哲学Ⅰ金４　石原孝二先生

堀口　圭

心の哲学の意義

心の哲学は「心とはなにか」という難解な問いに哲学的、神経科学的アプローチにより答えを提示していこうとする試みである。実態のよくわからない「心」について考察することの意義は私が思うに、他者理解のありよう、社会の成り立ち、さらには、脳死問題など心と身体、生命についての倫理的課題に直面した現代に生きる私たちが今後どうこの問題に付き合っていくべきかということについての考察の土台となることである。「心」について理解することはきわめて困難な作業であるが、この課題に対する努力を怠っては、今後の科学技術の発展に伴う倫理的問題にうまく対処できず、結果として不幸の種を生むことになるのだろう。

哲学的方法

① 論証

② 反論

一般的意味における反論とは異なるので注意

ⅰ　異論：ある論証の結論に対し、それと対立する結論を論証すること

ⅱ　批判：ある結論の論証に対し、その論証が正しくないことを論証すること

意識のカテゴリー

現象的意識：自分のまわりの出来事を感じ取るという経験、そして直面した出来事によって引き起こされる感情や情動のこと。クオリア（辛いとか痛いといった質的な知覚）が現れる意識

アクセス意識（反省的意識）：現在の心の状態や、記憶に結び付いた心の状態について考えたり描写したりする能力。さまざまな情報の断片を結び付けて心的イメージや心的表象へと統合し、自らの行動と決定を導く能力。

クオリア（現象的意識）

”*What it is like to be a bat.*”（コウモリであるということ）

「コウモリであるとはどういうことか」（ネーゲル）

コウモリであるとはどういうことかはコウモリしかわからない。この、「～である」という意識

は感覚（痛み等の知覚）という、「質感」によってもたらされ、これを「現象的意識」と呼

ぶ

知覚（表象）と意識

デカルト　心身二元論

現代の哲学者のなかに二元論をとるものはほとんどいない。

われ思う、ゆえにわれあり（je pense, donc je suis）はデカルト「方法序説」の有名な一節である。デカルトによれば、「私」以外の何物であっても、その存在を疑い、否定することはできるが、どんな「懐疑論者」でも私が存在しない、と仮定することはできない、とある。これは、私は存在しない、と思考すること自体、私の存在なしに行うことのできないことであるからである。すなわち、「私は存在しない」という仮定はむしろ「私」の存在を何よりも明白に証明しているということである。ここまでの議論には常識的にうなずけるだろう。ここからが問題となるのだが、デカルトは考える精神というものは、存在するために場所を必要とせず、いかなる物質的なものにも依存しない、非物質的存在であるとし、物質である肉体が有限のものであるのに対し、精神は不滅であるとしている。このように、精神を物理的な存在から一切切り離された存在とみなすことが二元論と呼ばれるゆえんである。しかし、この二元論は２０C の脳科学の解明にともない否定されるようになってきた。また、論理的にも次のような矛盾を示す。それは、心物因果の問題と因果経路の問題である。まず、心物因果の問題について説明しよう。因果とは、すべての事象を原因と結果を用いて説明するものである。ある人がカレーを食べたいと思ったとする。この、カレーを食べたい、という欲求はカレーのにおいによって引き起こされるとする。そしてカレーを食べたいという欲求が、カレーを注文するという行動を引き起こす。こう考えたとき（相当単純化してはいるが）カレーの発する化学物質を感覚器官で電気信号に変換しそれが脳に伝わる。ここまでは全て物理的存在である。その物理的存在が非物理的存在である心に働きかけることを論理的に説明することは非常に困難である。なぜなら、ここで想定されていることは、たとえば、念力のようなものだからである。心で「曲がれ」と念じる（非物理）ことによってスプーンが曲がる（物理）のとまったく同じモデルなのだ。（念力によってスプーンが曲がるということは普通認められないだろう）次に因果経路の困難について考える。同様にカレーの例で考えてみよう。カレーの刺激が電気信号に変換されて、ある人の脳に脳状態 A を生み出し、脳状態 A が脳状態 B を引き起こし、脳状態 B が神経に電気信号を送り、カレーを注文するという行動を引き起こすという因果経路を考える。この際、二元論の立場では、この神経生理学的因果とは別の因果経路、すなわち、刺激→心的状態 A→心的状態 B→行動の存在を認める。ここで、ある因果関係について、二つの因果経路が存在してよいのだろうか。二つの因果経路のうち、どちらか一方は不要なのではないか。しかし、心物因果の成立は心の基本的特徴であり、容易に否定することはできない。一方、神経生理学的因果は認知脳科学の発展により容易に否定することができない。よってこのように二つの因果経路が成立してしまい、二元論の立場からの説明は非常に困難になる。

以下にロック、ヒューム、カントの意識についての考えを紹介するが、いずれも二元論的思考を基盤とするものである。

ロック

記憶に残っていない意識、経験は人格の同一性を構成しないものである。すなわち、伝達できないまとまった意識、経験は別の人格を形成しているといえる。

ヒューム

ロックが以上の論で言外に想定していた人格の同一性の存在をヒュームは否定する。ヒュームによれば、我々が人格の同一性、あるいは自己の観念と思い込んでいるものは、知覚の類似性による錯覚にすぎないのである。ヒュームにとって存在するのは知覚のみなのである。ここで、知覚の類似性について若干説明を加えようと思う。たとえば、我々が太陽を見

