

内容と心的因果：思考は「力」を持てるか

NAKAGAMA, Koichi / 中釜, 浩一

(出版者 / Publisher)

法政大学文学部

(雑誌名 / Journal or Publication Title)

Bulletin of the Faculty of Letters, Hosei University / 法政大学文学部紀要

(巻 / Volume)

83

(開始ページ / Start Page)

1

(終了ページ / End Page)

19

(発行年 / Year)

2021-09-30

(URL)

<https://doi.org/10.15002/00024743>

内容と心的因果

— 思考は「力」を持てるか

中 釜 浩 一

0. 序

Bernard Williams はあるところで、「魔術」の特徴を、「物理的結果を生み出すと想定されるものが、記号 symbol の志向的意味内容 significance であって、当の記号を物理的に実現するものが持つ特定の〔物理的〕特質ではない、という考え」だと述べている⁽¹⁾。たとえば、テレパシーとは、他人の心の「内容」を、発言される言葉や書かれた文字のような何らかの外的・物理的媒介物（記号を物理的に実現するもの）によることなく、直接に感知する能力のことだろう。この想定の下では、自分の理解しない言葉を話す外国人や宇宙人の心の「内容」も、テレパシーによって直接知られると想定される。つまり、テレパシーの能力を持つ人とは、他者の心の中の「志向的意味内容」が直接にその人の心に因果的効果を持つような人だ、と考えられているのである。あるいは、テレキネシス（念力）が可能だと考える人は、たとえば「あの岩を持ち上げよう」というような心の中の欲求の「内容」（思念）が、身体の物理的活動を一切経由せずに、（時には物理的に可能な範囲をはるかに超えて）当の岩に直接因果的に働きかけることが可能だ、と想定しているだろう。

近代科学が、このようなオカルト的・魔術的能力の存在を否定したことは、周知の事実である。もちろん「魔術」の存在への信仰・願望は、いまでもわれわれの心の中に潜んでいる。映画や物語の中ではしばしば、「科学では説明できない特殊な能力」を保持する超能力者が活躍し、「肉体の死後にも死者の魂が残り、その魂がわれわれ生きている者に働きかけることができる」と信じる人は、当の魂が身体活動のような物理的媒介物によらずに、魂の持つ思考の純粋な「内容」によって、生きている者達に因果的な作用をする、と考えているのだろう。だがそうした願望や信仰は別として、「魔術的なこと」が現実にかかることは、われわれは普通考えない。現実には「魔術」の働く余地はもはやない、というのがわれわれ現代人の共通理解だろう。

「物理主義」とは、こうした願望や信仰の内に含意されている「魔術的力」の存在を否定し、そうしたものがわれわれが現に生きているこの現実世界に働くことを、徹底的に拒否するような形而上学的立場である。物理主義者は、「この現実世界に実在し因果的に働きかけうるものは、物理科学によって要請される存在者（素粒子、原子、分子等々）か、何らかの仕方ですれらに帰着されるものだけであり、

それら実在するものの間には、最終的に物理科学的法則で記述・説明できないような因果的關係は存在しない」と主張する。したがって物理主義では、「生命」や「心」のような生物学的現象、あるいは「会社」や「野球」のような社会学的存在者・制度など、一見して「非物理的」と思われるものであっても、その因果性（その力）を主張するためには、法則的還元・科学的同一性・付随性 supervenience 等々の関係によって、最終的には何らかの形で「物理的なもの」に帰着されなければならない、と主張される。逆に、そのような仕方では「物理的なもの」に帰着できないものの因果的実在性は、否定されることになる。後に見るように、物理主義がこうした仕方では「魔術」を否定する根拠は、「物理的世界の因果閉鎖の原理」や「因果的排除の原理」に基づいているだろう。

だが、このような「魔術的力」を拒否する「物理主義的世界像」のもとでその身分が危うくなるのは、魂や幽霊や超能力のようなオカルト的存在者ばかりではない。われわれが日常的に強く信じている「思考の因果性（思考の力）」自体もまた、危うくなってくるのである。この論文の目的は、こうした事情を確認するとともに、「思考の力」を認める余地が物理主義的世界像のもとでもなお残されているかを検討することである。

1. 「内容」と因果

われわれは普通、「自分の思考の「内容」は、自らの行動を決定するための、唯一ではないにしても、きわめて重要な一つの因子だ」と信じている。たとえば、われわれは「明日映画館に行きたい」という「内容」の欲求や「見たい映画は2時に開演する」という「内容」の信念がある人が持つことが、「明日何時に起きて、何時に家を出るか」というその人の決断を「因果的に説明する」と考える。あるいは、「1時に家を出る」という自分自身の行動（物理的身体動作）を決定したのは、上のような「内容」の欲求や信念を私が持ったことだと考える。こうした「民間心理学的・合理的な行為の説明」（デネットのいう「志向的構え」による説明）は、単に他者の行動を「予測」するための装置ではなく、自分自身や他者の行動を因果的に決定する「実在的要素」だとわれわれは強く信じている⁽²⁾。

だがその一方で、非物理的なものの因果的力を否定する物理主義の立場に立つなら、われわれの行為を因果的に決定する「思考」の正体は、当の思考を持つ人の脳の中で起こる生理的・物理的出来事（物理的存在者）の一つに過ぎない、と言わざるを得ないだろう。そしてそうであれば、ある人の「思考」がその人の行動にどのような影響を与えるかは、当の思考の「物理的出来事（脳現象）」としての在り方であって、それが表象する「内容」なのではない、と考える以外にないように思われる。

もう少し具体的な例で考えてみよう。たとえば、「あの道の角に、人によく吠えかかる犬がいる」という「内容」pの思考（信念）を私が持てば、（私が犬に吠えられることを好まないとすれば）私はその道を避けて通るという回避行動をとるかもしれない。だが、物理主義的な考え方によれば、私がそのような回避行動をとった真の原因は、当の思考を持った際の私の物理的な脳状態Bであって、Bが表象している「あの道の角に、人によく吠えかかる犬がいる」という事態、すなわち思考が表象する「内

容」 p なのではないことになる。もしもそうした「志向的内容」が、直接に私の身体的回避行動の原因になると考えるとしたら、それはまさに「念力」のような仕方で私の身体に働きかけると想定することになるだろう。物理主義者はそのような「内容」の直接的因果効力を否定しなければならない。つまり、思考の「内容」は行動の決定に関わらないことになる。

「物理主義は思考の「内容」の因果的力を否定することになるのではないか」という上の疑念に対して、物理主義的立場から答えることが不可能ではないと思えるかもしれない。たとえば、「われわれの思考の「内容」は、直接にわれわれの身体運動を決定するわけではないが、その「内容」を表象するようなある物理的状态が脳の中で「実現」されているという事実を介して、「内容」は身体の行動を間接的に規定している」と答えることができそうに思える。だが、こうした仕方で物理主義者が「思考の内容は「間接的」に因果的効力を持ちうる」と主張できるかは、はなはだ疑わしい。

