems are associated with the concept of “consciousness.” Good solutions to the easy problems will probably take a century or two of hard work in cognitive science and neuroscience.

The “hard” problem of consciousness is the problem of experience. Perhaps the most central fact about our mind is that, as Thomas Nagel has put it, *there is something it is like to be a conscious agent*. When we visually perceive the world, we do not just process information; we have a subjective experience of color, shape, and depth. We have experiences associated with other senses, with bodily sensations, with mental imagery, with emotion, and with the stream of conscious thought. This is the central fact about the mind, but it is also the most mysterious. Why should a physical system, no matter how complex and well-organized, give rise to experience at all? Why is it that all this processing does not go on “in the dark”, without any subjective quality? Right now, nobody has good

answers to these questions. This is the phenomenon that makes consciousness a real mystery.

Why are the easy problems easy? Because they are all problems about the explanation of functions. To explain phenomena such as (i)-(v), we need only give an account of how it is that brains do certain sorts of things in the control of behavior. And to explain the performance of functions, we need only find a mechanism that performs the function. This is the sort of thing to which standard methods are perfectly suited. Once we find a mechanism that performs the relevant functions, and show how it does its work, we have explained everything about these phenomena that needs explaining.

Why is the hard problem hard? Because it is not just a problem about the explanation of functions. To see this, note that even when we explain the performance of all the relevant functions there may still remain a further issue: why is the system that performs these functions conscious? Why should the performance of these functions give rise to conscious experience at all? This is a further question that an account of the performance of the functions does not answer.

Dennett は意識の世界が次の三つのものを含んでいると考えている。

(1) 私たちの外部の世界についての経験（光景、音、臭い等）(2) 私たちの内側の世界についての経験（幻想、独白等）(3) 情緒の経験（怒り、身体の痛み、飢え、渇き等）

(1)と(2)は表象内容をもっているように見えるが、(3)はそのような内容を欠いている。怒りは純粹経験であり、何かについての経験ではない。「ものが私たちにどのように現れるか」はしばしば「純粹経験」、あるいは「生の感覚」と呼ばれるが、以後は感覚質(qualia)とも呼ぶことにしよう。白いシャツをがどのように見えるかということを知ることによって「白」の感覚質がどのようなものかあなたは知っている。しかし、その白さはシャツの本当の性質であろうか。通常、シャツがどのように見えるかからシャツが本当に白いことを導き出すことはできない。ものの現れ方は私たちの心、心の構造に深く依存している。それに対して、ものの実在の仕方は私たちには何か重要な点で依存しているようには見えない。

私たちが現象的意識をもつということは、私たちであるような何かがあることである。ある心的状態が意識的であるとは、そのような状態であるような何かがあることである。黄色を見るときはどのようなことかをあなたは知っている。このような言葉遣いを使って Thomas Nagel は、世界の純粹に物理主義的な説明を超える、何かであることを経験するとはどのようなことかを知ることには特有のものがあることを論じた。

(#) Nagel の議論に入る前に、現象的意識、感覚質についての物理主義的立場からの典型的な対応を見ておこう。それは Dennett の主張にもっともよく表れている。

In *Consciousness Explained*(1991), Daniel Dennett sketches a theory of consciousness that he believes is consistent with current findings in cognitive science, and which provides a "first-approximation" explanation of the nature of consciousness and conscious experience. Dennett claims that cognitive science can in principle approach closer and closer to an understanding of the nature of consciousness, by extending and continuing current research. Dennett provides a sketch of a theory of consciousness, which he calls the "multiple drafts" model. In chapter 12 "Qualia Disqualified," Dennett argues that views of the mind are mistaken that hold that there are mental entities called "qualia," which are thought to be "intrinsic qualitative features of subjective experience" that lie beyond the explanatory reach of third-person, "objective" cognitive science. Putative candidates for "qualia" include experienced qualities such as the smell of a rose, the "look" of deep blue, the sound of a trumpet in the distance on a windy day, and so on. According to some philosophers, science can neither explain why qualia exist, nor explain why particular qualia are attached to particular experiences. Dennett, by contrast, argues that if qualia are defined to be things that play no functional role in cognition, then the most reasonable conclusion to draw is

that there are no such things as qualia. (既に Daniel Dennett という名前があちこちに登場した。彼は新しいタイプの随伴現象論を精力的に展開している。意識は「難解な問題」どころではなく、そのように見えるだけの、脳の随伴現象、副作用であるというのが彼の考えである。)

8.1 Thomas Nagel の物理主義に反対する論証 : What is it like to be a bat?

彼が物理主義に反対する論証を示したというのは正確でないかもしれない。彼の主張は、物理主義がどのようにして真であり得るのかわからないと述べる方が正確である。彼によれば、有機体が意識経験をもちという事実は、その有機体であるような何かがあることを意味している。これは経験の主観的性格と呼ばれる。意識経験があるところには常にその経験をもつ個体が存在し、その個体はその経験に一意的な関係にある。

(#) 上のことを別の形で表現すれば次のようになる。

It is not just that there is something that it is like to smell garlic but that there is something that it is like *for subject S* to smell garlic, which might be different from what it is like *for subject T* to smell garlic.

これに対して、物理的な状態は主観的な性格をもっていない。酸素は原子番号 8 であるがそれは違う酸素原子でも同じままである。つまり、物理的な状態は客観的である。もしそうだとすると、意識は物理主義に問題を投げかけることになる。意識は本質的に主観的現象であり、物理世界は客観的な現象であるから、私たちはどのようにして意識が物理現象に還元されるか、あるいは物理現象に関係しているか理解する術をもっていないことになる。純粹に物理的対象、出来事、状態、過程を使ってどのように主観性が説明できるというのか。彼が使った例はコウモリである。私たちはコウモリが意識経験をもちていることを疑う十分な証拠をもっていない。たとえコウモリに主観的な経験があるとしても、コウモリの意識経験の主観的性格は私たちにはわからない。ソナーを使って空を飛ぶとはどのようなことかは私たちにはわからない。私たちにあってコウモリのように行動することがどのようなことかが問題なのではなく、問題はコウモリにとってコウモリであるとはどういうことかである。(この問いは陳腐なのかもしれない。コウモリが意識をもっていないならば、端的に陳腐である。) Nagel の論証を論理的に述べるなら、次のように整理できる。

- (1) 主観的な経験が存在する。
- (2) もし物理主義が正しいなら、主観的経験についての物理的な説明がなければならない。
- (3) しかし、すべての主観的経験が本質的に単一の観点に結びついており、したがって、主観的対象の十分な概念はその主観的観点の採用によってしか得ることができない。
- (4) 客観的な物理理論は主観的観点を採らないので、経験の主観的性格を説明することができない。
- (5) 経験の主観的性格がどのようにして有機体の物理的な操作の中に実現されるかは謎である。
- (6) 物理主義は私たちが理解できない立場である。なぜなら、私たちはそれが現在どのように真であるかについて何の考えももっていないからである。

この論証から何が結果として得られるのか。まず、私たちの知識の限界である。他の動物の主観的経験について私たちには知ることができない。さらに、そのような事実は物理的事実ではない。(したがって、物理的事実と主観的事実は異なることになる。) それゆえ、物理主義的枠組みで扱うことができない事実が存在し、物理主義は誤りということになる。

これに対する哲学者たちの反応はさまざまである。Owen Flanagan は Nagel に賛成し、物理主義の誤りを認めたが、Martin Davies は Nagel が「事実」の曖昧さを無視したために物理主義が不完全であることでは

なく、私たちが物理的なものと心的なものを異なる仕方で考えるという、物理主義と両立することを示したに過ぎないと批判した。また、Dennett は三人称の、客観的な知識でもコウモリであるとはどのようなことかに対して実に多くのことを知ることができると Nagel を否定した。

8.2 Collin McGinn の物理主義反対の論証（後述の The Hard Problem も参照）

McGinn と Nagel の論証には多くの共通点がある。McGinn は物理主義者である。だが、彼は意識が脳にどんな役割をもっているか発見できないと思っている。彼の論証を、すべての出来事は物理的な出来事である、あるいは物理的な出来事によって説明できるという主張であるとするならば、物理主義に反対するというよりそれと両立可能である。彼は次のように議論を展開する。

- (1) 意識を説明する脳の物理的性質がある。
- (2) しかし、私たちの心の構造（特に概念形成能力）はその脳の物理的性質を理解することができない。
- (3) したがって、脳がどのように意識を生み出すか私たちにはわからない。

彼は(1)が正しいか否かより、私たちはそう信じていると述べるだけである。興味あるのは(2)である。脳や意識の研究を続けるなかで、私たちは意識を説明する物理的な脳の性質をどのように確定できるのか。意識を直接に内観によって調べることによってか、あるいは脳を探索することによってかであろう。しかし、内観は脳や神経機構については何も教えてくれない。したがって、脳を調べるほかなくなる。脳の直接観察によって何がわかるだろうか。身体のほかの部分の特徴と同じ点、異なる点がわかるだろうが、現象的意識の主観的性格はどこにも見出せないだろう。脳の直接観察の結果からそのような性格を間接的に導き出すことはできないだろうか。これもできない。というのも、説明したい主観的性格は理論的前提と観察結果のいずれにも含まれていないからである。脳の理論化（自然化）には意識は含まれていない。脳の性質が確かに意識を説明する。しかし、私たちが知識を得る基本的なメカニズム（内観、観察、そして理論化）はそれがどんな性質かを言い当てることができない。私たちのメカニズムは意識を知るには単純に不十分、不適切なのである。意識がどのように生じるかは原理上知ることができない。脳は、それがスマートである以上に複雑なのである。

8.3 Frank Jackson の物理主義反対の論証：The Knowledge Argument

Dennett によれば、私たちが色と呼ぶものは自然種に対応していない。にもかかわらず、私たちはバラは赤いと思うし、それを経験できると思っている。一方、科学は可視スペクトルは連続で、私たちが経験するような色の違いが予め組み込まれていないことを示している。さらに、私たちとは違う生物はスペクトルを違ったように区分することも知っている。すると、ある対象の色は何かと問われた場合、その答えに対応する科学的な事実はないことになる。しかし、色は幻覚や夢とは違って「主観的」ではない。人が異なれば異なるといったものでもない。では、色は一体何なのかということになる。あるスペクトルの範囲が生存と繁殖のために視覚システムによって識別される必要があったので、色を見るという能力が自然選択によって進化してきた。ある色と結びついた内部のシグナルが経験の質的な内容である。これが Dennett の答えである。感覚質はシステムの中で機能的な役割を演じる内部状態に還元されている。Dennett だけでなく、このように考えるのが正統的機能主義者である。しかし、感覚質は内部状態に還元できないと考える者もいる。機能的でない、感覚質の存在を擁護する論証の代表が逆転スペクトルの論証と Jackson のメアリーの論証である。既に言及した逆転スペクトルの議論(The Inverted Spectrum Argument)を復習しておこう。A と B は機能的に同型であるが、B の方に逆転が生じ、A が赤のものを見る時にもつ経験と同じ経験を B は緑のものを見る時にもつとしてみよう。そのような感覚質の逆転は調べようがない。B は「赤」と

