

## 書評

Jaegwon Kim,  
*Philosophy of Mind.*  
Westview Press, 2006.

坂井賢太郎

### 0. はじめに

本書の著者、Jaegwon Kim はブラウン大学の哲学教授であり、心の哲学、形而上学、科学哲学などで著名である。

世界は根本的に物質的なものであり、世界が物質的な粒子とその構造化された集合体でできていて、それらが全て正確に物理法則に従って動くということには説得力がある。本書の主な問いは、いかにして我々は心と心的状態 (mentality) をこの物質的な世界において位置づけることができるのかということである。本書の内容を私が特に興味を覚えた機能主義と心的因果(第 5, 6, 7 章)に重点を置いて紹介する。

### 1. 本書の具体的構成

第 1 章で Kim は、導入として、心の哲学で考えられていること、心を持つ主体であるとはどういうことか、そして心的状態間の関係、身体と心との関係などを概説する。その上で、心身スーパーヴィーニエンス、物理主義などの本書で扱う主要概念について説明がなされる。そして、人間が心的状態を持つこと、あるいはある状態が心的状態であることの様々な基準—認識論的基準、非延長性、志向性—について触れ、それぞれの内容が説明される。そして、心的状態を感覚的状态と志向的状态に分けて考え、例えばさきの基準のうち感覚的状态は認識論的基準をよくみたし、志向的状态はまさに文字通り志向性という基準をよくみたす。しかし、いずれの規準も両方の心的状態をうまく特徴づけられていない。ここで、私たちは、「どのような共通する性質」のおかげで、両方の種類の心的状態を「心的」と呼べるのかという問題につきあたる。心的状態はもちろん選言性質、すなわち感覚的または志向的な状態とすることは可能であるが、これは赤と丸の共通性を見出すのに「赤または丸」をみたすと言っているに等しいのである。ここから、Kim は私たちが心的状態であること一般の共通する性質を持っていないことを示し、それが心的状態について統一した概念を持つことを妨げていることを指摘する。

第2章では心を持つ、あるいは心的状態を持つことのある説明の仕方として実体二元論が取り上げられる。実体二元論は心と身体が本質的に異なる種類の実体であると考えられる。そして、とりわけデカルト主義的二元論のもとでは、心と身体は根本的に異なる実体であり、人間はこの二つの実体が統合されたものだとされる。

まず、心が身体と異なる実体であるという主張を検討するため、Kim はデカルトによる心と身体が同一でないという七つの議論（認識論的なもの三つ、形而上学的なもの四つ）を確認する。

次に、彼は実体二元論が直面する二つの問題、心身問題とペアリング問題を提起する。前者に対しデカルトは、心身の連合は最も原始的なものなので他の概念で説明できないと、十分に回答しなかったことはよく知られているので、後者についてのみ詳しく紹介する。

まず物理的出来事の因果関係である。ある物体 A があり、それと本質的な特徴を同じくする物体 B がある。A は C に物理変化をもたらすが、B はもたらさない。このとき、A と C はある関係 R（ペアリング関係）にあるが、B と C はその関係にないと考えると、何が基盤となって、この関係を成り立たしめているのかという問題が発生する。これがペアリング問題である。この基盤を Kim はそれぞれの物体の空間的關係でありそれ以外は思いつかないと述べる。すなわち、因果連鎖の中で A は C とある空間的關係にあるため、C の原因となり、B はそうでないため C の原因ではないのだ。これはそのまま心的因果に適用可能ではない。なぜなら、前述のとおり、心は非物質的であり、空間的なものでないからである。この意味で非物質的な心は因果関係を成り立たしめる基盤を空間的關係とは別に与えないとペアリング問題を超えられず、実体二元論は魅力を失うとキムは考える。

第3章では、行動主義が取り上げられる。行動主義の利点は他者の心のある意味で理解可能なものとして扱うことができるという点である。すなわち、私密的な心的表現を公的にアクセス可能かつ主観的に正当化可能な事実と条件で記述することにより心的表現の意味を解釈しようとするのである。この行動主義の立場として挙げられるのが「論理的行動主義 (logical behaviourism)」、「存在論的行動主義 (ontological behaviourism)」、「科学的行動主義 (scientific behaviourism)」である。この三つの立場は方法こそ異なるが、心的事象を行動という公共的に観察可能な次元に還元するという点で共通している。

