

意識の起源と進化

意識はエピソード記憶のために生じたのか

前野隆司

1. はじめに

進化は魅力的である。長い年月を経て生命の高度な構造が創発してきた。そのメカニズムは、一見、ランダム探索と似ているように感じられる。物質が私たち人間を構成する細胞の配置になるように偶然に並ぶ確率は天文学的な数字なので、ランダムな探索の結果、人間が生じるとは考えられない。進化も、突然変異やランダムな交配に依存しているのだ。一見、ランダム探索と同様に極めて時間のかかる最適化演算であるように思える。このため、進化により複雑な生命が作り出されたという事実は直感的には信じがたい気がする。しかし、複雑で巧妙な私たち人間は、現に、単なる物質から進化してきたのである。これは、進化が、ランダム探索とは異なり、高度な生命構造を生成するために適切な手法であることによる。ここで、進化は、生命の遺伝子の構造をわずかに修正することにより個体の構造や機能の大きな変化を創発できるような手法であることが重要である。

私たちの身体のみならず、心も、当然、進化の産物であると考えられる。心の中の最大の謎といわれる「意識」ももちろん同様である。では、「意識」は、なぜ、何のために生み出されたシステムなのであろうか。また、進化系統樹の中のどの動物から、意識を持つようになったのであろうか。本稿では、これらの疑問に関する私の仮説(1)について述べたい。

第2章では、まず、進化はどのようにして生物種に新規性と無駄を付加してきたかを簡単に述べる。次に、第3章では、運動制御と行動制御を行なうために必要な脳の制御モデルについて述べる。第4章では、生物が持っていた制御系を微修正することによってエピソード記憶をするための機能が生じるはずであることを、第5章では、それを生成するシステムの機能が意識の機能とほぼ等価であることを述べる。第6章では、意識の機能ではなく意識の現象的な側面の由来についてコメントする。第7章は結びである。

2. 進化とは何か

まず、進化とは何か。いまさら私が説明するまでもないことだが、共通認識のために簡単な説明を行おう。

進化について合意されている点は、遺伝子の交差および突然変異と自然淘汰の結果として、変動する環境に適応できた種が生き残るような最適化手法であるという点である。最適化と異なり、ある環境に最も適応した最適なシステム構造を求めるような手法ではない。与えられた環境に対してそれなりに適切な種が生き残るのであって、たとえば、人間が最適なわけではなく、ゴキブリだってミミズだってそれなりに繁栄している。いずれも最適ではないが、それぞれ、現在の環境に適応している。

また、進化は決して突飛な「構成要素」を生成するシステムではない。つまり、新しい機能は創発するものの、それを形成する個々の要素は以前の生物の代替または

拡張である。つまり、進化というのは、真っ白な設計図にゼロから新しい生物のデザインをするような、華麗で創造的な最適化ではない。むしろ、突貫工事のようなものである。

哺乳類の骨を見比べてみるとよくわかる。ネズミもトリも人も、骨一個一個の形は様々であるが、骨と骨のトポロジー、すなわち、つながり方の関係性はよく似ている。鳥の羽は飛ぶためにゼロから設計されたものではなく、その先祖である飛ばない動物の前足を設計変更して作られたものである。人の手は道具を器用に操るためにゼロから設計されたものではなく、その先祖である、もつを持たない動物の前足を設計変更して作られたものである。

つまり、図一に示したように、進化による生物のデザインとは、船を金つちでトントンたたいて無理やり形を変え、自動車を作ったり飛行機を作ったりするような場当たり的で強引な作り方なのである。

身体だけでなく、脳神経系も同様と考えるべきである。もともと下等な生物が持っていたニューラルネットワーク構造に多くの追加のたこ足配線を行うことによって、新しい情報処理ができるように設計変更しているようなものと考えるべきであらう。

