

# 自己欺瞞の静的パラドクスとフォーク・サイコロジー

塩野直之

Folk Psychology and the Static Paradox of Self-Deception

Naoyuki SHIONO

## I

この論文で私が出発点とするのは、自己欺瞞の問題は、人の行動や思考に関するフォーク・サイコロジーにその基盤を持つという論点である。自己欺瞞は、人の持つ命題態度、とりわけ信念や意図に、後述するような一定の整合性に欠けるパターンが見られる点に特徴がある。そして、自己欺瞞が哲学上の問題として注目されてきたのは、そのような特徴から何らかのパラドクスが導かれると考えられたからであった。ところで命題態度とは、フォーク・サイコロジーという一つの理論的モデルに基づいて、解釈という営みを通じて人に帰属させられるものである。ゆえに、フォーク・サイコロジーなくして命題態度は存在しない。したがって、自己欺瞞とはフォーク・サイコロジーを前提して初めて成り立つ問題である。そしてそれに対する解決も、そのような枠組みの中で、パラドクスの解消ないし軽減を図るかたちで探求されてきたのであった。

こうした理由により、自己欺瞞の問題を考察するに際しては、フォーク・サイコロジーとはそもそもどのようなものをかをふり返ることが有益である。これに関して非常に重要でありながら、しばしば十分な注意を払われないのが、規範理論と記述理論の区別である。フォーク・サイコロジーには、規範理論としての側面と記述理論としての側面がある。本論文における私の目的は、自己欺瞞がとりたてて哲学上の問題とされてきた理由、およびその問題を解決するためになされてきた提案のいずれもが、この規範理論と記述理論の区別の軽視、とりわけ規範理論の偏重の上に成り立っていると指摘することである。

規範理論とは、人がどのように行動したり思考したりすべきかを規定する理論である。たとえば論理学は、人が実際にどのように思考するかではなく、むしろ、人がなるべく真なる前提から偽なる帰結を導かないようにするにはどのように思考するのがよいかを述べるものである。意思決定論は、さまざまな信念や欲求をもとにして、とりうる選択肢の中からどれを選択するのが最善かを述べるものである。言い換えれば、規範理論とは、人が仮に理想的な合理性を備えていたら、どのように行動したり思考したりするかを描く理論である。

他方、記述理論とは、人が現実を示す行動や思考を描く理論である。人は必ずしも理想的な合理性を備えてはいないから、規範理論としてのフォーク・サイコロジーだけでは、そのすべ

てを記述し尽くすことはできない。しかしそのような場合でも、その人の行動や思考は必ずしも、「規範からの逸脱」「合理性の破綻」としてしか見ることができないわけではない。多くの場合、逸脱や破綻には何らかのパターンがあり、記述理論としてのフォーク・サイコロジーは、そのようなパターンも考慮に入れて、なるべく多くの現象をその理論のもとに捉えようとするのである。

具体的には、記述理論としてのフォーク・サイコロジーには、規範理論に含まれる諸概念がすべて含まれる。人の行動や思考が実際に合理的で、そのプロセスを論理学や意思決定論によってたどることができる場合には、当然、規範理論の概念がそのまま記述においても役に立つ。だが、記述理論の道具立てはそれだけではない。「愚かである」「軽率である」「ちょっと頭が混乱している」「感情に流されている」「目先のことしか考えていない」などの概念は、人の行動や思考が合理的でない場合に、その人にあてはまる。規範理論が適用できないとき、これら記述理論特有の諸概念がそれを補うのである。

記述理論にはさらに、それらの概念のそれぞれがとりわけどのような場面で適用可能か、つまり、たとえば人はどのような事柄に関して軽率さを発揮しがちで、どのような状況で感情に流されやすいかについての、大まかな一般化が含まれる。すると逆に、人はどのようなときには合理的だと期待できるか、すなわち、規範理論に含まれる諸概念がそのまま記述においても適用できるのはおおむねどのような場合かについての一般化も、記述理論には含まれることになる。こう考えると、全体として、規範理論の原理が合理性であるとすれば、記述理論の根本原理は合理性を含む人間本性、つまりヒューマン・ネイチャーだと言うことができる。記述理論は、人はその普遍的な本性からして、一般にどのような状況でどのようにふるまうものかを、合理的な場合も不合理な場合も含めて包括的に捉えるのである。

