機能する感情・幻想する感情

柴田正良（金沢大学）

真理と虚偽の判断を意味するような、厳格な意味における理性は、それ自体では意志に対するいかなる動機にも決してなりえないし、また何らかの情念(passions)もしくは情緒(affections)に触れない限り、何らの影響力も持ちえないことは明らかなように思われる。（Hume［1748］. p.161. 邦訳、p.273）

「情念論」第五節の冒頭で、ヒュームは、情念もしくは感情こそがわれわれを動かしている本体であって、理性もしくは計算はその企てに奉仕する下僕にすぎない、と主張しているように思われる。私は、ここでこのヒュームの主張を額面どおりに受け取るどころか、さらに、われわれのような因果的世界における有限な行為者にとって、感情がきわめて重要な機能を果たしているということを論じたいと思う。しかし、人工知能の研究が如実に示しているように、もっぱら計算という意味の理性の働きに関しては、多くの研究成果が得られているが、感情、このわれわれの生と認識と行為の営みのすべてをその中にたゆたわせている存在に関しては、まだその理論化は端緒についたばかりだ。ここでは私は、その遙か先に実現されるはずの感情理論に向けて、三つの感情の相貌を描きたいと思う。それらは、感情理論の一部というより、むしろ感情現象の広がりを示すことを意図したものである。その一つは心理学者の戸田正直がかつてアージ理論として解明しようとした＜機能する感情＞の姿であり（１節）、第二は脳神経科学者のA. ダマシオのソマティック・マーカー仮説が捉えようとしている＜身体状態の表象としての感情＞であり（２節）、最後は、行為者が因果的世界における自らの存在を自由だと思いこむための＜幻想する感情＞である（３節）。

１節　野生合理性

感情は、これまで、行為者の合理的な行為選択を誤らせる非合理的な撹乱要因と見なされるのが一般的であった。抑えがたい怒りが余計な一言を言わせたり、失敗に対する恐れが本当の失敗を引き起こしたり、過度の愛情がかえって恋人を去らせたりする、というのはよくある話だろう。では、感情は一体何のために存在するのか。感情は、人間の進化の中では、たまたま選択された別の機能Xの副産物として残ってしまった厄介なお荷物にすぎないのだろうか。しかし、感情が行動や認知に及ぼす影響の甚大さを考えるなら、感情がたとえ悪役であっても、自然淘汰を生き延びてきた主役ではなく脇役だ、ということは考えにくい。つまり感情は、認知機能に縫いつけられた単なる＜フリル＞ではないのだ。また、感情は脇役ほど軽いどころか大きな災厄を人類にもたらすのだが、もっと生存価値の高い別の機能Xのゆえにそれと共に選択されてしまった、という可能性は、その災厄が大きければ大きいほど、進化の長い歴史は「感情を伴わない仕方でのその機能Xの実現」を選んだであろうがゆえに、これも考えにくい。つまり感情が主役級の悪役なら、自然淘汰の矛先はほかならぬ感情にこそ向けられ、人類の初期の段階できれいさっぱり消し去られていたであろう。それゆえ、現在の人類に感情が現に備わっているということは、感情がその機能のゆえに淘汰を生き延びてきたということの証拠としては不十分かもしれないが、感情がトータルとして淘汰の対象となるほど悪しき結果を人類に及ぼし続けてきたわけではない、ということの証拠にはなるはずだ。

とすれば、ありそうな話は、感情はむしろその独自の機能のゆえに進化において選択されてきた、ということではなかろうか。ただその機能は、進化のある段階では生存価値を著しく高めたが、現在、その＜機能の一部＞はかえって人類の生活にとってマイナスの効果を時として与えるような具合に環境が変化してしまったのではないか。戸田はこのような観点から、感情の機能を捉えているように思われる。

戸田の感情理論を見る前に、感情が存在しない認知システム、つまり現在の人工知能おいて何が問題として生じているのかを確認しておこう。その問題とは、フレーム問題と呼ばれる情報処理における「計算」問題だ(1)。人工知能は、われわれ人間のような自然知性の人工的模倣であるが、この二つを分かつ決定的な点がある。それは、自然知性にとっては＜生き延びる＞ということが至上命題となっているの対し、少なくとも現在の人工知能にはそのような厳しい制約がかかっていないということである。例えば、自分の生存にとっての＜危険＞というのは、自然知性にとっては存在しても、人工知能にとっては存在しない。したがって、切迫した危険を避けるための緊急の情報処理は、自然知性にとっては生存という究極の目的のためになされるが、人工知能にとってはたんに「計算のための計算」にすぎない。フレーム問題はさまざまな側面を持つが、この対比の観点からすれば、それが生じてくる理由ははっきりしている。常識問題という、正解がありそうで正解のない計算問題を人工知能に与えてみよう。それは、考慮すべき項目が多すぎたり、あるいは答えを評価すべき文脈が多すぎたりして、一般に「計算量の爆発か、データ量の爆発か」と呼ばれるジレンマに人工知能を陥らせるだろう。なぜかというと、人工知能は、自分が正解を探している計算がそもそも何のための計算かを知らないからである。つまり、人工知能は、正解のない探索をどこかで打ち切るための＜理由＞をそもそも持っていないのだ。それは、生存のための資源が極めて限られている、という有限性の制約の下に人工知能が本当は置かれていないからである(2)。

その結果、デネット(Dennett [1987])が見事に描いたようなロボットにとっての苦境が生じてくる。自分にとって大事なエネルギー源、バッテリーを時限爆弾の仕掛けられた部屋から救い出すとき、そのロボットは自分の行為の副次的な結果を考慮に入れ損なったために、バッテリーと一緒に爆弾も運び出してしまう（そして、爆弾はロボットとバッテリーを吹き飛ばす）。そこで次のロボットは、行為結果との関連性を見過ごさないように推論機構を強力にしてもらうのだが、今度は、実際には無視すべき、関連性のないあらゆる推論（例えば「バッテリーを運び出しても白山の標高は変わらない」）に果てしなく巻き込まれてしまう（またもや時間切れで、爆弾は破裂）。さらに第三のロボットは、今度は、関連性の判断を可能にしてもらったおかげで、悲劇的にも、あらゆる推論の関連性の有無を果てしなく推論するという破滅の道を辿るのだ（もちろん、この場合もBang !）。