いう語を A が指示するのと同じ対象に用いるからである。もしこのようなことが可能であれば、機能主義は正しくないことになる。A と B の内的な機能体制は同じでありながら、二人の心的生活は（その感覚質が異なるゆえに）異なるからである。さらに、科学が内部状態の因果的、機能的な性質だけを研究するのであれば、色に代表されるような感覚的に経験される質を説明することはできない。これが逆転スペクトルの論証であった。前置きが長くなったが、Jackson の論証に話を移そう。

Jackson は次のような思考実験を考案した。メアリーは完全に白黒の世界で生まれ、育てられ、有能な神経科学者に成長した。彼女の専門は視覚の神経生理学であった。色の知覚に関するあらゆる物理的な事実を彼女は学んだ。物理世界の見た目の完全さは、物理的な出来事を引き起こすが、それ自身は物理法則に従わない、独立の心の領域など存在しないことを示している。これが正しければ、世界に関する物理的な知識は世界に関する知識のすべてであることになる。物理主義の主張によれば、色の知覚に関するあらゆる知識はメアリーのもっている知識に含まれていることになる。そこで、Jackson は次の問いを出す。メアリーが白黒の世界から開放されて普通のカラーの世界に身をさらした時、彼女は「色」がどのようなものか学ぶのではないか。もし、そのような学習によって新しい知識が得られとしたり、それは物理主義的な知識ではない。というのも、彼女は既に色に関する物理的知識をすべてもっているはずだからである。この思考実験を論証の形に整理すると次のようになるだろう。

- (1) メアリーはモノクロの世界で物理的な世界についてのすべての知識を習得した。したがって、色の経験はないが、色についての物理的な知識はもっている。
- (2) 物理主義は、世界についての知識は物理的な知識であると主張する。
- (3) しかし、メアリーは色の感覚質を知らない。
したがって、物理主義では知ることができない知識が存在する。

この反物理主義の主張に対してどのように物理主義者は返答するのだろうか。以下に四つの返答を挙げてみよう。

- 1 カラーの世界を見ても、メアリーは新しい知識を獲得しない。
- 2 彼女が獲得するのは新しい知識ではなく、別のタイプの知識（こつや技能）に過ぎない。
- 3 彼女は新しい事実的な知識を獲得する。しかし、それは既知の知識を新しい様式で獲得するに過ぎない。
- 4 彼女は新しい事実的な知識を獲得する。しかし、それは物理主義と両立する精妙な仕方では個別化した場合に過ぎない。

これらはいずれも Jackson の基本的な姿勢、つまり認識論的な問題として物理主義の欠点を暴くという姿勢をそのまま認め、反物理主義的な結論は避けようとするものである。まず最初のものを考えよう。Churchland、Dennett がその代表で、未来の神経科学は今以上に進歩し、色の質や主観的経験を明らかにしてくれるので、メアリーは初めて色を経験しても驚くことはない。私たちは現在の物理主義的知識とその未来がどのくらい隔たっているかを知らない。これに対する Jackson の返答は、知識論証は物理主義が誤りであるという論証ではなく、物理主義が誤りである可能性が高いことを示唆しているだけであるというものである。

2 番目の反論を Churchland は次のようにまとめている。

- 1 メアリーは部屋を出る前に他人について知るために必要なすべての物理的なことを知る。
- 2 メアリーは部屋を出る前に他人について知るために必要なすべてのことを知っていない。

3 それゆえ、物理主義的でない他人についての知識が存在する。

1と2の「知る」を考えてみよう。物理的なものの知り方は文によって表現でき、真か偽かの判断がつくような知識である。それに対して、「知る」の異なる解釈がある。それは命題で表現できないような知識で真でも偽でもないような知識である。このような知識が2に含まれていれば、この推論全体は正しくないことになる。別の2に対する反論は Lewis と Nemirov によるもので、メアリーは新しい経験によって獲得したものがあがる、それは事実に知識ではなく、コツや技能(know-how)に過ぎないというものである。

以上私たちは三つの論証を眺めてきた。いずれも自然主義的な見方では扱いにくい、扱うことができないものを使っての論証であった。これら論証は積極的に意識や感覚質が何であるかを主張しているのではない。意識や感覚質に関する常識的な特徴を使って、それらが自然主義的に説明できないことを主張している。哲学的な論証としては興味深い、これらの論証に答えることができたとしてもそれで意識や感覚質がわかるというものではない。現象的経験、主観的経験、主観的観点、感覚質、これらは私たちの心の特徴である。そのどれもが意識の理解には不可欠である。それらすべてを考えることはここではできないので、感覚質について以下に考えてみよう。

8.4 感覚質(Qualia)

私たちの感覚や経験は様々である。そのような感覚や経験をもつのは私であり、私がある経験をもつ状態(であるようなもの)が存在する。哲学では私たちの心的生活を内観的に知ることのできる、現象的な側面を感覚質と呼んできた。この広い意味では感覚質が存在しないというのは困難である。だが、どの心的状態が感覚質をもつのか、感覚質はそれをもつものの内在的な性質なのかどうか、感覚質はどのように物理世界に関係しているのか等に関して多くの意見の対立がある。現在、多くの哲学者が活発に議論を展開している。

(「感覚質」の使用法)

感覚質は、

- a 内観によってわかる
- b 経験の志向的内容が変わらなくとも変わることができる
- c 対象の直接に見ることのできる性質(例えば、色)の心的な対応物である
- d 経験の現象的性質を唯一決定するものである

のような性質を仮定されている。感覚質をもつ心的状態は多分次のような状態である。

- (1) 知覚経験 (2) 身体感覚 (3) 感情、情動 (4) 気分、気持

(感覚質は還元できないか)

感覚質の文献には思考実験が溢れている。既に見たメアリーの論証はその一つである。その思考実験の結果として、哲学者の中での一般的な説明は、色に結びついた主観的、現象的な性質の領域があつて、メアリーはそれを知ることになるというものである。無論、この説明は自然主義者には使えない。ここではゾンビについての反還元主義的思考実験を考えてみよう。

正常な人間が分子レベルで同じように複製され、しかし、現象的な意識は全く欠いているのが哲学のゾンビである。私が登山に成功し、頂上で水を飲みながら周りの風景を見る時、私はさまざまな視覚的、聴覚的な経験をしている。しかし、私の複製のゾンビにはそのような経験は一切ない。ゾンビは私と物理的に

全く同じにつくられているので、同じ環境におかれれば同じ内的な心的状態をもつはずである。彼は私と同じなので、同じ物理的な刺激に対して、同じ内部処理をし、同じ行動的な反応を示すだろう。現象的でない心的状態が機能的な状態であるという仮定のもとでは、私の複製であるゾンビは私と同じ信念、思考、欲求をもっている。この哲学的なゾンビは感覚質についての物理主義的見解に重大な問題を投げることになる。

まず、ゾンビが形而上学的に可能であれば、簡単な論証によって現象的な状態は内的、客観的で物理的な状態とは違っていることを示すことができる。そのような客観的で物理的な状態 P があるゾンビの中に現象的な状態 S なしに生じることができるとしてみよう。直観的に、 S は S なしには生じることができない。痛みはそれなしに生じないだろう。したがって、ライブニッツの法則から S は P と同じではない。次に、ある人が私に微視的なレベルで同一で、同じ環境におり、しかし、現象的な経験は一切ないならば、経験や感情についての事実、感覚質についての事実は微視的なレベルでの客観的事実によっては決定されないことになる。これは物理主義者が容認できることではない。

(感覚質の説明上の溝)

どのような現象的な経験をしているかは内観によってわかる。その時に脳の中で客観的に何が進行しているか不完全ではあっても把握できる。この場合の両者の関係はどうなっているのか。どのように客観的な知見が増えても、今まさに経験していることとそれについての脳の内部の知見の間には大きな隔りがある。この説明上の隔り(explanatory gap)をどのように埋めたらよいのだろうか。この溝は埋めることができないという立場は、経験や感情は還元することのできない主観的で、非物理的な性質であることを認めることである。(性質二元論の主張を思い出そう。Chalmers がその代表である。)これは物理主義から離反することではなく、物理的性質や状態には還元できない主観的なものがあるということを示しているだけであるとするのが Searle の立場である。他の人々はこの溝は将来埋まると考える。現在は主観的な観点と客観的な観点を結び合わせる概念を欠いているので溝があるように見えるだけに過ぎない。このように考える者の中でも特徴的なのは既述の McGinn の立場である。溝は埋めることが可能であるが、私たちや私たちと遺伝的に同じ子孫によっては不可能であるというのが彼の立場である。

別の立場で近年賛同者が多いのは、溝は本物で埋めることはできないが、そのことが意識の本性或機能主義的な理論に何の結果ももたらさないという立場である。この見解では、溝は経験や感情と物理的、機能的現象の間にある分岐、あるいは主観的なものと客観的なものの分岐を示しているのではない。溝は単に現象的な概念の特別な性質から出ているに過ぎない。この概念が世界を二つに分けるといった考えを誘導しているだけである。この見解の一つの形は、現象的な概念は内観を通じての現象的な状態に適用された指標的概念でしかないというものである。別の形は、現象的な概念は他のものと異なる特別の一人称概念であるとする。

(感覚質の表象説)

ものの見え方は内包的に語られる。紫色のものはすべてそしてそれだけが毒をもつような世界に私が住んでいても、紫色に見えるものが私の眼には毒をもつようには見えない。(毒をもつように見えることは普通の人にはない。誰にも砂糖は四角には見えない。)足が痛いと感じても、足をもっている必要さえない場合がある。私の痛みは幻肢の痛みかもしれない。このような事実は感覚質が表象的な性質であることを示唆している。もし感覚質が表象的なら、次の質問が待ち受けている。経験の表象的内容のどの側面がその現象的な性格に関連しているのか。明らかに内容のすべての側面ではない。私とあなたが携帯電話を同じ角度から見るとしよう。私にはそれが何かわからず、あなたはそれが携帯電話であることを知っている。二人の経験は明らかに異なっているが、二人に携帯電話がどのように見えるかは同じである。これを表象