記述理論としてのフォーク・サイコロジーは、少なくともつぎの点で自然科学の理論と似ている。それは、理論が現象にあてはまることによって、その現象が説明され、理解可能なものとなるという点である。一つの自然現象にある自然科学の理論や法則があてはまると、当の現象から神秘性がなくなり、それが説明されて理解可能なものとなる。同様に、人の行動や思考も、それが合理的なものである場合もない場合も、記述理論としてのフォーク・サイコロジーがあてはまると、われわれにとって理解可能なものとなる。解釈という営み、すなわち人に命題態度を帰属させるという営みは、その人を理解しようとする営みである。ゆえに解釈において、規範理論の諸概念のみならず、「愚かである」などの記述理論特有の諸概念もが積極的に活用されるのは当然のことである。

規範理論と記述理論の関係について、もう少し考察しよう。第一に、上で見たような記述理論にのみ現れる概念があることは、合理性が解釈の一つの中心原理であることを損なうものではない。人の行動や思考が合理性を一切欠いたものであったなら、それらは理解可能なものとならなければならぬか、そもそもその人に命題態度を帰属させることさえできなくなるだろう。この意味で、規範理論と記述理論はそれほど大きく乖離することは許されない。言い換えると、規範理論の諸概念は、記述においても必ず中核的な役割を果たす。

第二に、規範理論よりも記述理論の方が、現に生じるどれだけ多くの現象にあてはまるかという点で、適用範囲が広い。われわれがもし規範理論しか手にしていなかったならば、単にその理論が適用不可能で、そこに合理性の破綻が生じているとしか言いようのない多くの事例を、記述理論は理解可能なものとしてくれる。だが記述理論の諸概念にも、自ずと適用範囲の

制限がある。人はさまざまな愚かさを発揮するものであるが、そこには一定のパターンがあるはずであり、したがって限度がある。あまりに度外れた愚かさは、「愚か」という概念のスコープに収まらないものとなるだろう。この意味で、人の示すどのようなふるまいも記述理論によって理解可能なものになるとはかぎらない。

第三に、「規範性」の力は、記述理論によって理解可能な人に対してもはたらきつづけるものである。たとえば、人が愚かな思考パターンを示した場合、それは「愚かだ」という概念の適用によって理解可能とはなるが、同時にその人に対しては、そのような思考をすべきでないという規範の力がはたらく。言い換えると、愚かであるという状態は、人がそこに開き直って安住していられる状態ではない。「愚かである」などの概念の適用は、その人を合理性の観点から非難する含みを持ち、その行動や思考を改めることを迫るのである。

上述の諸点を、正しい計算と計算間違いという単純な事例に即して見てみよう。まず、「計算間違い」という概念は、規範理論には含まれず、記述理論にのみ含まれる。規範理論には、正しい計算の規則だけが含まれるのである。そして記述理論には、計算間違いのさまざまなパターン、間違いを犯しやすいのはどのような状況で、逆に間違いが起きにくいのはどのような状況かに関する一般化が含まれる。ここまでは、話の前提である。

さて、「足し算をしている」「計算間違いをした」などの概念が適用可能であるためには、当の人が足し算ないし計算の能力を十分に持っており、たいていの場合には正しい計算の規則に従うことができるのでなくてはならない。つまり、計算の能力を一切欠いた人には、計算に関わる概念はそもそも適用できない。これが上の第一の点である。つぎに、ふだんは正しい計算をする人が、 $68 + 57 = 115$ と答えたとき、この人のふるまいは正しい計算の規則だけに照らしても理解できないが、「計算間違い」とみなすことで理解できるようになる。その人は繰り返し上げを忘れたのであり、それはよくある種類の軽率さである。だが、もしその人が $68 + 57 = 5$ と答えたならば、これはもはや計算間違いではすまない。そのような事例を理解可能なものとしてくれるような記述理論の概念は存在しないであろう。以上が第二の点である。最後に、 $68 + 57 = 115$ と答えた人のふるまいは、計算間違いとして理解できるものの、同時に、その答えは間違っているから正しい計算の規則に従って正しい答えを出すべきだという規範の力ははたらき続ける。これが第三の点である。

規範理論と記述理論の上述の区別は、私にはきわめて当然のものと思われる。それはほとんど、「規範」という概念からアプリアリに出てきそうですらある。というのも、規範が人の従うべき何ものかである以上、現実の人がそれを完璧に満足するものでないことは当然であり、だとすると、現実の人を記述するには何らかの別の道具立てが必要になることが帰結するからである。しかしこの論点も、必ずしもすべての哲学者の合意を得るとはかぎらない。デネットはつぎのように書いている。

だが、信念の理論に誤謬の許容レベルを設定して、人間の可謬性を「公認」としようと試みるならば、それはチェスに「公式寛容ルール」というさらなるルールを追加するようなものである。つまり、チェスのゲームは、そのゲームの他のルールに対する違反となるような差し手がk回以下であるならば、ルールに則ったチェスのゲームだ、というルールである<sup>(1)</sup>。