このことを示すために、次のような状況を考えてみよう。この現実世界 W_A において、日本語の文を全く理解しない外国人 S も、別の言語（たとえばタガログ語）で「あの道の角に、人によく吠えかかる犬がいる」と表現される内容 p の思考を持つことができるだろう。そして、その内容 p の思考に基づいて、私と同様の回避動（これをタイプ A の行動と呼ぼう）をとることもできる。そうした思考を持つ際の S の脳状態 B_s は、私の脳状態 B とは、「物理的状态」として記述される限り、全くタイプを異にするかもしれない。実際、 B あるいは B_s が「あの道の角に、人によく吠えかかる犬がいる」という内容を表象するようになるに至るまでの私と S との「個人的履歴」は全く異なっている。われわれは、全く異なる環境で生まれ育ち、全く異なる犬に関して異なる経験を持ち、全く異なる言語を習得したのである。だがそれにもかかわらず、物理的に異なるタイプの脳状態 B と B_s とは、同様の物理的結果（タイプ A の行動）をもたらすことができる。

さて、 B と B_s の物理的タイプが異なる以上、 B から A に至る生理学的過程と、 B_s から A に至る生理学的過程ともまた、タイプとして異なる因果過程であるだろう。それら異なる状態（ B と B_s ）が、異なる因果過程（ $B \rightarrow A$ と $B_s \rightarrow A$ ）を通じて、同じタイプの物理的結果（ A ）をもたらすことの説明は、 B と B_s とが同じ「内容」 p を表象していることによる、と言いたくなるかもしれない。したがって、同一の「内容」 p の思考は、異なる脳状態 B と B_s とに「実現できる」という「間接的」な仕方で、 A の生起を因果的に規定している、と言えそうに思えるかもしれない。

だが、このことは、内容 p が行動 A を因果的に規定するということとは異なる。以下で示すように、物理主義者は B と B_s とが共有する「内容」 p を、 B と B_s との「共通原因」と考えることはできない。 B や B_s が「内容」 p を持つことには、因果的な必然性はないからである。 B と B_s の直接の原因は、それらとは別の脳状態でしかありえない。内容 p と脳状態 B や B_s との関係が因果的必然性を欠くなら、 p が B や B_s を介して「間接的」に A の原因となる、ということも言えはざである。したがって、内容 p は（魔術的力を持たない限り）「直接的」に身体に影響することがありえないのと同様に、脳状態を介して「間接的」に身体に影響することもあり得ない。

「p」が表す事態がBやBsの「共通原因」ではありえないことについては、以下のような仮想的状況を考えればさらに明らかになる。

Sの「個人的履歴」は現実の場合とほぼ同じだが、私の個人的履歴が現実のそれとは十分に異なっているような物理的可能世界（現実世界 W_A と同じ物理法則が成り立つ可能世界）Wを考えてみよう。そしてWにおいて、Sの脳状態Bsは現実世界 W_A と同じく「あの道の角に、よく人に吠えかかる犬がいる」という内容の事態pを表象するのに対して、私の脳状態Bは「あの道の角に、おとなしい猫がいつも眠っている」という内容の事態qを表象しているとしよう。このようなWにおいて、脳状態Bにいる私と脳状態BsにいるSとは、pとqという異なる「内容」を表象している。だがそれにもかかわらず、二人が行う行動は現実世界 W_A で行うのと同じ回避行動である。 W_A とWでは同じ物理法則が成り立っているから、BからAに至る因果過程（ $B \rightarrow A$ ）とBsからAに至る因果過程（ $Bs \rightarrow A$ ）とは、 W_A とWとで同じタイプの物理的過程のはずだからである。このとき、たとえある事態を表象する「内容」pがBsの「間接的原因」だと認めたとしても、それはqを「内容」として表象しているBの原因ではありえない。したがって当然、WにおいてはpはBとBsの「共通原因」ではありえない。 W_A において「思考が内容pを表象していなければ、回避行動Aは生じなかっただろう」という反事実条件文も成り立たないから、「内容」pはAの原因ではない。よって、「内容」pはそれを表象する脳状態を介して、行為Aの「間接的」に原因となる、と主張することもできない。

さて、上の議論が成り立つなら、私の行動を「合理的なもの」として説明するという試みは、物理主義とは両立しないことになる。合理的説明とは「内容」による説明だからである。Wにおいて私の脳状態Bは「あの道の角に、おとなしい猫がいつも眠っている」という「内容」qの思考を表している。qは私の回避行動Aを合理的に説明しない（おとなしい猫がいることは、私がある場所を避けて通る「理由」にはならない）が、それにもかかわらず、脳状態Bは回避行動Aを引き起こす。したがって、いかなる「理由」をもつか（いかなる内容の思考をもつか）は行動Aを説明しない（内容は説明的に無関連である）。よって、物理主義のもとで思考の「内容」が因果的力をもつことがないとすれば、「合理的説明」は「真の説明」ではなく、たまたまこの現実世界で成り立っている「相関関係」を示すだけの単なる「見かけの説明」以上のものではなくなるだろう。

それでは、こうした「内容」の因果的力を否定する結論を退けるために、上で想定した私の脳状態Bが内容qを表すような可能世界Wは「物理的に不可能」だ、と主張することができるだろうか。すなわち、脳状態Bは「あの道の角に、おとなしい猫がいつも眠っている」という「内容」qを持つことは決してなく、常に「あの道の角に、よく吠える犬がいる」という「内容」pまたはそれと十分近似した「内容」p*だけしか表象できない、と主張することは可能だろうか。

このような主張を行うためには、脳状態Bとその（外在主義的）「内容」pとの間に何らかの「必然的結びつき」があることが、物理主義的に説明されねばならないことになる。だが、これは極めて困難に思える。たとえば、われわれが持つ水の観念は、環境の違いによって、 H_2O を表象することもXYZを表象することもできる⁽³⁾。したがって、地上の私と双子地球上の私*が持つタイプ同一な脳状態Bは、

「まわりに H_2O がある」という「内容」をもつことも「まわりに XYZ がある」という「内容」を持つこともありうる。すなわち、脳状態のタイプと「内容」のタイプとの間には、物理的必然的關係はなく、その脳状態が置かれている環境等の特定の偶然的で「指標的要因」によってその「内容」は定まる。こうした標準的な「外在主義的」考え方を取るなら、内容「p」と脳状態 B との必然的結びつきを因果的に説明することはできず、脳状態タイプとその内容タイプとの間の「法則的必然性」あるいは「法則的付随性」を主張することは困難である。そうであれば、B が内容 p ではなく内容 q を表象するような物理的可能世界を想定することに、論理的問題はないだろう（この問題については最後にもう一度触れる）。

こうして、物理主義的な形而上学の枠組みのもとでは、「内容」と「理由」による行為の合理的説明は真の因果的説明とはならない可能性がある。以下では、物理主義の立場に立ちながら「内容」の因果性を考えることが可能なのかという問題について、Kim と Woodward という二人の代表的論者の見解を批判的に検討しながら論じることにする。特に Woodward らによる「介入説」は、「内容」が身体的物理的過程に介入するという事態を物理主義的に理解可能とするかを考察する。

2. 付随／排除論法と「内容」の因果性

Jaegwon Kim は *Physicalism, or Something near Enough* [2005] をはじめとするいくつかの著作の中で、「付随／排除論法」と称される論法を展開した⁽⁴⁾。「付随／排除論法」とは、「心的性質 M_1 が脳状態 B_1 に付随する」という物理主義的な「心の哲学」の基本的主張を認めたときに、「 M_1 が他の脳状態 B_2 や他の心的性質 M_2 に因果的に影響できる」という可能性を否定する議論である。もしも思考の「内容」を脳状態に付随する「心的性質」の一つと考えたとすれば、Kim のこの論法は「内容」の因果的効力を否定するものと解釈することができるだろう⁽⁵⁾。