主義者は次のように考える。対象の3次元の表面、端、表面の形等を私たちの経験は同じように表象するからである。私たちが共有する表象はDavid Marrの2 1/2-D sketchの内容に似たものである。この内容はおそらく非概念的なものである。それは初期の感覚処理の出力とより高次の認知処理の入力をつくっている。表象主義者はこのレベルの内容が感覚質が見出される場所であると主張する。DretskeやTyeはそうに考える。(McDowellはこれに反対する。)感覚質についての表象主義はその表象内容に関して外在主義的である。ある経験が表象するものは少なくとも部分的には外部環境にある要因によって決定される。したがって、微視レベルまで同じ双子はその経験の表象内容に関して異なることができる。表象主義を広く採用すれば、意味と同じように感覚質は頭の中にはない。これは反デカルト的である。感覚質は主体が直接に気づく内的観念の内在的性質ではない。むしろ、それらは個人と環境の間の外在的關係によって固定される性質である。表象主義は、感覚質はある表象的な性質と同一である、という感覚質についての同一性テーゼである。

表象主義への反対は反例でよく与えられる。同じ表象内容をもつが、異なる現象的性格をもつ経験がその例になる。Peacockeらはこのような経験を挙げている。逆転スペクトルも同じ例となる。ここでは逆転地球を考えよう。逆転地球は仮想の天体で、色は地球の補色に逆転している。空は黄色、草は赤色、熟れたトマトは緑色等々である。逆転地球の住人は空は黄色と考え、草の赤を見ている。しかし、彼らは空を「青」、草を「緑」、熟れたトマトを「赤」と呼ぶ。それは私たちと同じである。見かけ上は逆転地球の住人は私たちと変わらないことになる。Blockのものと話では、ある夜あなたが眠っている間に他の星の科学者が色逆転のレンズをあなたの眼に入れ、あなたを逆転地球に連れて行く。そこであなたは逆転地球であなたに対応している「あなた」と入れ替わる。目が覚めたあなたは色逆転レンズのお陰で何の違和感もない。あなたは以前の自分と同じだと考える。あなたが空を見上げるとき、それはあなたにとってどのようなものであるかというのは、地球上でどのようなであったかということと全く同じである。

しかし、あなたが逆転地球の生活に慣れ、言語に通じてくると、あなたの志向的内容はその住人と一致するようになり、空は黄色であると信じるようになる。同じように空が黄色であるという視覚経験をもちようになるだろう。したがって、以前のあなたの内的状態に対して志向的に逆転した内的状態にあなたは従うようになるだろう。一方、あなたの経験の現象的な側面は変化しないままである。このことから、外在主義的な表象主義は誤っているという結論になる。(この結論は外が同じなのに内が異なるという点でPutnamの双子地球と反対の例になっている。)

これに対する表象主義者の反論は経験が関与している限り色に関する正常の探索に何の変化もないというものだろう。「正常な」は目的論的、非目的論的の二つの意味をもっている。もし経験が正常に探索するものは自然がそのように探索するようにデザインしたものなら、地球から逆転地球への環境の変化は正常な探索に何の相違ももたらさないだろう。あなたの種が進化してきた設定で青を探索するようにあなたの種にデザインされた感覚状態は、たとえ逆転地球でそれが黄色いものを見て引き起こされるとしてもそのことを続けるだろう。

(この反論に対してさらに反論が続くというのが現状である。)

8.5 The Hard Problem (Case Studies of McGinn and Penrose)

easy problemとhard problemの区別については既に言及した。ここでは困難な問題について哲学のケーススタディを行ってみよう。意識を扱った二人のケーススタディと考えてもらえればよい。意識は曖昧な言葉であるが、簡単に研究できる特徴をもっている。しかし、同時に既存の考え方や道具立てでは研究できそうにもないように見える特徴をもっている。

意識の問題で真に困難なのは経験の問題である。思考や知覚は情報の処理であると共に、主観的な側面を

もっている。この主観性については既に多くを論じてきた。有機体が経験をもつことは否定できない。私たちの認知システムが視覚や聴覚の情報処理に携わっている時、私たちは視覚経験や聴覚経験をもつのはどうしてなのか。心的イメージを楽しみ、情動を経験するようなことがなぜあるのかをどのように説明したらよいのだろうか。経験は物理的なものから生じると広く認められているが、どうして生じるのか、どのように生じるのかは誰も知らない。物理的な情報処理がなぜ豊かな内的生活をもたらすのか。

8.51 意識は私たちの理解を超えたものか？(Colin McGinn のケース)

私たちは本当に意識について理解することができるのだろうか。意識の謎は理解するには余りに深いと考える代表が Colin McGinn である。私たちの脳は進化の産物であり、したがって、認知に関して限界をもっている。認知に関して完全に適応しているわけではない。ネコやネズミが量子力学を理解できないように、人間は心とものの関係を理解できない。意識は永遠に人間には理解できないものである。これが McGinn の考えである。この悲観論をもとにそこから反自然主義的な立場を取るという方向に彼が進まないのは Searle の場合と同じである。以下に、彼の悲観論を意識が空間的でないという点を中心に詳しく吟味してみよう。

意識はどこにあるのか。デカルトによれば意識は空間には存在しない。空間に場所を占めるとするのは心の存在様式ではない。したがって、心的状態を認識する仕方は知覚を通じて知る際にどのような空間的概念を使っても把握することができない。物理学には観察できない対象が多く存在するが、それと同じように心的状態を考えることはできない。原子が肉眼で観察できないのと同じ意味で意識は観察できないのではない。単に観察できないどころか、それを空間的な対象としてどのように考えればよいかさえ私たちは知らない。したがって、技術的に知覚できないのではなく、原理的に知覚できないのである。意識が延長をもたないというデカルトの直観は今でも私たちの心的なものについての日常的な概念の核心部分を占めている。この意識の特徴は空間的な扱いを拒絶することになる。

もし意識が空間的でないなら、それはなぜこの空間をもつ世界に生まれ出ることができたのか。意識も進化の過程の中で生み出されたものだとしてみよう。すると、それは宇宙や生命の進化と違って、意識でない空間的なものから空間的でない意識への進化であったろう。これは大変神秘的に見える。これが意識に関する空間問題(the space problem with respect to consciousness)である。この問題には歴史的に二つの主要な対応があった。一つは、意識は物質から生まれたという前提を否定することである。意識は物質から独立したものであると考えれば、デカルトの立場、伝統的な二元論になる。他は、意識が非空間的であるという前提を否定することである。意識状態が脳状態でしかないなら、それは古典的な唯物論になる。この二つの立場に対して第三の途が考えられる。それは次のように述べられる。物質的な発現を保ちながら、通常の非空間的な意識概念を否定しないことである。脳の行うことを説明しようとする、脳は現在の物理学で認識された空間的な性質だけでは十分ではなく、脳はそれ以上の性質をもたなければならない。この性質は今の科学では捉えることができない。特に、意識の発現に関してはそうである。したがって、意識は科学にとって端的に謎となる。心身問題を解決するには少なくとも空間の新しい概念が必要である。空間がどのように構成され、構造化されるかを考えなければならない。これは単に脳生理学の問題ではなく、物理学まで及ぶ問題である。

科学史の中での空間概念、日常生活での空間概念は現在の物理学での空間概念とは大きく異なっている。したがって、物理空間は人間の知覚 - 行動システムに与えられた空間とは大きく異なっているに違いない。意識はこの世界のどこかに局在してはいないが、この世界の外にあるわけでもない。この意識と空間の特異な関係を記述できるような理論を私たちはもっていない。だが、意識と空間は自然主義的仕方で結びついているはずである。ここに空間問題を解く鍵がある。

人間の認知システムが特定の構造と操作様式をもつことは、その能力を超えて存在するものがあることを

示唆している。実際、Chomsky も私たちの認知システムが特定の目的をもったモジュールの集合であり、その目的以外のものに対しては働かないと考えている。言語能力はそのような代表例である。言葉で絵を描くことはできない。これは科学的な知識の能力についても同様であると彼は考える。それらも生物学的な基礎をもつ、偶然的な認知構造の集合であって、進化の過程で私たちに備わるようになったものであると考える。これが正しいとすれば、自然の世界に関してそのすべてを理解する能力があるとは信じがたいことになる。私たちの認知システムは私たちが生存するのに必要な事柄に関してだけ感知するが、関係のない事柄に関しては無力であることになる。

このような一般的な考察から、私たちの認知能力を超えるような問題がある、最初から私たちに解決できない問題があるということになる。上述のデカルト的な空間性を思い出すなら、私たち人間の世界は基本的に空間的である。Strawson はこの空間の役割に注目した。個別と普遍の区別、つまりは主語と述語の区別は空間的な違い、区別という考え、経験に基づいているというのが彼の主張であった。 x と y が普遍 P の異なる例であるとみなされるのは x と y が異なる場所にあるからである。したがって、命題という概念そのものが空間的な差異という概念を前提にしているのである。私たちの思考の構造全体が空間という概念に基づいている。Strawson の上の主張が正しければ、空間的でない意識は思考システムで扱えないことになる。

8.52 量子的な意識 (Roger Penrose のケース)

Roger Penrose によれば、意識は量子世界と古典世界の連結である。量子力学と意識が連結しているという洞察は観測行為が量子的な出来事に効果を及ぼすという原理に基づいている。観測行為は究極的に意識的な観測者を含んでいる。ここに意識と量子力学の結びつきがある。Stuart Hameroff との共同研究で、彼は量子状態から古典状態への転換は microtubules と呼ばれるタンパク質内部で生じるという理論を提唱している。脳の microtubules はこの転換を実行しながら、時間の流れとともに意識的思考の流れを起こす経験をもたらす。これが彼の積極的な意識についての考えであるが、彼の研究には(1)人間の心は機械によっては捉えることができない、(2)意識は量子力学を適用しなければわからない、という相反するように見える主張がある。(私は実際この二つの主張は相反していると思う。)まず、その消極的な主張を見てみよう。彼は *Shadows of the Mind* においてゲーデルの不完全性定理を使って「人間の思考は計算不可能」であるという結論を導き出した。実際には二つの論証があるが、ここでは最初のものを取り上げよう。ちなみに、彼の証明に関してはいくつかの難点が指摘されている。