周知のとおり、デネットによれば、規範理論が十分にあてはまらない場合にわれわれのなすべ

きことは、志向的スタンスを放棄してデザインスタンスや物理的スタンスに移行することである。つまり、フォーク・サイコロジーの諸概念を用いた記述理論は存在しないのである。だがわれわれは、かくも安易に志向的スタンスを放棄せねばならないのであろうか。現にわれわれは、志向的スタンスにとどまりつつ、合理性の破綻を理解可能なものとして描く記述理論を手に行っているのではないだろうか。

他方、近年の心の哲学では、合理性を欠いた心の諸状態に関して、それをフォーク・サイコロジーの枠組みの中で記述的な理論を介して理解しようという試みも盛んになりつつある。たとえばボルトロッチェは、自己欺瞞やとりわけ妄想などの諸状態が、適切な証拠による根拠づけを欠き、それゆえ信念形成の規範的なプロセスを経っていないにもかかわらず、なぜ信念たりうるかを説得的に論じている<sup>(2)</sup>。ボルトロッチェの著書をおおむね好意的に評した論文で、マーフィーはつぎのように述べている。

自己欺瞞や妄想のような状態は、信念に関する哲学的理論のうち、信念の特定と帰属を本質的に合理的なプロセスとみなし、そこにしばしばかなり高い水準の合理性を課すものにとって、厄介なものとなる<sup>(3)</sup>。

私のいま提起した考えの方向性が正しければ、われわれの信念帰属の実践は、そのような厳しい哲学的制約とはほとんど関係のない、あらゆる種類の事情を考慮に入れるのである<sup>(4)</sup>。

このように、「合理性の破綻」は、人の行動や思考の理解という記述的な営みにおいて、必ずしも格別に問題視すべき事柄ではないと考えることができる。すると、自己欺瞞の問題を検討するに際しても、そのどこが問題なのか、再考する必要があるだろう。というのも、冒頭で述べたように、自己欺瞞をはじめとする不合理性の諸形態が近年の哲学で論じられてきたことには、それらにおける合理性の破綻がただちに何らかのパラドクスを招くという懸念が背景にあるからである。そしてその再考に基づいて、この問題の解消のためにしばしば提案されてきた、「心の分割」という論点の妥当性も再評価すべきである。

## II

自己欺瞞のパラドクスには、静的パラドクスと動的パラドクスの二種があると言われる。静的パラドクスとは、人が  $p$  と  $\sim p$  という互いに矛盾する信念を同時に抱くことから生じる問題である。他方、動的パラドクスは、人が自己欺瞞の結果、 $\sim p$  という偽なる信念を抱くに至るプロセスに関わる。人を欺くという意図は一般に、それが欺かれる本人に気づかれてしまうと成功することはない。しかし自己欺瞞においては、欺く者と欺かれる者が同一の人である。すると、自己欺瞞は意図的行為ではありえないことになるのではないだろうか。これが動的パラドクスの問題である<sup>(5)</sup>。

本論文では、主に上の静的パラドクスについて論じる。動的パラドクスは、おそらくより大きな困難を提示するものであるが、その本格的な検討は別の機会に譲ることにしたい。まず、言うまでもなく、理想的に合理的な人は、 $p$  と信じかつ同時に  $\sim p$  と信じることではない。合理的な人は、両立不可能な二つの命題があったとき、そのどちらがよりよい証拠に支持されているかを検討するはずであり、その結果、多くともそのいずれか一方しか信じるに至らないか

らである。

しかし、自己欺瞞に関するわれわれの日常的な理解によれば、矛盾する信念の同時所有という状態は、おそらくそれほど珍しいものではない。そして、記述理論としてのフォーク・サイコロジーは、その人はたぶん感情に流されてちょっと頭が混乱しているのだ、といった言い方でその状態を理解可能なものとしてくれるであろう。ごく概略的に述べると、その人は  $p$  という事態の成立に気づき、 $p$  という真なる信念を抱くのだが、そのことがもたらす感情的な不快感などのせいで、偽なる信念  $\sim p$  をも抱くに至る。さらに、おそらくはその感情のもたらす混乱状態のため、不注意にも、これら両者の信念をともに保持し続けるのである。