キムは行動を (i) 「心理学的反応と応答」(ex. 血圧の上昇)、(ii) 「身体の動き」(ex. 窓を開けること)、(iii) 「身体の動きを含む行動」(ex. 書類にサインをすること)、(iv) 「身体の動きを含まない行動」(ex. 推論) の4つに分類する。行動主義は行動の公共的な観察可能性に主眼を置くため、iii と iv を i と ii のもとで考えようとする。しかし、これはうまく

いかない。なぜなら、行動主義は心的事象を行動に還元してしまうゆえに、心的事象を行動における因果的説明から締め出してしまうことになるからである。これでは、私たちの心理学的な説明は不完全なものにならざるをえないであろう。ここから、心身問題に取り組む別の理論が求められる。

第4章では、心脳同一説が検討される。私たちが心の働きは脳と関連していると考えるのは十分な理由がある。では、心は脳であると言えるのであろうか。脳が私たちの心的生活の座であるとする考えは、心的現象と脳内プロセスの相互関係が広く知れ渡っていて構造化されているからだ Kim は考える。では、この相互関係はいかにして正当化されているのだろうか。そこで、Kim は私たちが異なるものの相互関係を認めている例として七つの具体例をとりあげ、そこから、因果的相互関係 (ex. 動物精气)、心身の予定調和、機会原因説、二重相 (double aspect) 説 (ex. 雷鳴と稲光は同じ現実の別の相である)、随伴現象説、創初説 (emergentism)、心身同一説 (Mind-Body Identity Theory) (ex. 痛みと C 繊維の発火は同一の現象である) などの相互関係を説明する立場を挙げ、それぞれ特徴づけを与える。そしてこれらの説を評価する基準としてオッカムの剃刀を用いる。オッカムの剃刀のもとでは、先の七つの理論のうちで心脳同一説は最も有望であると思われる。そこで、心脳同一説における二つの立場が紹介される。

心脳同一説は通常、「出来事 (event)」について語り、心的出来事は脳の中の物理的出来事であるとする一種の物理主義である。ここで、この出来事を形而上学的にどのように考えるかで二つの立場が区別される。一方は心的出来事を世界の基礎的で具体的な個物とみなし、物質と同じように性質を持ち、ある種に分類されるとみなすトークン物理主義であり、他方は心的出来事のある対象による性質例化だと考えるタイプ物理主義である。

しかし、このいずれもうまくいかない。トークン物理主義はいかにしてある出来事トークンの心的性質が物理的に基礎付けられている、あるいは説明されるかについて語っていない。すなわち何が心的性質であるかについて語っていない。ここから、石や木が意識を持ち、犬や猫が意識を持たない世界は思考可能である。これはもはや物理主義とはいえない。そして、タイプ物理主義は、どうしてある対象がある性質を例化するかについて語っていない。ここから、タイプ物理主義は多重実現可能性からの反論にさらされる。例えばタイプ物理主義者は痛みが C 繊維の発火だと言う。しかし、人間とはかなり異なった神経システムを持つ爬虫類や軟体動物といった生物は痛みを感じられない有機体であるのだろうか。そして、私たちはいかにして全ての痛みを感じられる動物が C 繊維を持つと確認できるのであろうか。そして、私たちと非常によく似た複雑で豊かな精神生活を持つ地球外生命体であるがその生体は炭素でできていないというような生命体は存在し得ないのだから

うか。このような反論によってタイプ物理主義も引退に追い込まれる。

そして第5章では機能主義、とりわけ機械機能主義 (machine functionalism) が取り上げられる。機能主義は、行動主義、タイプ物理主義がうまく説明できなかった心的性質あるいは心的状態を機能的状態、すなわちそれが引き起こされる原因と引き起こす結果から解釈される内部状態とすることでこれを乗り越えようとする。例を挙げて確認しておく、機能主義的理解では、痛みを例えば細胞損傷の検知と考える。すなわち、それが行う因果的役割から特定化する。この役割を果たす以上、これを実現する仕組みはどんなものであってもよいと考える。すなわち、タイプ物理主義が越えられなかった多重実現の問題を越えられる。また、機能主義は行動主義が単なる行動、あるいは行動的状态 (Behavioural disposition) としか考えられなかった心的状態を「内部状態」として考えることでその実在を擁護することができる。また、因果的特徴から内部状態を考えることで行動主義では説明できなかった心的状態同士の関係を考えることができる。