(ゲーデルの結果をを利用した論証)

Gödel の第 1 不完全性定理は、算術を形式化するのに十分で、かつ健全と仮定される形式システム F において、真ではあるが証明できない算術の文が存在する、ということを主張している。ここでは「形式化する」とは「完全に正確にする」と考えておいてよい。「真」、「証明可能」は論理学での通常の意味である。システム F が健全とは、 F で証明可能などのような文もモデルにおいて真になることである。第 1 定理は証明と真理の概念の間に違いがあることを明らかにしている。

Gödel は次のようにしてこの結果を得た。まず、Gödel numbering と呼ばれるコードによって「私は証明可能ではない」ことを述べた算術の文を F において構成する方法を示した。この文を $G(F)$ と呼ぼう。この文が偽であると仮定する。すると、 F が健全であるという仮定に矛盾して、 $G(F)$ は証明可能になる。したがって、それは真で、それはまさに証明不可能である。Gödel の第 2 不完全性定理は、もし F が無矛盾であれば、Gödel numbering のコードを通じて F の無矛盾性を主張する F の文 $CONSIS(F)$ は F において証明可能ではない、ということを主張している。

Penrose は第 1 不完全性定理から、私たちの推論能力はどのようなコンピュータの推論能力をも必然的に

超えるという結論を導き出そうとした。もし Penrose が正しいならば、強い AI (前出を参照) の企ては放棄されなければならない。つまり、私たちの推論能力をコンピュータによってコピーすることはできないことになる。Penrose は次のように論じる。 F を私たちの推論能力を捉えることのできる形式システムとする。すると、次のように論証できる。

主要論証

- (1) 私たちは $G(F)$ が真であることを決定することができる。
- (2) F は $G(F)$ が真であることを決定することができない。
- (3) それゆえ、私たちの推論能力は F の推論能力と同じではない。

(1)の論証

(a) 私たちは Gödel の第 1 不完全性定理が真であることを決定することができる。それゆえ、次の文が真であることを決定することができる。

もし F が健全であれば、 $G(F)$ は真である。

- (b) 私たちは F が健全であることを決定することができる。
- (c) それゆえ、私たちは $G(F)$ が真であることを決定することができる。

(2)の論証

Gödel の第 1 不完全性定理によって、 F は $G(F)$ を証明することができない。しかし、 F にとって何かが真であることを決定するにはそれを証明する以外にない。

以上のことから、Penrose は「人間の心は機械によっては捉えることができない」という結論を出すことになる。

(#) 以下に Penrose の論証の誤りを示す論証を挙げておく。

(Preliminaries)

A provability predicate P for a formal system F is a predicate satisfying the following conditions:

- (1) If a sentence A is provable in F , then so is the sentence $P([A])$ (Here $[A]$ is the gödel-number of A .)
- (2) It is provable in F that: if $P([A])$ and $P([A \text{ implies } B])$, then $P([B])$.
- (3) It is provable in F that: if $P([A])$, then $P([P([A])])$

Löb's Theorem: For any sentence A of F , if it is provable in F that: $(P([A]) \text{ implies } A)$, then A itself is provable in F . In fact, the provability of the implication $(P([A]) \text{ implies } A)$ is a necessary and sufficient for the provability of A itself.

Corollary: Any formal system in which the implication $(P([A]) \text{ implies } A)$ is provable for every sentence A , is inconsistent!

Now look back at Penrose's argument. Notice that it makes use of the notion: "We can determine that ...is true". Abbreviate this "WCD" (for "we can determine that"). So the notation "WCD(x)" means "we can determine that the sentence with gödel number x is true." Question: Is "WCD" a provability predicate? This is not a precise mathematical question and it can't be answered with mathematical certainty, but it is reasonably clear that the answer is "yes." In fact, however, for our purposes, WCD does not have to meet condition (3). So we can ignore it. Penrose makes direct use of (2) when he argues that since we can determine that: soundness of F implies that $G(F)$ is true, and we can determine that: F is sound, it follows that we can determine that: $G(F)$ is true. As for (1), remember that Penrose is assuming that in principle we can determine the truth of any mathematical truth even if we can't always do that by the method of proof. So WCD does satisfy condition (1). Now, finally, Penrose also

assumes that whatever we can determine to be true is true. So he has it that: $WCD([A])$ implies A , for every sentence A . It follows by the corollary to Löb's Theorem stated above, that the system of reasoning Penrose is using to mount his Gödelian argument is inconsistent.

次に Penrose の二つ目の主張である量子力学と意識の関係に話を転じてみよう。Penrose は実に丹念に古典力学による脳の研究はうまく行かないことを示していく。その結論に賛成して神経活動の研究には量子力学を用いるというのではなくとも、Penrose の量子的な考えには興味深いものがある。それは量子力学と観測者の役割である。まず、量子力学と脳の活動の関係を、次に量子力学的な形而上学と主観的意識を考えてみよう。

(1) 量子力学と脳の活動

量子力学の実験においていつも問題になるのは波束の収縮である。彼はこれを R 過程と呼び、それが実在的なものかどうか検討している。波束の収縮は微視的な原子と巨視的な脳の間の中間的なサイズの領域で起こる量子重力効果によって支配されているというのが彼の信念である。この解釈を脳に適用するとどうなるのか。ニューロンが状態 A (発火している) と状態 B (発火していない) の重ね合わせの状態にあるとしてみよう。これは $a|A\rangle + b|B\rangle$ のように書ける。 $|a|^2$ と $|b|^2$ はそれぞれニューロンが発火する、発火しないの確率である。 $|A\rangle$ は状態 A を表している。古典力学と量子力学の重要な違いは量子的な干渉である。古典的な予測と異なる観測可能な量子効果があるかどうか という問いは状態の重ね合わせと直接に結びついている。ニューロンが発火している、いないの間の重ね合わせを可能にするには発火によってつくられる化学的相違が大きすぎるという考えには疑問の余地があるが、ニューロンの発火の前に観測可能な量子効果が起こると考えられる。だが、発火前の神経活動の量子効果を見出すことは困難だろう。脳は高温で活動している。そのため、状態 $|A\rangle$ と状態 $|B\rangle$ は環境と異なる仕方で絡み合っているだろう。したがって、二つの状態が実質的に重なり合うように発展することは考えにくく、量子効果は無視できるようになるだろう。

この反対にたいして Penrose は microtubules の性質についての仮説を立てる。細胞内に見出せるこれらのタンパク質はニューロンの中での計算に有用な性質をもっている。 microtubules がニューロンの中で量子的コヒーレンスを保持するとしても、ニューロン間でもそれが保持できるとは考えにくい。Penrose の答えは次のようになる。

“.. the quantum coherence must leap the synaptic barrier between neuron and neuron. It is not much of a globality if it involves only individual cells!” (p. 372)

彼の話に欠けているのはどのように量子的コヒーレンスがシナプスの壁を飛び越えるかの説明である。進化はこの不可能に見える壁を乗り越える術を見出したのだらうというのが彼の暫定的結論である。

(2) 物理学と形而上学

量子力学と意識を結びつけようという Penrose の試みはいくつかの難点をもっていることがわかった。しかし、その試みとそこで果たす量子力学の役割は意義があると思われる。特に、量子力学における観測者の役割は心の科学に欠けている主観性に重要な関連をもっている。量子力学の本性を Penrose は巧みに描いてくれるが、それは心と脳の関係に密接に結びついている。主観的意識としての観測者と量子力学の観測者を結びつける満足いく方法を見出すことはまさに形而上学の仕事である。

私たちは二つの極めて異なるが、しかし、哲学的な論証の吟味だけでない実質的な心や意識の例を見てきた。これらの例から何か結論を引き出すことは早計であるが、将来の心の科学が現在の認知科学の範囲を超えて進むであろうことは実感できたと思われる。

9 別の自然化の仕方

9.1 「心とは何か」の別の捉え方

「心はなにか、どのようなものか」という問いは、人間にとってのもっともポピュラーな問いである。誰もが何度かはこの問いを考え、そして満足のいく解答を得られなかった経験をもっているはずである。うまく解答できない理由を私たちは知っていると思っている。心は多くの異なる働きをもつ、きわめて複雑な装置であり、私たちは心を解明するのに心を使わなくてはならない。そのような複雑な心を捉える十分な知識がどのようなものかを私たちは知らないし、自らを対象にすることは原理的にできないのではないかとも思っている。このような理由が確かに心が他の対象とは異なる、独特のものであるという理解を生み出し、心を人生と同じように神秘的で不可解とする態度をつくってきた。

心の振舞いは今世紀に入り、論理学の進歩によって一転する。「合理的に考える」仕組みと、その説明は論理学の革新によって先鞭がつけられる。それを具体化したのはコンピュータの発明である。合理的な思考の仕組みが一段落すると、当然ながら私たちの関心は心の別の側面に移行する。それは意識と一般に呼ばれている心の働きである。これを「困難な問題」とすれば、合理的な思考の働きは「容易な問題」であった。

「容易な問題」を解く基本的な枠組みは心の働きを表象と計算の二枚看板で考えようとする機能主義にある。この枠組みのなかで意識をはじめとする「困難な問題」も解けるかどうかで意見が分かれることになる。「困難な問題」は解けたと思っていた「容易な問題」にも影響を与え、全面的な心の理解の見直しを迫るところまで広がりつつある。

私が今まで述べてきた基本的な姿勢は自然主義である。自然主義を満足行くように定義することはできないが、議論の組み立てにおいて超自然的な要素が入らないように議論を展開することと考えてよい。そのような姿勢で機能主義に基づく心の理解を考えてきた。ここでは自然化のもう一つの方法として、生物学的な自然化についてその実際の姿と哲学的問題を考えてみよう。自然化に加えて実在論を想定する。私たちがその時その時にもつ視点や観点はしばしば対象を異なるように見せるかもしれないが、そのような対象は私たちがそこにいなくても存在している。これがここでの実在論のもっとも包括的な捉え方である。

9.2 「意識が捉えがたい」とはどのようなことか

Nagel の背後にあって、Searle、McGinn らに共通するもの 意識はわからないもので、現在の科学的探求を超えたものという共通理解 は確かに私たちが意識についてもつ認識である。端的に、私たちは意識を知っていないという認識である。この認識は次のような危険な論証を生み出しやすい。