図一を見るとわかるように、進化は、トントンたたいたりたこ足配線をしたりといった突貫工事であるにもかかわらず、時として、新たな機能を生み出す。つまり、手足が羽根になり、空を飛べるようになったことにより行動範囲が大きく広がったり、ヒトが立ちあがって手が自由になることにより道具を使えるようになったり、脳の新皮質が大きくなることにより複雑な思考ができるようになる。進化は、遺伝子の変化は意外と小さいにもかかわらず、出現した身体機能という意味ではカタストロフィックな変化が生じるような非線形性の強い方法なのである。もちろん、よりマイナーな進化も含めて、生み出された新たな機能と考えるとよいが、ここでは明らかに新たな機能が生み出されたと考えうるような大きな変化に着目するものとする。

進化において生じる新たな価値には必ず「無駄」が含まれる。最適化では、新しい機能を発現するために必要な機能のみの付加が実現されるのに対し、進化はより冗長であり、言い換えれば無駄を含んでいる。しかし、そのことこそが、個体の適応や次の進化のために重要なのである。つまり、無駄は余裕ともいえる。冗長さは口バストさ(頑強さ)につながる。進化は、環境変動がドライブイングフォースとなっているため、ある静的な目的関数を満たすような最適化ではなく、目的関数自体の変動や拘束条件の変動に対して適応可能なような方向への準最適化となっているのである。

進化によって生じる「無駄」または「余裕」は、前の生物の形見である場合も多い。盲腸も尾てい骨も、以前の動物が持っていた機能が、人間では不要であるにもかかわらずたまたま失われずに残ったものであると考えられる。

つまり、重要なことは、進化によって生物に「無駄」と「余裕」が生じるのは、進化的に獲得した新たな機能に付随する「無駄」または「余裕」であるか、過去の生物が持っていた形状や構造の微調整の結果であるか、いずれかであるということである。もしも何らかの無駄や余裕が新たに生じたとなると、これまでになかった全く新しい構造によって、付加された有効な機能とは独立に無駄や余裕が生じることはないと考えるべきであると

いうことである。

以上のような考え方は、生命においては基本的に進化の淘汰圧に適合した機能のみが獲得されたと考ええる目的論的機能主義の一種(あるいは、無駄と余裕を包含するレベルへの拡張)と言えるであろう。本稿ではこれ以降、目的論的機能主義に、以上のような考え方が含まれるものと考ええることにする。

ここで私が何を言いたいのかということ先走って述べると、人間が持つ「意識」という機能も、それまでの生物が持っていた何らかの機能の微調整に基づき機能としては創発的に生じたのであって、これまでそれに類似したものがどこにもなかった生物から、ある種が進化した際に、突然、飛躍的な構造変化によって生じた機能であるとは考えにくいということである。

そこで、意識という機能が発生するに至る必然的な道筋を、以下に私なりに説明してみたい。

3. 生物の制御系進化の道程

「意識」が出現する前の生物の制御系はどのように進化したのであろうか。極めて大雑把に我田引水の工学的な議論をすると、フィードバック制御から、フィードフォワード制御へ、である。

図2をこらういただきたい。図2は生物の行う制御の流れ、すなわち、以下の四つの制御または制御系の学習を表している。

- (1) 身体と外界の変動に場当たり的に対処するフィードバック制御
- (2) こうなる・こうするためにはこうするべき、というように結果から原因を予測するための逆モデル(フィードフォワード制御器)のフィードバック誤差学習
- (3) 逆モデル(フィードフォワード制御器)を用いたフィードフォワード制御
- (4) こうしたらこうなる、という、身体・外界(あるいは脳のある部分)の振舞いの順モデルと、前述の逆モデルとを循環型に接続したイメージまたは思考(という内面的行動)

まずは、運動制御の例に基づいて、これら四つの制御について説明しよう。皆さんが腕を伸ばして何か物体をつかむリーチング動作を行うとき、腕の運動制御はどのように行なわれるであろうか。