Kim の議論を要約しよう。Kim によれば、「付随／排除論法」とは、(A)物理的因果閉鎖原理、(B)因果的排除原理、(C)心身付随テーゼ、の三つを前提とする論法である。これら三つの前提は、物理主義者であれば当然受け入れなければならないものとされる（すなわち、物理主義的態度をとるということは、上の三つの前提にコミットするということである）。より詳しく述べれば、

(A)物理的因果閉鎖原理。Kim の定式化によれば、これは「時刻 t で物理的出来事 e^p が何らかの「十分な原因を持つなら、 e^p は t で「物理的に十分な原因 c^p を持つ」という原理である。

この原理は、「物理的出来事の因果的説明に非物理的存在者がかかわると考える必要はない」ということを示している。ただし「因果閉鎖原理」は、原因を持たない物理的出来事が存在しうることを否定せず、また非物理的存在者が存在することを排除するものでもないから、この原理自体は物理的決定論や物理的還元主義にコミットするわけではない。それが主張するのは、ある出来事が何らかの「十分な原因」を持つような「物理的」出来事として記述される限り、その出来事には「十分な物理的原因」が存在する（非物理的原因を挙げることは不要だ）、ということだけである。

この原理は、物理領域の「因果的自足性」を主張するものであり、物理主義者は物理的領域においてこの原理が成り立つことにコミットしている。これによれば、物理的現象を引き起こすために「魔術」は必要ないことになるだろう。ある物理的現象が十分な原因を持つなら、それには十分な物理的原因がなければならないから、「魔術」を使わなくてもそれは起こったはずである。

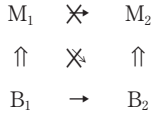
(B) 因果的排除原理。これは「出来事 e が時刻 t で「十分な原因」 c を持つなら、(過剰決定のケースでない限り) e は t で c 以外の十分な原因を持つことはない」という原理である。この原理は、 e が c によって因果的に決定／生成されるなら、 e が c 以外のものによって決定／生成されることはない、ということを実証する。

この原理もまた、それ自体は必ずしも「(物理的であれ非物理的であれ) あらゆる出来事が十分な原因を持つ」という決定論にコミットするものではなく、また「物理的原因」以外の何らかの原因の存在することを排除するものでもない。この原理のポイントはむしろ、同一の出来事に対して二つ以上の「十分な原因」が存在するような「過剰決定」が、きわめてまれな偶然以外に生じないことを示すことにある。二つの出来事 c_1 と c_2 とが同一の出来事 e の十分な原因であるような「過剰決定」overdetermination の事態 (たとえば、ある人の死(e)が、その人の脳と心臓に同時に二発の弾丸が命中すること(c_1 と c_2)によるような場合) においては、通常の因果判断が含意するような「 c_1/c_2 が起こらなければ e は起こらなかっただろう」という反事実条件文が成立しなくなるという因果的変則が生じる。因果排除原理は、こうした過剰決定的な事態が、上の例のような奇跡的偶然以外には、一般に生じないことを主張するものであり、われわれの現実的世界の因果構造に関する一般的形而上学的前提である。

この原理によって、「魔術」は物理的出来事の生起にとって「必要」でないばかりでなく、物理的出来事の説明から積極的に「排除」されることになる。もしも「魔術」が物理的出来事 e^p が十分な原因であれば、因果閉鎖原理により e^p は十分な物理的原因 c^p を持つから、 e^p は c^p と「魔術」とによって「過剰決定」されることになってしまう。だがそのような「過剰決定」はこの世界では基本的に起こらず、「魔術的原因」は物理的原因によって「排除」されることになる。

(C) 付随テーゼ。これは、「すべての心的状態は、何らかの物理的な脳状態に付随する」というテーゼである。「付随性」supervenience の解釈には強弱様々のものがありうるが、ここでは「物理的脳状態の変化なしに、心的状態の変化はない」というごく一般的意味で理解することにしよう。Kim によれば、付随テーゼは、「科学定同一性」や「法則的還元」よりも弱く、心を脳状態で説明しようという物理主義的な心の哲学が前提すべき最小限のテーゼであり、このテーゼを認めないような心の哲学は「物理主義的」なものとは言えないことになる。

さて Kim は、これら三つの前提を認めるなら、ある心的状態 M_1 が脳状態 B_2 の物理的原因になったり (下向き因果)、別の心的状態 M_2 の物理的原因になったり (水平因果) することは否定されると主張する。Kim の考えはしばしば、以下のような図形 (Kim の図形) で例示される。



ここで、「 $\uparrow$ 」は付随関係を、「 $\rightarrow$ 」は因果関係を表す。この図形が表現するのは、「 M_1 が B_1 に付随し、 M_2 が B_2 に付随し、 B_1 が B_2 の十分な原因であるとき、 M_1 は B_2 の原因でも M_2 の原因でもありえない」ということである。

Kimはこの結論を、以下のように論証する。

- ①心的出来事 M_1 が、他の心的出来事 M_2 の原因である [$M_1 \rightarrow M_2$] か否かが問題である。
 - ②心身付随テーゼより、 M_2 は、 M_2 と同時に生起している物理的な基盤の出来事（脳状態） B_2 によって決定されている [$B_2 \rightarrow M_2$]。
 - ③因果的排除原理(B)より、 M_1 か B_2 の両方がそれぞれ独立に M_2 の原因であることはない [$M_1 \rightarrow M_2$ か $B_2 \rightarrow M_2$ かいずれか一方しか成り立たない]。
 - ④③より、 M_1 が M_2 の原因となりうるのは、 M_1 が B_2 の原因となる [$M_1 \rightarrow B_2$] 場合（「下向き因果」が存在する場合）のみである（心的性質間の「水平因果」の否定）。
 - ⑤心身付随テーゼより、 M_1 には物理的付随基盤 B_1 が存在する。
 - ⑥因果閉鎖原理(A)により、 B_2 には因果的に十分な物理的原因が存在する。 B_1 はその候補である [$B_1 \rightarrow B_2$]。
 - ⑦同一説をとらないなら、 M_1 と B_1 は異なる出来事であり、④と⑥より、 B_2 は M_1 と B_1 という異なる二つの十分な原因を同時に持つことになる。
- だが、
- ⑧因果排除原理(B)より、 M_1 か B_1 のいずれかは B_2 の独立な原因として排除されねばならない。
 - ⑨ B_1 が排除され、 M_1 が B_2 の原因だと仮定すると、因果閉鎖原理(A)より、 B_2 は何らかの物理的原因 B_3 を持たねばならず、 M_1 と B_3 との選択に関して⑧と同様な問題が生じ無限後退に陥る。よって、 B_2 の原因として排除されるのは M_1 でなければならない。

結局、 B_2 や M_2 の原因であるのは、 B_1 であって M_1 ではない。 M_1 と M_2 、 M_1 と B_2 との間には「規則的随伴関係」が存在し、かつそれらの関係によって「 M_1 が起こらなければ M_2 は起こらないだろう」のような反事実条件文が含意されるとしても、本来の因果過程は、 M_1 と M_2 の物理的基盤である B_1 と B_2 との間でのみ成り立っているのであり、 M_1 が因果的に働くことはない。