続いて彼は Putnam が定式化したチューリングマシンを用いた機能主義のモデルを紹介する。チューリングマシンの紹介は省くが機能主義では心的状態を内部状態と考えるゆえに、心を持つとは次のようなチューリングマシンだと考えることができる。すなわち、あるものが心的状態を持つとは、物理的に実現された十分な複雑さを持ったチューリングマシンを持つことであり、その心的状態はその内部状態と同一である。そしてある有機体の心理は次のように考えることができる。ある心を持つ対象  $S$  の全ての可能なインプットとアウトプットのリストを  $V$  とし、 $S$  におけるインプットとアウトプットの対応を示したものを  $C$  としよう。このとき、 $S$  にとっての「心理」を構成するには、 $S$  の内的状態の集合を仮定して  $C$  を明確に体系化する理論を与えればよい。このように考えれば、心的状態あるいは心を持つことは、適切なチューリングマシンを持つことに本質があると言える。心的状態にとってある程度の構造的な複雑さを持つ脳を持つことは重要である。しかし、脳はたんぱく質をベースにした生物学的なものであることが重要なのではなく、それが物理的なチューリングマシンであることによって重要なのである。

ところが、この機械機能主義の心的状態の説明は、うまくいかない。すなわち、ある心的状態を持つことがあるインプットとアウトプットのリスト  $V$  を持つことにとって十分であったとしても、必要ではないかもしれないからである。それを示すのが John Searle によって考案された「中国語の部屋」という思考実験である。その概要を述べると、ある部屋に中国語を全く理解しない人が閉じ込められている。彼は「ルールブック」を持っており、それには文字列を別の文字列に機械的に変換する規則が書かれている。これは単純にその文字列の意味ではなく形に関しての変換に適用される規則だ。従って、その人は中国語を

全く知らなくてもこの規則を適用できる。ここで、彼は部屋に入れられた紙に書かれた中国語の指示に従って、中国語を返すというタスクを与えられる。彼は、ルールブックに従ってそのタスクをこなす。これは、中国語話者と全く同じ反応だ。しかし、彼は全く中国語を理解していない。ここから、インプットとアウトプットからその内部状態を同定することはできないという帰結が導かれる。

第6章では機械機能主義とは違うもう一つの理論、因果理論機能主義 (Causal-Theoretical Functionalism) が取り上げられる。両者の主な違いは、前者が心的状態をチューリングマシンの内部状態として考えるのに対し、後者は心的状態はそれが持つ因果的關係から定義されると考える点である。それは具体的には次のようなことである。次のような「痛みの理論」を考える。ここで  $x$  は心を持つ対象である。

(T) いかなる  $x$  にとっても、 $x$  が細胞にダメージを負い、かつ普通に気づくような状態であるならば  $x$  は痛みを持つ。そして、もし  $x$  が起きているならば、 $x$  は普通に気づくような状態にある傾向にあり、もし  $x$  が痛みを持つならば、 $x$  は身をよじり、うめき、不安な状態になる。そしてもし  $x$  が普通に気づくような状態にない、または苦しい状態にあるならば、 $x$  はいつもより多くタイプミスをする傾向がある。

ここで、傍点を付した表現は非心的な述語であり、観察可能な性質である。他方、太字の表現は心的な述語であり、観察不可能である。ここで、次の議論の準備として(T)にあらわれる心的述語を述語変項に置き換えて「ラムゼイ文」にする。