フレーム問題に象徴されるような人工知能の困難は、一言でいえば、＜生存の管理＞を行う機能が欠けているということを示している。そして、冒頭のヒュームの言葉が示唆しているのは、感情こそがこの＜生存の管理＞という機能を果たしているのではないか、ということだ。ロボットが感情のありがたさを実感するとすれば、それは皮肉なことだが、「今ここ原理」による感情の＜文明非適合性＞を強調する戸田 [1992]の記述の中に見いだされるかもしれない。ともかく、以下では、感情機能の一つの具体的な姿を、戸田が描く野生人の切実とはいえ一種ユーモラスな＜逃げっぷり＞を通して見てみよう。もっとも、ここでの私の意図は戸田理論の正確な評価にはない。むしろ、ここに描かれた感情の積極的な機

能、野生の合理性を読者にも共有してもらうことにある。

戸田によれば、感情は、遺伝的に基本枠の設定された行動選択･行動遂行用の心的ソフトウェア、「アージ･システム urge system」の一部である（「アージ」の原義は、衝動：駆り立てる、追い立てる）。感情は遺伝的に人間に組み込まれた行動選択機構であり、野生環境であればどこでも通用するが、いわゆる本能ほど融通のきかないシステムではない。野生環境では、生じた出来事の影響範囲はその周囲に限られており、時間空間的に少し遠くなるだけでその影響はほとんどゼロになるのがふつうである（戸田の言う「弱システム」）。したがって、きわめて切迫した出来事が身近に生じた場合には、遠くのことがらや過去のことがらを一切省略して、すべての認知能力を＜今のこの出来事＞に投入するのが合理的だろう（「今ここ原理」）。アージ･システムは、この「今ここ原理」を、余分な情報処理の自動的排除によって実現するシステムである。このシステムのおかげで、人間は、時間圧がかかると情報処理の精度を犠牲にしてでも何とか早く答えをだす、というトレードオフができる(3)。これが、実は、デネットのロボットたちには不可能な事柄だったのだ。そして、戸田は気づいていないように見えるが、感情は、フレーム問題に対して自然が用意した解答なのかもしれない。しかし、太古の自然選択の報いとして、「今ここ原理」を身上とする感情は、文明環境では不都合を起こしやすい。というのも文明環境では、生じた出来事の影響は、ルールや組織や情報網を通して時間空間的に遠くまで及んでしまうからである（彼の言う「強システム」）。たとえば戸田の言うように、野生環境では見つけた食べ物を黙って頂戴しても何の問題もないが、文明環境では、まず「泥棒！」ということになる。つまりアージの働きが強くなれば、文明環境では当然考慮すべき遠くの事柄が「見えなく」なり、その結果、今ここの利害に駆られて、迷惑やルール違反や犯罪を引き起こしてしまうのだ。

以上が、戸田による感情の＜文明非適合性＞の説明だが、それで、野生人においてはそのアージはどう働くのか？　例えば、ライオンなどの敵が急に目の前に出現すれば、＜恐れアージ＞が起動される。すると技能化された状況認知評価モジュールによって、瞬時に「恐れ」の度合いが計られ、＜恐れアージ＞の強度が決定される(4)。それにしたがって、いくつかの「恐れ活動プラン」のなかからその場に適切なものが選ばれる。たとえば、恐ろしい状況への対処はそこから逃げるのが基本だが、逃げ方にもいくつかのプランがあるだろう。余裕時間がほとんどなく、事態が本当にのっぴきならないものであれば、一か八かのジグザグ逃走、パニック行動くらいしか取りうるプランはない。パニック行動もそれなりに合理的なのだが、しかし少しでも余裕時間があれば、適切な逃走経路を選択するかどうかは死活問題となる。野生人でも（だからこそ？）熟知した場所なら、「尖った大きな岩のところで左に曲がると川があるから、その川に飛び込んで逃げる」といった「条件つき行動プラン」が可能である。敵が自分よりも足が速いと分かっている場合には、ただ走るより木に登ったり、水に飛び込んだりする方が有効かもしれないからだ。さらに余裕があれば、その大きな岩を見過ごしたらどうするか、といった「逃走サブプラン」も用意されているかもしれない。このように、逃走プランの実行はそう単純な話ではない。しかも野生人の十八番である走る技能に加え、逃走の途中でつくわすさまざまな障害に対する「緊急対処行動サブプラン」のようなものも必要である。大きな穴を飛び越えたり迂回したり、垂れ枝をくぐったりへし折ったり、崖をよじ登ったり駆け下りたりと、野生人もなかなか忙しいのである。

野生人は走る。しかしこれらの奮闘にもかかわらず、逃げても逃げきれない場合には、「窮鼠猫を噛む」、闘争プランが逃走プランに取って代わるだろう。死

に物狂いの闘争は、場合によっては敵をひるませ、偶然、敵に勝つこともあるかもしれない。しかしさらに、相手が強大すぎて逃走することが問題にならないときには、死んだふり、諦めプランしか残されていないこともあろう。もちろん、野生人のような＜逃げ＞のプロの場合は、時間圧が極大（余裕時間がゼロ）になった場合には、無行動的「諦め」が意識にのぼる前にすべてが終わっているかもしれない。しかし、それでも、相手との力の差がすべてにおいて圧倒的に大きい場合には、「下手に動くよりも始めから腰を抜かしていた方が生き延びる確率が高い」(ibid. p.94)かもしれない。もっともこういう話の展開では、せっかく登場した野生人があえなくジ･エンドとなってしまうので、その前に、戸田が描く＜食欲アージ＞の割り込み状況に目を転じよう。割り込みは、戸田がコンピュータ用語から借用したもので、「新しいデータ入力が割り込みを掛けてきたら現行の情報処理を一時中断し、その入力処理をすませてからもとの処理に戻れ」というようなことを意味する。この機能もまた、複雑な環境で生き延びている人間のような行為システムにあっては、あらかじめプログラム化しておくこともリストアップしておくことも不可能な情報処理であり、「割り込みのたびに意思決定が必要になる」(ibid, p.95)。ここで、認知と行動においてもっとも重要な意思決定が戸田の描く感情の機能（アージ・システム）によって担われていることに注目しておこう。まさに感情は、この意味ですべての認知能力の支配者なのだ。