さて、この Kim の付随／排除論法によれば、思考の「内容」を脳状態に付随する性質と見なす限り、当の「内容」の因果的無効性が直ちに帰結するだろう。われわれは、「あの道の角に、人によく吠えかかる犬がいる」という内容 p を持つ私の信念 M_1 が、「あの道を通るのを避けよう」という内容の思考

M_2 を持つことの原因であり、 M_2 に基づいて実際に別の道を通るという行動 A をとる、と信じている。だが、こうした一連の過程において、真の因果過程は上の思考 M_1 や M_2 の付随基盤である脳状態 B_1 や B_2 の間の生理学的物理的過程において生じている。たとえば、 B_1 に付随するのが「あの道の角に、人によく吠えかかる犬がいる」という「内容」p の思考ではなく別の「内容」q の思考であったとしても（さらには、B にいかなる心的状態や内容も付随しなかったとしても）、私の一連の行動には何の変化もなかったことだろう。すなわち、私が B_1 において「あの道の角に、おとなしい猫がいつも眠っている」という「内容」q の思考を持ったとしても、あるいは私が「ゾンビ」であっていかなる思考内容の意識も存在していなかったとしても、私の実際の行動は全く変わらず、やはりその道を回避して別の道を通っていただろう。これはわれわれの民間心理学的直観や「合理的存在者」としての自己理解に著しく反するように思えるが、物理主義の立場に立つ限り、心的性質を含む付随性質が因果的効力を持つことを認めることは困難に思える。Kim が指摘するように、物理主義者は「還元主義」か「随伴現象説」かのいずれかの選択を迫られるのである。

さて、上のような Kim の議論は、まさに「魔術を物理的世界の説明から排除する」ための議論と同型の構造をしていることが分かる。すなわち、思考内容を含む心的性質がわれわれの行動に因果的に影響すると考えることは、「魔術」が物理的出来事に影響すると考えることと同じだけの不当性を含むのである。そして「魔術」の否定が、まさに近代科学がわれわれにもたらした肯定的成果の一つに数えられるべきものだとなれば、「内容」の因果性（思考の力）を否定することは、同様に科学的「成果」の一つと考えられなければならないことになる。

だが繰り返すなら、こうした考え方は自らや他者を「合理的存在者」として考えるわれわれの持つ自己認識と真っ向から背馳する。もちろん、いくつかの心理学的証拠によって、われわれの瞬間的判断に基づく行動は、それを意識する以前に決定されていることが示されると言われることがある。しかしわれわれの「長い目で見た long-term」判断が、思考の「内容」に基づく合理的推理に基づくということを否定することは、きわめて困難に思える。たとえば、「将来ある技能を必要とする職業に就きたい」という「内容」の思考（欲求）を持つ故に、私が本屋からある本を購入するとき、その行為の決定には購入直前の非意識的な脳の状態だけではなく、「将来の自分の職業」に関する「長い目で見た」意識的判断が関わっているだろう。そして、私が意識しているのはその判断・思考の「内容」であって、まさにその内容によって私の「本を買う」という行動が決まる、と私は考える。私は、そうした「内容」に関する意識的合理的計算をなすことで自らの行為を決定しうるものとして、自分自身を認識しているのであり、「内容」の因果性を否定するということは、そうした自己認識を完全に否定することである。

そこで問題は、物理主義に立ったうえで、思考の「内容」が行動に影響する」という合理的自己理解を「魔術的」なものとはしない方策はありえないのか、ということになる。この可能性を、心的因果を「介入」の概念によって考えようとする Woodward の考え方を参照しながら、次に検討してみよう。

3. 介入説と「内容」

3-1. 介入説

Woodward は、2003 年の Making Things Happen などの著作や論文で、「操作と介入 manipulation and intervention」に基づく因果性の分析を提案した⁽⁶⁾。これは、「因果」の概念および「因果的説明」の概念に対して、従来の Hume 的な「規則説」や Salmon 的な「保存量説」、あるいは Lewis らによる「反事実的条件法的分析」とは異なる分析を与え、あるいはそれらを補うものだとされる。そしてこの介入説 interventionism に基づいて、Kim の付随論法を批判し、「心的因果」が物理主義的仕方での理解可能性であることを主張した⁽⁷⁾。以下では、このような Woodward の介入説が、思考の「内容」の因果性に関して、どのような帰結を持つかを検討することにする。

まず、介入説の基本的考え方を解説しよう。介入説では、「因果関係」とはある種の量を表すいくつかの変数間の「関数的依存関係」の一種として理解される。現実の世界には関数的に依存する事象間の様々な相関関係が観察できるが、因果的主張とは、そうした関数的依存関係のクラス全体の中で、「一定の「介入」の下で、ある事象に関わる変数の値の変化が、他の事象に関わる変数の値を変化させる」ような部分クラスについての主張だとされる。

簡単に要約すれば、「X が Y の原因である」とは「しかるべき（可能な）「介入」のもとで、X を表す変数の値をある値に変化させるならば、Y を表す変数の値も変化するだろう」ということを意味する主張である。因果関係はしばしば、直接的因果関係を示す矢印「→」によって、いくつかの変数を結合する「因果有向グラフ」で表示される。ただし、この「因果グラフ」には、非因果的（論理的／概念的／数学的／付随的 etc.）依存関係にある変項は含まれないことが仮定されている。

「介入説」の特徴を、代表的な因果説である「規則説」と「保存量説」と対比して示すと、以下のようになる。ヒュームの「規則説」によれば、「因果関係」は「出来事タイプの間の恒常的随伴関係に関する受動的観察」に基づいて理解される⁽⁸⁾。タイプ F の出来事にタイプ G の出来事が①時間的に先行し、②時間・空間的に隣接し、③恒常的に随伴する、という「タイプ規則性」が観察されたなら、「F の原因は G だ」とわれわれは判断するようになる。単称因果言明「Fa の原因は Gb だ」は、一般的規則性「F は G に法則的・規則的に結合する」という裏付けを持たなければならない。ヒュームの「規則説」においては、因果判断の持つ客観的内容は上のような「タイプ規則性」に尽きており、それに加わる「単称的因果関係の必然性」の観念は、われわれの側の主観的反応（われわれが因果推理を行う際の心の被決定感）を客観世界に投射したものに過ぎない。

これに対する「介入説」の批判のポイントは、「規則説では、因果的判断を行う際のわれわれの側の「積極的介入操作」の役割が見落とされている」ということである。たとえば、出来事タイプ間の恒常的随伴関係に関する受動的観察では、しばしば「真の因果性」と「共通原因の存在によって生じる出来事間の相関関係」とが区別できない場合が生じてくるが、これを解決する際には当の因果過程に積極的

に「介入」することが必要になる。さらに、われわれの単称的因果判断は、必ずしも「法則的規則性」の存在を前提するわけではなく、またそれにコミットするわけでもない。「適切な単称的因果判断を下せる」ということは、生物が環境の中で生存していくための必須の要件の一つであり、一般的規則性の理解に達しない動物や幼児でも、しばしば正しい単称的因果判断を下すことができる。こうしたことは、われわれの持つ「因果」概念や、われわれが下す単称的因果判断が、「法則」や「規則性」の概念よりもより原初的であること、またそれが客観的・実践的意味を持つことを示している。