まず考えられるのは(1)のフィードバック制御である。手を伸ばしてみ、物体に届いていなければもう少し手を伸ばす。行き過ぎていれば少し戻す。届いていないとか行き過ぎていくという情報は、視覚などのセンサ情報を用いて取得される。これがフィードバック情報である。このようなフィードバック制御は場当たり的である。

反射がまさにフィードバック制御である。昆虫は基本的に反射によるフィードバック制御主体の運動・行動を行なっていると考えられている。

蛾は、一方の目の前が明るく他方の目の前が暗いとき暗いほうの目のある側の羽はたきの量が大きくなるような反射行動を行う。その結果として、光の方に向かうことになる。つまり、蛾は、簡単な神経回路の宿命として虫除け灯に突入するのであって、なんらかの高度な意思決定をした結果なのではない。

これに対し、物体までの距離はこれくらいだから、筋肉にこれくらい力を入れれば手がちょうど物体に届くだろう、ということ脳内で計算して、計算結果に基づいて腕に力を与えるような制御を、内部モデルを用いたフィードフォワード制御という。フィードフォワード制御器は小脳にあり、(2)に示したようにフィードバック制御をした際の誤差を利用して学習されることが知られている(2)。「腕の筋肉に力を入れたら腕が動く」という世の中の原因と結果の流れとは逆に、「腕をここまで動かすためには筋肉にこれくらいの力を与えればいい」という逆問題を解くための制御器であるので、逆モデルとも言われる。フィードバック制御では行き過ぎたり戻りすぎたり、振動が生じる可能性がある上、情報を脳までフィードバックしてくるために時間がかかってしまうのに対し、内部モデル(逆モデル)を用いたフィードフォワード制御(3)では、予測したとおり腕がすつと目的地まで行くので、高速に適切な制御を行なえるというメリットがある。ただし、物体や腕の位置の変化を視覚や体性感覚などのセンサデータを用いて逐一モニタしているわけではないので、物体が動いてしまったり、何らかの原因で腕が思うようには動いていなかったりした場合に対処できないという欠点がある。

フィードバック制御を行なっていた生物から、フィードバック誤差学習を行うことによってフィードフォワード制御器(逆モデル)の学習を行なえるような生物へと構造的な進化が進みうることは、容易に想像できる。なぜなら、フィードバック制御器は、センサニューロンと脳のニューロン、運動ニューロンをつなぐニューロンネットワークであり、フィードフォワード制御器やその学習経路のニューラルネットワークも同様に簡単なニューラルネットワークの接続により実現できるからである。つまり、身体を金づちでトントンたたいてモディファイしたように、ニューラルネットワークの配線を多少複雑化させることによって、逆モデルを用いたフィードフォワードモデルを作ることができる。

先ほど、昆虫の制御はフィードバックが主体だと書いたが、多少の学習を行えることは周知の事実である。ある場所には食べ物があ、どこに自分の巣がある、といった情報を、昆虫も何らかの形で記憶することができ、これは一種の逆モデルであると考えられる。「ある場所に行く」と食べ物がある」という事実の逆モデルを学習した昆虫は、逆モデルを用いて、「食べ物を食べるためにはその場所へ行けばいい」という実世界と逆方向の行動指針を計算することができる。したがって、昆虫のような単純な生物は、フィードバック制御のみならず、逆モデルを用いたフィードフォワード制御も行なっていると考えられる。

もちろん、逆モデルを用いたフィードフォワード制御を行なえる生物は、フィードバック制御という場当たり的な行動しかできない生物に対し、環境変化への適応という面で有利である。(有利だから進化するのである。)(前述のように、逆モデルは一種の記憶であり、「何々をするためには何々をすればいい」という行動指針を計算できるからである。過去に経験した何らかの環境情報をフィードバック誤差学習という形で記憶し、それを次の運動・行動の際に利用できるからである。)