さて、あるアージが起動中に他のアージが割り込みをかけてくるようなとき、二つのアージにはいくつかの関係がありうるが、単純な場合には、例えば＜恐れアージ＞の下で逃げている最中に＜食欲アージ＞が割り込み、たまたま手にもっていた食べ物を食べながら逃げるというようなこともあるだろう。あるいは、逃走経路の近くに、うまそうな木の実がなっているのを見つけ、＜食欲アージ＞が起動する場合があるかもしれない。この場合、＜恐れアージ＞が＜食欲アージ＞の割り込みを許すなら、木の実を食べたあとで逃走行動を再開しても逃げられるという「計算」が働いたことになる。このように現アージの再開が可能であるような新アージの割り込みは、戸田によれば「遅延割り込み」である。しかし逆に、＜食欲アージ＞の割り込みが許されない場合もあろう。その場合の「計算」は、木の実を取っていると敵に追いつかれる、といったものであり、新アージは、それを実行すれば現アージの遂行が不可能になる「阻止割り込み」をかけていたことになる。しかし、いったんこの場は木の実をあきらめても、あとで戻ってきて木の実を食べることが可能かもしれない。その場合の＜食欲アージ＞は、「遅延可能割り込み」であり、戻ってくるコストを考えるとそれが無意味であるような場合は「遅延不可能割り込み」である。また、木の実取りのプランが実行されなくとも空腹感は消えないだろうから、＜食欲アージ＞は、＜恐れアージ＞の強度が下がるまで＜待機アージ＞の状態になるだろう。しかし、それは必ずしもおとなしく自分の出番をまっているわけではなく、少し余裕ができれば、逃走中に「何か手間をかけずに食べられるものはないか」と辺りをきょろきょろ見回すことくらいのことはするだろう。もちろん、これらの各アージの相互関係を決めるための「計算」が、アージ強度の体感性（感じ）といったアナログ型の仕掛けによって可能となっているのは言うまでもないことだ。つまり人間は主観的、意識的には期待値や確率などの計算をせずに、実際には頭の中での場面の想像によって「状況認知評価モニターがアージ強度を体感させてくれる」(ibid., p.101)のである。

ともかく、野生人は走る。自分の生存を賭けて･･･

２節　身体表象としての感情

ところで、戸田が描いた感情の姿は、行動制御のためのソフトウェアとして捉えられたものであり、それを支えるハードウェアのメカニズムは論じられていない。戸田が述べる、固定プログラムでない、技能化された、自由度の高い処理過程や、価値の体感的評価などは、いかにして人間という身体物理的な仕掛けにおいて実現されているのか。実は、感情を支える基盤は脳、とくに計算機械としての脳の実質である神経回路だけではない。例えば、感情や情動と関係している神経系は脳の他の部位と身体の全体に信号を送るが、その一つの経路は血流であり、信号は化学物質の形で送られ、それが細胞の受容体に作用する。もう一つの経路である神経系では信号は電気化学的な形を取るが、それが他のニューロンや筋肉や器官に作用し、それらがまた化学物質を血流に放出する。こういった全身体的なダイナミックな過程において実現される感情は、たんに行動制御のソフトウェアとしての側面だけからでは、その機能の全貌を捉えることができないだろう。神経生物学的な実現メカニズムからの解明がない限り、むしろ、感情のソフトウェアとしての機能すらもが一つの謎と化してしまう恐れがある。

そこで以下では、感情の機能を、身体と脳の相互作用レベルにおける＜生命の管理＞と捉えるダマシオの感情理論、ソマティック・マーカー仮説を見てみることにしよう。もちろん、ダマシオの理論もまだ完成されているとは言い難い。しかし、来るべき感情の理論は、彼が目の前にしている感情という現象の広がりと深みを捉えなければならないだろう。

ダマシオは、通常ほぼ同義に用いられる「情動 emotion」と「感情 feeling」を区別して用いる。彼によれば、この両者の違いは、情動が基本的に内的もしくは外的な刺激に由来する身体状態の変化であるのに対し、感情は、その変化の知覚だということにある。情動は、生得的なもの（一次の情動）であれ、学習によるもの（二次の情動）であれ、脳の制御下にある神経細胞の末端が多数の器官に引き起こす身体状態の一連の変化である。情動を経験しているとき、われわれの「･･･内臓（心臓、肺、腸管、皮膚）や骨格筋（骨についている筋肉）や内分泌腺（例えば、下垂体や副腎）の機能のいくつかのパラメータに変化が起こる。いくつかのペプチド調節物資が脳から血流に放出される。免疫系も急激に変化する。動脈壁の平滑筋の基本的な活動が増加し、血管を収縮させ、細くする（その結果、蒼白になる）。逆にその活動が減少すれば、平滑筋は弛緩し、血管が膨張する（その結果、紅潮する）。全体として、一連の変化は機能的バランス、すなわちホメオスタシスと対応する平均的状態の範囲からの逸脱を示している」(Damasio［1994］p.135. 邦訳、p.220)。

このような身体状態の急激な変化は、今度は逆に、神経的信号と化学的信号の二つのルートによって脳に伝えられる。神経的な信号は、「･･･皮膚、血管、内臓、随意筋、関節などからのインパルスを脳に伝える神経末端を介して、連続的に脳に送られている。神経学的用語でいえば、この旅の復路は、頭、首、胴、四肢のそれぞれに端を発し、脊髄と脳幹へと進み、網様体（とくに覚醒や睡眠の調節に関わっている脳幹内の核群）と視床を抜け、視床下部、辺縁構造、島領域と頭頂領域にあるいくつかの別個の体性感覚皮質へと進んでいく回路に依存している。とくに体性感覚皮質は、身体で何が起こっているかの説明を受け取っている」(ibid. pp.143-4. 邦訳、p.230-1)。では、もう一つのルートである「化学的な旅」は、ダマシオによればどうなっているのか。「情動が生じているとき身体に放たれるホルモンやペプチドは、血流を介して脳に至り、いわゆる血流－脳関門を通って、あるいはもっと簡単に、関門を持たない脳領域（例えば最後野）や脳のさまざまな部位に信号を送る装置を持つ領域（例えば脳弓下器官）を通っ