次に、サモンらの主張する因果の「保存量説」によれば、因果的説明とは単なる出来事タイプ間の統計的相関規則を示すようなものではなく、実在する物理量やエネルギーの連続的伝播に言及することが必須の要件である⁽⁹⁾。単なる統計的相関は何事も説明しない。すなわち、「保存量説」によれば、原因である出来事トークンと結果である出来事トークンとの間に時空的に連続的経路が存在しており、その連続的経路を通じて何らかの実在的物理量が伝播される際に、初めて上の二つの出来事トークンの間に因果関係が成立する。因果関係の本来の関係項は、時空的位置を持って存在する具体的個別的「出来事トークン」であり、恒常的随伴関係や統計的相関のような「出来事タイプ」間の規則性は、この連続的経路によって実在的物理量が保存されていることの証拠となったり、それによって説明されたりするが、そうした規則性自体が因果関係を成立させるものではない。真の因果的説明には、こうした連続的経路への言及が不可欠である。

これに対して Woodward は、「このような「保存量説」はわれわれの因果判断の持つ「実践的役割」を軽視している」と批判する。われわれが何らかの因果判断に基づいて実践的行為をなす際に、そこに「実在的物理量が不変な仕方で伝播される連続的経路」が存在するか否かの知識は、常に可能なわけではないし、また特に重要なわけでもない。幼児や動物はそうした「連続性」や「不変性」に関する概念を持つことはないだろうが、それにもかかわらず正しい（実践的に適切な）因果判断をなしうる。

一方介入説によれば、因果判断をなす際に重要なのは、原因と結果の間に時空的に連続的経路が存在することではなく、「原因への適切な介入によって、結果が何らかの仕方で変化するか否か」ということである。こうした「介入による変化」を正しく把握できたなら、それを支配する法則や連続的因果的経路の存在を何ら知らなくても、われわれは因果関係について適切に理解・判断できたことになる。介入説によれば、こうした「介入の可能性」こそが、因果判断の核心にあるのである。

3-2. ランダム化実験

上で見たように、「介入説」は「規則説」や「保存量説」のこうした難点を免れるために、「積極的介入」の概念に訴える。「介入」とは「単なる受動的な「観察」ではなく、積極的に「実験」的に介入することによって、因果関係の存在を確かめる」という通常の科学的操作を哲学的に精密化したものと考えられることもできる。このような「介入」が因果の判定においてどのような役割を果たすかの一定の「イメージ」を得るために、Woodward の例を加工した人工的模型を使って説明してみよう⁽¹⁰⁾。

喫煙者の集団の間に、指が黄色く変色する傾向と、ある種の肺の病気にかかる傾向とが存在するとし

よう。ここで、喫煙に関する変数を S 、変色に関する変数を Y 、肺の病気に関する変数を D とし、 S 、 Y 、 D はそれぞれ二値の値 $\{0, 1\}$ をとると考える（1 は「喫煙者である／指が変色する／肺の病気にかかる」をあらわし、0 はそれらの否定を表す）。つまり $S=1/S=0$ 、 $Y=1/Y=0$ 、 $D=1/D=0$ である。このとき、仮に以下のような総数 400 の統計データ（介入前のデータ）が得られたと想定する。

$\langle Y=1, D=1, S=1 \rangle$: 80 名, $\langle Y=0, D=1, S=0 \rangle$: 20 名, $\langle Y=1, D=0, S=1 \rangle$: 30 名, $\langle Y=0, D=0, S=0 \rangle$: 270 名

このとき、 Y と D の間には以下のような相関がある。 D/Y 相関は、

$P(D=1/Y=1)=80/110=0.73$, $P(D=0/Y=1)=30/110=0.27$,

$P(D=1/Y=0)=20/290=0.07$, $P(D=0/Y=0)=270/290=0.93$

すなわち、「指が黄色い人が肺の病気にかかっている確率」73%、「指が黄色い人が肺の病気にかかっていない確率」27%、「指が黄色くない人が肺の病気にかかっている確率」7%、「指が黄色くない人が肺の病気にかかっていない確率」93%、である。

また、 Y/D 相関は以下になる。 $P(Y=1/D=1)=80/100=0.8$, $P(Y=0/D=1)=20/100=0.2$, $P(Y=1/D=0)=30/300=0.1$, $P(Y=0/D=0)=270/300=0.9$ 。すなわち、「肺の病気にかかっている人の指が黄色い確率」80%、「肺の病気にかかっている人の指が黄色くない確率」20%、「肺の病気にかかっていない人の指が黄色い確率」10%、「肺の病気にかかっていない人の指が黄色くない確率」90%、である。

これらのデータから、 D と Y との間に規則的相関（恒常的随伴関係）があることに基づいて、「肺が病気になれば指が黄色くなる」あるいは「指が黄色くなれば肺が病気になる」という因果的主張がなされるかもしれない。だが、この因果的主張の成否を判定するためには、上の 3 つの変数からなる因果構造に対して「実験的介入」を行う必要がある。

たとえば、以下のような因果的な介入（ランダム化実験）が考えられる。爪の色を変色させる人為的操作（変数 Y に介入）をして、前被験者 400 名からランダムに選んだ半数 200 名を $Y=1$ に、残り半数 200 名を $Y=0$ となるように設定する。この介入によって、 S （喫煙）と D （肺病）との関係が「遮断」され、 Y （爪の色）と D （肺病）との関係だけが評価されるようになる。仮にその結果が以下になるとする。

$\{Y=1, D=1\}$: 48 名, $\{Y=0, D=1\}$: 52 名, $\{Y=1, D=0\}$: 152 名, $\{Y=0, D=0\}$: 148 名

このとき、介入後の D/Y 相関は $P(D=1/Y=1)=48/200=0.24$, $P(D=0/Y=1)=152/200=0.76$, $P(D=1/Y=0)=52/200=0.26$, $P(D=0/Y=0)=148/200=0.74$ 。すなわち、「指の黄色い人が肺の病気にかかっている確率」24%、「指の黄色い人が肺の病気にかかっていない確率」76%、「指の黄色くない人が肺の病気にかかっている確率」26%、「指の黄色くない人が肺の病気にかかっている確率」74%となる。

また介入後の Y/D 相関は、 $P(Y=1/D=1)=48/100=0.48$, $P(Y=0/D=1)=52/100=0.52$, $P(Y=1/D=0)=152/300=0.51$, $P(Y=0/D=0)=148/300=0.49$ 。すなわち、「肺が病気の人の指が黄色い確率」48%、「肺が病気の人の指が黄色くない確率」52%、「肺が病気でない人の指が黄色い確率」51%、「肺が病気でない人の指が黄色くない確率」49%となる。

い人の指が黄色くない確率」49%, となる。

さてこのような「介入」によってSのDへの影響を排除した後では、Yの値を1から0に変化させても、Dの確率分布がほとんど変化しないことが判明する。このとき、この介入によって、もともと観察されていたYとDの相関は、共通原因Sによるものであって、YとDの間には真の因果関係は存在しない、と判定される。すなわち、指が黄色く変色することは、肺の病気にかかることの原因でもなければ結果でもないと思なされる。それらは、喫煙という共通原因の随伴する二つの結果だ、と判定されるのである。