もう少し進化した生物は何ができるようになるであろうか。それは、順モデルの獲得と、順逆モデルを用いた運動・行動のエミュレーション(4)である。順モデルは、読んで字のごとく、逆モデルとは逆に、実際の世

中の原因と結果の流れと同じ方向に現象を脳内で再現するモデルである。リーチング動作の話に戻ると、腕の筋肉に力を入れると腕が物体まで伸びる、という現実の腕の動きを脳内で再現するモデルであり、ミラーニューロンがこれに関わっているといわれている。このようなニューラルネットワークが脳内にあると、脳内で運動の準最適化計算を行うことができる。つまり、脳内の逆モデルはもととは順方向に時間が流れる現実世界に接続されていたのに対し、図2に示したように両端のスイッチを切り替えれば、逆モデルと順モデルが循環的に接続される。これは、「こつする」ためにはこつすればいいはず」という逆モデルの予測を、実際にやってみる代わりに脳内の順モデルをやってみて、うまくいかなければそのずれの分だけ逆モデルの計算を修正して再び順モデルに流し込んでみる、という脳内シミュレーションが行なえるということである。川人はこれを順逆緩和型神経回路モデルと呼んでいる(2)。

順モデルも、逆モデルやフィードバック制御器と同様、ニューラルネットワークにより形成できる。したがって、進化というたこ足配線により脳内の順モデルを設計できることは容易に想像できる。

順モデルを獲得した生物は、場当たり的なフィードバック制御と、もう少し気の利いた「こつするためにはこつすればいい」という計算を行えるフィードバック制御のみならず、「こつするためにはどうすればいいか、頭の中で少し考えてみよう」というような順逆緩和型計算を行えるということだ。前の二つの制御よりも高度である。単に、逆モデルという、記憶に心じた運動・行動を行なえるのみならず、記憶を思い出してよりよい運動・行動は何であるかを脳内でシミュレーションすることができる。

脳の中には多数の順逆モデルが直並列に階層接続されており、外界および内部からの情報の流れに従って、寄せては引いて行く波のように、多数のニューロンが波動のように発火と停止を繰り返しているものと考えられる。つまり、図2では、順モデル、逆モデル、フィードバック制御器がそれぞれひとつずつしか描かれていないが、実際にはこれらが多数直並列に接続されていると考えられる。

ここで重要なことは、「こつ」までの話には「意識」は出てこないことだ。いや、「こつ」までの話の中で、「記憶」「考える」といった、意識と関連する単語を使ったではないか、と言われそうだが、そうではない。「記憶」には身体が記憶する暗黙知または非宣言的記憶と呼ばれるものがある。無意識に手を伸ばしても取る場合のように、体が覚えたやり方で、意識を介在せずに記憶が使われるケースは多々ある。また、順逆緩和型計算の説明の際に、「こつするためにはどうすればいいか、頭の中で少し考えてみよう」というようないい方をしたが、これは比喩である。自分は考えている、と意識するときの思考は意識下にあるといえようが、ここで、「考える」と言ったのは、無意識下で働くニューラルネットワークモジュールの間で順逆計算が繰り返される場合に、あたかも考えているかのように脳内で情報がぐるぐるともわるため、例えて言ったまでである。

つまり、場当たりのフィードバック制御にも、「こつすればこつなる」という記憶を用いたフィードフォワード制御にも、「こつするためにはどうすればいいか」を計算する順逆緩和型計算にも、意識は不要である。すなわち、餌があれば食べる、眠くなれば眠る、誰かが襲

つてくれば戦う、または逃げる、といったフィードバック制御主体の生き方をしている生物には、「意識」はなくてもかまわないということである。また、あそこには餌がある、あそこには自分の巣がある、こんな顔をした生物からは逃げるべきである、といった情報を逆モデルとして記憶できる生物も、「意識」を持つている必要はない。さらに、「餌があるのはあっちかこっちか」「巣に戻るにはこっちの道とあっちの道どちらが近いか」「こんな顔をした生物からは逃げるべきか戦うべきか」といった単純な逆問題を順逆緩和型計算によって解く生物も、特に「意識」をもつ必要はない。

