# 信念の実在性について


| メタデータ | 言語: Japanese<br>出版者:<br>公開日: 2017-10-03<br>キーワード (Ja):<br>キーワード (En):<br>作成者: 柴田, 正良<br>メールアドレス:<br>所属: |
| --- | --- |
| URL | http://hdl.handle.net/2297/36300 |

# 信念の実在性について

柴田正良

フランク：あなたの機械なんかどうでもいいんですよ！あなたは、私たちが「信ずる」という語で同じことを意味していると信じますか？

認識論学者：私がそう信ずるかだって？　機械を見るからちょっと待って。そう、私はそう信じているということだ。

フランク：何ですって。機械を見なければ、あなたは自分が何を信じているかをも言えないと言うのですか？

認識論学者：もちろん、その通りだ。

フランク：しかしたいていの人は、何を信じているかと聞かれたら、すぐに言ってくれますよ。どうしてあなたの場合、自分の信念を見つけるために、脳読取機を自分の脳に使って機械の出す結果から自分が何を信じているかを見つけるといった、途方もなく回りくどいやり方をするのです？

認識論学者：私が何を信じているのかを見つけるのに、これ以上に何か別の科学的で客観的な方法でもあるかね？

――「認識論的悪夢」、R・スマリアン(1)

## I　エイリアンとの遭遇

冒頭に引いたスマリアンの悪夢が実現するとすれば、われわれは、他人に対する信念の帰属どころか自分自身に対する信念の帰属に関してすら、ありうべからざる訂正を強いられることになりそうである。歯痛に悩む青年が、脳生理学者の診断を受け、自らの痛みを錯覚だと言い聞かせながら悶々と日を送る、ということすらあるのかもしれない。スマリアンの悪夢は、われわれの日常的な直観、つまり自分自身の信念の誠実な表明には誤りの余地がないという直観に

対するあからさまな挑戦である。哲学の議論において訂正不可能性(incorrigibility)という尊称を戴いているこの直観は、通常は、自分自身の心的な内容に対する直接的な接近可能性というデカルト的なアイデアを盾に擁護されるが、その論法が様々な戦線で綻びを見せ始めているのも事実である。

しかし、この論文で私が示そうと試みるのは、訂正不可能性という概念についての全面的な検討でもなければ、スマリアンの悪夢を可能とするような心のモデルの探究でもない。むしろ、例の直観が含んでいる何がしかの真理を、日常的にわれわれが帰属させている信念の実在性として描いてみせることである。そのためには、そもそも何らかの信念を帰属させることが意味あることなのかどうか、という疑いの生じうるような境界事例から話を始めるのがよいであろう。なぜなら、心を問題とする領域においては日常的な直観のもつ吸引力が余りにも強いため、論議をそのようなものの支配する地盤からいったん引き離す必要があるからである。簡単に言えば、SFでつとにおなじみの場面、つまり、われわれが人間ならざる未知のエイリアンに遭遇して、彼らにどう対処するのかという問題に直面している場面から出発することにしよう。そのとき、そのエイリアンには自らの信念に対する訂正不可能性があるのだろうか。あるとしたら、それはいかなる意味においてだろうか。

ここで、いわゆる他我認識の問題からの教訓を確認しておくのは有益なことである(2)。この状況でまず強調しておかねばならないことは、エイリアンと他人とが占める位置の本質的な類似性である。すなわち、われわれが他人の感ずる痛みや欲求や意図について何の困難も感ぜずに語り、そしてそのことが全く正当であるとしても、もちろん、われわれは、「その他人がそれであるところの」心の状態を「その他人がそれであるところの」知り方で知っているわけではない。要するに、仮にわれわれがロボットやエイリアンに特定の信念や欲求を帰属させることにためらいを感じるとしても、逆に、同じ人間同士である他人に対してならネーゲルが強調するような体験への特権的な接近法がある、というわけではないのである(3)。そこで、他我認識の問題に深入りせずに、この類似性をいささか独断的な仕方で表現すれば、次のようになろう。まず第一に、他人に心的な状態を帰属させる点でわれわれが正当化されているなら、ある種の振舞い上の諸条件を満たす限り、それが機械であれエイリアンであれ、その他なる存在に心的状態を帰属させる点でわれわれは正当化されうる。また第二に、いかなる振舞い上の条件を満たしても人間以外の機械やエイリアンには心的状態を帰属させる点でわれわれが正当化されないのであれば、他人に心的状態を帰属させる点でも、われわれは正当化されえない(4)。しかし、このことは、心的なものを行動による定義によって消去しようとした行動主義を主張するものではない。それは、たとえ或る種の物理主義や機能主義によって心的状態がいかに定義されようとも、定義されるべき心的状態の同定は、それに先だって振舞い上の諸条件を手がかりに行なわれていたのでなければならない、という事情を言わんとするものである。とはいえ、あらかじめ議論を先取りして言えば、タイプ物理主義はもちろんのこと、タイプ機能主義という形でさえも、心的なもののいわゆる「定義」というのは不可能であろうと思われる。