この人が信念  $\sim p$  を抱くに至るプロセスの詳細については、動的パラドクスに関わる問題であるからここでは検討しない。静的パラドクスに関連する事柄に絞って考察すると、私はつぎの点が重要だと思う。すなわち、信念  $p$  と信念  $\sim p$  の両者を保持しているあいだ、この人には「混乱している」「軽率」「不注意」といった記述的な概念が必然的にあてはまる。これは先に述べた、合理性を欠いた状態は人がそこに安住していられる状態ではないということと関連する。 $p$  と  $\sim p$  の両者を信じている人は、それら両者を証拠に基づいて冷静に検討した結果、どちらも信じるに値すると考えて信じているわけではない。そのようなことはありえない。さらに、人は自らの信念に矛盾がないかどうか注意を払い、万が一その気配があったときは、証拠に基づく冷静な検討によってただちに少なくとも一方を破棄すべきである。これは、どの人にも常に及び続ける規範性の力である。したがって、 $p$  と  $\sim p$  の両者を信じている状態は、人が規範性の力に従い損ね、それゆえ非難されて然るべき状態に置かれているときにのみ生じる。このように、記述理論としてのフォーク・サイコロジーは、矛盾する信念の同時所有という状態をたしかに理解可能なものとはしてくれるが、それが修正されるべき不安定な状態だという理解を必ず伴うのである。

では、信念  $p$  と信念  $\sim p$  を同時に抱くことは、なぜとりわけパラドクシカルだと考えられるのだろうか。この問題を順次検討してゆこう。まず注意すべき点として、矛盾する信念の同時所有は、矛盾命題を信じることは区別されねばならない。つまり、 $B(p)$  かつ  $B(\sim p)$  という状態は、 $B(p \wedge \sim p)$  という状態とは異なる。率直なところ、私は後者も可能だと考えているが、そのことをここで主張するつもりはない。自己欺瞞者といえども決して  $B(p \wedge \sim p)$  という状態にあるわけではないというのが、この点に関する通説である<sup>(6)</sup>。

するとまず、 $p$  と信じかつ  $\sim p$  と信じている人は、なぜ  $p \wedge \sim p$  と信じるに至らないのかという疑問が生じる。というのも、その移行に必要とされる推論はたんなる連言の構成であり、自己欺瞞者がそのような基本的な能力を一般に欠いているとは考え難いからである<sup>(7)</sup>。だがここでも、規範理論と記述理論の区別に訴えることは有効である。人は自らの抱いている信念の論理的帰結をすべて信じるべきだというのは、規範的な合理性の原理である。現実には、人がそのような論理的万能性を示すことはない。そして、現実の人が推論をするには、当の推論を行う論理的能力に加えて、その推論をすることへの動機が必要である。このことは、連言の構成のような単純な推論についても当然あてはまる。自己欺瞞に陥っている人は、 $p$  と  $\sim p$  から  $p \wedge \sim p$  への推論をすることへの動機を持ち合わせていない。むしろ、そのような推論をしない動機があるはずである。これが、 $p \wedge \sim p$  という信念が形成されない理由である。

ただし上の状況で、自己欺瞞者は  $p$  という真なる信念と  $\sim p$  という偽なる信念を同じ仕方

じていることを、頑なに否定するであろう。他方、 $\sim p$ であること、そして自らが $\sim p$ と信じていることは、積極的に肯定するであろう。このように、当人の言い分は、偽なる信念に一方的に肩入れし、真なる信念はその存在すらを否定するものとなっている。そしてこのことは、 $p \wedge \sim p$ という信念が形成されないことの一因であるにちがいない。さて、この人の置かれた心の状態がいったいいかなるものであるかは、たしかに解明を必要とする。だがそれは、信念  $p$  が原因となり、どのようなプロセスを経て信念  $\sim p$  が形成されるかの全体的な説明がなされて初めて可能となることである。つまり、これは静的パラドクスよりもむしろ動的パラドクスに関わることであるから、その検討は別の機会に譲る。

さて、デイヴィドソンは、自己欺瞞の静的パラドクスは、ただちに心の分割という考えの導入を必要とすると考えているようである。

重要なのは、密接に結びつきつつ対立する複数の信念を人は別々にしておくことができるし、実際しばしばそうするということである。われわれはこのかぎりでは、心の諸部分のあいだに境界線を引くことができるという考えを受け入れねばならない。私は、(明白に) 対立する信念のあるときはつねに、それらの信念のあいだのどこかにそのような境界線を設定する<sup>(8)</sup>。

この主張の論拠の一つとして、デイヴィドソンはつぎの論点を提示する。すなわち、もし「 $A$  は  $\sim p$  と信じている」から「 $A$  は  $p$  と信じていない」を導出することができるならば、矛盾は当の自己欺瞞者の信念のみならず、その記述のうちに現れてしまう。というのも、「 $A$  は  $p$  と信じている」と「 $A$  は  $p$  と信じていない」は、 $A$  の命題態度の内容のレベルではなく、その記述のレベルで矛盾する命題となっているからである。

われわれは信念のいかに荒唐無稽なパターンであれ、それを自己欺瞞者に帰属させることを厭わないかもしれない。しかし、彼の混乱を記述しようとして、われわれ自身が矛盾に陥ることがあってはなるまい<sup>(9)</sup>。