一見、餌を食べている自分のクオリア、逆モデルの結果を用いて意思決定する自分のクオリア、あっちかこっちか思考している自分のクオリア、といったものを意識できた方が何らかの利点を持っていそうであるが、考えてみるとそのような種が「意識」を持つことの進化的な利点は特にない。脳のニューラルネットワークが同じ計算をして同じ結果を出してくれるのであれば、べつにその計算をしている自分自身の存在を意識のクオリアとして感じていなくてもかまわないのである。私たちは無意識的にバランスをとって立っていたり左右の足を順に出して歩いたりするが、すべての行動がそのように無意識的である生物を想像してみればよい。そして、前に述べたように、進化は、その種が進化する前の種よりも何らかの点で環境適応のために有利な機能が付加されるような準最適化であるから、意識という機能を付加することによってその生物が有利にならないのであれば、「意識」は生じないはずなのである。

4・エピソード記憶の進化的獲得

次に、エピソード記憶について考えてみよう。エピソードには、自分自身の旅行から帰ってきた、今朝は目玉焼きを食べた、といった自己の個人的なエピソードと、聖徳太子が大化の改新を行った、自民党が大勝した、ニュースで悲惨な事件を報道していた、といったような他人のエピソードがある。ここでいうエピソード記憶は、前者を指す。

エピソード記憶ができる生物と、できない生物とでは、何が同じで何が異なるのであろうか。

いずれも、意味記憶ができるという意味では似ている。ある場所において餌があったとしよう。エピソード記憶のできない生物も、ある場所に行ったときに餌を見つけた経験から、餌を食べるためにはあの場所に行けばよい、という逆方向の行動を導くための記憶を行うことができる。もちろん、腹が減ったときには記憶に基づいてその場所に行って餌を食べるといった行動を起こすことができる。エピソード記憶のできる生物は、ある日その場所に行ったらおいしい餌を見つけた、という体験を記憶できる。腹が減ったときにはその体験をもとにその場所に行くことができる。また、これらをもとに意味記憶を作成するような脳内処理を行うこともできる。したがって、単に餌の場所を覚えるだけであれば、両者は大差ない。

では、餌のある場所が多数あって、餌の量が変化する場合や、餌が腐り易い場合のように、環境が変動する非構造化環境に置かれた場合はどうであろうか。エピソード記憶ができない生物も、体験をもとに高度な逆モデルを生成できれば、それを用いて行動の場合分けをすることはできる。ある場所の餌が少ないときには別の場所に

行けばいいとか、ある場所の餌は腐り易いのでその場合は食べないで別の場所にいた方がいいといった行動の選択を行うことはできる。しかし、エピソードを記憶できないので、自分はきのうどの餌場に行つてどれくらい食べたかとか、おとといは新鮮だったがきのうは腐りかけていたといったような、スケジューリングの込み合った詳細な情報は把握できない。一方、エピソード記憶のできる生物は、きのうはAという餌場でおいしいものをすべて食べた、今朝はBという餌場に行つたところ腐りかけていた、といったエピソード記憶に基づき、今日はCに行つてみよう、という行動を起こせる。したがって、エピソード記憶をできる方が、より高度な行動を行えそうである。

コミュニケーションについて考えてみれば、両者の違いはより明白である。

意味記憶だけができる生物は、誰が自分の親戚で、誰が敵で誰が味方かという意味は記憶できる。したがって、親が近づいてくれば嬉しい、敵が近づいて来たら逃げろ、といった簡単な行動を行うことはできる。親がえさを口に運んでくれたら喜んで食べる、というようなことも行える。