さて、およそ訂正不可能性なるものが出現するとすれば、それは志向的話法で語られた限りでの他なる存在のうち以外にはありえな

い。というのも、因果的-物理的な記述には、真偽決定の方法に関して完全な外延性が期待できるからである。それゆえ、私の当面の問題は、志向的なものが因果的なものとの対比でいかに理解されるのかを示すことである。その際、私が採用したいのは、自明な前提というよりそれ自体問題のある幾つかの前提であって、それらは、(1)物質にならぶ心という「モノ」は存在しないという存在論的な意味でのマテリアリズム、ないし物理主義、(2)物的なもの相互の因果関係についての法則論上の決定論、(3)志向的なものについての話法は、心の内面性の報告が主たる役割なのではなく、むしろ行為の説明を構成するためのものである、という一種の外在主義である。急いでつけ加えなければならないが、ここで志向的話法とは、意図や知覚や信念や希望や欲求のように命題的態度(propositional attitudes)を表現する話法のことだと大まかに理解しておきたい(5)。さて、このような前提からすると、われわれが未知なる存在に遭遇したとき、志向的話法は、いかなる役割を果たすものとして現われるのだろうか。以下では、この問題を、まず大筋としてデネットの立場を擁護しながら考えていくことにし(II節)、ついで、それに対する反論への解答をデイヴィッドソンの非法則論的一元論(anomalous monism)の中に求め(III節)、最後に、信念内容の真理条件的解釈が及ぶ範囲を限定することによって、訂正不可能性の成立する意味を確定しようと試みる(IV節)。

## II　志向的態度(intentional stance)

デネットに従えば、われわれが何らかの未知なる存在の振舞いを予測し説明しようとする場合、基本的には三つの異なる態度が可能であり、われわれはそのいずれの態度をも自由に取ることができる(6)。それらを簡単に逐次枚挙すれば、次のようになる。まず第一は物理的態度(physical stance)であって、それは、当の対象の物理的状態と自然法則についての知識を基に振舞いの予測を行なうものであり、いわばその対象をたんなる物理的システムとして扱う態度に他ならない。第二は設計的態度(design stance)と呼ばれるものであり、それによれば、当の対象の振舞いは何らかの目的ないしプログラムに従ったある種の機能(function)を果たすものと見なされ、その予測には、当の機能的システムがいかなる物理的状態によって実現されているかは無関係なこととなる。最後の第三番目の態度は志向的態度(intentional stance)であり、これは、予測すべき対象の振舞いを信念と欲求によって説明されるべき合理的な行為と見なすことであり、つまりは、相手をわれわれと多かれ少なかれ同じ志向的システムとして扱うことである。

この三つの態度の特徴は、デネット自身が用いている具体例を参考にするとより分りやすくなるであろう。例えば、われわれの前に一つの目覚し時計があるとき、その振舞いを予測するために最も効率のよいやり方は、それを時刻を知らせるという目的のために設計されたシステムと見なす態度である(7)。つまり設計的態度を取るなら、われわれは、その時計がゼンマイ式であろうと、振子式であろうと、またその動力源がなんであろうとおかまいなしに、それが果たすはずの機能にのみ従って、その振舞いを予測し説明することができる。しかし、この設計的態度がうまく行くのは、当の対象がその機能を故障なく果たしている限りであって、その時計が一部熱で溶けるとか、水で錆び付くというようなことが生じれば、そのとき