て、脳の中に浸透していく。脳は、そのいくつかのシステムの中で、他のシステムが誘発した身体風景の複雑な神経的眺めを構築するが、その眺めの使い方、そして眺めの構築そのものが、身体により直接の影響を受ける」(ibid. p.144. 邦訳、p.232)。ダマシオの印象的な言い回しによれば、脳は、自分の身体風景の眺め(a view of the body landscape)を手にしており、それに従って、最終的な生命の調節、ホメオスタシスの維持を行っているのだ。つまり脳は、時々刻々と変化する身体状態の変化を、神経的および化学的な基盤に支えられた＜表象＞を通して眺めているわけである。したがって感情とは、このような連続的なプロセスを追いながら、身体がしていることを＜経験＞することにほかならない。（図を参照）。

図と説明（ダマシオ『生存する脳』講談社の231頁にある図とその説明をここに挿入）

では、感情は受動的な現象であり、認知や行為に積極的には関わらないのか？

そうではない。まったく逆に、感情は、生命の管理のためのメカニズムを用いて生存の管理を行う装置である。もっと具体的に言えば、ダマシオが例に挙げているように推論と決断において、感情は決定的な役割を果たす。そしてここで、感情が形式的な論理能力よる計算では解決不可能な選択問題を＜合理的＞に解決する、と強調されていることは、戸田のアージ理論との単なる偶然の一致ではない。要するに、問題解決に使うことのできる注意と作動記憶は限られている。しかも「費用便益計算」を適用すべき選択肢のシナリオは無数にある。われわれが計算の脇道に迷い込む可能性は大きいし、そもそも推論戦略の基本にある確率論と統計学に対して人間はひどく無知であり、推論はその誤用だらけだ(ibid. pp. 165-73. 邦訳、p.260-70)。ダマシオが診ていた患者の例が、ちょうど感情のこの機能と関連する脳の部位、前頭前・腹内側部(ventromedial prefrontal region)に損傷を受けていた。彼はある冬の日、凍結した道路を自動車で運転していて、目前の女性の車のスピンに巻き込まれそうになった。誰もがパニックになり、急ブレーキを踏みそうになるその恐ろしい状況で、彼は少しも動揺することなく冷静にその凍結部分をやりすごし、自分の事故を回避した。ところが、異常だったのはその次の出来事である。彼がダマシオらにその出来事を語ったあと、次の診察の日取りを決める段になった。一月先の二つの候補に関して、彼は三十分もの間、どうでもよいような理由をあれこれと並べ立てて、いつまでたってもこの簡単な選択を行うことができなかった。凍結部分を冷静沈着に運転した彼は、「そのときと同じくらいに平静に、退屈な費用便益分析、果てしない話、実りのないオプションと帰結に関する比較でわれわれをうんざりさせた」(ibid. p.193. 邦訳、p.298)。ところが、ついにたまりかねて、二番目の日はどうかとダマシオらが提案すると、彼はあっさりと、しかも静かにそれを承諾したというのである。ダマシオは、これを純粋理性の限界を示す好例だと考えている。

ソマティック・マーカーとは、まさしくこのようなバカげた理性の浪費、計算の無駄を省いてくれる自動化された意思決定メカニズムである。あるいは、われわれが漠然と「直観」と呼んでいる、判断や行動に関する無意識的なバイアスのメカニズムである。例えば、ある問題の解決に向けた推論を始めようとすると、時には、そのさまざまなシナリオの忌まわしい結果が頭に浮かび、ある不快な直観(gut feeling)を経験する。ダマシオがこの感情を「ソマティック」と呼んだのは、それが身体（ソーマ）の状態に由来するからだ。彼の分析では、このとき、その感情、つまり内臓(viscera)もしくはそれ以外の身体部位から発するある種

の感覚が、あの悪しき結果のイメージと結びつき、それをマークする。すると、そのソマティック・マーカーは、推論に関連したそのネガティブな結果にわれわれの注意を向けさせ、自動化された危険信号として機能する。「＜この先にある危険に注意せよ。もしこのオプションを選択すればこういう結果になる＞。この信号は、われわれがネガティブな行動を即座にはねつけ、他の選択肢から選ぶように仕向ける。この自動化された信号により、われわれは将来のごたごたを回避することができるだけでなく、少数の選択肢から選ぶことができるようになる(ibid. p.137. 邦訳、p.270. 傍点は引用者)。もちろん、ソマティック・マーカーは、ネガティブなイメージだけでなく、ポジティブなイメージもマークする。

さて、感情が余分な選択肢の幅を自動的に狭めることによって情報処理のコストを削減する、という以上のストーリーは、それだけのアイデアとして見れば、ド・スーサ(5)や戸田と本質的には変わらないだろう。あるいはむしろ、認知の理論としては、戸田のアージ理論よりも肌理が粗いかもしれない。しかし、ダマシオのソマティック・マーカー仮説の優れている点は、この認知システムを実現するメカニズムを神経基盤の側から描こうとしているところであり、したがって逆に、そこが彼の仮説のまだ未完成な点でもある。だが、現在のところ、誰にこれ以上のことを望むことができるだろうか。いずれにせよ、今後に展開されるべき感情の理論は、認知のストーリーとそれを支える実現メカニズムのストーリーとを、メダルの裏表として含んでいなければならない。では、ソマティック・マーカーを実現する神経基盤のストーリーはどのようなものなのか。