ているとしよう。それは、昨日は丸くて、今日も丸く、一年前も、おそらくは明日も丸いだろう。そうした知覚の類似性は「太陽」という存在の同一性を導出する。ここからは少し理解しがたいと思うものもいるかもしれないが、「太陽」の同一性は「自己」の同一性という「幻想」を創出するのである。詳しく説明すると、「太陽」の同一性は同一な「太陽」を同一に見ることができる観察者である「自己」の存在を前提としており、「太陽」から同一性を導出できるほどに同一な視点、感性をもった「自己」の知覚に類似性を見出し、その同一性を導出するのである。この考え方は「知覚」という精神を物質から隔絶した存在としてみなしている点でデカルトの物心二元論的思考の枠内ではあるが、デカルトが精神の不滅を主張したのに対し、ヒュームの主張は自己の同一性の根拠を知覚にのみ求めるものであるので、「私」（精神）の不滅を否定する立場にある点でデカルトと異なる。なぜなら、ヒュームの自己同一性はあくまで「知覚」に基づいている以上、知覚を可能にする肉体が滅びたのちに、精神のみが存在し続けることはありえないからだ。この点で、デカルト的二元論は若干の変異を遂げたといえる。

カント

カントといえば大陸合理論（デカルトなど）とイギリス経験論（ロック、ヒュームなど）を批判的に統合した「超越論的哲学」で有名である。これは客観的認識の条件を考察したものである。カントは直観の多様（コップがある etc）すなわち、感性的な表象は「私は考える」（私はコップがあると考える etc）という表象を必ず伴うとし、すべての表象は「私の表象」によって可能となる、と考えている点で、「私」以外のあらゆるものを疑って「私は考える」のみを真としたデカルトとは異なる。カントはこの「私は考える」という表象をあらゆる表象の基礎になっているという意味で、根源的統覚（純粋統覚）と呼んでいる。これと反対のものとして、経験的統覚がある。経験的統覚とは、自己の具体的な表象のことであり、「暑い」は経験的統覚である。（なぜなら、「暑い」の基礎には、「暑い」と「私は考える」という根本的統覚が隠れているからである。）

意識と無意識

フロイト

フロイトの意識についての理論は３つの時期で区分することができる。フロイトは１８８１年ウイーン大学医学部を卒業、１８８５年パリ留学でヒステリー研究で有名だったジャン＝マルタン＝シャルコーのもとで催眠療法を学んだ。フロイトはユダヤ人であったため、大学教授にはなれず、１８８６年開業医として多くのユダヤ系の弟子たちと臨床研究を進める。当時フロイトはヒステリー患者の治療を通して、外傷的な体験が苦痛な記憶となり、無意識的な力として神経症を引き起こすと考えた。この理論にもとづいて、フロイトは「お話療法」と呼ばれる治療法を生み出す。これは、患者の無意識的領域にある苦痛な体験を身体的症状に表出させるのではなく、意識化、言語化することによって症状を消失させようとするものであり、当時流行していた、「エネルギー保存則」をも参考にしたものであった。その後、フロイトはもっぱら無意識の考察に没入し、精神分析についての理論を構築していく。精神分析は古典理論と後記理論にわけることができる。まず古典理論について説明しよう。古典理論は第一局所論とも呼ばれる。局所論という呼称はフロイトが脳の働き（部位）を「意識」「無意識」「前意識」に分類したことによる。フロイトはヒステリー患者の誘惑的な外傷体験のほとんどが妄想であることに気づき、ヒステリーの原因を無意識（抑圧された願望）に求めた。そして、無意識的表象を意識的表象にもたらすことでヒステリーは治療されるとした。ここで前意識について軽く説明しておこう。前意識とは確かに知っていると思えるが、具体的にどういうことであったのか思い出せない記憶とか、知識などである。思い出そうとする努力を通じて、そのような記憶や知識が意識に甦り、思い出されるという経験も多数ある。あるいは、まったく忘れ去っていた、またはそんな経験などしたこともないと思っていたことが、思いがけない何かのきっかけで記憶に甦り、思い出すと言うようなこともしばしばある。無意識の領域にあったと考えられるが、何らかの努力や契機において意識に昇るような記憶や

知識、感情等は、「意識の領域」と「無意識の領域」の両方の領域に存在することになる。そこで、このような「心の領域」の特定部分を、「前意識の領域」と称し、略して、前意識と呼ぶのである。ここまでがフロイトの古典理論であるが、ここからは後期理論の紹介に移る。後期理論においてフロイトは脳を自我、エス、超自我という３つの働きに分類する。このため、後期理論はしばしば「第二局所論」とも呼ばれる。以下それぞれの働きを説明していこう。まず自我という言葉に注意しなければならない。自我とは日本語の一般的用法では、意識とほぼ同義で用いられるがここでの意味は異なる。自我とは、エスという人間の本能的欲求とそれに対する倫理的、理性的な抑制を与える超自我との葛藤と調整という機能を指す。フロイトはヒステリーの原因をエスの性欲動と攻撃性（死の欲動）にあるとしている。この後期理論は古典理論（第一局所論）で説明不能な現象を説明するために構築されたものであり、古典理論と相互排他的なものではない。