の時計の振舞いを予測するためには、われわれは物理的態度にまで後退せざるをえない。この設計的態度が、一定の目的を果たすために作られた人工物に最もうまく適合するのは見やすい道理であるが、また同時に動物や植物のある種の振舞いにも適していると言うことができよう。また、この時、物理的態度が、正常に作動している機能的システムに対して原理的に有効であることも注目すべきことである。それに対して、他方、チェスゲームを行なうためのプログラムを組み込んだコンピュータの振舞いについてはどうだろうか(8)。そのコンピュータの対戦相手であるわれわれは、チェスに勝つためにいかなる態度で敵の指し手を予測すればよいのだろうか。実際のところ、このコンピュータがかなり上手にチェスを行なえるのだとすると、設計的態度を取るのは得策とはなりえない。というのも、このようなコンピュータのプログラムは、それを設計した当人にとってさえも設計的態度で振舞いを予測するには複雑すぎるからである。また、もちろんのこと、原理的には可能ではあっても、この場合に物理的態度を採用することは、デネットの言葉を借りれば、実際には効果の期待できないヘラクレス的難事業となるであろう。そこで、われわれがチェスで勝ちを収めるための最善の手段は、相手がチェスのうまい人間であるかのように見なす態度、つまり、志向的態度を取ることであり、このことは、とりもなおさずコンピュータが最も合理的と思われる手を指してくると想定することに他ならない。しかし、もちろんこのことは、そのコンピュータを人間そのものと見なすということではなく、それが設計通りに機能し、しかもその設計が所定の目的にとって最適のものであるがゆえに、最も合理的な指し手をそれが選択すると考えることである。とはいえ、ここには志向的態度にとって最も特徴的なことが表われている。それは、このコンピュータにチェスの規則や駒の配置についての情報すなわち「信念」と、ゲームに勝つという目標すなわち「欲求」を帰属させることによって、その指し手を合理的な行為者の行なう「行為」として扱う、という一般的なパターンに他ならない。すなわち、志向的態度とは、信念・欲求・行為の三つ組という周知の志向性のネットワークによって、相手の振舞いを記述し説明する態度だということになる。

さて、われわれが未知なるエイリアンに遭遇した場合、この三つの態度のいずれを採用するのが正しいのであろうか。実は、以上のデネットの戦略の中で特筆すべきことは、これら三つの態度の選択がプラグマティックなものであり、それ自体で真偽を問いうるようなものではないという点である。

ここで、こうしたデネットの考え方に対する、志向的なものの実在性という点での反論を考察することによって、志向的話法の独自性というわれわれの当面の主題を明らかにしたいと思う。まず第一に問題とされるのは、デネット自身が認めているように、先の三つの態度の選択は基本的に真や偽を争えるような事柄ではなく、たかだか説明における効率性というきわめてプラグマティックな制約がそこにあるだけであり、したがって、そもそもデネットと同様に物理主義を仮定するなら、志向的態度で語られる「事実」というものはないのではないかという点、つまり志向的話法とは、世界に対応するもののない一種のフィクションではないかということである。志向的態度の選択の任意性に関するこの問いを、志向的話法の外部問題と呼んでおこう。さらに第二の問題は、かりにわれわれが志向

的態度を採用したとしても、そこには心理学的説明における志向的概念の周知のいかがわしさがすべて残されているということであり、これを志向的話法の内部問題と呼んでおこう。例えば、その最大の問題は、志向的説明のパターンを形成する信念・欲求・行為の三つ組が悪循環をなし、その環のいずれの部分から説明を始めても「客観的事実」の確定に恣意性が残るということである。というのも、このパターンの特徴は、信念を欲求と行為によって説明し、その欲求を再び信念と行為によって、そして最後に行為を信念と欲求によって説明することにあるがゆえに、相互に極端な調整をしあえば、論理的にはいかなる信念をも整合的に帰属させることができるからである。

これらの外部問題と内部問題は表裏一体のものとなって志向的なものの実在性に対する懐疑的な問題を引き起こすわけであるが、私は、むしろここにこの話法の、他に還元しえない存在理由を見いだしたいと思う。例えば、デネットに対するノジィクの反例を見てみよう(9)。ノジィクの論点は、かりに物理的な世界についての「完全な知識」を所有し、われわれ人間の振舞いをすべてその知識を用いて予測しうる火星人がいるとしたら、彼らにとっては設計的態度も志向的態度も共に不用なものであり、彼らの目から見ればわれわれは、単純なサーモスタットと同様、信念や欲求を保持する者ではなく、たんなる物理的システムに他ならなくなる、というところにある。もしこのノジィクの論点が正しければ、志向的なるものは、たかだか観察者の見方に相対的にのみ現われるフィクションだということになるであろう。この論点は、同じく、先ほどの内部問題を用いても構成することができる。すなわち、志向的態度を取ることが最終的にはわれわれの恣意的な決断に任されているのだとすれば、いかなるものも志向的システムだと見なすことが許されるように思われるのである。例えば、目の前の机は、人間に変身したいという欲望を持ち、じっと佇んでいれば人間に変身できると信じているがゆえに、じっと静かに立ちつくしているのであろうか。また、われわれを乗せている大地は、人々をしかじかの地まで運びたいと願い、自ら動くことによってそれが叶うと信じているがゆえに、ゆっくりと地球の上を這っているのであろうか。問題がここまでくれば、志向的態度とは、ただ擬人的な話法によって世界を描写することにすぎないことになりはしないであろうか。そこで、われわれは次節で、心的なものに関するデイヴィッドソンの存在論的な議論を交えながら、この問題を考えることにしよう。