私たちは意識を知っていない。

それゆえ、私たちは意識を知ることができない。

この論証は誰がみても誤っている。論証としては誤っているが、経験的に正しいことがわかるかもしれない。意識が捉えがたいものであることは論理的にも経験的にも証明されたわけではない。この論証を一つの命題にまとめ、その対偶にあたる論証を考えてみよう。それは次のような論証となる。

私たちは意識について知ることができる。

それゆえ、つくられた仮説が正しいかどうか経験的に証拠を集め、意識について知る。

これは仮説を作り、それを経験的に確認するという通常の科学的な方法を意識について述べたものに過ぎない。これは全く穏当な研究スタイルをそのまま述べたものである。これは経験科学がどのようにそのパ

ラダタイムを変えようと、それとは無関係に私たちが基本的と考えてきたことである。

しかし、これら二つの論証は私たちに随分異なる印象を与える。前者の論証の演出効果は、極めて消極的である。心に関する事柄は哲学的で非経験的な分析が必要であり、それが結果として私たちに心の経験的な追求を阻止するように働く。一方、後者の論証は、心も他の対象と同じように経験科学の対象であり、私たちに経験的な心の探求を促すように働く。

ここではこのような概念分析をすることが目的ではない。しかし、上述の分析を踏まえて見直すならば、Nagel、Dreyfus、Searle、McGinn らは十分な証拠、つまりは上の論証を成立させるのに十分な証拠を示しているとは思えない。彼らを消極派と呼んでおこなうなら、それより一歩進んで、具体化した形で問題を取り上げる典型として Penrose を考えることができる。彼は経験的に意識が（現在の科学で）捉えがたいことを科学的に証明しようとしている。Penrose の場合、意識が現在の計算主義では捉え切れない具体的な証拠としてゲーデルの不完全性定理が取り上げられる。類似の場合として、Stapp を取り上げることもできる。彼らは意識を量子力学と結びつけて考えるグループにつながる。

一方、積極派と呼んでもよいグループは科学の常套手段を信用する。意識が科学的に解明できるという仮説のもとで意識を考える。彼らにとって問題となり、心の哲学で激しく議論されてきたのは、どのような仮説が適切であるかという問題である。つまり、意識を解明する際にどのような仮説の構成のもとで考えるかに関して哲学的な議論が積極派の間で戦わされてきたのである。

9.3 消去的唯物論（Paul Churchland の場合）

(Paul Churchland, Professor of Philosophy at the University of California, San Diego, in his many articles and books such as *Matter and Consciousness*, and *The Engine of Reason, the Seat of the Soul*, concerns topics such as mind, knowledge, science and language. His wife, Patricia Churchland, is also a philosopher, and she writes about the same kinds of issues.)

心の状態や過程は脳の状態や過程であるという主張は心についての唯物論である。これが Churchland の立場で、心的なものは物理的なものに還元でき、したがって心は存在しない、消去されるべきである、というのがその主張である。彼の立場は神経科学的な証拠に基づき、常識心理学に対立している。Churchland によれば、常識心理学は誤っている。人間や動物の理解の基本単位は言語的に表現できる状態（「P を信じる」、「P を欲する」、「P を知る」等 - P は文 - ）ではなく、脳のニューロンの集合の活性化であるというのが彼の主張である。この主張は connectionism という認知科学の考えに基づいている。

脳のニューロンの結合は並列的（＝感覚入力順番にはなく、同時に処理される）で、通常のコンピュータが直列的（＝順番に処理される）に情報を処理するのとは異なっている。直列的だと、結合の一部が損傷を受けると全システムが動かなくなってしまうが、並列的な場合はすべての結合が互いに機能的に結びついているということはないので、システムは一部が故障するだけで済む。私たちが対象を見ると、脳は入力を脳の中のニューロン活性のパターンに変換する。視覚野のニューロンは特定の仕方刺激され、その結果パターンが生じる。このパターンは対象毎に異なる。違った顔つきのパターンは石のパターンとは大きく異なるが、違った顔つきのパターンの間ではその違いは小さい。Churchland によれば、さまざまな顔つきの平均である原型パターンがあって、それは他の対象の場合も同じである。原型パターンを使って顔つきを認知するのであって、そこに意識は必要ないというのが彼の考えである。（デカルト風の考えだと顔つきはどのように認識されたらだろうか。）純粋に物理的な基礎に基づいて私たちは対象を認知できる。（この主張は並列コンピュータシステムを使った実験で確かめられている。カリフォルニア大学の Garrison Cottrell のグループによる顔貌認知の実験のコンピュータは意識も意味の認識もできないが、顔つきの認知には 100% 成功した。顔の一部を隠した場合でも同じように認知に成功し、帰納的な推論の原始的な形もやはり意識や意味の助けを借りずに実行されることが示された。）

Churchland の還元主義的な立場に反対する一人が Thomas Nagel である。世界についてコウモリが気づいていることはコウモリの主観的な感じであり、これはコウモリの物理的な記述によっては捉えることができない。私たちはコウモリであるとはどういうことか(what it is like to be a bat)を知らない。この考えを使うと、あなたはあなた自身の経験をもっているのに、私はあなたであるとはどういうことかを知らない、ということになる。したがって、経験、それゆえ意識は物理的なものに還元することはできない。Frank Jackson と John Searle も類似の反消去主義的な議論を述べている。(それぞれの該当箇所を参照) 人の物理システムとしての知識はその人がもつ経験についての知識をすべてカバーすることはできない。意識現象は物理的な説明では与えることができない。これが彼らの結論であった。

Churchland は、それぞれの人はそれぞれの脳のネットワークをもち、みな異なっている、それゆえ、その違いが説明できない違いを生むだけであって、科学的な理解を超える感覚的、認知的状態が存在するわけではない、と返答する。私はコウモリであるとはどういうことかを知ることができる。しかし、コウモリ自身を知る仕方でそれを知ることができない。

では、消去主義の主張の理由に何か誤りはないのか。これに対して次の Stich の疑問はより具体的である。彼は消去主義の代表者であったが、近年その立場を微妙に変えつつある。その経緯は彼の最近の論文に詳しく告白されている。それは常識心理学が誤りであることから、そこに登場する概念、特に信念、欲求は消去されるべきであるという主張が正しいかどうかにもっぱら向けられている。つまり、誤りから消去への推論がどのように保証できるかである。その疑問は語の指示に関する考え方に凝縮されている。従来の記述説を取るならば、Churchland の主張は正しくなるが、因果説を取るなら正しくないということになる。したがって、問題はすべて指示理論をどのように考えるかに依存することになる。彼は deflationary idea を採ろうとしているが、Millikan は対応説を採っている。

(問い) 消去主義的な主張と Nagel の議論は本当に矛盾するだろうか。人の認知や意識には脳状態を使ってわかる部分も、経験のように予めわからない部分もあるという意見ではだめなのだろうか。すべてを脳状態で説明するのはどのような意味で用いられているのだろうか。まだしていない経験を物理的に説明することはできないが、既にした経験は物理的に説明することができる。「説明」の二様の意味が議論には使われていないだろうか。

Churchland は信念や欲求に基づく心理現象の常識的な見解を経験的な folk theory と考える。彼はこれを常識心理学と呼び、経験的な理論としては不十分であり、したがって、最終的には廃棄されなければならないと論じる。信念や欲求といった命題的態度は存在しないというのが彼の主張である。彼はまず常識心理学を経験的な理論として考えることができるとする。信念や欲求を仮定することによって人の行動が説明・予測される、そしてそれら命題的態度は行動に結びつく形で法則的であると常識心理学は考える。もし常識心理学が経験的理論であれば、それは誤り得る。もし誤りならば、それは消去されるべきである。では、常識心理学が経験的理論として誤っている理由は何か。

1. 常識心理学は私たちの行動の多くを説明し予測することができるが、できないものも多く含んでいる。例えば、心の病、創造性、知能、睡眠、運動能力、知覚、記憶、学習等は説明できないが、それは常識心理学がせいぜい表面的な理論に過ぎないことを意味している。
2. 常識心理学の適用範囲は次第に減少している。これは原始社会の世界観、人生観を現在の私たちのものと比較すれば明らかである。それは人間中心主義として今日では否定されている。
3. 常識心理学を科学的な枠組みに組み込むことはできない。生物学、進化論、有機化学、神経科学は人間の認識と行動に関して包括的に理解するために統合され始めている。この自然科学的枠組みに常識心理学をうまく組み込むことができない。それは常識心理学が意味、命題、志向的状态を根幹にもっているから

である。

多くの哲学者、特に機能主義者は信念と欲求というカテゴリーを消去することに反対する。第一に、常識心理学は規範的な理論、つまり、私たちがどのように合理的に考え、行動するべきかについての理論であり、それがなくては私たちの思考や行動を反省することができないゆえに不可欠であると機能主義者は考える。第二に、常識心理学は私たちの心的な状態を背後の物理的な構成を気にすることなく特徴づけることができる点で極めて有用である。(これは複数実現可能性から得られる。)一部の機能主義者にとっては常識心理学は経験的理論ではなく、概念的装置とさえ考えられている。

Churchland は第二の反論に対して、機能的状態は神経科学の発展とは無関係であるという考えは誤っており、それは科学史上錬金術を新しい化学理論から守ろうとして考えられた機能性と同じ誤りであるとする。第一の点に関しては、彼は常識心理学の規範的役割は別の科学理論によっても果たされると考える。その科学理論では心を記述するのに別のカテゴリーを用いるだろう。神経科学から生まれる認知の科学はまだ全く想像に過ぎないが、常識心理学のカテゴリーがそこで何か役割を演じるとは考えにくい。

(#) Churchland は常識心理学の経験的法則を次のように考えるべきであると言っている。

- (1) もし x が p であることを恐れるなら、 x は p でないことを欲する。
- (2) もし x が p であることを欲し、かつ q が p を引き起こすと信じるなら、 x が q を引き起こすことができ、 p を引き起こすより好ましい方法を知らないか、 p への欲求より強い欲求をもっていない場合、 x は q を引き起こすであろう。

これらの言明は命題の態度が互いにどのように関係しているか、行動にどのように関係しているかについての経験的な一般化である。

9.4 意識を考える上でいずれが適切か

常識心理学を信じて理論を組むか、常識心理学を消去して脳科学に還元するかで方向は大きく分かれた。経験主義の「経験」がどのようなものかがここには如実に登場する。私たちは意識を経験している。それを疑う者はいない。しかし、それはあくまで私の意識であり、他人や他の動物の意識ではない。他者の意識を自分の意識と同じように認めたとしても、それが同じ、違うといった点に関しては経験主義は何も具体的には語ってくれていない。(ここで経験主義とはどのようなものかという厄介な問題を提出するのは得策ではないが、意識の問題は確かに経験主義とは何かを見直すよい機会であることは確かである。経験主義の内容、つまり、何を経験とみなすかはもっとゆすぶられてもよい。)