デイヴィドソンは、この問題を自己欺瞞の最大のパラドクスの一つと考えているようである。そして彼は、これを回避するためには心の分割が必要になると議論を進めるのである。

たしかに、解釈を行うことによって解釈者が矛盾に陥ることがあってはならない。だがこの困難は、件の導出が記述理論ではなく規範理論に属するものであることを確認すれば容易に解消される。それは、 $A$  が合理的だという条件のもとでのみ認められるものだからである。実際には、自己欺瞞のような混乱状態にある人は、 $\sim p$  と信じつつ  $p$  と信じてもいるかもしれない。つまり、「 $A$  は  $\sim p$  と信じている」から「 $A$  は  $p$  と信じていない」の導出は、記述理論においては無条件には許容されないのである。そしてこのことを認めるにあたって、心の分割の論点を持ち出すことがただちに必要になるとは思えない。

矛盾する信念の同時所有には、つぎのような問題も指摘されるかもしれない。すなわち信念とは、主観確率が少なくとも 0.5 を超えた所定のレベルに達している状態のことである。ところで、 $p$  の確率と  $\sim p$  の確率の和が 1 になることは、確率論の公理系の定理である。そして、主観確率が確率論の諸定理を満たすことは、ダッチ・ブックの論証が示すとおりである。する

と、 $p$ の主観確率と $\sim p$ の主観確率の和は1になるはずであり、 $p$ と $\sim p$ の両方に対して0.5を超える主観確率を持つことはありえない。したがって、人が $p$ と $\sim p$ の両方を信じることはありえない。

この指摘に対しても、なすべき返答は明らかだと思う。ダッチ・ブックの論証が示すのは、人の主観確率が確率論の諸定理を満たさないと、その人はある巧妙な賭けに引き込まれて確実に損をするおそれがあるということである。そして、人は損をすることが確実な賭けをするほど愚かではないはずだという前提を認めれば、 $p$ と $\sim p$ の両方に対して0.5を超える主観確率を持つことはありえないという結論はたしかに出てくる。だがもちろん、人が現実にその程度の愚かさを示すことはいくらでもある。したがってこれもやはり、規範的な合理性をそのまま現実の人に当てはめようとする誤りである。

浅野は、つぎのような点が、矛盾する信念の同時所有の可能性に対する最大の困難だと論じている。すなわち、信念にはそれに伴う感情や行為があるはずであり、人が矛盾する信念を同時に抱いていると、その人がどのような感情を抱き、どのような行為をするかが理解不可能になるというのである。

問題は、欺瞞的主体は矛盾する二つの信念を同時に得ているとする想定が、それぞれの信念が惹き起こすべき行為や感情という観点から見て、果たして整合する主張なのかどうかである。・・・そうすると、主体はまったく同一の話題ないし命題 $p$ に関して、同一の時点で、不安を抱きかつ安堵もしていることになるが、これはいったいいかなる精神状態になるのか、私には理解できない。また、主体はいったいどちらの信念にしたがって行為すればいいのかという観点から見ても、デイヴィドソンの考えは不都合を生じるだろう<sup>(10)</sup>。

ここにどのような困難があるのか、私にはよくわからない。人がある事態に対して、不安と安心が混淆した両義的な感情を抱くことは、よくあることであろう。またその結果、その行為が支離滅裂なものとなって、その人が $p$ と信じているようにも見えるし $\sim p$ と信じているようにも見えるという状態も、きわめて容易に想像可能である。

浅野は、そのような両義的な感情を抱いたり支離滅裂な行為をしたりする人は、そのときどきの感情や行為と整合するように、 $p$ という信念状態と $\sim p$ という信念状態のあいだを急速に行ったり来たりしているとみなされるべきだと考えているようである。後述するように、私もそのような見方が可能であることを否定するつもりはない。だが、静的パラドクスを回避するために、どうしてもそのような記述を採用しなければならないと考える根拠は、特にないように思われる。

このように、静的パラドクスがパラドシカルだとみなされてきた理由を一つずつ検討すると、どれも決定的な論拠と言うに足るものではないことがわかる。同時に、それらの理由の多くが、人に備わる合理性を過度に高く評価する誤り、つまり、規範理論としてのフォーク・サイコロジーが、人の行動や思考の記述にあたって十分だと考える誤りに根差していることがわかる。

### III

ところでデイヴィドソンは、矛盾する信念の同時所有のあるところ、必ず何らかの境界線を設定して心の分割をせざるをえないと論じたのであった。私はそれに対して、矛盾する信念の