## III　余りに人間的な原理としての志向的話法

いったい志向的概念によって捉えられるものは世界の実在に関する事実なのか、と真正面から問うのは止めることにしよう。むしろ、志向的話法はいかなる意味で因果的-物理的話法に置き換え難いのか、と問うことにしよう。この置き換え難さの一つの意味を与えてくれるのは、心的な概念が物的な概念へ還元不可能であることを示したデイヴィッドソンの議論である。デイヴィッドソンは、心的なものと物的なものについてのかなり緩やかな同一説(identity theory)を取るが、物的なものによる心的なものの定義も、また心的なものと物的なものを結ぶ厳密な法則も、さらにまた心的なものの領域での厳密な法則も否定する(10)。その根拠は、心的な述語が、物的な述語とは全く別な仕方でこの世界を組織化し、しかも心的な述語による一

般化は、同じ領域内の概念による改良を加えても厳密な法則には到達しないような異種法則的(heteronomic)な一般化だということである。言い替えると、デイヴィッドソンの見方からすれば、いわば因果的-物理的話法は、厳密な決定論的法則への改良の可能性を自らに宿した仕方で世界を組織化するのに対し、志向的話法は、因果的-物理的話法が与える一般化とは本来的に無関係なパターンによって、つまり行為への関心という、因果的法則性からすれば勝手きわまる都合のいい観点から、世界をいわば斑に記述するわけである。例えば、実際に行為を説明するはずの信念や欲求はそれだけで閉じた系

▶ノジィク

をなすものではなく、その内容確定のためには常識や習慣といった、当面の説明には直接関連のない心的要因を次々に参照しなければならないが、こうした志向的概念の持つ全体論的性格(holistic character)は、心的なものを物的なものに法則論的に還元することの不可能性をはっきりと証しているものとなる。それゆえ、デイヴィッドソンは、たとえ宇宙の全物理的歴史を知る者がおり、実際に心的なものと物的なものの同一性が成り立つとしても、その者には一つの心的出来事も予測し得ないだろう、と主張するわけである[11]。

志向的話法が因果的-物理的話法に還元しえないものだとすれば、両者の相違は、むしろ志向的話法の存在理由を構成するものと解釈することができるであろう。まず、志向的話法が有効なのは、それがいかに大まかにであっても、物的なものとの因果関係の中に心的なものを置き、そのことによって、最終的にはわれわれにとって関心のある物的なものの振舞いの効率のよい予測を可能にするからだろうと思われる。つまり、フォーク・フィジィックス(素朴物理学 folk physics)というものがあったとすれば、それは事物の大ざっぱな因果的な予測をわれわれにもたらしたはずだと思われるが、それ以外に独特なタイプの存在者の振舞いを予測するためには、はるかに迅速で効率の良いやり方が必要であったであろう。つまり、人間や動物の振舞いの予測には、フォーク・フィジィックスでは間に合わず、フォーク・サイコロジー(素朴心理学 folk psychology)が必要であったはずである。そのために志向的話法は、因果的な連鎖の網の目をいわばスキップするためのパターンを必要とするが、これがためにこのパターンは説明の対象を単に記述するばかりでなく、ある場合には、むしろ記述の規範をそれに押しつけることも不可避となる[12]。

というのも、心的な概念による記述が常に物的なものと正確に対応することは望めないからである。このことは、志向的話法による説明が本質的には因果的説明の一種であることを否定するものではなく、むしろ、志向的話法が因果性と規範性の二つの要素を含まねばならないということである。その意味で、志向的話法が成功する保証は実在性との対応に求めることはできず、効率性の獲得はいわば一つの賭の結果とならざるをえない。