ダマシオが見るところ、ソマティック・マーカーの機能を支える重要な神経システムは、前頭前皮質(prefrontal cortices)にある(ibid. p.180-4. 邦訳、pp.281-6)。その理由は、一言でいえば、前頭前皮質は感覚、記憶、運動に関連した身体と脳からのあらゆる情報の集積地点だからだ。これは、まるで、哲学者のフォーダーが二十年以上も前から探し求めていた＜中央演算システム＞のように聞こえる(6)。まず第一に、（ｉ）前頭前皮質は、外界の知覚、外界についての思考、身体内の出来事、これらすべてに関する情報を受け取っており、この点で、心と身体のすべての風景を眺める特権的な脳部位となっている。というのも、前頭前皮質は思考のためのイメージを作り出しているすべての感覚皮質からの投射を受けているだけでなく、その領域には、過去と現在の身体状態を表象する体性感覚皮質も含まれているからだ。また第二に、（ii）前頭前皮質は、生存に関する生得的価値体系の情報を握っている。という意味は、それが脳の中で生体調節に関わっている部位からの投射も受けている、ということだ。後者はすなわち、生命の管理としてダマシオが強調する機能を果たしている部位であり、そこには「脳幹内にある神経伝達物質放出核（例えば、ドーパミン、ノルエピネフリン、セロトニンをばらまく核）、前脳基底内の神経伝達物質放出核（アセチルコリンをばらまく核）、それに扁桃体、前帯状回皮質、視床下部がある」(ibid. p.181. 邦訳、pp.281-2)。さらに、第三に、ダマシオによれば、（iii）前頭前皮質そのものが、個人の来歴に関わる偶発的な出来事の類別機能を果たしている。類別された個人的な出来事に関する知識は、信頼のおける予測や計画を立てるのに必要不可欠だが、それが内容別にこの領域のさまざまな部位に貯蔵されているようである。「例えば、生体調節に関する知識は腹内側部のシステムと密接な関係をもつ一方で、背側領域のシステムは、外界に関する知識（物体や人間、それらの時間空間における振るまい、言語、数学、音楽など）と連携しているように見える（ibid. p.183. 邦訳、p.284)。そして最後に、（iv）前頭前皮質は、運動支配と化学反応に直接関わっている。具体的には、前頭前皮質の「背側部や上部内側部は前運動皮質を活性化することができるだけでなく、そこを通路として、いわ

ゆる一次運動皮質（M1)、補足運動野(M2)、第三運動野(M3)がオンライン化されている。また、皮質下運動機構の基底核も同じように前頭前皮質とつながっている」（ibid. p.183. 邦訳、p.284）。さらに加えて、自律神経系の効果器に信号を送り、情動に関係した化学反応を促進させることができるのもここである(7)。

しかしながら、われわれは、以上のような説明が見慣れない専門用語で武装されているからといって、そう簡単に納得してはいけない。ダマシオの知見は、多くの実験上の、もしくは診察上の事例から得られた手堅いものではあろうが、残念ながら、われわれが十分な説明と考えているものからはまだほど遠い。例えば、脳の領域のどの部位が身体のどこからの神経投射を受け、さらに脳のどの領域へ神経投射をしているか、といったことや、神経ホルモンの放出を左右する脳の部位がどの特定の脳領域と密接に関係しているか、といったことが分かったとしても、それは情動や感情を実現するメカニズムについての説明の＜始まり＞にすぎないだろう。＜脳領域相互の密接な関係＞や＜神経回路に対する特定の化学物質の作用＞などというものが、神経基盤のレベルと認知機能のレベルにおいて実際には何を意味し、また両者の対応関係はどうなっているのか、それらに関するもっとずっと詳しい説明が手に入るまでは、情動や感情の全貌が分かったとは言えないだろう。とはいえ、ダマシオが描く感情は、＜生命の管理＞という生物体にとって不可欠のメカニズムを利用して、＜生存の管理＞という行為者にとっての不可欠の機能をダイナミックに果たす感情の姿を、われわれに雄弁に物語っているように思われる。

３節　幻想する感情

しかし、感情は、実在の真の姿をわれわれに告げ知らせるだけではない。ある場合には、むしろ、実在とは異なる姿をわれわれに幻想として与えるように思われる。私がここで取り上げるのは、そのような＜幻想する感情＞の一つの姿、行為決定における行為者の「自由だ」という感情である。そこで、行為を始動させようとするわれわれの意図、つまり意識を伴った自由意志が実はそれに先立つ無意識の脳活動の結果にすぎない、という驚くべき事実を示した脳神経科学者のリベットの実験を簡単に見てみよう（Libet ［2004］pp.123-5. 邦訳、pp.143-84)。つまり、この実験によれば、われわれが意識を伴って意図的に何かをするような典型的な行為、自発的な行為（voluntary act）の実行の際には、常にその550ミリ秒も前に脳は活動を開始しているというのである。

この実験は、手首を曲げるような簡単な運動を被験者がやりたいときに自由にやってもらい、そのときの脳における準備電位(RP: Readness Potential)の発生の時点と、「今、やろう」という自分の意志に対する最初の気づき（アウェアネス）の時点をそれぞれ計測し、それらのタイミングを比較するのが目的である。実験に関するリベットらの細かい工夫は省略するが、そこでは、ある時間内でならいつでも自由に手首を動かしていいという場合（第一の場合）と、まったく何の時間的制約もなく自由にそうしていいという場合（第二の場合）の二つが報告されている。いずれの場合も、タイミングが比較されるのは、準備電位の始動時点と、気づきの時点（これはある仕掛けによって被験者が事後に報告できる）と、実際の行為時点の三つである。およそ、40回ほどの試行が行われ、それぞれの時点の平均が比較される。ここで、行為時点とは筋肉の活性化の時点を意味している。またここで計測されている準備電位は、大脳皮質の補足運動野を発信源とすると考えられている。

さて、この結果、第一の「時間内での自由」の場合では、行為時点のおよそ80

0～1000ミリ秒前に準備電位の始動が計測され、第二の「無制限の自由」の場合では550ミリ秒前にそれが計測される。しかし、いずれの場合でも、行為意志の気づきの時点は、行為時点よりもおよそ200ミリ秒前だった。すると、三つのタイミングの比較を総合するとこうなる。

第一の場合：準備電位の始動→（600～800ミリ秒）→行為意志の気づき→
（200ミリ秒）→行為開始
第二の場合：準備電位の始動→（350～400ミリ秒）→行為意志の気づき→
（200ミリ秒）→行為開始

これらの出来事の時間的な生起順序を素直に解釈するならば、行為をなそうとする意志は、少なくとも意識された限りでの存在としては、行為を開始する役割を果たしていないということだ。しかし、それとして意識されていない自由意志とは、もはや、ふつうの意味で理解されているいわゆる自由意志ではないだろう。リベットはこう結論する。「まず、自発的な行為を結果するプロセスは、行為しようとする意識的な意志が現れるずっと前に、脳によって無意識に起動される。これは、もし自由意志というものがあるとしても、自由意志が自発的行為を起動しているのではないということを意味している」（ibid. p.136. 邦訳、p.159. 強調は原著者）。こうして、自由意志の存在と役割に関する哲学上の争いに決定的な経験的証拠が得られたように思われる。いわゆる自由意志は錯覚でしかないのだ・・・