同時所有は可能だが、その場合は必ず、「不注意」「軽率」「混乱」といった状態を同時に当の  
人に帰属させることを伴うと述べた。この両者は、たかだか些細な言葉の違いにすぎないと思  
われるかもしれない。つまり私の言う「不注意」とは、心の中に何らかの注意の及ばない状態  
が生じていることを含意し、したがってそこには注意の障害となる何らかの障壁、つまり心の  
境界線があるはずだということにならないであろうか。

たしかに、これを言葉の違いとして片づけることもできよう。だが私の考えでは、ここには  
むしろ、解釈やフォーク・サイコロジーの本質に関する根本的な見解の相違を見て取るべきで  
ある。デイヴィドソンには、解釈という営みは、解釈対象が規範理論に完璧に従うことを必要  
とするという発想があるように思われる。というのも、心の分割という考えを採用すると、分  
割された心のそれぞれに帰属させられる命題態度は完璧な整合性を示すからである。デイヴィ  
ドソンは、そのような整合性があるのはじめて、心というものが成立すると考えているのでは  
ないだろうか。

私の見解では、記述理論としてのフォーク・サイコロジーは、そのような完璧な合理性や整  
合性を示すものとして心捉えるのではないところに特徴がある。というのも、その中心をなす  
原理は、合理性ではなく人間本性だからである。この点を考察するに際して、フォーク・サイコ  
ロジーというものがそもそもどのように成立したかについて、想像をめぐらせることは有益であ  
ろう。フォーク・サイコロジーは、現実に生きているさまざまな人の行動や思考を理解しようと  
いう営みとして、何千年、何万年の歳月の中で発展してきたものである。それは、一人ひとりの  
人間、つまり一個の身体を持つそれぞれの人間を理解しようとする中で生じたものにちがいな  
い。それは決して、規範理論を構成する中核部分が先に整備され、それに記述理論特有の諸概  
念が後で付け加えられることによって成立したのではなかった。またそれは、哲学に特有の奇  
妙な思考実験を積み重ねることによって成立したのではなかった。それはむしろ、人間の本性  
に対する深い洞察が、歴史の淘汰にさらされ、次第に洗練されつつ一つの理論に収斂してゆく  
中で生じたのである。したがって、それはまず記述理論として、規範的な諸概念と、規範性の  
破綻を理解可能なものとする諸概念とが区別されずに混在するかたちで成立したであろう。

このことから、二つの帰結が導き出せそうである。まず、「心の分割」、つまり一個の身体を  
持つ一人の人間を複数の「心」ないし「人格」を持つものとみなすという考えは、解釈の営み  
においてはなるべく回避するのが望ましいということである。たしかに、ある理論が、本来は  
一個の身体を持つ一人の人間を理解するために成立したものでありながら、ある特定の場  
合にはそれを適用することに著しい困難があり、そのとき、一見したところ一人の人間と見える  
ものがじつは二人の人間なのだと考えるとその理論がうまくあてはまる、ということはいえ  
よう。その場合、しかし生物としてはやはりそこに人間は一人しかおらず、じつは二人の人間が  
いると強弁することが無理であるならば、人間とはカテゴリーの異なる「心」「人格」とい  
った存在者を設定し、一人の人間が二つの心を持つと考えることになるだろう。実際、分離脳患  
者や病的な多重人格者のような、きわめて特殊な人間の行動や思考を記述するにあたっては、  
こうした「心の分割」という措置がおそらく有効である。だが、自己欺瞞のような日常的な不  
合理性を描くに際して逐一そのような手段に訴えることは、フォーク・サイコロジーによる解  
釈や理解という営みの当初の目的に反することである。

第二に、憶測に憶測を重ねることにはなるが、規範理論と記述理論の関係について、つぎの  
ことが言えそうである。すなわち、規範理論は、先に成立した記述理論としてのフォーク・サ



イコロジーの本質をなす諸原理のうち、真理保存性や期待効用最大化といったすぐれて合理的なものに着目し、それらを論理学や意思決定論のような抽象的な理論体系として整備することによって構築されたのであろうということである。

デイヴィドソンは、もともと意思決定論の専門家であった。実際、彼はダッチ・ブックの論証とよく似たマナー・ポンプの論証の考案者の一人である。したがって、解釈に関する彼の理論が、合理性の想定に基づく規範的な理論を出発点とし、それを現実の人の行動や思考に適用しようという発想から生じているとしても不思議はない。だがこれは、規範理論と記述理論のあいだに成り立つであろう歴史的な経緯から見れば、おそらく逆立ちした考え方である。自己欺瞞という現象を理解するに際して、一人の人間を、それぞれは完璧な整合性を保った複数の心に分割するという提案は、偏った奇妙な考えであり、記述理論としてのフォーク・サイコロジーが持つ豊かな諸概念に着目すれば、必要のない発想である。