還元主義

２０ｃ以降の意識についての哲学はすべて認知脳科学の発展に多少なりとも影響を受け、そして、認知脳科学あるいはその発展に寄与したところの物理学、脳神経科学の成果を否定する立場の者は存在しなくなった。こうした立場はもはやデカルト以来の物心二元論によるものではなく、むしろそれを完全に否定し去る唯物論的立場をとるものである。唯物論的な発想（心脳同一説、構成主義、機能主義 etc）を理解するためには還元主義、とりわけ物理的還元主義の手法について理解しなければならない。まず、還元主義の辞書的な意味から紹介しよう。還元主義とは本来的には生物学の用語である。科学の扱う領域の中でおそらく生物学ほど複雑怪奇なものは存在しないだろう。なぜなら、生物の営みは我々の理解をはるかに超える複雑な相互作用の結果だからである。このような複雑な生命現象を理解するにはそれを根源的な事象に分解する必要がある。そしてそのように分解、還元された要素現象を一つ一つ理解することで、一見すると不可解に見える生命現象が理解可能なものとなる。これが還元主義の本来的用法であるが、意識の哲学で用いられる還元主義的手法はその対象を生命現象ではなく、意識という理解不可能な現象に向けられていることに注意されたい。意識についての還元主義的手法について説明する前に、まずは還元主義的手法を用いる際、一般に注意しなければならないことを述べておこう。それは、還元による説明の認識論は、そのままの形で付随性の形而上学にかなっていて、自然現象はそれがまさしく下位レベルの何らかの特性に論理的に付随しているとき、その特性を用いた還元によって説明可能だということである。つまり、ある自然現象を事例化する特性が下位レベルの特性に付随していた時初めて、その現象は当の下位レベルの特性を用いた還元が可能になるということである。（例えば水 H2O は水素と酸素の原子の特性に基づいてしか還元されえない）。さて、意識を還元主義的手法で説明しようとするとき、ある困難に直面する。それは意識が容易に物理的単位要素に分解可能なものではないからである。例えば水ならば、それは還元主義によれば H と O に分解することができ、それを単位要素と呼んで差支えないだろう。（もちろん中性子やニュートリノまで考慮すると果たしてそういってしまっていいのかは極めて怪しいものにはなるが）しかし、意識をそのような物理的単位要素に分解できるであろうか。すなわち、意識を H や O といった原子レベルまで分解したところでそれは意識についての説明になると言えるであろうか。還元主義の目的は、事象を根源要素に分解することによって、説明を試みることを目的としている。よって、物理的単位要素に還元したからと言って、それは意識についての説明には何ら役に立たないのとは明らかである。ここで、意識の説明に役立つ還元主義的手法とはなんであろうか。それは、機能への還元である。しかし、ここでまず確認しておかねばならないのは、今議論しているのは、現象的意識であるということである。なぜならば、心の二つの側面である心理学的側面と現象的側面のうち、私たちが今問題としているのはクオリア、すなわち現象的側面だからである。というのも、心理学的側面は行動と心理の因果関係で説明される、極めて単純な問題であり、それ自体は機能主義的に心の働きを説明したにすぎず、意識について最もその性格を本質的に示すのは現象的側面だからである。それでは、還元主義についての本題に入ろう。意識を還元主義的手法によって理解しようとするには、その機能的側面に着目して還元しなければならない。先ほどの説明と矛盾

するように思えるかもしれないが、心の現象的側面を理解するには、その機能に着目しなければならない。これは心理学的側面における「機能」が、行動と心の因果関係を意味していたのに対し、現象的側面における「機能」は、心的表象の因果的説明であるからだ。ここまでの説明で、意識を説明する上での還元主義的手法の意味は示したことになるので、以後の現代哲学における還元主義的手法は上述の意味であると理解されたい。

志向性と現象学

１現象学（マイペディアより引用）

原義は感覚的経験に与えられる現象、仮象を扱う学。狭義にはマッハの「物理的現象学」に学んだフッサールの哲学的立場をいう。すなわち、意識にも「客観的世界」の内部過程と見る実在論的、自然主義的前提を排し（現象的還元、エポケー)、あくまで意識に与えられる現象とその構造の記述のみが目指され（『イーデン』第一巻　１９１３年)、さらには「生活世界」での経験の記述が企図された。

２心の志向性

心の志向性にいて説明する前に、まずは志向性という概念を説明する必要があるだろう。志向性とは、あるものが何かを表しているときや、何かに向けられている、あるいは何かについてのものであるという性質である。志向性を持つものを表象といい、志向されるものを志向的対象と呼ぶ。たとえば「黒板がある」という言葉は目の前の黒板の存在を表す。言葉自体は黒板という志向的対象の表象であるということである。なお、ここで注意すべきは、志向的対象が実際に存在するか否かは表象にとってなんら本質的ではないということである。さて、心の志向性について考えて行こう。心的状態の志向性については二つの考え方がある。一つは、心の状態のうち、信念、欲求、感情、知覚といったものが志向性を持ち、なかには感覚のように、志向性を持たない心的状態も存在するという考え方である。もう一つは、すべての心的状態は志向性を持つという考え方である。(表象主義、志向説）この段階でどちらの考え方が正しいか議論するのは差し控えさせていただくとして、ひとまず両者の共通認識として、志向性を持つ心的状態が存在するということを前提に考えよう。志向性を持つ心の状態は、表象内容とそれに対する、信ずる、欲するといった態度の在り方という二つの要素によって区別されるものである。すなわち、同じ表象内容のものでも、態度が異なれば、心的状態は異なっているということができるのである。