しかし、この賭を成功させるために、志向的話法は、その規範性に関わる二つの本質的な制約を代償に支払ったように思われる。その第一のものは、志向的話法の適用条件に関する制約であり、それは、この話法が振舞いを単に記述するためのパターンにとどまらず、振舞いをそれへと準拠させるべき規範のパターンでもあるということである。一言で言えば、われわれがこの話法によって説明できるのは、まさにわれわれによく似た信念・欲求・行為のパターンを所有している存在者でしかないのである。例えば、ある存在者に対してこの三つ組のどれか一つだけを用いて他の二つを適用不可能なものと見なす、というようなことはできない。原則として、欲求と行為なしの信念だけの存在という怪物は、おとぎ話のうちにしか現われないであろう(13)。しかし、そればかりではなく、信念・欲求・行為のそれぞれは、互いに相当程度整合的であらねばならない。というのも、余りの合理性の欠落は、そもそも説明すべき事態そのものを失わせてしまうからである。さらに加えて、信念・欲求・行為それぞれの内容は、われわれに周知のものでなければならない。というのも、われわれに理解できない内容のものとは、まさに信念でも、欲求でも、行為でもないからである(14)。われわれは、実は、余りに異様な信念や欲求を説明する志向的パターンを持っていないが、しかしそのために手持ちのパターンを改良すべきであるとか、あるいはそれによる説明が反証されたとか見なさないのは、志向的話法による説明がたんなる理論による事例の説明ではないからである。

もちろん、以上の制約はすべて程度問題ではある。しかし、この種の制約が本質的であることには変わりがなく、このことは、志向的話法がわれわれ人間という存在に固有な原理であり、人間は人間に似せてしか世界を理解することができないという、逃れ難い状況を証し立てているのである。根源翻訳における寛容の原理(principle of charity)が示すように、相手の存在に対する最低限の真なる信念内容を伴った合理性の仮定は、その存在が志向的存在であることの可能性の条件に他ならない。それゆえに、志向的話法の成功が実在性との対応を必要としないのは、ちょうど帰納法の正当化と同様に、成功した場合しか志向的話法を適用すべき事例だと見なされないからである。

さて、規範性に関わる志向的話法の第二の制約は、この話法の適用結果における制約である。つまり、この話法は、効率の良さとある範囲の的中率の代償に、それ以上の正確な予測の本質的な不可能性を引き受けたように思われる。すなわち、ある範囲の逸脱事例の存在と正確な予測の不可能性は、このやり方、つまり志向的話法による予測がいかに成熟したものになっても、避け難い限界となって残ることになったのである。デイヴィッドソンが示したように、志向的話法の用いる概念を定量化し、数量的な予測を行なうことは、概念上不可能なことと言わざるをえない(15)。それゆえ、われわれの知的能力が有限である限り、ある種の複雑すぎる振舞いに対処する

には、本来的に予測不可能な範囲を残した説明のネットワーク、つまり、そのメカニズムについてのわれわれの無知が本質的に除去しえないような理解の方法を取らざるをえないわけである。われわれの無知を既知の正確なメカニズムに際限なく置き換えていくことのできないような話法で自らを語らざるをえないこと、ここに、この話法が含む本来的に規範的な要素、つまり倫理的性格を解く鍵があると思われる。というのも、規範性(規則に従うこと)をどのように解釈しようとも、それが意味をなすのは、決定論的なメカニズム(法則性)を免れた領域でしかないからである。メカニズムの記述によって原理的に同定される振舞いは、そのままでは規範性の主体という概念と相容れるものではない。それゆえに、志向的話法は、因果的なメカニズムの記述によっては尽くされえない規範性の主体という概念を可能とするのである。

ここで、われわれは、先ほどの外部問題と内部問題によるデネットへの批判に答えることができるように思われる。デイヴィッドソンによる経験主義の第三のドグマの議論を思い出してみよう(16)。志向的話法によって語られる事実は実はイリュージョンではないかと疑うことは、われわれの全く理解しえない概念図式からみれば自分たちが実在世界と思いこんでいるものがイリュージョンではないか、と疑うことに類比的である。違うのは、因果的-物理的話法も志向的話法も共にわれわれにとって不可避のものであり、しかも両者の間には独特な還元不可能性があるという点である。それゆえに、物理的なものに定義的に還元されうるものだけを「実在的」と呼びたいのであれば、確かに志向的なものは「実在的」ではないであろう。しかし、そのことによって、志向的話法のもつ「実在性」はいささかも損なわれることはないと言わねばならない。言い替えれば、因果的-物理的話法と志向的話法は、共に争うべきような「実在性」を共有しているのではない、というのが事の真相なのである。

## IV 信念内容の真理条件的解釈と訂正不可能性

さて、以上の議論のあらましは、訂正不可能性の意味をクワインの根源翻訳、ないしデイヴィッドソンの根源解釈の状況の中で計り直してみようという提案に他ならない。私の考えでは、訂正不可能性が出現するのは、信念の主体が自分の心的な状態に対して哲学的に特権的な意味で接近しうるがゆえにではなく、志向的話法が信念主体を規範性のパターンにおける自律的な主体として含意するゆえにである。