金杉は、心の分割というデイヴィドソンの発想をおおむね継承するかたちで、自己欺瞞の静的パラドクスと動的パラドクスを論じている。金杉によると、これらのパラドクスを回避することは可能であるが、その代わりに一つの困難を招きよせることになる。それは、心の分割を認めると、分割された心はもはや二つの心にほかならず、その一方が他方を欺くことは、もはや他者を欺くことでしかないのではないかと、つまりそれを「自己欺瞞」と呼ぶことの根拠が消滅するのではないかという問題である。

しかし、心の分割の議論に対しては、一般に、次のような批判が容易に思い浮かぶ。心の分割が生じた主体は、二人の主体に分裂していて、もはや一人の主体として認めることはできないのではないかと。そして、二人の主体の間での欺瞞であれば、それはもはや「自己」欺瞞ではないことになる。それゆえ、心の分割を導入することは、自己欺瞞の理解として不適切なのではないだろうか<sup>(11)</sup>。

金杉は、この困難を緩和するために大いに議論を展開しているが、それを完全に克服するには至らないことを認めているようである。

ここでは金杉の議論を立ち入って検討することはできないので、以下はいささか外在的な批判とならざるをえないが、私には、このような困難が生じることこそ、合理性を唯一の原理として心の存在と同一性を認めるという戦略がどこかで道を踏み外していることの証左であるように思われる。金杉は、もし合理性以外に心の同一性の基準となりうるものがあるとしたら、それは身体の同一性だと考えているようである。そして、身体の同一性がやはり求める基準として不十分である以上、命題態度の織りなす合理的な関係に依拠する以外に残された選択肢はないと考えを進め、その結果、上述の困難を抱え込むに至るのである。さて、身体の同一性が心の同一性の基準として十分でないことには、私も同感である。だが、命題態度のあいだに成り立つ関係は、合理的な関係だけではない。われわれはむしろ、さまざまな命題態度が、人間本性という観点から見て理解可能なパターンをなしているかどうかに着目すべきである。そして、そこに理解可能なパターンがあるならば、われわれはそこに一つの心の存在を認めることができる。デイヴィドソンや金杉には、この可能性が盲点となっているのではないだろうか。そしてその背景には、本論文のテーマである、規範理論と記述理論の区別の軽視、ないし規範理論の偏重という傾向があるのではなかろうか。

## IV

先に論じたように、矛盾する信念の同時所有という状態は、それ自体、とりわけ回避すべき哲学的問題をはらんでいるとは思えない。つまり、自己欺瞞の静的パラドクスは、パラドクスではない。しかし私は必ずしも、そのような状態があることをわれわれは積極的に認めるべきだと主張しているのではない。たんに、静的パラドクスはそれを否定する根拠にならないと言いたいだけである。そもそも私は本論文で、動的パラドクスについて論じることをしていないから、自己欺瞞の状態にある人の心的状態の特徴を十分に立ち入って検討したとは到底言えない。たしかに私は、「不注意」「軽率」「混乱」といった概念がそのような人の理解に有用でありうるとは主張したが、もとよりこのことをもって自己欺瞞の考察として足りるとみなしているわけではない。人が自己欺瞞に至るプロセスについて、そしてそのプロセスを経て到達した状態について、より積極的な特徴づけが与えられないかぎり、自己欺瞞の解明として満足すべきものとならないことは当然である。

ただここには、われわれの手にしているフォーク・サイコロジーは、たしかに長い歴史的過程を経て洗練されてきたものではあるが、まだ発展途上の「フォーク」なものであり、決して十分に緻密なものではないということが関係している。すなわち、矛盾する信念の同時所有という状態が本当にあるかどうかについて、現状のフォーク・サイコロジーは必ずしも白黒の明確な回答を用意しているわけではない。私は先に、自己欺瞞の状態にある人は、 $p$  という信念状態と  $\sim p$  という信念状態のあいだを急速に行ったり来たりしているものであり、したがっていずれの時点においても両方を同時に信じていることはないという見方を否定はしなかった。また、メレのように、自己欺瞞者はそもそも  $p$  という真なる信念を一度も抱いておらず、信念として帰属させる必要があるのは  $\sim p$  だけだという見方も可能である<sup>(12)</sup>。私のこれまでの議論は、そのような立場を排除するものではない。では、信念  $p$  と信念  $\sim p$  の同時所有という記述と、これらの代案のいっただれが正しいかは、どのようにして考えたらよいであろうか。それはおそらく、現状のフォーク・サイコロジーの中では決着のつかない問題である。