２命題的態度

上述のように、心的状態の中には、志向性を持つものが存在すると考えられよう。このことが「心とは何か」という本講義の問いに対していかなる帰結をもたらし得るであろうか。ここで、一旦心から離れてもっと身近な表象について考えてみよう。私たちにとって目にすることができるもっとも単純な表象に言語と絵画がある。言語は、ある文章は何らかの現実（もしくは虚実）としての志向的対象を表象しているといえる。同様に絵画もその絵に描かれているものが現実に存在しているかどうかは問題とせずに、それ自体、何らかの志向的対象を表象しているといえよう。しかし、言語と絵画の表象の仕方は、哲学的議論によるまでもなく、何らかの相違があることは常識的に想像がつくだろう。そのことを、構文論的構造の有無という観点から論理的に考察していきたい。まず構文論的構造について説明する必要があるだろう。まず、表象によって表象されている特徴は、「志向的特徴」と呼ばれる。それに対して、表象自体に備わる特徴は「内在的特徴」と呼ばれる。ここで構文論的構造が存在する、というのは、表象が後者を備えていることを意味する。ところで、言語がある対象を表象するとはどういうことだろうか。たとえば、「私は今パソコンを操作している」という言葉が表象する志向的対象は、パソコンを操作する私である。このとき、先ほどの言葉は言語（日本語）の文法規則に則っている。つまり、言語には、語が構成規則に従って組み合わされ、文が形成される、という特徴がある。ここで、先ほどの言葉の「パソコン」という語は、たとえば、「電気屋さんでパソコンを買った」という文章で表されるパソコンの意味は同一である。(現

実に私が今使っているパソコンと電気屋さんで買ったパソコンは異なるかもしれないが、内在的特徴として同一であることに、それが表すところのものが同一である必要はない。）このように、言語には文脈独立性（構成要素である語がさまざまな文脈を通して共通の意味をもつこと）が存在し、文脈独立性のある構成要素が構成規則に従って組み合わされた、構文論的構造が存在する。一方絵画については、それは存在するのだろうか。赤いリンゴの絵と青いリンゴの絵はどちらもリンゴを表象するという点で共通するにもかかわらず、一方のどの部分をとっても他方には含まれていない。絵画において、構文論的構造が存在しないことを理解するためには、まずは、表象によって表象されている特徴と、表象自体に備わる特徴の違いに着目する必要がある。複数の異なるリンゴの絵が共有するのは、その志向的特徴である。なぜならば、それらの絵は、同じ「リンゴ」を表象してこそはいるが、絵の具の種類、量、キャンバスの質感など、ありとあらゆる内的特徴が異なるからである。よって、言語には構文論的構造があるのに対し、絵画にはそれは存在しないということになる。さて、ここで心は絵画と言語どちらの部類に入るのであろうか。この問いに対して、多くの哲学者の意見は、「～ということ」という形で言語化することができる信念、欲求などは、言語的な表象である、あるいは、言語的表象の一部であると考える。この「～ということ」は、「命題」と呼ばれ、信念などは命題に対する心の状態という意味で、「命題的態度」と呼ばれている。心の命題的態度が言語的表象であるゆえに構文論的構造をもつということは、のちに紹介する「心脳同一説」や「機能主義」の説明に大きくかかわることになる。少しだけ議論を先取りすれば、心脳同一説や機能主義が正しいとするには、心を生み出す脳が構文論的構造を有していなければならないということを意味するのである。しかし、心のもう一つの構成要素であるクオリアはどう位置付けるべきだろうか。クオリアは意識されるものではない、「私が私であること、現象的質感」であることは、冒頭に付した。だとすれば、これは心の内在的特徴と言えるのではないか。たとえば、痛みの感覚などはそれ自体何も志向していないように思える。しかし、クオリアのうちでも知覚などは志向性を持つように思える。実際、クオリアを、志向的特徴を持つものとするかどうかはしばしば議論の対象となる。すべてのクオリアを志向的な存在とする見方は「クオリアの志向説」あるいは「クオリアの現象主義」と呼ばれる。クオリアの志向説の立場から、痛みのような現象を説明すると、それは、痛みという心的状態によって身体の何らかの不調、損傷が表象されていると考えるのである。クオリアが心の内在的特徴であるか、志向的特徴であるかは物的一元論にとって極めて重要である。なぜなら、物的一元論に反対する立場をとる論者の根拠が、心の内在的特徴で非物理的存在であるクオリアの存在であるからだ。ここで、クオリアを心の志向的特徴であると考えれば、物的一元論の立場からのクオリア問題は解決するように思える。（なぜなら、痛みの知覚のようなクオリアが、身体の刺激および損傷に対する表象であるとすれば、非物理的存在であるクオリアの存在を根拠に精神の物理的世界からの分離を唱える物心二元論者を論破できるからである。）だが、先にもみたように、クオリアを心の志向的特徴とする見方にはなおも議論の余地があるし、仮にクオリアを志向性を持つものとしたところで、クオリアの志向説によってクオリアを物的一元論のうちに位置づけられるかどうかは、とりわけクオリアを志向的特徴とするような心の状態の志向性を物的一元論の中に位置づけることができるかどうかにかかっている。この点についても十分に明らかであるとは言えない。

以上までが心の哲学についての基本となる考え方、いうならば「序論」である。以下に上述の考え方をすべて理解したことを前提に、心と意識について、現代の哲学者がいかに考えてきたかを簡単に説明する。彼らの立場をすべて理解しようというのはこの講義のみでは不可能なので、各自参考文献を参照されたい。