3-3. 介入説と付随／排除論法

すでに見たように、介入説では因果関係は特定の出来事に関わる変数間の関数的依存関係の一種として理解される。Woodward〔2003〕によれば、各種変数の集合をV、Vに含まれるある変数をXとYとしたとき、以下のような仕方で「直接原因」「寄与原因」の定義が与えられる。

① XがYの「直接原因」であるのは、VのX以外の変数Zの値をすべて一定値に固定する時、Xに対する可能な介入IがYの値またはYの確率分布を変化させる場合である。

② 変数集合Vに関して、XがYの「寄与原因」であるのは、次の(i)(ii)が成り立つ場合である。すなわち、(i)因果矢印「→」で結ばれたXからYへの有向経路E（すなわち、直接原因の連鎖）が存在する、(ii)E上にないVの他の変数Zの値をすべて一定値に固定する時、Xへの可能な介入がYを変化させる。

③ XがYの「原因」であるのは、XがYの「直接原因」であるか、「寄与原因」であるかの場合である⁽¹¹⁾。

さて、「XがYの原因であるか否か」を「介入」によって正しく評価するためには、Xの値を変化させる操作がVの他の因子への操作と「混同」されないように、X以外のYの原因を「制御・固定する必要」がある。そこで、Xのみに対する操作を特定するために、次のように「介入変数」Iが定義される。

④ Iが、Yに関して、Xに対する「介入変数」であるのは、次の(i)～(iv)が成り立つ場合である。(i)IはXの直接原因である、(ii)Iがある値iを取るとき、Xは他の変数の値に依存せず、iにのみ依存する（IはXの原因である他の変数を遮断する「スイッチ」として機能する）、(iii)IはYの直接原因ではなく、X以外のYの原因の直接原因でもない、(iv)Iは、Xを通る有向経路上にないようなYのいかなる原因変数Zからも、独立である。

IによるXへの介入によって、Xは完全にIの制御下に入り、I以外Xの原因とXとの因果関係が断ち切られるが、因果構造の他の部分は影響を受けない。この場合、因果グラフの各変数の値は、(a)他の変数の値から独立に変化することが可能である、(b)各変数の取りうるあらゆる値の組み合わせが可能である、(c)どの因果の矢印も断ち切ることができる、ことが仮定されている。

変数集合Vに属する変数の間の因果関係は、それらの間の「定義的關係」や「論理的関係」とは区

別されねばならない。このために、以下のような「独立な固定可能性」の仮定が置かれる。

⑤変数の集合 V が、それに含まれる変数の値の「独立な固定可能性」を満たすのは、 V のどの変数 X も、それに対して介入することによって、他の変数 Y 、 Z 等から独立に、それが取りうる任意の値に設定できる場合である。

さて、「因果的依存関係」は独立性の仮定を満たすが、定義的關係や論理的関係のような「非因果的依存関係」はそれを満たさない。すなわち、非因果的依存関係は、変数の取りうる値に対して、様々な仕方で制約を課する。「付随関係」は定義的關係や論理的関係と同じく「非因果的依存関係」の一例であり、ある変数の値が別の変数の値を「非因果的」に制約する。Woodward によれば、Kim の「排除論法」の誤りは「因果的依存関係」に関して妥当であるような推論を「非因果的依存関係」に対して適用することによるのである。

Woodward の Kim に対する議論を理解するには、Woodward [2015] で論じられている例を使うのが一番分かりやすいだろう⁽¹²⁾。

Kim の図形 ($P_1 \rightarrow P_2$, $P_1 \Rightarrow M_1$, $P_2 \Rightarrow M_2$) が成り立つような状況を考える。

$$\begin{array}{ccc} M_1 & & M_2 \\ \uparrow & & \uparrow \\ P_1 & \rightarrow & P_2 \end{array}$$

ここで、(a)変数の値を、 $M_1=\{m_{11}, m_{12}\}$, $M_2=\{m_{21}, m_{22}\}$, $P_1=\{p_{11}, p_{12}, p_{13}, p_{14}\}$, $P_2=\{p_{21}, p_{22}, p_{23}, p_{24}\}$, (b)付随関係を、 $\{p_{11}, p_{12}\} \Rightarrow m_{11}$, $\{p_{13}, p_{14}\} \Rightarrow m_{12}$, $\{p_{21}, p_{22}\} \Rightarrow m_{21}$, $\{p_{23}, p_{24}\} \Rightarrow m_{22}$, (c)因果関係を、 $\{p_{11}, p_{12}, p_{13}\} \rightarrow p_{21}$, $\{p_{14}\} \rightarrow p_{22}$, とする。

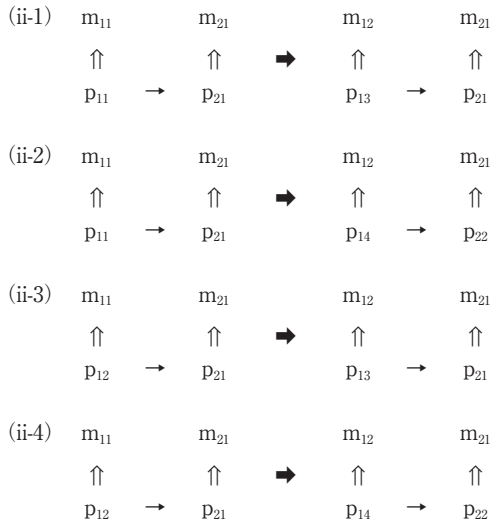
この設定のもとでは、

(i) 値 p_{11} を p_{14} へと変化させる P_1 への介入 I_p は、 P_2 の値を p_{21} から p_{22} へと変化させるから、 P_1 は P_2 の原因である。

(ii) M_1 の P_1 への付随関係によって、値 m_{11} を m_{12} へと変化させるような M_1 への介入 I_m は、 P_1 の値を (p_{11} か p_{12}) から (p_{13} か p_{14}) へと変化させることを介してなされる。すなわち、「介入 I_m によって m_{11} を m_{12} に変化させる」のは、(ii-1)「 p_{11} を p_{13} に変化させる」、(ii-2)「 p_{11} を p_{14} に変化させる」、(ii-3)「 p_{12} を p_{13} に変化させる」、(ii-4)「 p_{12} を p_{14} に変化させる」、のいずれかによってもたらされる。

(iii) P_1 と P_2 の間の因果関係(c)から、(ii)の介入 I_m によって、 P_2 の値は p_{21} か p_{22} の値をとるようになる。一方、(ii-1) P_1 の値を p_{11} から p_{13} に変化させた時 P_2 の値は p_{21} のままで変化しない、(ii-2) P_1 の値を p_{11} から p_{14} に変化させた時、 P_2 の値は p_{21} から p_{22} に変化する、(ii-3) P_1 の値を p_{12} から p_{13} に変化させた時、 P_2 の値は p_{21} で変化しない、(ii-4) P_1 の値を p_{12} から p_{14} に変化させた時、 P_2 の値は p_{21} から p_{22} に変化する。

分かりやすくするために、上の関係を図示してみよう。



ここで、 $\Rightarrow$ は「左の Kim 図形であらわされる事態に対して、 M_1 に対する介入 Im の結果、右の Kim 図形で表される事態が生じる」ということを示す。

さて、これらの介入のいずれの場合も、 M_1 の値は m_{11} から m_{12} に変化するが、 M_2 の値は m_{21} のままで変化しない。すなわち、 M_1 への介入は M_2 の変化をもたらさない。よって M_1 と M_2 の付随基盤 P_1 と P_2 との間には因果関係があるにもかかわらず、 M_1 と M_2 との間には因果関係はない、と判定される。