しかし、「いつどこで誰と何をした」というエピソードを記憶できない。つまり、親といっしょにどこかおいしいものを食べた経験も、仲間といっしょに何かを成し遂げた経験も、敵と争った経験も、エピソードとしては記憶できない。「誰々はこのようだ」「何々はどのようだ」といった記憶しかできないから、「親は優しい」「友達と遊ぶと嬉しい」「あいつは嫌いだ」といった単純な意味しか記憶できない。思い出がないのである。

つまり、認知症(痴呆症)のような状態である。今朝は誰と何を食べたのが、友人とどんな文脈でどんな会話をしてどんな新しい情報を聞いたのが、読みかけの本の前半には何が書いてあったのか……、を思い出せないいや、今朝食事をしたかどうかが、友人と会ったかどうかが、本の前半を呼んだのかどうかもわからない。エピソード記憶ができなくなつてから二十年経つた人が、いつも自分は記憶を失う直後の自分、二十年前の自分、であるかのように振舞うというが、そのような状態だ。過去に起こつた様々なエピソードを何も記憶できないような状態なのである。

このような状態の生物に対し、エピソード記憶のできる生物の方が、環境適応上有利であることは容易に想像できよう。エピソード記憶をできるということは、「いつどこで誰と何をした」、そして、その結果を受けて、「次に、どこで誰と何をした」といったような、複雑な状況変化に伴い時間発展する原因と結果の連鎖を脳内に順モデルとして記述できるということである。「何々はこれこれである」というような意味のみを記憶できる場合に比べ、理解し表現できる内容がより入り組んでいて高度である。

エピソード記憶のできない生物の生活は、今朝何時に起きたか、あるいは、今朝食事をしたかどうかが、といったエピソードを思い出せないが、腹が減つたら何かを食べるし、眠くなつたら眠る、気ままな生活である。エピソード記憶のできる生物は、起きた時間も場所も、食事の内容も、食事をしながら読んだ新聞の内容も覚えていられる。気ままな生物よりも知的な生き方をできることは明らかであろう。

さらに、エピソード記憶を行える生物は、順逆緩和型計算により、様々な因果関係をシミュレーションできる。

複数の異なる経験結果をもとに、様々な行動の可能性について吟味を行う、思考という行為を行えるのである。以上より、エピソード記憶を行える生物は、進化的に有利であるから、エピソード記憶を獲得したと考えられる。

5. 意識とエピソード記憶の関係

では、生物がエピソード記憶を行えるためには、どのような機能が必要であろうか。

エピソード記憶ができない生物の脳の中では、フィードバック制御器、順モデル、逆モデルを司るニューラルネットワークが直並列に階層接続されており、外界および内部からの情報の流れに従って、寄せては引いていく波のように、多数のニューロンが波動的な発火と停止を繰り返していると考えられることを前に述べた。

著書『脳はなぜ「心」を作ったのか』「私」の謎を解く受動意識仮説(筑摩書房、二〇〇四)では、脳の中のニューラルネットワークのモジュールを、せつせとそれぞれ役割をこなしている小びとたちに例えた。稚拙な例で恐縮だが、「小びと」とは単純な機能を達成する脳内のモジュールの比喩である。

フィードバック制御器、順モデル、逆モデルを司る小びとたちは、自律分散的に情報処理をこなしているのであつて、そこに「意識」は介入する必要がなかった。なぜなら、人の行動は、「無意識」の小びとたちの民主主義の結果、自己組織的かつボトムアップ的に決定されるのであつて、「意識」という独裁者がトップダウンに決めるのではないからである。「 h は x である」という意味を記憶するだけの生物にとっては、「今、自分は何を行っているのか」を観測するための意識のようなシステムは必要ない。

これに対し、エピソード記憶とは、「自分は何を行つたか」の記憶である。自分は何を行つたか、を記憶するためには、それに先立って、「自分は今何を行つている」ということをエピソード記憶に送り込む機能が脳内に存在している必要がある。