心脳同一説

心脳同一説のテーゼ：各タイプの心の状態は特定のタイプの脳状態と同一である。

心脳同一説の立場の利点は、上述の因果経路の問題を克服できることにある。心脳同一説においては心物因果と神経生理学的因果は異なる二つの因果経路ではなく、まさしく同一の因果経路だったのである。さらに、心脳同一説においては心物因果の不可解さは一切生じない。(心物因果の不可解さはデカルトの二元論の項で説明) なぜなら心脳同一説においては、働きかけの主体であり客体である心は物理現象そのものであり、この点で一般の物理現象における因果関係となんら変わりないからである。このように考えると、心脳同一説は心の説明としてもっともなように思えるが、実際には心の問題はそれほど単純に解決するものではない。

機能主義

機能主義のテーゼ：各タイプの心の状態は、特定の機能で定義される状態である。

心脳同一説に対する、物的一元論の立場からの反論が機能主義である。心脳同一説が非難される点は、以下のような場合を想定することであらわになる。たとえば、科学技術が極めて進歩した近未来において、ある人は脳腫瘍の結果脳の全部を摘出し、その脳とまったく同じ機能を果たす人工脳細胞を摘出箇所に移植されたとしよう。心脳同一説が正しいとすれば、この人はもはや心を持つことはできないということになる。なぜならば、心脳同一説において心的状態は脳状態そのものであり、もはや脳を持たない人間にとって、脳状態は存在せず、よって彼には心的状態が発生しえないからである。しかし、これはあまりにも排他的な見方すぎないだろうか。同じ心を持つ人間を、その人の物理的構成要素によって心的状態を持つか否かを判断するというのは、かなり妙な考え方である。その人が感情を持ちうるのであれば、その人に心的状態が存在するといってよいのではないか。この点を改善しようとした立場が機能主義である。機能主義は、心の各タイプの状態は機能で定義されるとする見方である。ここで注意しなければならないことは、「機能」とは因果的役割だということである。つまり、各タイプの心の状態の本質は、それがどのような因果的役割を果たすか、もっとわかりやすく言えば、心的状態がいかなる原因で生じ、いかなる結果をもたらしたか、ということである。以下では、特定の機能で定義される状態を「機能的状態」と呼ぶ。機能主義によれば、同一の機能的状態を実現するのに物理的状態が同一であることは必ずしも必要とはされない。このように、同一の機能的状態がさまざまなタイプの物理的状態によって実現可能であることを、「多型実現可能性」もしくは「多重実現可能性」と呼ぶ。授業ではパトナムの機能主義を扱いそこではパトナム自身が気づいた機能主義の限界としてのクオリアについての説明がなされたが、上述のクオリアの志向説をとればこの問題が解決されることは明らかであろう。

生物学主義（サール）

この考え方は、いわば心脳同一説と機能主義の合いの子のようなものなので、説明は簡略でもよいと思われる。生物学的自然主義において、心は脳を原因とする（caused in）と同時に、心は脳において実現されている（realized in）のである。つまり、心は脳そのものではないが、脳において成立し、それは機能的状態を脳において実現するというものである。

指示の魔術説

―水槽の中の脳―

今、あなたが邪悪な科学者によって脳摘出手術をうけて、あなたの脳は巧妙な手術によりすべての神経がコンピュータに接続された状態で水槽に入れられたとする。あなたは普段通りに知覚するようにコンピュータから電気パルスが送られていると仮定しよう。このとき、あなたは、「私は水槽の中の脳である」という命題を証明することができるだろうか。確認

するが、あなたの脳のうちで起こる心的表象はすべてあなたにとっては真実にしかみえないのである。仮に「私は水槽の中の脳である」という命題が真であると仮定すると、あなたにとってのすべての心的表象は偽であるという前提に矛盾する。よって「私は水槽の中の脳である」という命題は偽である。それでは、このような命題が一見すると真になりうるように思えてしまうのはどうしてだろうか。それは、二つの誤謬から生じる。一つは、物理的な可能性をあまりに真面目に受け取ってしまうこと、もう一つは無意識的に指示の魔術説、すなわち、ある心的表象は必ず何か外在的な実在を指示する、という誤謬である。私が今問題とするのは、後者の誤謬である。これは上述の志向性に関する記述で述べたことだが、志向的対象が現実に存在するかどうかは、心的表象にとって何ら本質的なものではないのである。

消去主義

消去主義のテーゼ：心的表象および命題的態度等の心に関する現在の理論（常識心理学をも含む）は全て誤りであり、未来の科学の発展により、心理現象は物理的、化学的現象の一部として解釈される。

消去主義は物的一元論の一部でありながら、二元論はおろか、物的一元論のあらゆる心的状態についての説明、さらには常識心理学の正しさを全否定する。これは、相当ラジカルな立場ともいえるが、「将来の科学の発展によって」という但し書きは現在の私たちにとっての不可知性を想定する点で、神学的要素と同じにおいを感じる。どこか投げやりで胡散臭いのである。だが、消去主義の主張には現代神経科学の成果を反映したもっともな批判（おもに脳の構文論的構造に関する）を提示している。それは、ある命題的態度をとる際にそれに対応する脳もしくは脳と同じ機能を果たすものの構文論的構造の存在を否定することである。脳状態の構文論的構造は次のようにして否定される。脳細胞はシナプスの接続により神経系としての役割を果たしている。ところが、このシナプスの接続は１ルートに対して１つのものしか表さないというわけではない。一つのシナプスのネットワークは複数のことを全体に分散して重ね合わせることができる。この辺の理論は脳神経科学の分野であるから深入りはしないが、要するに、脳神経の構造は決して構文論的構造ではない、すなわち、文脈独立性を持つ要素からなっているわけではなく、したがって命題的態度が存在しえないということである。