(T<sub>R</sub>) 状態  $M_1, M_2$ 、そして  $M_3$  が存在する。  $M_3$  は、いかなる  $x$  にとっても、 $x$  が細胞にダメージを負いかつ  $M_1$  にあるならば、 $x$  は  $M_2$  にあるようなものである。そして、 $x$  が起きているならば、 $x$  は  $M_1$  にある傾向がある。そして、 $x$  が  $M_2$  にあるならば、 $x$  は身をよじり、うめき、そして  $M_3$  になる。そしてもし、 $x$  が  $M_1$  にないまたは  $M_3$  にあるならば、 $x$  はいつもより多くタイプミスをする傾向がある。

(T<sub>R</sub>) を次のように短縮する。

(T<sub>R</sub>)  $\exists M_1, M_2, M_3 [T(M_1, M_2, M_3)]$

ここで「痛みを持つこと」は次のように定義できる。

痛みを持つ<sub>def.</sub>  $\exists M_1, M_2, M_3 [T(M_1, M_2, M_3) \wedge x \text{は} M_2 \text{にある}]$

ここで注意しなくてはならないことは、 $(T_R)$ によって、「痛みを持つこと」の概念はそれが持つ因果法則的關係によって与えられたことであり、 $M_1, M_2, M_3$ は $x$ が持つある状態としか言っていないため、「心的状態で心的状態を定義する」という循環に陥っていないということである。このように考えると、機械機能主義は述語変項 $M$ をチューリングマシンの内部状態と取る、因果理論機能主義の特殊なケースなのである。

このように二つの機能主義を紹介した上で、Kimは機能主義が相手にしなければならない三つの反論——クオリアの問題、交叉接続された脳の問題、機能的性質・選言性質・そして因果的力の問題——を挙げる。本稿では特に重要な第一の論点に絞って紹介する。

機能主義は例えば痛みの例化に共通するものをその因果的役割とするが、Kimはその痛みの例化ということにはもっと明らかな答えがあるのではないだろうかという疑問を投げかける。すなわち、痛みは「痛む」から痛みなのであり、それが共通しているおかげで「痛み」であるのではないか。より一般化すると、知覚 (sensation) はその現象的性質、すなわちクオリアによりその知覚であるのではないか。そして、機能主義者の知覚的な出来事をその因果的役割からとらえようとする試みは、その現象的な側面をすべてとらえこなってしまうのではないか。「逆転クオリア」の思考実験より、因果的役割と現象的性質が分離すると考えることは可能である。ここから、機能主義は全ての心的状態と性質を説明する理論ではないという批判にさらされる。

第7章では心的因果の問題が扱われる。心的出来事を含む因果的な関係は、私たちの毎日の経験でよくあることである。心が物理的な出来事やプロセスに何らかの仕方につながれていなければ、私たちはほとんど経験というものを持ってないだろう。そして、心と心間の因果関係も人間の知識にとって本質的なものであるだろう。推論を行う場合を考えよ。しかし、この心的出来事間の因果関係はそう簡単にはいかない。取り上げられる反論の一つが Davidson による心の変則性 (anomalism of the mental) であり、もう一つが Kim による「スーパーヴィーニエンス論法」である。より強力なスーパーヴィーニエンス論法を紹介する。

「スーパーヴィーニエンス論法」は物理主義を取りながら心的因果を維持することへの反論として用いられる。まずスーパーヴィーニエンス論法では、物理主義であるために維持しなければならないこととして「物理的因果の閉方性」を前提とする。物理的因果の閉方性とは因果関係に関して物理的世界は閉じているとするものであり、それ以外の、例え

ば心的領域から物理的領域へ因果関係が与えられることはありえないとするものである。スーパーヴィーニェンス論法の具体的な議論はこうである。

心的状態Mは物理的状態Pにスーパーヴィーンする。心的状態Mは結果としてM\*を引き起こし、このとき、M\*が例化されるのは物理的状態P\*が例化されたからとする。このとき、M\*が例化されたのは 1) Mが例化されたから、2) P\*が例化されたから、の二つの原因を持つことになる。これは過剰決定である。過剰決定を避けるためには、MはP\*を例化する形で因果関係に割り込んでいると考えることはできるがこれは物理的因果閉方を破ってしまう。残る道はMがPに還元可能であると考えよう。