常識心理学というと大袈裟すぎて事態が返って捉えられなくなる。問題は心や意識を考える際にどのようなデータを出発点にするかである。わたしたちが日常生活でいつも使っている心理的な事柄、常識的に通用している心に関する共通理解と表現方法を経験主義的に認めるか、それともそれらを一度括弧に入れて、それとは違う科学的な枠組みでそれらを捉え直すかの違いである。心を考える際に、常識心理学を出発点にするか、それともそれを結果の確認だけに使うかの違いである。いずれにしる、常識心理学はどこかでそれに頼らなければならない。

物理学に基づく世界観が静的であるのに対して、常識心理学における心的状態や過程は一般に動的なものと考えられている。その端的な例は私たちの心が創造的な能力を持ち、変化する環境に的確に反応でき、ダイナミックなものという理解の仕方である。それは近代の人間観の重要な要素の一つにさえなってきた。この心の振舞いの動的な理解の仕方は自然と比較しても劣るどころか、それ以上にダイナミックなものと考えられてきた。なぜこのような理解が生まれたのであろうか。科学的心理学が批判的に人間の心理を扱ったのとは違って、常識心理学は心理状態をそのまま受け入れ、それを都合の良いように解釈するという特徴をもっている。最初に受け入れるデータとそのデータに対する扱いが科学的心理学とは異なっている。

変化する知覚的世界をそのまま承認する。その上で、それを再構成し、理解できるような形式によって解釈する。常識心理学は科学理論の一つであるとする考えに無理に反対する気はないが、確かにそれは科学的心理学とは異なっている。

デカルトの方法的懐疑であれ、ヒュームの懐疑であれ、それは経験的なものが信頼できる知識になることを保証しない。その限りでは常識心理学的知識は未だに十分な根拠を獲得していない。むしろ、できないというのが長年の経験的結果である。すると、常識心理学の立場は根拠付けられていないものということになる。私たちは、しかし、この常識心理学の知識を根拠が与えられていないゆえに括弧に入れるということを日常生活ではしていない。それどころか、常識心理学を信頼して生活している。これに対して、科学的心理学は全く異なった立場を取る。それが行動主義的な心理学であれ、認知心理学であれ、経験科学が認めるものだけを信頼する。常識心理学の結果はすべて再検討され、実験にかけられ、そのデータが吟味される。科学的とは、ここでは与えられるデータへの態度の違いとなって現れている。このような違いは両者を理論とみなすことに影響を与えるはずである。

常識心理学は経験的データに対して寛容である。それは吟味の対象ではなく、吟味の前提である。この違いは極めて明確である。データへの吟味がない場合、私たちの知的追求はデータの形式や相互の論理的関連へとシフトする。常識心理学はこの意味で、非経験的、論理的、数学的な理論として把握される。これは皮肉な結果である。最も経験的な、生のデータを扱いながら常識心理学はそれを吟味の前提にすることによって、非経験的な要素を必須の性質としてもつことになる。実際、常識心理学の理論化の例を眺めてみれば、それは言語表現を中心にした推理や推測の形式化である。科学的心理学はデータを吟味する。吟味が強すぎると理論に至ることができない。そのような傾向は行動主義的心理学に顕著に見られる傾向である。データに関わり、その獲得の過程にばかり執心すると、出来上がるのは悪しき経験主義的研究ということになる。

科学的心理学は常識心理学と比べはるかに臆病である。データのないところでは沈黙する。この沈黙は心を神秘的なものとみなす伝統を長らえさせて来た。科学は革新的であるはずなのに、心理学の分野では相変わらず常識心理学が人間の未来を担っている。これは科学の皮肉どころか汚点と言ってもいいだろう。常識心理学がアプリアリな性格を強くもつものに対して、科学的心理学は経験的である。この奇妙な結論は誤ってはいない。「経験的」という概念が如何に常識心理学的な概念でないかが如実に現れている。

常識心理学はどのような真理の部分のカバーしているのだろうか。従来、認識論と呼ばれてきた分野は基本的に常識心理学であると考えることができる。それは知識に関する経験的でない理論という点では常識心理学である。前提としての心的能力とそれに基づく知識の体系化は優れて常識心理学の特徴を具現している。常識心理学はまた、私たちの世界観、人生観をも含んでいる。ここでも基本は与えられたデータを前提にして組織的な体系を考えることである。観念論的なシステムのほとんどは常識心理学の恩恵に浴している。

10 行動の表現：進化論的表現と心理学的表現

10.1 進化論的な利他的行動とはどのようなものか

私たちが日常生活の場面で遭遇する利己性や利他性は心理学的なものであり、進化論の対象ではない。私たちの経験する範囲内での利他性、利己性は、進化論によって考えられてきた利他性、利己性ではない。既述のように、進化論の対象は集団の進化であり、利他性や利己性も集団的な特徴として理解されてきた。それに対して、心理学的な利他性や利己性は個人の行動についての性質である。したがって、進化論的な利他性、利己性と心理学的な利他性、利己性を同一視することはできない。系統発生と個体発生は明らかに異なる自然的な過程であるが、それと同じように集団レベルと個人レベルの利他性、利己性も明らかに異なる過程なのである。

人間の行動が利他的か利己的かの二つに排反的に分類できる、つまり、どのような行動も利己的か利他的のいずれか一方である、というのは明らかに誤っている。むしろ、私たちの行動のほとんどは利己的でも利他的でもないように見える。食事をするという行動は利他的でも利己的でもないと考え人が多いだろう。しかし、食事の種類や仕方についての規則、そして、誰にどのように優先的に与えるかといったことを考え出すと、そこには明らかに利己的、利他的と形容できる食事に関する行動が浮かんでくる。この例が示唆しているのは、利己性、利他性という行動の性質が特定の行動やそのパターンを指しているのではなく、行動という物理的な過程に随伴する性質であるということである。

さて、一般的な話から、進化論での利己性、利他性の具体的な議論に眼を転じてみよう。唐突だが、ここで Simpson の paradox の印象的な例を挙げておこう。それは University of California at Berkeley での入学試験についての男女差別の疑いの例である。男女同数の受験者に対して、合格者全体を比べると男のほうが女より多い。これは男女差別ではないかという疑いがかかった。しかし、学部ごとに調べ直してみると、二つの学部はいずれも男女の合格者数に関して全く公平であったことがわかった。（下の表はこれをわかりやすくしたもので、実際の学部や数ではない。）

	学部 1	学部 2	総計
応募者	90女; 10男	10女; 90男	100女; 100男
合格率	30%	60%	
合格者	27女; 3男	6女; 54男	33女; 57男

この例と同じことが進化に関しても存在する。それぞれ100人ずつの群があり、そこには二つの行動パターンを示す個体が存在している。利己的な行動をする個体と利他的な行動をする個体の二つである。100人の集団におけるそれぞれのタイプの個体の適応度の変化とそのタイプの個体の比率は Sober の論文のグラフが役に立つ。グラフの内容を数式で表現することは困難ではないが、それよりはそのグラフの理解に基づいて考えたほうがわかりやすい。グラフは集団の変化を直観的に表しているとはいえ、私たちの議論にはそれで十分である。グラフでは個体数は変化せず100人と固定されている。実際の集団遺伝学のモデルは個体数を制限しない。

利己的な行動をする個体は集団内に1しかない場合、適応度4をもつ。その場合、残りの99は利他的な個体である。この集団に利己的な個体が増えるに連れ、利己的な個体の適応度は次第に減少していく。なぜなら、一人なら集団内でわがままいっぱいに行動できたのが、同じ行動をする利己的な個体が増えるにしたがって、そのわがまは制限されていくからである。一方、利他的な個体の適応度は利己的な個体が少なければ、比較的旨く行動していけるが、利己的な個体が増えるに連れ、やはり次第に減少することになる。また、この集団全体の平均適応度は利己的な個体が増えるに連れ、集団の統一が次第になくなっていくということから、減少していく。

Sober の論文にはグラフの他に表が載っている。この表こそ Simpson のパラドックスを適用してつくったもので、利他的な行動のほうが利己的な行動より適応度が高い場合があり得ることを示したものである。そして、このことが利他的な行動が利己的な行動より適応度が低いという一般的にはもっともと思われる仮定の下で、グループとグループの集まりが存在するならば、利他的な行動のほうが利己的な行動よりは適応度が高くなる場合があり、したがって、利他性が集団内に保持され、選択的に有利であることが可能であることになる。これが利他性の存在についての基本的な考え方である。

グループ1	グループ2	総計
1S; W = 4	99S; W = 2	100S; W = 2.02

99A; $w = 3$ 1A; $w = 1$ 100A; $w = 2.98$

(この表と最初の表は異なることを表しているように見えるが、上の W, w を 1 以下の重みづけにして考えるなら、最初の例とそれほど異なる印象を与えない例となる。1 以下の重みが合格率に対応すると考えればよい。)

このような結果を別の仕方でまとめてみると、次のように言うこともできる。もし、下の推論が与えられた場合、それは正しいだろうか。

どのような部分集団においても、利己主義は利他主義よりも適応度が高い。

適応度の高くない性質はその頻度が低下する。

それゆえ、利他主義はその頻度が低下する。

明らかにこの推論は誤っている。上の表の総計の数値 ($W = 2.02, w = 2.98$) がこの推論の結論の反例になっているからである。どのような部分集団においても頻度が低下する利他主義者は、全体ではその頻度が高くなることもあり得るのである。

10.2 血は水よりも濃し?