信念が言語を有する存在にのみ固有な現象であるという保証はどこにもないし、それどころか信念が言語や命題に対応するということも少しも明らかではないが(17)、まず、訂正不可能性とは、他者の発言に関する問題であるという自明な事実を思い起こそう。つまり、言語をもたぬ動物の信念(があるとして、それ)を確定しようとする場合、その信念の内容は、その動物の行動に対するわれわれの側からの説明の項として登場するにすぎない。われわれは、自分たちの志向的話法のパターンを多少修正して、その動物のおかれた因果的環境に当てはめ、例えば、「近くに水がある」というような信念をそれに帰属させるのである。しかしながら、このような方法は言語を有する存在の信念にとっては余りに目の粗いものであり、こうした場合にわれわれが彼の信念を知るためには、それを彼の発言の単純な関数とみなすのが最も効率的なのである。すなわち、われわれ

は、「pである」と発言する者に対して、pであるという信念を帰属させることをおおむね正しいとみなすのである。それゆえ、訂正不可能性は、さしあたりは信念の伝達と帰属における一種のプラグマティックな要請であるように思われる。つまり、信念の表明に関する正確さは、信念の伝達と帰属を安全に効率よく行なうための根拠として承認された一種の規約なのである(18)。

しかし、このことの正確な意味は、何らかの経験的証拠があれば、いくらでも他者の信念の誠実な表明を訂正させることができる、ということではない。それは、この訂正不可能性が原理的に成立しない状況を考えてみれば明らかである。そもそも、他者の発言を訂正させる理由は、われわれの側からして真なる信念を彼に帰属させるために、その信念に対応する発言を彼に要求することに他ならないが、そのことはとりもなおさず、その理由が手持ちの経験的証拠を解釈するわれわれの「意味の理論」の側にあることを意味する。ところが、それが根源翻訳という事情にあっては、このことはまさにその理論による翻訳が不成功であることに他ならない。というのも、際限のない訂正が生ずるということは、他者の信念の最大限の部分を真となすように当該の経験的証拠を解釈せよ、という寛容の原則に根源翻訳が合致していなかったからである。したがって、この場合には、発言を訂正させるべき理由がそもそも消えたのである。しかし、他方、根源翻訳そのものの成功は認め、彼の発言をそのまま彼の信念の内容に対応するものとみなし、われわれの側からすれば余りに不合理な信念を彼に帰属すべきであろうか。残念ながら、この方策は、われわれ自身によく似た信念・欲求・行為しか理解しえない、という志向的話法の原理に反する。それゆえ、余りに異様な志向的システムはもともと志向的システムではありえず、少なくとも彼は、われわれの側からすれば志向的システムであることを止めたのである。したがって、この場合には、そもそも彼の「発言」を翻訳すべき理由が消えたのである。

以上の事情は、もちろん、他者の発言の訂正可能性が様々な文脈で実際生じうる、ということを少しも否定するものではない。そうではなく、訂正不可能性がどのような範囲内で生じうるかを、最も粗く描こうとしたものである。すなわち、志向的システムは、われわれの「意味の理論」から見て真なる信念の保持者でなければならないと同時に、われわれに理解可能な行為のシステムとして自律的なシステムでなければならない。それゆえ、訂正不可能性は、これら両者のバランスの上に生じると言ってよい。つまり、一方では真なる信念の帰属が第一戦略となる幼児の言語習得過程での訂正可能性と、他方では他者を理解可能な志向的システムとして扱う戦略が挫折し始める精神病者の訂正不可能性のはざまに、それは危うく位置しているのである。こうして、訂正不可能性は、志向的話法が可能である限り信念主体はこの両極の内部に位置しなければならない、という第一の弱い意味で出現する。

ところで、訂正不可能性がこのような揺らぎを示すのは、根源翻訳で用いるわれわれの「意味の理論」が実はすべての科学の諸成果を含んだ「全-理論」であり、そこで次第に精緻さを増していく因果的-物理的説明の進展と、およそ何世紀もの間何の進展も見せていない志向的説明のパターンとの間に、時としてギャップが生ずるからに他ならない。しかし、「意味の理論」による説明は、その因果的-物理的説明の部分を徹底させることによってはこのギャップを埋め

ることはできない。というのも、志向的システムに関してわれわれが最終的に求めるものは、志向的話法による行為の理解であり、その際の説明は、行為の主体によって再構成されうる限りでの、行為と結びつけられた原因(行為の主たる理由)によるものでなければならないからである(19)。しかし、このことは同時に、行為説明の第一適格者として行為者当人を承認することであり、ここに訂正不可能性が、志向的システムを自律的なシステムとして成立させる根拠として、第二の強い意味で出現することになる。