哲学的ゾンビ

これまでの議論は、立場上の相違こそあれ、すべて物的一元論を支持し、意識を何らかの形で還元可能とみなすものであった。しかし、この考え方にも反論が想定される。仮に私と姿かたちが全く同じ、すなわち物理的性質が全く同一で、私と同じように知覚し、私と同じように反応するが、意識を持たないゾンビが存在するとしよう。（このゾンビは、映画で出てくるような理性を欠いたもの、すなわち、心の心理学的機能を持たない心理学ゾンビではなく、現象的意識を持たない現象学ゾンビである。）そもそもこのようなゾンビは存在しえないと反論する者もいるだろう。しかし、そのような反論は、このゾンビの現実世界においてそのようなゾンビは作れないということを意識的にせよ無意識的にせよ前提にしている。しかし、哲学的議論における仮定は、それが論理的に可能であるだけで十分なのである。そして論理的な可能性は記述描写が一貫しているときに存在する。この哲学ゾンビは意識の機能的説明ができなくなる。さらに言えば、意識は物理的なものに論理的に付随しないので、その還元は不可能なのである。同様の反証は何もゾンビを想定せずとも可能である。たとえば、スペクトルの反転を想定しよう。私と物理的構成が全く同一だが私が赤いと思うものを青と知覚するなど、現象意識がすべて反転した人間の存在を仮定すれば、先ほどと同様の結論が得られることは明らかである。

意識と物―人口知能―

強い AI と弱い AI

AI（Artificial Intelligence ）には「強い AI」と「弱い AI」という区分がある。（サール）強い AI とは「適切にプログラミングされたコンピュータは心にほかならず、プログラムそれ自体が人間の認知状態の説明である」と考える立場であり、弱い AI は「コンピュータは人間の心をシュミレーションするための有益な道具に過ぎない」とする立場であり、サールは後者の弱い AI を認めつつ、前者の強い AI に否定的な立場をとる。

中国語の部屋

強い AI に否定的なサールによる有名な思考実験がある。中国語を理解しない人が密室に閉じ込められ、中国語で話しかけられたときの手引き（中国語の A という音の連続に対して B という音の連続で答えればよいというようなことがその人の母語で具体的に示されたもの）をその人に渡す。ここで、中国人を部屋の外に連れてきて、中国語で中の人間と会話させる。このとき、一見すると会話は成り立っていて、機能面からは中の人は中国語を理解しているように見えるが、実際には手引きを読んで機械的に応答しているだけであって、何一つ理解していない、という結論が得られる。サールはこのようにして、機能面では完璧に人間の心的因果を実現していたとしても、それは実際に AI が心を持っていることを意味しないということを示そうとした。

強い AI の立場からの反論

構成不変性

意識が物理的なものとすれば、意識は脳の機能構成によって生まれるということになる。機能構成とは、システムの様々な部分の間にある、これらの部分と外部が交わす入出力の因果相互作用の抽象パターンである。シナプスなどを思い浮かべていただきたい。機能構成は、抽象的な構成要素の数、それぞれの構成要素がとりうる異なる状態の数、各々の構成要素の状態がすべての構成要素の先行する状態とシステムの入力にどう依存するか、システムからの出力が構成要素の先行する状態にどう依存するかを明細に示す依存関係の系統を明らかにしていくことによって決定される。すなわち、これを人間に関して言えば、物理的構成が全く同一の二人の人間は全く同一の意識をもつということである。このことから、意識は物理的状態から生ずるが、物理的状態そのものではないように、意識は機能構成から生じるが、機能的状態そのものではない、という結論が導かれる。

強い AI の擁護

チューニングマシン、プログラミング、有限オートマンなどは専ら、数式という抽象的な概念を扱う。一方に、人間の意識、実世界にある認知システムはもっと具体的なものであって、物理的な形態を備え、物理的世界にあるほかのものと因果的に相互作用している。しかし、私たちは、そうした具体的な物理的世界にあるものを理解するのにしばしば数学的実体である計算理論を使いたがるものだ。そのためには、抽象世界と具象世界の間にかかる橋、すなわち「インプリメンテーション」が必要となる。AI はコンピュータの一種であるのだから、その扱う対象は抽象的な数学的実体とならざるを得ない。ここで、強い AI にとって必要となるのは、このインプリメンテーションであるが、これは実際相当容易ではない。しかし、このインプリメンテーションとして「CSA」というシステムが想定される。CSA については詳しく説明することはできないが、その概要だけ説明することにしよう。CSA とは簡単に言えば FSA（有限オートマン）の集合で、機能的構成をなすものである。FSA とは、入力の有限集合、内部状態の有限集合、そして出力の有限集合を与え、それに結びつく状態推移関係を与えることで指定される、いかなる内的構造も持たない単純な要素 S1 である。神経細胞の一つ一つに対応するものである。CSA はベクトル｛S1,S2,S3・・・・Sn｝で規定される。つまり FSA の総体であり、脳に相当する。このように、AI が機能構成を持てば、上述の機能不変性からすれば、AI も意識を持ちうるという結論に達する。