血縁集団は最初から所属のグループが決定している。生得的に所属が決まっているという点で、グループ固定型といってもよい。個体の適応度の代わりに血縁集団の包括的な適応度 (inclusive fitness) を使って、血縁集団に関する利他主義の存在を論証したのは Hamilton であった。

一方、グループ選択のほうは Williams の還元主義的傾向の強い遺伝子選択説 (genetic selectionism) の考えによって否定され、ずっと無視されてきた。Simpson's paradox などを使ったグループ選択の見直しはまだ始まったばかりである。血縁選択の場合と異なり、遺伝子の類似性に依存せず、グループ流動型である。血縁選択は Hamilton によって考えられたが、その選択の単位としての立場は定まっていない。生物学者の中には、血縁選択をグループ選択の一つと考える者も、(子育てと結びついていることから) 個体の適応度であるとする者もいる。利他主義の進化を握る鍵は集団の構造にある。以下の論証は、グループ選択、血縁選択、互助性 (reciprocity, a kind of conditional altruism) のどれにも適用できる。

利他主義者は他人に利益 b を与え、そのための費用 c がかかるとする。利他主義者は自ら与えることはなく、利他主義者がいれば利益 b を受け取るだけである。利他主義者の適応度を $w(A)$ 、利己主義者の適応度を $w(S)$ とすれば、利他主義が進化するのは、 $w(A) > w(S)$ の場合である。いま、集団が n の利他主義者と何人かの利己主義者からなっているとしてみよう。その場合の利他主義、利己主義の適応度は、

$$w(A) = (x - c) + (n - 1)b \quad w(B) = x + nb$$

ここで x は利他主義、利己主義という性質がなかったときの適応度である。すると、次のことが簡単な計算からわかる。

- (1) もし全員が相互作用していれば、 $w(A) > w(S)$ if and only if $c + b > 0$.

この関係は他人への供与が利益 b と費用 c をもたらすならば、利他主義が進化できないことを示している。ここで、次の表に表されているような二人の対についての関係を考えてみよう。

受託側

A S

本人	A	$x - c + b$	$x - c$
	S	$x + b$	x

この表から、A と S の適応度は次のようになる。

$$(2) \quad w(A) = (x - c + b)P(A/A) + (x - c)P(S/A) \quad w(S) = (x + c)P(A/S) + (x)P(S/S)$$

ここで $P(A/S)$ は、対の一方が S のとき、他方の個体が A である確率である。(2) から、簡単な計算によって、

$$(3) \quad \text{対になっている個体が相互作用していれば、} w(A) > w(S) \text{ if and only if } P(A/A) - P(A/S) > c/b$$

対のグループが勝手につくられるなら、 $P(A/A) = P(A/S) = P(A)$ 。この場合、 $c/b < 0$ 。利益と費用があるならば、これは矛盾であるから、利他主義は進化しない。もし $P(A/A) = 1, P(A/S) = 0$ ならば、 $b > c$ となる。利他主義者どうしの相互作用において、利益が費用を越えるならば、利他主義が進化することを示している。こうして、集団が対に分割できて、しかもその対が似た者どうしの場合、利他主義は進化できることがわかる。

では、利他主義者の対はどのようにしてつくれるのか。一つの可能性は血縁者である。これが Hamilton の血縁選択の基本的な考えである。血縁者どうしは似ている。その対としての組み合わせは $P(A/A), P(S/S)$ のいずれかである。Hamilton の不等式は $r > c/b$ のときに利他主義が進化すると述べている。 r は血縁係数で、完全な親子の場合、 $r = 1/2$ である。その場合、 $P(A/A) - P(A/S) = 1/2$ となる。

p を A の頻度、 q を S の頻度とすると、交配による子どもの割合は次のようになる。

$$P(AA) = p^2 + pq/2 \quad P(AS) = pq \quad P(SS) = q^2 + pq/2$$

したがって、 $P(A/A) = P(AA)/P(A) = p + q/2, P(S/S) = P(SS)/P(S) = q + p/2$ となり、これを(2)に代入し、

$$w(A) = (x + b - c)(p + q/2) + (x - c)(q/2) \quad w(S) = (x + b)(p/2) + x(q + p/2)$$

$$(4) \quad \text{相互作用が完全に親と子供の間であるなら、} w(A) > w(S) \text{ if and only if } 1/2 > c/b$$

10.3 集団間のネットワークと選択

「他の個体のため」という要素をもつ行動を指すのはわかるが、それは「自分の属するグループの個体のため」なのか、それとも単に「他の個体のため」なのか、という疑問である。普通の生物は自分の属するグループを識別している。そして、利他的な行動は無制限にではなく、一定のグループ内のメンバーに対して発揮される。血縁選択の場合は、明らかに遺伝的なつながりをもとにして利他行動がその範囲を規定されている。血縁選択とグループ選択の基本的な違いはここにある。グループ選択には集団内のメンバーの間に遺伝的なつながりはないのである。ある特定のグループ（その典型は家族）をつくるように指令された遺伝子はあるかもしれないが、その個体が属するであろうようなすべてのグループについて、そこに属することを指令する遺伝子はない。したがって、グループは非遺伝的な事柄によって形成されており、その直接的な生物学的形成理由はないことになる。少なくとも、何か生得的なグループ形成要因は遺伝的にあるかどうかは、現在のところわかっていないのである。

どのようなグループに対しても利他的の行動できる場合と特定のグループに対してだけ利他的な行動をする場合では利他性の度合いが異なる。実際の場合は、どのような個体もあるグループに属しており、すべてのグループに属するといったことはないので、極端に利他的であるというのは可能性でしかあり得な

い。

- (1) あるグループが存在して、もしそこに入ったら、そのグループのメンバーのために行動する。
- (2) もしあるグループに入れられたら、そのグループのメンバーのために行動する。
- (# 両者の違いは普量化と存在量化の記号の違いでもある。)
- (3) もしあるグループに入れられたら、そのグループのメンバーのために行動しない場合もある。
- (# ここで (3) は後に述べる選択の可能性を表した命題となっている。)

もし (1) のような場合であったとすると、二つのグループが一緒に考えられてもそれは名目的なものに過ぎず、単に物理的に交流があると言うだけである。二つのグループのメンバーはグループが分離されていた場合と全く同じよう行動し、表にあるように適応度が逆転するということは起こり得ない。二つのグループが一緒になるとき、そこで大きなグループ内での適応度が考えられているのであり、それがグループごとの適応度と異なるのは、生物学的には新しい集団内での新しい行動に関する評価なのである。生物学的に意味があるためには新しい集団は何らかの効果をもつ。もしそれが分離していたときの行動と同じであるなら、それは適応度の計算上、古い行動と新しい行動の区別が付かないという、計算上の不備に過ぎない。新しい集団内の適応度の変化は個体の行動の平均であるから、(1) と (2) の二つのタイプの行動に関してはどちらか一方か、両方とも認めるかである。両方とも認めるとは単なる計算上の不備であって、生物学的には無意味である。したがって、いずれか一方ということから、(2) の場合と言うことになる。これは「他の個体のため」というのが特定のグループのためというのではなく、任意のグループに対して適用できることを示している。「他のメンバーのため」というのは、グループに関して特定されていないのである。

二つの異なるグループに同時に属し、かつ「他の個体のため」が異なる行動である場合。これは相当に複雑である。そのようなとき、上の group-flexibility は私たちにどちらのグループの基準で行動するかを選択を迫る。そこで選択に苦慮するといった行動を私たちに強いるのが group-flexibility の帰結である。group-fixed の場合には選択に苦慮すると言ったことは起こらないのである。

10.4 人為選択のコントロール

心理的な利己主義は最適化理論の適用によって説明できる。例えば、子どもの世話はそれ自体両親にとって目標なのか、それとも彼ら自身の満足のための手段なのか。Sober によれば、4 つの基本的な構造が考えられる。極端な利己主義者、極端な利己主義者、穏健な利己主義者、穏健な利他主義者である。これら分類は次のような表に基づいている。

	other-directed	
self-directed	a	b
	c	d

極端な利己主義者(a=b=4, c=d=1) 極端な利己主義者(a=c=4, b=d=1) 穏健な利己主義者(a=4, b=3, c=2, d=1) 穏健な利他主義者(a=4, b=2, c=3, d=1)はそれぞれの評価基準にしたがって行動することになる。このような評価は進化上の適応度とは直接に関係ない。このような行動パターンを、それがどのように起るかを信念や欲求と組み合わせれば行為論の問題となる。また、それがどのように進化の構造に結びつくかは自然主義的な行動の問題となる。では、このような枠組みと進化的な利己主義、利他主義はどのように関係するのか。

進化と心理の異なる利己性、利他性は、そのような形質を担う遺伝子を想定することで一本のシナリオに

纏め上げることができる。ここで Dretske の言う引き金原因と構造化原因の区別を使えば、どうなるだろうか ([Dretske])。遺伝子は個体の行動の引き金原因になれるか。環境(あるいは刺激)と遺伝子の共同作業が行動の原因であるとすれば、遺伝子は引き起こされる行動の引き金原因の一部として働いている。そして、その同じ遺伝子はそれをもつ個体がどの程度生存し、子孫を残せるかという適応度を担うものとして、進化の場面でも働いている。これは引き金原因と構造化原因が遺伝子を介して一部重なり合っていることを示している。これは比喩的に近くの原因と遠くの原因と言い換えてもよい。そのどちらにも利己的、あるいは利他的遺伝子が顔を出すことになる。

ある計算結果が正しいかどうかは、それを習得した学校教育のせいにする場合よりは、数そのもののしくみを使う場合が多い。しかし、それは学校教育による説明がいつも誤りということの意味ではない。場合によっては学校教育のせいにする方が正しい場合もある。 $2 + 3 = 5$ の正当化の文脈と $2 + 3 = 5$ の説明の文脈は異なる。近くの原因、遠くの原因はこの説明の文脈が二つ以上あることを示している。求められているのはこの説明の文脈における進化と個人の役割の解明である。進化は遠くの原因、心理は近くの原因である。遠くの原因の存在は近くの原因を無視してよいということにはならず、逆に近くの原因は遠くの原因を忘れてはならない。因果的な責任は両方にある。(因果的な責任を一部に含む倫理的な責任に関しても同じように主張できる)。正当化は知識内容の正当化であって、その知識の対象に関する正当化ではない。対象に関する正当化は説明の適切さである。知識の対象としての進化的、心理的な利己性、利他性は上の二つの説明文脈を注意深く見分けることによって、統一的に扱うことが期待できる。しかし、同時にこれは次のような厄介な問題を含んでいる。

「氏が育ちか」を考えてみよう。氏は遠くの原因であって、その人の正当化の理由ではない。育ちは近くの原因である。実際その人となりは遠くの原因と近くの原因の混合である。では、この混合は二つの原因に分解できるか。力学の変化はその原因を局所的に分解できるが、変異の原因は分解できない。したがって、ある人をその氏と育ちに分解することはできない。これは「氏が育ちか」という問いそのものが無意味なことを意味している。実際の人為選択の問題は自然と人為の混合である。したがって、そこに自然と人為の区別ができなくなることを不可避的に想定しなければならない。心理的な利己性、利他性と進化的なそれらの間での区別ができないことは、人為選択に対して基本的な事柄である。このことをまとめると次のようになる。