しかし、リベットは私にとっては驚くべき反応をする。彼はどうあっても自由意志の存在とその実効的な役割を否定したくはないのだ。この実験で（第二の場合）、意識的な意志は最大で脳活動（準備電位始動）より400ミリ秒ほど遅れるが、（50ミリ秒の報告誤差を補正すれば）少なくとも行為開始の150ミリ秒前には現われる。したがって、この150ミリ秒の間に意志が行為の開始や持続に影響を与える可能性は残されている。もっとも、リベットによれば、最後の50ミリ秒は一次運動皮質が脊髄運動細胞を活性化し、それを通して筋肉を活性化するのに使われるので、実際に活用できるのは100ミリ秒の間だけである。しかし、それでも実際に、われわれが予定していた行為の100～200ミリ秒手前のところで、その実行を拒否することができるのを、リベットは完全な意味においてではないが実験的に確かめている。そして、ここからリベットは自由意志についての新しい概念を提案する。「･･･意識的な自由意志はわれわれの自由で自発的な行為を起動していない。その代わり、自由意志は、行為の結果や実際の遂行を制御することができる。それは、行為を進行させたり、行為が起こらないように拒否することができる。･･･このようなケースでは、自由意志はたんなる受動的な傍観者にはとどまらないのだ」(ibid. p.139. 邦訳、p.162)。

しかしながら、当然、リベットがすがりつくこの＜拒否の意志＞すらもがそれに先立つ無意識の脳活動の結果にすぎないのではないか、という疑念が湧いてくるだろう。というよりも、そのように考える方が彼の実験全体とうまく符合するように思われる。しかし、リベットは、ここで断固として＜拒否の意志＞の突然の出現に固執する。つまり、この＜拒否の意志＞に関する気づきはそれに先立つ無意識の脳活動プロセスを必要とするが、その気づきの対象たる＜拒否の意志＞そのものは、それに先立つ無意識の脳活動プロセスを必要としない可能性がある、というのだ(ibid. p.146. 邦訳、p.171-2)。とはいえリベットがその根拠としているのは、その可能性を否定するだけの実験的な証拠がないという極めて消極的な理由にすぎない。しかも、少し考えれば分かるように、たとえ＜拒否の意志＞

がそれに先立つ無意識のプロセスなしに突然に出現したとしても、それは、それ自体が意識される前に出現したことにしかならない。つまり、われわれが拒否の発動を行っていると自ら意識したときには、それはすでに意識されずに発動してしまっているのだ！　というのも、意志の意識的な気づき（アウェアネス）には、その意志の内容にかかわらず常におよそ400ミリ秒必要だからである。とすれば、意識はやはりこの場合でも、傍観者たらざるをえないであろう。

もちろん、われわれはリベットに対して、その突然の＜拒否の意志＞や＜制御の意志＞は脳の中でいかなる先立つ原因もなしに生じたのか、とさらに問うことができよう。彼の用意する答えは新たな心脳の二元論であり、それを彼は「意識的精神場の理論 Conscous Mental Field Theory」と呼び、それを実証するための実験すら提案している(ibid. pp.157-184. 邦訳、pp.185-218)。しかし、それを論じるのは別の機会に譲るとして、ここでは、むしろ物理主義的な決定論の立場から、問題を眺めかえしてみよう(8)。もしこの立場が正しいなら、リベットが恐れたように、物理的原因をもたないような自由意志は存在しない。また、その意味での「自由だ」というわれわれの実感(feeling)は幻想である。実は、自由意志の存在に対するリベットの主張には、「積極的」な動機が二つある。その一つは、もし自由意志が存在しないならわれわれの倫理的責任の体系はどうなるのか、という帰結に関する危惧であり、もう一つは、われわれは自分たちの自発的な行為を実際に起動しているという実感である。どちらも、存在を立証するための十分な根拠とはなりえないが、存在を擁護するための十分すぎる動機にはなる。しかし、物理主義の立場からすれば、この後者に関して進むべき道は、「自由だ」という実感が実は真相を捉えていたという空しい論証ではなく、むしろ、真相を捉えていないどころかなぜ感情はわれわれに「自由だ」という幻想を持たせてしまうのか、ということの説明にある。

リベットも「実感(feeling)」と表現しているように、私も「自由だという感情(feeling)をもつ」という言い方の方が、「自由だという感覚(sensation)を持つ」という言い方より正確だと考える。なぜか。自由だという感じは、ある種の感覚器官（自由検知器官？）がわれわれに教えるものではなく（たぶん、そんな器官は存在しない）、それこそダマシオが描こうとしたような感情のダイナミックな仕掛けの総体だ、というのがその要点である。つまり、裏を返して言えば、感情はその機能を実現するための構造全体のゆえに、高次の行為者に、「自由だ」という幻想を実感させざるをえないようになっているのだ。それがどういう意味かを説明するために、ここで再び、ダマシオに登場してもらおう。

まず、人間のような高次の行為者においては、熟慮による行動に際して多くのの選択肢、多くののシナリオがイメージとして浮かんでくる。「イメージがあるから、すでにわれわれが手にしている行動パターンのレパートリーの中から選択し、選択した行動の仕方を最適化することができる（われわれは、いくぶん意図的、いくぶん自動的に、さまざまな行動オプション、さまざまなシナリオ、さまざまな行動結果を表象するイメージを、心の中で評価することができる」(Damasio ［1999］p.24. 邦訳、p.44-5)。ところで、ニューラル・パターンによって形成され、支えられるこのイメージは、視覚、聴覚、触覚、内臓、身体表面などから来るあらゆる事象のイメージを含んでおり、したがって、いま自分が直面している外的対象や、過去の出来事の記憶や、未来の出来事に関するイメージも含んでいる。しかし、その行為者（ダマシオの言う有機体）は、それらの事象のイメージを形成するだけではない。イメージの所有者である自己のイメージを形成し、事象のイメージが誰に属するのかという関係性のイメージも形成する。そしてダマシオは、＜認識の感情 the feeling of knowing＞、つまり「あらゆる種類の