仮に、「自分は今何を行つているか」をエピソード記憶に送り込むための脳内のシステムをモジュールAと呼ぼう。このシステムは、自律分散的に行われた小びとたち(脳内のニューラルネットワーク群)の仕事の結果として、自分の身体が行っていること、歩いているのか、しゃべっているのか、食べているのか、をモニターし、「私は何々をした」というエピソード記憶として記憶するためのシステムである。

このようなシステムは、図2に示したシステムの延長線上に考えられる。つまり、金づちでトントンたたくあるいは追加のたこ足配線を行うところの準最適化計算である進化によって、容易に作り出せると考えられる。すなわち、原因と結果の関係を記憶する順モデルを多少変更すれば、時間とともに発展する因果関係を記憶するモデルであるところのエピソード記憶を生成することはさほど困難ではないであろう。そのために、自分のサブシステムが行つた結果、今の自分が行つていることをまとめ、それをエピソード記憶に送り込むためのシステムであるモジュールAは、短期記憶に関連するシステムであると想像できる。なぜなら、短期記憶は、現在から少しさかのぼつたできごとを記憶するシステムである。自分が今行つていることは、今の瞬間のみのできごとではなく、前後の文脈を構築するに足るだけの時間間隔を

持った現象を表すはずであるから、短期記憶が行っていることは、まさにそのような現在の近傍のできごとの記憶であると言える。

以上、エピソード記憶を行うためにはモジュールAというシステムが必要であることを述べた。

話は変わって、「意識」について考えてみよう。既に「意識」には何度かふれたが、ここでは、意識とは何であるか、改めて確認しておこう。

意識には、モノやコトに注意を向ける働きであるアウェアネスと、自己意識であるところのセルフコンシャスネスがあるといわれる。いずれも、注意の対象がモノやコトなのか自分なのかという違いがあるだけで、現在自分が何をしているのかに注意を向ける働きであるという意味では同様である。

チャーマーズ(3)以来、心の哲学者は、「意識」を機能的な(ファンクショナルな)側面と現象的な(フェノメナルな)側面に分けて考える場合が多い。機能的な側面とは、「意識」というシステムは、「現在自分が何をしているのかに注意を向ける」という機能を実現しているモジュールである」といつとぎのように、「意識」がもつ機能面を表す。一方、現象的な側面とは、「現在自分が何をしているのかに注意を向けている」というその生き生きとした意識現象そのものことである。意識の質感であるところのクオリアについてもよい。多くの心の哲学者に合意されていることは、意識の機能的な側面をいかに明らかにしようとも、意識の現象的な側面はいかならず謎のままなのではないか、というところである。つまり、もしも図2と似たような図によって意識の機能を説明できたとしても、私たちが今生き生きと目の前のリンゴをみずみずしく思うクオリアや、今まさに私たちは生きていると感じるそのクオリア自体を、紙の上に図示して説明することはできない。

意識の現象的な側面については後で考察することにして、ここでは機能的な側面について考えよう。

意識の機能的な側面とは、「現在自分が何をしているのかに注意を向ける」機能である。

現在自分が何をしているのかに注意を向ける、ということとは、無意識の小びとたちの自律分散的な情報処理の結果、自分が現在体験していることのうち、最も重要で注目すべき点に意識のフォーカスを当てる、ということである。

ところで、モジュールAは、自分が行ったことをエピソード記憶に流し込むための機能を果たしていたのであった。ということは、意識の機能的な側面は、モジュールAの一部であると言えるであろう。つまり、自律分散的な小びとたちの振舞いの一部に注意を向け、体験し、それをエピソード記憶に流し込むのがモジュールAの働きであり、意識の働きとは、そのうち、注意を向けて体験する部分なのである。

つまり、エピソード記憶のためにはモジュールAが必要であり、意識はその前半部分なのである。言い換えれば、「意識」の機能は、エピソード記憶を行うために必要である。

ただし、「意識」の機能は、エピソード記憶を行うために十分であるかどうかは吟味する必要がある。もしも他の可能性、すなわち、意識の機能はエピソード記憶以外の何かのためにも価値があるのではないか、という可能性を否定できれば、意識はエピソード記憶の必要十分条件であるといえることになる。つまり、意識の機能は、エピソード記憶をするために進化的に生じた、と言える