- (1) 進化的、心理的な利己性、利他性は概念的には独立している。
- (2) 人為選択の際の利己性、利他性は心理的な性質である。
- (3) その性質が自然、人為の選択の過程を通じて進化的な利己性、利他性に影響する。
- (4) その結果、進化的、心理的な利己性と利他性は実際には互いに区別できなくなる。

この結論に一概に悲観する必要はないのかもしれない。区別は重要でないという統一的な視点を採用するならば、当然の結果であろう。

10.5 目的論について

認知動物行動学(cognitive ethology)は人間以外の動物の思考過程、意識、信念、合理性について比較し、それらの進化を研究する。この新しい分野への反対、懐疑、賛成はさまざまである。反対の典型は常識心理学的な説明と人間中心主義である。

自然科学の中でも生物学はデザイン、目的、適応的機能に関する目的論的言明を許すという点でユニークな地位を占めている。伝統的に目的論は二つの種類がある。その最初はアリストテレス的な見解で、対象の運動は外的な干渉がない限り実行される、その対象の内在的な目的によって説明される。二番目は超自

然的な意志が働くために現象が生じるというものである。今の生物学は両方とも拒否した。

にもかかわらず、目的論は生物学に欠くことができないように見える。有機体の複雑な形態や行動を理解するには「それらは何のために」あるのかと問わなければならないように見える。これを説明する強力な説明が自然選択を使った統計的な説明である。生物学内部には目的論の必要性に関する強い緊張がある。したがって、目的論の有無は生物学の哲学にとって基本的な問題である。そして、現在でも哲学者の間で目的論的な説明の特徴づけに関しては一致は存在していない。

(#) There are essentially two schools of thought regarding functions in philosophy of biology. There are those who follow Wright's (1976) line and argue that the function of an object explains the *continued existence of that object*. An example is that the function of the heart is to pump blood. Pumping blood is one thing that the heart does, and the fact that it pumps blood (as opposed to making thumping noises) explains why we have hearts. The fact that organisms evolved hearts means that blood circulates more efficiently. More efficient circulation means better distribution of chemicals and nutrients to the system and quicker waste disposal. This in turn leads to better survival rates, hence, more progeny. This view of functions requires an *historical approach*.

These functional explanations explain why something is there in natural systems by picking out the most valuable property of that thing to previous generations. Usually these sorts of explanations appeal to evolution by natural selection. However, these sorts of functional explanations can also explain an artifact's function by appealing to the intention of the designer, or the use to which the object has been put that keeps it around. In either case, these explanations turn on the *etiology* of the object or trait in question.

The second school of thought rejects the notion that functional explanations concern the origins or continued existence of something. Instead, following Cummins, functional explanations point out *how a trait or property or capacity contribute to the overall capacities of the system* in question. The function of birds' feathers is to help them fly, despite the fact they evolved to aid in thermoregulation. According to this perspective, functional explanations tell us what something does, why that thing is important (now), not why it exists or continues to exist.

10.51 起源論としての目的論

Francisco J. Ayala の分析を通じて進化生物学での目的論の役割を見てみよう。生物学での目的論は次の三種類の場合に正当なものとみなされる。

(i) when an end-state or goal is consciously anticipated by the agent, (ii) in connection with self-regulating systems (e.g. regulation of body temperature), and (iii) in reference to structures anatomically and physiologically designed by natural selection to perform a certain function.

最終的にすべては最後の場合に帰着する。したがって、彼の目的論の説明は歴史的である。自然選択は因果的な過程であるから、生物学での目的論は内在的となる。彼の主張は Larry Wright によってさらに哲学的に分析される。彼は「機能」に着目し、生物に限定されない一般的な分析を行った。それに依れば、X の機能が Z であるとは、

- (a) X が Z を行うゆえに、X はそこに存在する、
- (b) Z は X がそこに存在することの結果である。

この図式には自然選択は何も登場しないが、自然選択はこの図式を満足する。Wright にとって、目的論的説明はある事態の存在することの説明であり、因果的説明の一種である。

Robert Brandon は特定の生物学理論と独立の目的論的説明に批判的であるが、生物学での目的論的説明は認める。彼は機能よりは適応を考える。機能生物学は「どのように」という点に注目するが、進化生物学は「なぜ」に注目する。Brandon は「何のために」という問いに対する対象は適応だと主張する。表現型の特徴は自然選択の直接の産物である。形質 A が適応である時、目的論的説明の構造は「形質 A の存在

はAの過去の例の結果によって説明されるが、任意の結果ではなく、Aをもつものの適応性に関連した結果だけが言及される。」適応概念は「何のために」という問いに答えるために使われるゆえに目的論的であるというのが彼の主張である。

10.52 非歴史的な目的論

形態学者 Martin Rudwick は化石の機能の研究を通じて、機能の追求は工学的な原理が必要で、それは起源とは論理的に異なると考えた。これは機能には二つの異なる意味があることを示唆している。物理・化学的な性質に基づき構造がどのような性能をもつかということとそれが有機体の生活で果たす役割の二つである。前者は、有機体の環境に言及しない、その形態から生じるあらゆる物理・化学的性質であり、後者は、有機体の環境に言及する、その歴史の途中での有機体の行動や性質の使用すべてを含むものである。生物学的な役割から機能を区別すると、生物学的な適応がより明確になる。独立に機能を理解することによって、それが特定の環境においてどのような生物学的な役割をもつようになったかわかる。

Robert Cummins は上の考えに近い機能の説明を考えた。彼の見解は、有機体の形質の機能は「能力(capacity)」によって定められなければならないというものである。ある能力に関する形質の機能は有機体が能力をどのようにもつかを説明する助けになる形質の効果である。例えば、ハトの飛翔能力に関する翼の機能は、空に浮かぶようにすることである。というのも、これが翼が飛翔にどのように貢献するか説明するからである。彼の機能分析の記述は正に工学的な分析によってなされるものである。形態学者と同じように彼は非歴史的である。Cummins はそれまでの機能的な説明の理解 (e.g., Nagel や Hempel で、自然科学での説明として機能的説明は特別のものではないとする) は誤っていると考える。その誤りは二つの仮定にある。一つは、機能的説明は形質の存在についてのものであるとする点である。二つ目は、形質による生物学的な機能の行使はそれをもつ有機体の適応度を増加させるという仮定である。

Nagel は Mayr を批判し、機能と目標追求性(goal-directedness)の区別を注意深く論じた。生物学の機能的言明はそうでない言明に同値であると言うのが彼の見解である。彼の考えはさらに前向きに機能を考える John Bigelow, Robert Pargetter に受け継がれる。現存のものを未来の状態や出来事を使って記述するというのがその考えである。これは歴史的な機能の捉え方の反対である。したがって、ある形質の機能とはそれをもつ有機体に生存上の有利さを与える性向をもつ形質の結果、効果である。これは塩の可溶性の説明と同じような説明として機能的説明をかんがえることである。

Bedau は機能に異なる考え方を示す。彼は三つの異なる段階を考える。単なる良い結果が生まれる段階、それに基づく選択の結果の段階、意識的に良い結果を予期する段階である。生物学での目的論はこの二番目の段階で、三番目の段階を持ち込むことは自然主義と両立しないと考えられている。彼はこれが両立可能と考える。価値は実在的な消去できない性質で、倫理学のあり方とも結びついている。しかし、これはまだ緒に就いたばかりの問題である。

主要参考文献

- Block, N. (1980) 'Troubles with Functionalism', in *Readings in the Philosophy of Psychology*, Volume 1, Ned Block, ed., Cambridge, MA: Harvard University Press, 268-305.
- Churchland, P.M. (1981) 'Eliminative Materialism and the Propositional Attitudes'; reprinted in his *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press, 1989.
- Churchland, P. (1984) *Matter and Consciousness*, Cambridge, MA: MIT Press, 1984.
- Churchland, P. S. (1995) 'Can Neurobiology Teach us Anything About Consciousness?', in *The Mind, the Brain, and Complex Adaptive Systems*, ed. by H. J. Morowitz and J. L. Singer, Reading, MA: Addison-Wesley.
- Churchland, P. (1995) *The Engine of Reason, the Seat of the Soul*, Cambridge, MA: MIT Press.

- Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.
- Dennett, D. (1991) 'Real Patterns', *Journal of Philosophy*, 87: 27-51.
- Dennett, D. (1996) *Kinds of Minds*, New York: Basic Books.
- Dretske, F. (1987) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.
- Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Fodor, J. A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.
- Goldman, A. I. (1993) 'The Psychology of Folk Psychology', *Behavioral and Brain Sciences*, 16: 15-28.
- Gordon, R.M. (1986) 'Folk Psychology as Simulation', *Mind and Language*, 1: 158-171.
- Horgan, T. and G. Graham. (1991) 'In Defense of Southern Fundamentalism', *Philosophical Studies* 62:107-134.
- Jackson, F. (1986) 'What Mary Didn't Know', *Journal of Philosophy*, 83, 291-5.
- Kripke, S. (1972) 'Naming and Necessity', in *Semantics of Natural Language*, ed. by D. Davidson and G. Harman, Dordrecht, Holland: Reidel, 253-355.
- Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
- McClelland, J. and D. Rummelhart (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, 2: Cambridge, MA: MIT Press.
- McCulloch, W. and W. Pitts (1943) 'A Logical Calculus of the Ideas Immanent in Nervous Activity', *Bulletin of Mathematical Biophysics*, 5: 115 – 133.
- Millikan, R. (1984) *The Language of Thought and Other Biological Categories*, Cambridge, MA: MIT Press.
- Millikan, R. (1993) *White Queen Psychology and Other Essays For Alice*, Cambridge, MA: MIT Press.
- Minsky, M. L. and S. A. Papert (1969) *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press.
- Nagel, T. (1974) 'What Is It Like to Be a Bat?', *Philosophical Review*, Vol. 83, no. 4, 435-50.
- Searle, J. (1992) *The Rediscovery of Mind* Cambridge, Cambridge, MA: MIT Press.
- Sellars, W. (1956) 'Empiricism and the Philosophy of Mind', reprinted in *Science, Perception and Reality*, London: Routledge and Kegan Paul, 1963.
- Stich, S. P. (1996) *Deconstructing the Mind*, Oxford: Oxford University Press.
- Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.
- von Eckardt, B. (1993) *What is Cognitive Science?*, Cambridge, MA: MIT Press.