Woodward によれば、 M_1 と M_2 の依存関係は、 M_1 から P_1 への写像 ($F(M_1) \Rightarrow P_1$)、 P_1 から P_2 への写像 ($G(P_1) \Rightarrow P_2$)、 P_2 から M_2 への写像 ($H(P_2) \Rightarrow M_2$) からなる合成写像 ($H(G(F(M_1))) \Rightarrow M_2$) によってあらわされる。この合成写像がいかなる性格を持つかは、 F 、 G 、 H の具体的形に依存する。たとえ、心的性質と脳状態との間に Kim の図形で示されるような関係があったとしても、それだけからは、心的状態の間に因果関係があるか否かは判定できない。さらに、たとえ M_1 と M_2 の間に何らかの相関があるとしても、それらの間に安定した依存関係があるとは限らない。

こうして Woodward によれば、心的性質の因果効力の問題は、その基盤性質の間の因果関係の問題から独立な問題であることが示される。心的性質の因果性は、基盤性質間の関係がどうであるかではなく、「当の心的性質への介入が他の心的性質や脳状態の性質に関わる変数に変化をもたらすか否か」によって判定されなければならない。ここから、Woodward は Kim の「付随／排除論法」が、介入の際に固定すべき変数に関して誤った前提に立つ故に、誤った結論を出す、と批判する。すなわち、Woodward によれば、

①変数 M_1 のある値によって表示される心的性質が、変数 M_2 や P_2 のある値によって表示される心的性質や脳状態の原因であるのは、 M_1 に対する介入 Im によって M_1 の値に生じる変化が、 $M_2 \cdot P_2$ の値の変化と結びつく場合である。

②「説得」や「だまし」のような通常心的因果と見なされるものは、心的性質の変化が他の心的性質の変化を引き起こすケースの一例である。こうした場合、 M_1 から M_2 や P_2 に向けて因果的矢印を引く

ことは適切である。

③だが因果排除論法は、②の直観に反して、「 M_1 が M_2 に対して因果的効力を持つためには、 P_1 から独立に（すなわち、 P_1 を固定しても） M_1 が M_2 に作用しなければならない」と前提している。これは、「 M_1 の因果的効力をテストするために、どの因子を固定すべきか」に関する誤解に基づく。 M_1 の因果効力を評価する際には、その実現基盤 P_1 を固定してはならない。

すなわち Kim の議論は、「 M_1 の変化が M_2 の変化と相関する場合でも、 M_2 の変化の十分な原因となる物理的变化（ M_1 の基盤性質 P_1 の変化）が必ず見出される故に、 M_1 は M_2 の原因から排除される」というものと解される。確かに Kim の指摘する通り、物理主義の立場に立つ限り、 M_1 の M_2 に対する因果的効力は完全に P_1 に由来し、 P_1 の効果に付け加わる「付加的効果」を M_2 に対して持つわけではない。物理主義者はこれを認めなければならない。だが Woodward によれば、このことは、 M_1 への介入が M_2 の変化をもたらさないことを示すものではなく、したがって M_1 が M_2 の原因ではないことを示すわけでもない。

Kim が M_1 への介入の可能性を考慮しなかったのは、 M_1 が P_1 から独立に変化できないからである。だが、 M_1 は P_1 から因果的に独立な変数ではなく、それらの間には付随的制約関係が存在する。したがって、 M_1 への介入に際して P_1 を固定することを要求するのは不当なことであり、その要求に基づく推理は正しい因果判断に導かない。繰り返せば、付随性質の因果性は、付随基盤の因果性に「付加される効果」を持つか否かではなく、「付随性質への介入がそれと独立な他の変数にいかなる変化を引き起こすか」によって判定されなければならない。心的性質の因果性が物理的基盤性質の因果性によって排除されるというのは、間違った結論だ、というのが Woodward の論点である。

4. 「内容」への介入と「内容」の因果

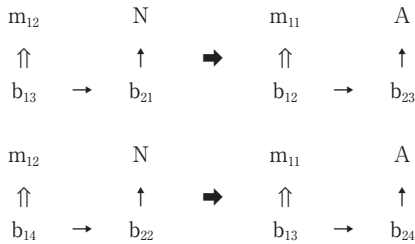
さて、それでは果たして Woodward の議論によって「思考の「内容」がわれわれの行動に因果的影響を及ぼす」というわれわれの直観は、物理主義者も受け入れ可能なものとなるだろうか。これを考えるために、前に論じた例にもう一度戻ってみよう。

先と同様に、「あの道の角に、人によく吠えかかる犬がいる」という内容を p 、「あの道の角に、おとなしい猫がいつも眠っている」という内容を q とし、内容 p の思考を m_{11} 、内容 q の思考を m_{12} とする。 m_{11} は私の脳状態 b_{11} と S の脳状態 b_{12} で実現しており、 m_{12} は私の脳状態 b_{13} と S の脳状態 b_{14} で実現する（すなわち、 b_{11} と b_{12} はどちらも p を表象しそれを「内容」とする思考であり、 b_{13} と b_{14} はどちらも q を表象しそれを「内容」とする思考である）。また、「いつもと同じよう道を通る」という行動 N を行う際の私の脳状態を b_{21} 、回避行動 A を行う際の私の脳状態を b_{23} 、また N を行う際の S の脳状態を b_{22} 、 A を行う際の S の脳状態を b_{24} とする。このとき、変数 M と B の取る値は、 $M=\{m_{11}, m_{12}\}$ 、 $B=\{b_{11}, b_{12}, b_{13}, b_{14}, b_{21}, b_{22}, b_{23}, b_{24}\}$ となる。

心的状態は脳状態に付随しており、 $b_{11} \Rightarrow m_{11}$ 、 $b_{12} \Rightarrow m_{11}$ 、 $b_{13} \Rightarrow m_{12}$ 、 $b_{14} \Rightarrow m_{12}$ が成立している。脳

状態の間の因果関係は、 $b_{11} \rightarrow b_{21}$, $b_{12} \rightarrow b_{23}$, $b_{13} \rightarrow b_{24}$, $b_{14} \rightarrow b_{22}$ である。また脳状態と身体運動との間の因果関係、 $b_{23} \rightarrow A$, $b_{24} \rightarrow A$, $b_{21} \rightarrow N$, $b_{22} \rightarrow N$ も成り立っていると仮定しよう。

このような設定の下で、内容 q の思考 m_{12} から内容 p 思考 m_{11} への変化は、行動 N から A への変化をもたらすだろう。すなわち、以下のような Kim 図形の変形が書ける⁽¹³⁾。



この図形は、私の場合も S の場合も、何らかの「介入」によって内容 q の思考 m_{12} から内容 p の思考 m_{11} へと変化することが、異なる脳状態を介して、N から A への同様の行動の変化を伴うことを示している。仮定によって「内容」は脳状態に付随するから、「内容の変化」は「脳状態の変化」に伴っており、したがって基盤性質の変化に付け加わる因果的効果を持つわけではない。だが、それにもかかわらず、内容への介入が行動の変化を伴う以上、Woodward によれば、「内容が行動の原因だ」という主張は成り立つことになる。