AI のその後

AI 研究では AI にいかにしてフレーム（世界に対するインタープリテーション、解釈）を与えるかという困難に直面した。なぜなら、私たちの世界観は常識と呼ばれる知識（表象）の総体であると考えられたからである。しかしやがて AI 研究開発の過程で、ブルックスは「表象なき知性」という考え方に至り、フレーム問題の解決を試みた。それは、私たちは世界の明示的な表象やモデルは不要である。つまり世界をそれ自身をモデルとするのが一番よいという考えである。要するに従来の AI は知識（表象）をインプットすることで世界観を構築しようと試みたが、ブルックスは環境とのインタラクションの中で知性を実現できると考えたのである。

以降の内容は、今までの議論よりずっとわかりやすいと思うので、簡潔に紹介するに止める。

身体性認知

（1） 概念化：有機体の性質により制約を受け、その性質によって固有のものがなされる。
置き換え：環境と相互作用する有機体（身体）は従来認知の核とされていた表象に置き換わるものである。これはブルックスの「表象なき知性」と同じ考え方である。

（2） 構成
従来想定されていたのは、脳という認知システムに対し、身体の状態や外部環境が影響を与え心的状態を形成するという因果的プロセスであった。一方、身体性認知においては脳だけでなく、脳を含む身体とその周りの外部環境そのものが認知システムなのである。

（3） 拡張された心
Embodied:心的プロセス＝神経外の身体構造や身体プロセスにより構成
Embedded:心的プロセス＝主体の脳と外部の提携により構成
Extended:心的プロセス＝外部環境を含む
外部環境を認知システムに含むとはどういうことだろうか。これは認知システムの神経的プロセス（脳神経）とそれを表現する外部メディア（道具、記号、文字 epistemic action）の両方が認知システムという考え方である。この考え方はあまり実感がわかないかもしれない。たとえば、記憶障害のある人について考えよう。記憶障害のある人と、健常者の友人はともに共通の知り合いから「美術館は５３番通りにある」と聞いた。健常者の友人はそれを記憶して美術館にたどり着いた。一方、記憶障害のある人は、「美術館は５３番通りにある」という言葉をそのまま紙にメモしてポケットに入れて持って、適宜確認しながら、無事美術館に到着し友人と落ち合った。ここで、身体性認知の考え方によれば、健常者は「美術館は５３番通りにある」ということを頭の中で記憶していたのに対し、記憶障害者はその情報を紙に書いて保持していた。これは方法に違いこそあれ、両者ともに「美術館は５３番通りにある」という信念を持っていたという点では同じだということになる。よって、身体外のメディアも認知システムに含めることができるのではないか、というものだ。

ダマシオ

ダマシオは有機体と外部環境の相互作用の中で、外的対象（環境）に対する「表現」（脳内表現、内的表現）という「神経的表象（neural representation）が生じ、これが行動に影響を与えるとしている。しかし、彼の想定する、ニューロン回路に生じる生物学的変化がイメージを生む過程は、クオリアの「ハードプロブレム」（チャーマーズ）である。
彼はまた、身体と脳の関係について、身体の状態や外的環境の状態をモニタリングするためのものとして「心」が進化的適応の必要性から生じたとする。
さらに、人間には「文化的認知」と呼ばれる、他個体を意図主体として相互認識する能力があるとする。これにより、記号的人工物をもちいた社会的学習が可能となる。同様の概念として、「心の理論（theory of mind）」がある。これは他の

個体に心的状態を帰属させる能力のことである。

ヘテロ現象学

ここで再び「コウモリであるとはどういうことか」に戻って考えよう。しかし、今度は他者（人間）の意識についての考察である。ヘテロ現象学は、行動主義の客観性と現象学の主観性の中立的なもの、すなわち、「客観的現象学」である。これは被験者の体験の主観的性格を被験者の行動、様子から客観的に判断し、被験者にとっての「ヘテロ現象学的世界」つまり、被験者が持っていると信じている世界を理論として形成、一人称複数仮説的にもとづいて人間一般の意識について考察する手法である。一人称複数仮説とは「私」にとっての意識が「あなた方」にとっての意識と共通の性質を持っているとする仮説である。

ハンフリーの理論心理学

ハンフリーは「盲視」と呼ばれる、視神経に視覚情報が伝達されているにもかかわらず、「見えない」病気の患者に対し、さまざまな色を見せると、彼らが一定程度正答することに着目した。盲視患者は知覚の意識、すなわち知覚というクオリアがないのに、実際には知覚している。これは、知覚にとって、クオリアは本質的ではない、ということを示している。クオリア（現象的意識）とは、それを持つことにより、進化適応的観点からいえば、その生き物の考え方や望むもの、信じることに影響を与えるもので、その生き物は意識なしではなしえなかったような適応性のある形でこの世界で行動するようになるのである。また、ハンフリーは意識とは幻想であるとする。ハンフリーにとって、クオリアはグレガンドラム（実世界の源）の志向的対象（心の中の対象）に作用し生み出される幻想なのだ。なお、感覚刺激からの反応として誕生し、幻想を生み出すものを「イプサンドラム」と呼ぶ。外部表現を伴う反応が外部表現を消失する段階まで行くと（脳神経のイメージを作る部分を指すか）それはイプサンドラムとなる。こうして現象的意識を楽しむ、いやむしろ、享受することが生存にとって有利に働く。というのも、現象的意識、自己を存続させる行為を選択しようとする傾向がその生物内に生まれるからである。