この第二の意味での訂正不可能性は、行為説明に関する以上のような一般的な要請のゆえに、信念内容に対して意味論的な解釈が施される場合でも破棄されることはない。この間の事情は、例えば、パトナムの有名な思考実験を借用すると話が分かりやすいであろう。パトナムは、そこで、「水」と呼ばれるものがXYZという化学式の物質で構成されているという点を除いて、原子の配列に至るまですべて地球と同じ惑星を「意味」の考察の舞台にしたが、この双生地球(twin earth)の思考実験は、自然種語の外延がその語の使用者の心理状態によっては決定されないということを鮮やかに示すことによって、「意味」が頭の中にはないという事情を描いてみせたとされている(20)。しかし、このことは逆に、当人の心理状態の同定に対して、信念内容たる命題の真理条件が決定的であることを示唆している点で重大である。例えば、パトナムの想定にしたがえば、地球上の人物aと双生地球上のそのドッペルゲンガーa′が、どこかの美しい渓流に手を入れながら「この水は何と冷たいことか」と感嘆の声を上げるとき、両者はまったく同一の心理的状態にありながらも、aは「この水」によって$H_2O$を指示し、a′はXYZを指示していることになる。したがって、この場合、aは$H_2O$に対する事象関与的態度(de re attitude)をもっているのに対し、a′はそれを欠落させ、その代わりにXYZに対する事象関与的態度をもっているのである。それゆえ、厳密に言えば、事象関与的信念に関する限り、両者の心理状態はパトナムの意に反して同一ではないのである。それどころか、バージの議論に従えば、この事情は言表関与的信念(de dicto belief)に関しても同じである。というのも、例えばaが「水は酸素を含まないということ」を信ずる時このaの信念は偽であるが、それに対応するa′の信念は真である(21)。それゆえ、これらの信念も真理条件

▶パトナム

の異なる命題に関するものであるゆえに、両者の心理状態は同一ではないのである。

こうして、われわれは、スマリアンの悪夢とは別の意味ではあるが、またしても、自分自身が本当はいかなる命題を信じているのかを知らないのだ、という厄介な事態に直面するように思われる。行為を理解するための志向的説明と信念内容に関する真理条件的解釈とのこのギャップに対し、われわれは何と言うべきであろうか。真理条件的説明が世界の因果的説明の進展と手を携えて行くことには何の問題もないであろう。しかし、先に見たように、肝心なのはそれが行為の説明にとってどれくらい有益なのかという点である。というのも、信念に言及する際にわれわれが求めているのは、水が $H_2O$ であれ XYZ であれ、喉の渇きをいやす行為に対する「近くに水がある」という信念の説明力だからである。例えば、これまで一つの自然種と思われていたものが実は二つの自然種の外延から成るということが判明したとしても、われわれは、それ以前の行為の説明を二種の異なった信念による説明へと細分化させる必要を認めないであろう。それゆえ、志向的説明にとって必要なのは、私と私のドッペルゲンガーがともに「近くに水がある」と信じているときに、その両者を同一の心的状態として扱う概念のレベルなのである。マギンの言うように、信念に関してその認知的役割の理論と真理条件的理論を相対的に独立させて含む、一つの十全な意味の理論が可能だとするなら、前者の部分はその概念のレベルに照準を合わせねばならないであろう(22)。そのとき、そのレベルは、行為説明に自らが用いた概念の意味論的再解釈を無用なものと宣告する、志向的なシステムの自律性によって確保されるであろう。しかし、マギンの言う二つの部分のギャップが常に再生されることを考えるなら、デネットのように、信念分析における命題という概念の不備と、信念の事象関与・言表関与という区分のいかがわしさを率直に認め、信念主体の想念的世界(notional world)を分析の手法として導入する方がよいのかもしれない(23)。そこでは、私と私のドッペルゲンガーはまったく同一の想念的世界に住み、私と友人はほぼ同じ想念的世界に住み、私と精神病者はほとんど絶望的なまでに異なった想念的世界に住んでいることになる。そのとき、一つの想念的世界の成立は、自律的な信念主体のもつ訂正不可能性をその根拠とする、とわれわれは言うことができるであろう。

いずれにせよ、これまで見てきたように、訂正不可能性には主観性の神話にまつわるような神秘的なものは何もないし、またそれによって確保される信念の実在性も、言うなれば、志向的話法による行為の説明をわれわれが捨て去らない限りにおける実在性以上でも以下でもないのである。