フォーク・サイコロジーは、決して現時点で完成されたものではなく、今後、より緻密なものへと発展させてゆく余地を残したものである。そしてその発展の作業には、哲学的考察と経験的考察とが手を携えて取り組むことになるであろう。ボルトロッチィは、さまざまな不合理な信念状態をフォーク・サイコロジーの枠組みの中で説明するという哲学的プランを進めるにあたり、それがやがては脳科学を含む経験科学とうまく接合するだろうとの楽観的な見通しを持っているようである。つまり、パラドクスの回避のような哲学上の課題に応えつつ、経験的にも十分に裏付けのあるフォーク・サイコロジーを整備することが、きっと可能だということである。私自身は、正直なところそこまで楽観的にはなれないが、その可能性に賭けるしかないと考えている。

フォーク・サイコロジーを発展させる作業には、自己欺瞞の動的パラドクスに関する哲学的考察も、もちろん寄与するであろう。動的パラドクスが引き起こす困難は、静的パラドクスのそれよりも深刻かもしれない。私は動的パラドクスに対する十分な考察をまだ持ち合わせていないが、議論なしに結論だけを述べれば、メレと同様、自己欺瞞者に「自分を欺こうという意図」を帰属させるべきではないという考えに傾きつつある。そのかぎり、自己欺瞞を他者を欺くことと類比的に捉えることには限界があると思う<sup>(13)</sup>。ところで、自分を欺こうという意

図の存在を否定すると、自己欺瞞者が矛盾した信念を同時に所有しているという想定は、もしかするとともはや不必要なことになるか、ないしは辻褃の合わないことになるかもしれない。仮にそうだとすれば、私が先に「現状のフォーク・サイコロジーの中では決着のつかない問題」と呼んだものに対しては、動的パラドクスをめぐる哲学的考察によってフォーク・サイコロジーを発展させる過程で、決着をつける可能性が開かれることになる。

フォーク・サイコロジーはこのように、人の行動や思考を理解するためのものとして現に使用されつつも、哲学や経験科学によってさらに発展させることのできる理論的モデルである。そしてその発展に際しては、当然、それを整合的な理論とすることが求められる。だが求められるのは、あくまでも理論の整合性である。その理論によって記述される対象が整合性を示すことは、必要とされる要件ではない。解釈の対象となる人ないし心が完璧な整合性を示すべしということこそ所与として、合理性を原理とする規範理論を突き詰めるかたちでフォーク・サイコロジーを発展させることは、控えめに言ってもその発展の唯一の形態ではないし、おそらく実りある形態でもないであろう。

#### 註

- (1) Dennett (1978), p. 21. なおボルトロッチェは Bortolotti (2010), p. 19 において、この箇所を引用している。
- (2) Bortolotti (2010).
- (3) Murphy (2012), p. 19.
- (4) Ibid., p. 22. じつは、ボルトロッチェとマーフィーには、妄想がフォーク・サイコロジーの概念で記述できるかどうかに関して見解の相違があり、そのため厳密にはこのマーフィーの引用は適切でないのだが、本稿の目的にとっては、これらの論者とデネットのあいだにある、より根本的な対立が理解できれば十分である。
- (5) 静的パラドクスと動的パラドクスの概略については Mele (2001), pp. 6-8 を参照。
- (6) Davidson (2004), p. 217. 邦訳 p. 351.
- (7) 柏端 (2007), p. 34.
- (8) Davidson (2004), p. 211. 邦訳 p. 341.
- (9) Ibid., p. 217. 邦訳 p. 350.
- (10) 浅野 (2012), p. 80.
- (11) 金杉 (2012), p. 55.
- (12) Mele (2001).
- (13) もちろん、他者を欺く場合には欺く意図が存在するだろうからである。周藤 (2011) によると、他人に嘘をつくことが定義的にそのような意図を含むという考えは、アウグスティヌスに遡ることができるであろう。

#### 文 献

- 浅野光紀 (2012). 『非合理性の哲学：アクラシアと自己欺瞞』, 新曜社.
- 柏端達也 (2007). 『自己欺瞞と自己犠牲：非合理性の哲学入門』, 勁草書房.
- 金杉武司 (2012). 「自己欺瞞のパラドクスと自己概念の多面性」, 『科学哲学』 45-2: 47-63.
- 周藤多紀 (2011). 「二種類の嘘：アウグスティヌスによる『嘘』の定義」, 『アルケー』 19: 111-122.
- Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.
- Davidson, D. (2004). *Problems of Rationality*. Oxford: Oxford University Press. [D・デイヴィッドソン著, 金杉武司・塩野直之・鈴木貴之・信原幸弘訳 (2007). 『合理性の諸問題』, 春秋社.]
- Dennett D. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Mass.: MIT Press.
- Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Murphy, D. (2012). "The Folk Epistemology of Delusions", *Neuroethics* 5: 19-22.