心の理論（theory of mind）

社会は自己―他者関係で成り立っている。他者と関係を結ぶには他者の考えていること、思っていること、欲求など心の動きを多少なりとも理解できなければならない。それでは、人はどのようにして他者の心を理解し、社会を形成するのだろうか。「自己と他者」の問題は第二次大戦前後の行動主義全盛期にはほとんど議論されることがなく、１９５０年代後半の認知革命ののちもこのテーマの研究が直接なされることは少なかった。「心の理論」についての本格的な考察が始まるのは１９６０年代以降の実証主義的手法、脳科学の発展によってである。「心の理論」とは、他者の心の動きを読み取る心の働きについての理論である。そしてこれはモジュール性（機能としての独立性）を持つものとされた。「心の理論」は１９７８年にアメリカの動物心理学者デイビィット・プレマックが「チンパンジーは心の理論をもつか」という論文で提示した概念で、チンパンジーの社会性、すなわち、仲間に食物を分け与えたり、仲間を欺いたりという高度な社会的行動に着目し、他者に心的状態を帰属させること（imputing mental sates to others）を「心の理論」と定義した。この「心の理論」は１９９６年、霊長類学者のリゾラッティーらによって神経科学的なヒントを与えられることになる。リゾラッティーらは、マカクザルの前運動皮質の F5 領域（腹側前頭皮質）で、マカクザル自身が手指の運動を行うときと、ほかのマカクザルや人間が同様の動作をしているのを見ているときに、鏡に映したようにまったく同じニューロンが発火することを発見した。これは「ミラーニューロン」と名付けられ、「心の理論」における他者理解の方法の解明に大きく貢献するものである。結果、視覚運動ニューロンには、対象物の提示に対して反応を示すカノニカルニューロンと、対象物志向的行

為に反応するミラーニューロンの二種類があることが明らかとなった。しかし、このミラーニューロンの発見が直ちに「心の理論」の仕組みの解明につながるわけではない。というのも、視覚運動ニューロンは、「対象物の認識」と「対象物に対する行為の認識」を区別するが、「ミラーニューロン」は「心の理論」と「物の理論」を区別するものではないという結論が導かれるからである。なぜなら、マカクザルは鏡像自己認知獲得以前の段階の動物だからである。鏡像自己認知とは、字のごとく、鏡に映った自分の姿を自分だと認識する能力のことであり、鏡像自己認知によって鏡に映った自己の姿を自分だと理解できるという能力は、人間においても発達初期から備わっているものではない。人間は生後６か月～１２か月にかけては鏡に中の自分の姿を、だれか他者がいるように思ってしまう。１３か月から２４か月では子供は鏡の中に慎重に対処しつつ、鏡を避けるような反応を示し、２０か月から２４か月の子供はでは自己認知機能がしめされた。この自己―他者意識によってのみ間主観性の問題が意味を成す。すなわち、或る出来事が一人の主観においてのみ知覚される（事故においてのみ知覚される）のではなく、複数の主観、つまり、自己―他者関係の中で知覚ら意味が共通に成立しうるようになるのである。先の議論につなげれば、子供にとって最初の他者は母親であり、通常の母子間では発達早期に相互同期的で一体的な関係が成立するのであり、このことを第一次間主観性と呼ぶ。しかしこれは子どもの側に主観が存在しない状態で、子供の主観はいうならば母親の主観と同期しているのであるから、ここに自己―他者意識は存在しなく、よって「心の理論」も存在しえない。やがて、子供は養育者と視線を介したインタラクションを行うことができるようになる。このように自己とは異なる主観性を持った他者との主観性の共有が始まることを第二次間主観性という。「心の理論」が問題とするのはこの段階の間主観性である。間主観性とは他者の外的行動との同期的調整の模倣、他者の内的心理におけるそれである共感とつながるものである。ここで、先ほどのマカクザルには鏡像自己認知能力がないことが分かっている。よってマカクザルには他者―自己意識が存在しないと考えられる。自己―他者意識が存在しないマカクザルにおいてミラーニューロンの作用が発見されたことは、ミラーニューロンが「物の理論」と「心の理論」を区別しないということを示すものであり、ミラーニューロンの発見が直接に「心の理論」の解明にはつながらないという難点を提示しているといえよう。なお、「心の理論」の他者理解の様式には（１）シミュレーション説と（２）セオリー説がある。（１）シミュレーション説とは、或る出来事についての他者の心的状態を自己の経験に即して、自らが体験したもののように仮想することによって推測するとする説である。このシミュレーションは必ずしもミラー＝シミュレーション（ミラーニューロンによるシミュレーション）である必要はない。ミラー＝シミュレーション以外のシミュレーションとして、恐怖（扁桃体）などから生じる自他の共通活動部位は脳幹や体性感覚皮質など、身体感覚にまで及んでいた、という実験データがある。（ダマシオ　２００７）また、（２）セオリー説は他者理解のために、自己の経験に即してシミュレーションすることなく、他者の考え、信念といった心の動きを理解するには、論理的推論がなされているという考え方である。この例としてなじみ深いものが「偏見」や「ステレオタイプ」であり、これらは明らかに我々が「社会的知識」その他の知識を用いて他者理解を試みていることを示している。

参考文献

金杉武司『心の哲学入門』（勁草書房　２００８年）

デイヴィッド・J。チャーマーズ『意識する心』（白揚社　２００９年）

子安増生・大平秀樹　『ミラーニューロンと＜心の理論＞』（新曜社　２０１１年）