事象のイメージの形成に伴う感情」(ibid. p.26. 邦訳、p.47) こそが意識だと主張するが、われわれは、いまは彼の意識の説明にまでつき合わなくともよいだろう(9)。われわれにとってのポイントは、行為者が効率のよい行為選択を行い、よりよき生存を維持するためには、感情というメカニズムを介してイメージの操作と生命の管理が結合される必要があること、そしてさらに、イメージの効率のよい操作のためには、＜事象＞と＜自己＞と＜その二つの関係性＞のイメージを形成する必要があるということだ。とくに重要なのは、＜自己＞を意味するイメージが他のイメージ群から区別され、それらが感情という舞台の上で関係しあうという構造が生まれたことである。

さて、このようなダマシオの道具立てから、われわれは、自由という幻想についての一つの説明を組み立てることができるだろう。いまや行為者にとって、さまざま行為のシナリオや行為結果のイメージと＜自己＞が登場する、いわば仮想的な＜自由の劇場＞が目前に展開される。その劇場において、高く価値づけられたある行為オプションと＜自己＞が重ね合わされるとき、複数の行為オプションからの最善の行為選択というイメージが形成される。このとき、この選択が自由意志によってなされる、と言う必要はない。行為者の行為選択は、決定論的であれ確率遷移的であれ、因果法則的になされる。しかし行為者は、その選択が、最も高い価値評価によって誘導された＜自己＞の欲求にしたがって行われているのを知っている。その欲求は、最終的には生命の管理と直結した感情のメカニズムが与えているだろう。そして行為者は、多くの場合、＜自己＞の選択した行為が実際に遂行されるのを知るようになる。つまり、このような＜自由の劇場＞において行為者の＜自己＞は、＜知る自己＞という傍観者でありながら、自らがその意志によって行為を選択し、実行したかのように事態を見るだろう。存在しているのは、行為オプション、自己、およびそれらの関係を表象するニューラル・パターンと、実際の行為を引き起こしているニューラル・パターンであり、これが、基本的には、行為者が自分には自由意志があると感ずる幻想の内実である。

このように見るなら、われわれが過去の行為に関して、＜他のようにも行為しえた could have done otherwise＞という独特の感じをなぜ持つようになるのかも説明できる。この場合、行為者が手にしているのは、過去の行為オプション、行為結果、自己、およびそれらの関係を表象するニューラル・パターンと、選択されなかった行為を擬似的に引き起こしている仮想的なニューラル・パターンである。このとき、行為者は、自分の欲求どおりに、あるいは意図どおりに行為をなしたといういつもの感じを、選択されなかった過去の行為オプションに対して持つことになるだろう。しかし、くり返して言えば、＜他のようにも行為しえた＞という感じをいかに強く実感したとしても、われわれが実際に現になしたのとは別のように行為しえたことが保証されるわけではない。むしろ、私が提供する構図にしたがえば、それは幻想である。

まとめよう。たんなる反射や本能的な行動だけに制限されているような低次の行為者ではなく、複数の行為シナリオを所有するわれわれのような高次の行為者の場合、感情は、それらのシナリオと行為を関係づけるために生命の管理という内的な使命を用いる。そして感情は、その内的な使命から導かれる欲求や衝動にしたがって、＜事象＞と＜自己＞と＜それらの関係性＞のイメージが演じる＜自由の劇場＞の芝居を、行為者の眼前に展開する。これは、因果的な世界のなかで、因果的なメカニズムとしての感情が高次の行為者のために果たす機能である。ここには、いわゆる自由意志の出番はない。しかし、行為選択のためのこの感情の機能、つまり＜自由の劇場＞の副産物として、行為者は、自己が自分の欲求どおりに選択し、実際に行為していることを知る。それが、われわれの「自由だ」と

いう実感の中身だ。つまり、自由とは、複数の選択肢のイメージを制御する感情というシステムが行為者にもたらす、ある意味で必然的な幻想にほかならない。あるいは、こう言った方がいいかもしれない。高次の行為者である限り、われわれは＜自由の幻想＞という刑に永遠に処せられている、と。

なお、本稿は、平成19～21年度科学研究費研究課題「認知ロボティクスの哲学」（代表者：金沢大学・柴田正良）の研究成果の一部を含んでいる。

注

(1)　フレーム問題、とくに「哲学者にとってのフレーム問題」と言われるものの詳細に関しては、柴田 [2004a]を参照されたい。そこでは、フレーム問題の源泉の解明と共に、自然知性においては感情こそがフレーム問題をやわらげている、という論点が提出されている。
(2)　この状況をチャーニアク[Cherniak 1986]は「有限性の苦境 finitary predicament」と呼ぶ。彼によれば、この制約のゆえに、われわれ人間は形式論理的に完全で健全な推論規則の代わりに、素早いがいい加減な発見法(quick but dirty heuristics)をしばしば用いる。
(3)　「これは人間の情報処理が原則的には固定プログラム（アルゴリズム）によってなされておらず、処理過程に自由があって、答えの見当をあまり大きく狂わせないような処理部分を省略することができるからと考えられる」(戸田 [1992]. p.53)。そして、それを可能にしたのは進化だ、と戸田は考える。「つまり、アージの活動プランが余裕時間などを考慮に入れて自動的に情報処理の仕方や行動プラン決定方式などを変更するのは、自然淘汰的に獲得されたトータルな合理性の近似解ではないか、ということである」(ibid. pp.85-6)。
(4)　この状況認知が極めて早く行われなければ、死が野生人を待っている。しかし、＜恐ろしい＞状況認知には無限のヴァリエーションがあるのでそれをすべて記憶することはできないし（データ量の爆発）、未経験の新しい＜恐ろしさ＞も次々に出現しそれとしての認知を迫るだろう（計算量の爆発）。そこで戸田はこう述べる。「だから認知された状況から“恐ろしさ”抽出するためには「情報処理」が絶対必要であって、しかもその情報処理は相当高度のものでなくてはならない。その高度の情報処理が非常に早く結論が出せるということは、それが完全に「技能」化されたものであることを示している。」(戸田 [1992] p.44)。
(5)　「行為と信念の決定における感情の機能は、＜純粋理性＞（＋単なる欲求）によっては埋められないギャップを、知覚の情報閉鎖を模倣することによって満たすことである。これは、哲学者のフレーム問題を扱う母なる自然のやり方の一つだ」(de Sousa [1987] p.195)。
(6)　フォーダーは『精神のモジュール形式』で、領域限定的な認知モジュールと中央演算システムに関して次のように述べている。「たとえ入力系が領域限定的であったとしても、そうでない何らかの認知メカニズム[中央演算システム]がなければならない。・・・入力系が与える表象は、どこかで相互に接触しなければならず、その相互接触を生み出すメカニズムは、それを生み出すというまさにそのために、一つ以上の複数の認知領域から得られる情報にアクセスできなければならない」 (Fodor [1983] pp.101-02)。
(7)　ここまでダマシオのソマティック・マーカー仮説を紹介したのだから、そ