註

(1) R. M. Smullyan, 'An Epistemological Nightmare', in D. R. Hofstadter and D. C. Dennett (eds.), *The Mind's I*, Basic Books, 1981, p. 421.(『マインズ・アイ』、坂本百大監訳、TBSブリタニカ、下巻、二八四頁)

(2) ここで言うところの他我認識の問題とは、本誌一九八九年三月号で永井均氏が問題にしているような「他者の問題」とは異なる、「頽落形態におけるデカルト主義」の問題である。これは、深い哲学的驚きに浸ってばかりはおられず、目前の他者とすぐにも交渉せねばならない者の問題である。

(3) Cf. T. Nagel, 'What is it like to be a bat', in *Mortal Questions*,

Cambridge University Press, 1979, p. 169.（「コウモリであるとはいかなることか」、『マインズ・アイ』、下巻、二三八頁）

(4) もちろん、こうした教訓を引き出すことは一つの哲学的立場の表明であるが、ここではそのための議論の代わりに、さしあたり、SFを読むときに駆使しているわれわれ自身の直観に訴えたいと思う。

(5) 志向的なものを命題に対する心的な態度として理解するのは今や一般的になりつつあるが、そもそも命題というものの存在論上の身分がはっきりしていない以上、この方策もまたあらゆる危険から免れているとはとても言い難い。

(6) D. C. Dennett, 'Intentional Systems', in *Brainstorms*, Bradford Books, 1978, pp. 4ff.

(7) D. C. Dennett, 'True Believers', in *The Intentional Stance*, The MIT Press, 1987, p. 17.

(8) D. C. Dennett, 'Intentional Systems', in *Brainstorms*, pp. 4ff.

(9) D. C. Dennett, 'True Believers', in *The Intentional Stance*, pp. 25ff.

(10) D. Davidson, 'Mental Events', in *Essays on Actions and Events*, Oxford Univerity Press, 1980, pp. 212ff.（『デイヴィッドソン行為論集』、服部・柴田訳、勁草書房、近刊予定）

(11) D. Davidson, 'Appendix: Emeroses by Other Names', in *op. cit.*, p. 224.

(12) 例えば、われわれはp→qとpを信じている者はその帰結qをも信じているとみなすが、このことは、われわれの大部分が事実そう信じているということに留まらず、そう信じていない者はわれわれのような合理的システムとはみなさないということ、したがって、合理的システムたる限りそう信ぜよということなのである。この点については、例えば、大沢秀介、「AI・素朴心理学・合理性」、『理想』第六四〇号、一九八八年、六七頁以下を参照。

(13) 同様に、欲求と信念なしの機械も文字通りには行為しない。志向的話法のこのような部分的適用は機械を人間に少しも近づけない、という点については、伊藤笏康、「ロボットは考えるか」、『放送大学研究年報』第五号、一九八七年を参照。もっとも、このような読み方は氏の本意に適うものではないかもしれない。

(14) D. Davidson, 'Psychology as Philosophy', in *Essays on Actions and Events*, pp. 231ff.

(15) *ibid.*, p. 233.

(16) D. Davidson, 'On the Very Idea of a Conceptual Scheme' in *Inquiries into Truth and Interpretation*, Oxford University Press, 1984, pp. 183-198.（「経験主義の第三のドグマ」、土屋俊訳、『現代思想』七月号、一九八七年）

(17) 信念の内容を命題とする解釈の困難性については、例えば、土屋俊、『心の科学は可能か』、東京大学出版会、一九八六年、八九頁以下を参照。

(18) Cf. D. C. Dennett, 'Intentional Systems', in *Brainstorms*, pp. 19f.

(19) 行為主体によって再構成された行為の原因は、もちろん、われわれにとって理解可能であるという条件を満たしている。というのも、それは前述の訂正不可能性の両極の内部で生ずるからである。なお、行為の原因としての主たる理由(primary reasons)に関しては、D. Davidson, 'Actions, Reasons, and Causes', in *Essays on Actions and Events*, pp. 5ff. を参照。

(20) H. Putnam, 'The Meanig of 'Meaning'', in *Mind, Language, and Reality*, Cambridge University Press, 1975, pp. 223ff.（「『意味』の意味」、大出晃監修、『精神と世界に関する方法』、紀伊國屋書店、一六〇頁以下）

(21) T. Burge, 'Other Bodies', in A. Woodfield (ed.), *Thought and Object*, Clarendon Press, 1982, p. 111.

(22) C. McGinn, 'The Structure of Content', in *op. cit.*, pp. 229ff.

(23) D. C. Dennett, 'Beyond Belief' in *The Intentional Stance*, pp. 151ff.

（しばた　まさよし・哲学）