だが、物理的可能世界 W ではどうなるだろうか。W は現実世界 W_A と同じ物理法則が成り立っている世界と想定されているから、因果関係は上の設定と同じであるだろう。W においては、S の付随関係は変わらないが、私に関して心的性質と脳状態の間の付随性の関係が変化する。すなわち、 $b_{13} \Rightarrow m_{12}$, $b_{12} \Rightarrow m_{12}$, $b_{11} \Rightarrow m_{11}$, $b_{14} \Rightarrow m_{11}$ となる。すなわち、以下のような Kim 図形の変形が描かれることになる。



W においても心的変数 M の変化が行動の変化をもたらすから、「心的状態が因果的効果を持つ」ということは成り立つ。だがこのとき、私に関する心的状態変数は内容 p を持つ状態 m_{11} から内容 q を持つ状態 m_{12} に変化することになる。すなわち、「あの道の角に、人によく吠えかかる犬がいる」という内容 p の心的状態 m_{11} から「あの道の角に、おとなしい猫がいつも眠っている」という内容 q の心的状態 m_{12} への変化が、「いつもと同じ道を通る」という行動 N から「あの道を避けて通る」という回避行動 A の変化を伴うことになる。

だが、このことは、Wにおける私の行動を「合理的」に説明しないだろう。私がいつもと違う行動をとる「理由」は、「よく吠える犬がいる」と信じるからであって、「よく眠る猫がいる」と信じるからではない。つまりWでは、「内容」と行動との間の因果的関わり方が逆転してしまうのである。もしもこのようなWが可能であれば、思考の持つ特定の「内容」が特定の行動を因果的に説明する、ということとはやはり成立しないように思える。「物理主義的立場に立てば、内容に言及する合理的説明は、真の因果的説明ではなく、たまたまこの世界で成り立っている偶然的相関（付随関係）によるのに過ぎない」という結論を否定するには至らない。

したがって、「内容」の因果的効力を主張することによってわれわれの合理的存在としての自己理解を物理主義的枠組みと調和させるためには、単に心的性質変数への「介入」によって自分や他人の行動変数に変化が生じる、というだけでは不十分である。上で見たように「内容」と「行動」との関係が「逆転可能」であれば、「内容」による合理的説明は真の因果的説明ではない。

合理的説明が物理主義的に可能な因果的説明でもありうるためには、「内容」と物理的脳状態との間の関係を、これまで想定してきた以上に「本質的關係」だと考える必要が出てくるだろう。たとえば、この世界でpを内容とする脳状態が全く別の内容qをもつような物理的可能世界Wは、やはり不可能であり存在しないと考える必要はなくなるかもしれない。だがすでに見たように、ある脳状態がある内容を持つようになることが、自分がどのような因果過程を経てある内容を表象するかに関する様々な偶然的・指標的条件に基づいていると考える限り、「内容」と物理的脳状態の間の本質的關係を与えることは困難に見える。

だが、ここまでの考察で一つ欠けていた要素がある。それが「意識」である。「意識」の因果性を認めることで、合理的説明と因果的説明を結びつけらるかもしれない。この可能性を論じるためには、「意識」についての物理主義的説明の検討が必要になる。だが、そのためには、稿を改めなければならない。

《註》

- (1) Williams [1978] p. 288
- (2) Dennett [1989], [2001]
- (3) Putnam, *The meaning of 'meaning'*, in Putnam [1975]
- (4) Kim [2005], [2011], *Mental Causation* in Mclaughlin et. al [2009]
- (5) 志向的内容は関係的性質であって心的状態に内在する性質ではない、と主張することもできるが (cf. Tye, M, *Representationalist Theories of Consciousness*), 思考「内容」の因果性を問題にするときには、「内容」を内在性質と考える必要があるだろう。
- (6) Woodward [2003]
- (7) Woodward [2015]
- (8) Hume [1978], Part 3
- (9) Salmon [1984]
- (10) Woodward [2015] pp. 310-311 の例を加工した。
- (11) Woodward [2003] chap.2. ただし、そこで与えられている「総原因」total cause の定義は、ここでの議論

に直接かわらないので省略した。

(12) Woodward [2015] pp. 319-321

(13) この図では、脳状態から行動への直接的因果関係を書き込んでいるので、Kim の図形と完全にパラレルな形ではないが、論点は変わらない。

参考文献

Davidson, D [1970], *Mental Event* in Davidson [1980], *Actions and Events*, Clarendon Press

Dennett, D [1989], *The Intentional Stance*, The MIT Press [2001], *Kinds of Minds, The Origins Of Consciousness*, Poenix

Gbhartner, A [2017]. *Causal Exclusion and Causal Bays Net* in *Philosophy and Phenomenological Research*, Vol. XCV No. 2

Hume, D [1978], *A Treatise of Human Nature*, ed. by Selby-Bigge, Oxford Univ. Press

Kim, J [1993], *Supervenience and Mind*, Cambridge Univ. Press [2005], *Physicalism, or Something Near Enough*, Princeton Univ. Press [2011], *Philosophy of Mind (Third Edition)*, Westview Press

McLaughlin, B.P. et.al, eds [2009]. *The Oxford Handbook of Philosophy of Mind*, Oxford Univ. Press

Putnam, H [1975], *Mind, Language and Reality*, Cambridge Univ. Press

Salmon, W.C. [1984], *Scientific Explanation and the Causal Structure of the World*, Princeton Univ. Press

L.A. Shapiro [2010], *Lessons from Causal Exclusion*, in *Philosophy and Phenomenological Research*, Vol. LXXXI No. 3

Tye, M, *Representationalist Theories of Consciousness* in McLaughlin [2009]

Williams, Bernard [1978], *Descartes, The Project of Pure Enquiry*, Penguin Books

Woodward, J. [2003], *Making Things Happen*, Oxford Univ. Press [2015], *Interventionism and Causal Exclusion*, in *Philosophy and Phenomenological Research*, Vol. XCI No. 2

Contents of Thought and Mental Causation

NAKAGAMA Koichi

Abstract

In this paper, I discuss critically the idea that contents of thought can causally influence our bodily behavior. If we take the physicalist stance, we should say that a mental state is actually a brain state or supervenes on a brain state and that contents of our thoughts are causally irrelevant to bodily behaviors or physical changes in general. If we reject 'magic' or the idea that intentional contents directly cause physical changes, it seems that we should also reject the causal efficacy of our mental contents. But this rejection is totally against our intuitive self-recognition as rational agents, so it is a serious problem for physicalist whether the rational explanations of our behavior are genuine causal explanations or not.

First I examine Kim's supervenience/exclusion argument. On Kim's argument, there can be no Horizontal Causation between Mental States and no Downward Causation between Mental States and their physical bases. So if intentional contents are interpreted as some characteristics of mental states, those contents have no causal efficacies. Against this argument, Woodward's Interventionalism asserts that the intervention to mental property variables can produce a change of their physical bases variables, so that the mentality can be a causal element in the Interventionalist sense of causation.

But I argue that contents of beliefs can be inverted but with still exactly the same physical effects in another physically possible world, so that contents cannot be genuine physical causes but only accidental concomitants of physical events in this world. If we want to recover the causal efficacies of contents, at least we should take account of another missing elements, i.e. the consciousness and its causal power, and this will be a difficult project on the part of physicalist (but still a project well-worth trying, I suppose).