のうちの重要な要素、仮想ループ（"as if " loop)を落とすわけにはいかないだろう（Damasio［1994］pp.155-8. 邦訳、pp.245-9）。これは、ソマティック・マーカーを実現するもう一つのメカニズムであるが、これまで述べてきた身体ループとは異なり、成長の過程で獲得されてきた。それは、もっぱら時間とエネルギーを節約するために、身体全体の本物の反応をバイパスして、感情の類似物を脳内に作り上げる。そのとき、われわれは、身体反応が実際には生じていないにもかかわらず、あたかも情動的状態を有しているかのように、あるいはあたかも身体が活性化されているかのように感じる。つまり、＜脳(a)→身体状態→脳(b)→脳(a)＞の身体ループの代わりに＜脳(a)→脳(b)→脳(a)＞の仮想ループが働き、それが推論や意思決定を支えるのだ。ダマシオにとって感情現象の本体はあくまで身体ループなのだが、仮想ループは、感情現象を複雑にしている「頭で感じられるだけの感情」を説明する上で重要な要素だと私には思われる。

(8)　ここで、私が取っている立場は、まさしくリベットが回避しようとしている立場、すなわち、実体一元論的な物理主義であり、性質に関してはスーパーヴィーニエンスを認めることでエピフェノメナリズムを積極的に承認する、いわゆる自由意志否定の立場である。心的因果とも関連して展開されるこの主張については、柴田［2004b]、柴田［2006a]、柴田［2006b]、柴田［2008]を参照されたい。

(9)　「意識がもたらす先駆的で斬新なやり方は、生命調節という内なる営みをイメージ処理と結びつけることだった。言いかえれば、有機体の内と外に存在する対象や事象を表象するイメージの処理に、生命調節システム----脳幹や視床下部など、脳の奥まったところにある----を関わらせることだった」(Damasio［1999］p.24. 邦訳、p.45)。「奇妙な言い方だが、意識は、われわれが見たり、聞いたり、触ったりするとき、事象(what happens)の感情として始まる。･･･そして適切な文脈におかれると、その感情は、それらのイメージにわれわれのものというレッテルを貼り、それよりわれわれは、まさにその言葉どおり、われわれは見る、聞く、触る、などと言う」(ibid. p.26. 邦訳、p.47.)。

なお、邦訳が存在する場合はできるだけそれに従ったが、一部、勝手に訳文を変えたところもある。ここで、それぞれの訳者に感謝の意を表したい。

参考文献：

Cherniak, Ch., 1986, *Minimal Rationality*, The MIT Press.（チャーニアク『最小合理性』柴田正良監訳、勁草書房、近刊予定）

Damasio, A., 1994, *Deacartes' Error: Emotion, Reason, and the Human Brain*, Penguin Books(2005).（ダマシオ『生存する脳』田中三彦訳、講談社、2000）

--------, 1999, *The Feelings of What Happens: Body Emotion in the making of Consciousness*, Vintage(2000).（ダマシオ『無意識の脳　自己意識の脳』田中三彦訳、講談社、2003）

--------, 2003, *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, Harcourt Inc.（ダマシオ『感じる脳』田中三彦訳、講談社、2005）

Dennett, D.,1987, ″Cognitive Wheels: The Frame Problem of AI″, in Pylyshyn [1987].
（デネット「コグニティブ・ホイール」信原幸弘訳、『現代思想』1987年4月号）
de Sousa, R., 1987, *The Rationality of Emotion*, The MIT Press.
Fodor, J. A., 1983, *The Modularity of Mind*, The MIT Press.（フォーダー『精神のモジュール形式』伊藤笏康・信原幸弘訳、 産業図書）
Hume, D., 1748, “A Dissertation on the Passions” in: *The Philosophical Works*, Vol. 4,
Th. H. Green and Th. H. Grose(eds.), Scientia Verlag Aalen(1964).（ヒューム『人間知性の研究・情念論』渡部峻明訳、哲書房、1990）
Libet, B., 2004, *Mind Time*, Harvard University Press.（リベット『マインド・タイム』下条信輔訳、岩波書店、2005）
Pylyshyn, Z. W. (ed.), 1987, *The Robot's Dilemma*, Norwood, NJ: Ablex Publishing Co.
柴田正良, 2001,『ロボットの心』講談社現代新書
--------, 2004a,「ロボットがフレーム問題に悩まなくなる日」『シリーズ心の哲学：ロボット篇』（信原幸弘編）所収、勁草書房、pp.119-174.
--------, 2004b,「The Exclusion Problem とエピフェノメナリズム」『理想』No.672、pp.69-82.
--------, 2006a,「よみがえったソクラテス-----物理主義と心的因果の問題を理解するために-----」『思想』No.982、pp.4-15.
--------, 2006b,「機能的性質と心的因果-----キム的還元主義を越えて-----」『思想』No.982、pp.53-76.
--------, 2008,「感情のクオリアと可能世界」『感情とクオリアの謎』（長滝祥司・柴田正良・美濃正編）所収、昭和堂、pp.3-30.
--------＆月本洋, 2007,「感情の作り方（ニューラルネットワークの場合）」『中部哲学会会報』第39号、pp.1-18.
戸田正直, 1992,『感情』東京大学出版会