第8章では意識という概念の整理と意識に関わる様々な議論が紹介される。Kimは、意識の語法として「主体意識 (subject consciousness)」と「状態意識 (state consciousness)」の二つを挙げそれぞれ特徴づけを与える。次に意識が持つ側面からの分類を行う。すなわち、意識は特有の質的特徴、すなわちクオリアを持つあるいは伴うものとそうでないものの二つに分けられる。そしてクオリアを持つ意識は、それが持つクオリアによってタイプ分類されるもの（痛みなど）と、クオリアを伴うが、第一義的にはそのクオリアによってタイプ分類されないもの（感情など）に分けられる。

そして、続く節では、現象的意識に関する「説明ギャップ」と「ハードプロブレム」の問題が取り上げられる。「説明ギャップ」とは、次のような問題である。ある神経的な状態Nが痛みを引き起こすとする。このとき、なぜNはかゆみやすぐたさではなく痛みを引き起こすのか、そしてなぜ、痛みはN以外から引き起こされないのか。この痛みとNの間に存在するギャップが説明ギャップである。「意識のハードプロブレム」とは質的な意識の状態はそうでない意識（機能的意識）と同じ仕方で取り組むことはできないとするものである。すなわち、例えば痛みは機能的な性質を持つが、それが全てではない。むしろ、私たちが持つ痛みが痛みであるのは、その機能的特徴によるというより、その質的特徴、すなわち「痛む」からと信じる傾向があるからである。この点で、質的な意識の状態は機能的意識とは異なる特徴づけが必要となる。

第9章では、心的内容 (mental content) が取り上げられる。私たちの「志向的状态 (intentional state)」は「内容 (content)」 (ex. 今日は寒いということ、明日はもっと暖かいということ etc.) とそれへの「態度 (attitude)」 (ex. ～と信じる、～と疑う etc.) からなる。心的内容はある状況 (a state of affairs) を表象する。ここで、異なる主体は同じ状況について異なる志向的態度を持つことができ、同じ主体がその状況について異なる態度を持つことも可能である。しかし、何のおかげで、このような志向的狀態は、すなわち命題的態度は、その意味を持ち、そしてそれらが意味する状況を表象するのか。本章では、これを説明する理論とし

て「解釈理論 (interpretation theory)」、「情報意味論 (information semantics)」、「目的論的アプローチ (teleological approach)」、「内容外在主義 (content externalism)」が検討される。

第 10 章では心の還元可能性について論じられる。結論としては、意図的-認知的な心的状態は機能化可能であり、従って物理的状态に還元可能かつ因果的力をもつことができる。また、その他のクオリアのような心的状態も経験において差異を示すという側面に注目すれば機能化可能である。この点で、両方の心的状態を随伴現象説から救うことはできる。しかし、質的な心的状態としてのクオリアそれ自体は随伴現象説から救うことはできない。このことは物理主義の限界を象徴していると Kim は結論付ける。

## 2. 最後に

本書はその名の通り、心の哲学に関する様々なトピックとそれに関する議論を幅広く扱っている。それぞれについて丁寧な解説と評価、具体例があげられており相互に参照可能なようにそれぞれのトピックの間関係についても述べられているため、現代の心の哲学を学ぶために非常に良い入門書となると思う。また、心の哲学について一通り学んだ人にとっても、それぞれのトピックの確認とそれぞれの章末に付された **Further Readings** からの更なる研究に進んでいくことができるという点で有益な書であると思う。

しかし、私は本書の結論には納得していない。というのは、クオリアの質的特徴なるものについて、私たちがその本質的なところにはまだ議論が届いていないため説明できないというだけで、物理主義の限界を象徴しているとは私には思えないからである。それは、現代の心の哲学全体の問題であると私は考える。そもそも、クオリアとは極度に私秘的なものであるため、Kim の考えているクオリアと私が考えているクオリアといったものが共通点をもっている保証はないのである。クオリアはしばしば物理主義への反論に使われるが、それを使う者たちはそのクオリアなるものが存在 (あるいは例化) されているということをしちんと言うことができるのであろうか。このような点を踏まえると、物理主義の限界を見定めるのはまだまだ早いと私は考える。

[京都大学大学院修士課程・哲学]