ことになる。

意識があったほうが、なかった場合よりも、環境適応のために有利な局面があるだろうか？ 私は、何年も考え続けているのだが、エピソード記憶ができるという利点以外に、意識の利点を全く思いつかない。今の自分が行っていることに注意を向け、体験するシステムであるところの意識は、その体験を記憶するところのエピソード記憶の前処理としての利点以外に、なんら利点を持つとは思えない。

たとえば、言語を用いていることを意識するために意識という機能は発生したのだ、ということを主張する心理学の一派がいる。意識がなければ、少なくとも人間が操るような高度な言語は存在し得ないかもしれない。そうならば、意識は言語の必要条件であると言える。しかし、エピソード記憶はできるが言語は持たない生物はいるので、意識は言語の十分条件であるとは言えない。また、私たち人間は言語化されない体験をも意識できることから、意識は言語のため、とは言えないことが直感できる。

意識はあるがエピソード記憶ができないという生物もいるのではないかと考える人も多い。つまり、意識はエピソード記憶の必要条件ではあるが、エピソード記憶のできない生物、すなわち、昆虫や、もともと原始的なアメーバなどが意識を持っていたとしたら、十分条件とはいえないのではないかと、いつわけだ。昆虫が意識を持つているかどうか聞くわけにはいかない。証明は困難であるが、2章と3章で既に延々と述べたことは、その可能性の否定であった。つまり、進化とは環境に適応するための機能の付加であり、環境適応性に寄与しない機能が進化的に付与されることはない。もちろん、「無駄」または「余剰」が付与されることはあるが、それは、何らかの別の機能が付与されたことに付随して付与される場合であって、例えば昆虫という種が生じたときに、昆虫の環境適応のために必要とは思えない意識が付与されるとは考えにくい。一方、エピソード記憶ができる生物は進化的に有利であり、エピソード記憶を行うために意識の機能は必要なのであるから、エピソード記憶を行うために意識の機能が進化的に付与されることは理にかなっているのである。

したがって、私は、エピソード記憶のできる生物だけが意識の機能を持っていると考えている。

今わかっているところでは、鳥類や哺乳類の一部がエピソード記憶を行えるといわれているから、それらの種が意識を持っている種である可能性が高い。

ただし、たとえば、鳥類のエピソード記憶の実験に、以下のようなものがある。鳥に、新鮮な餌をAという場所とBという場所に隠させた。Aの方がおいしいのだが数日で腐ってしまう。ある鳥は、何日間かはAの餌を食べに行っていたのだが、腐ったところにはBを食べへに行っていたというのである。つまり、Aの餌はいつ置いたのだったのになら腐るといつ時系列情報を鳥はエピソードとして記憶できるというのだ。しかし、これは、意識ないしはモジュールAを用いないエピソード記憶である可能性もある。今私は餌を隠している、という体験を行わなくても、餌を隠した日から腐るまでは3日、という意味と、無意識的なタイマーを持っている、以上のエピソードの利用は可能である。したがって、人間のような高度なエピソード記憶を行うためには意識の機能が必要だが、単純なエピソードの場合には他のやり方もありえると考えられる。

したがって、厳密に言つと、他のやり方へ意識を用いないやり方では記憶できないほどの高度なエピソード記憶を行うために、意識の機能は必要十分である、というのが私の考えである。

6. 意識という現象は進化的に必要なか

以上の議論は心の哲学者の興味を満足させるものではなくろう。なぜなら、意識の機能的な側面に関する議論であったからだ。心の哲学者が知りたいのは、意識の現象的な側面には進化的な必然性があるのか、という点である。

意識の現象的な側面が進化的に意味をもつような状況には二つの可能性が考えられる。

ひとつは、意識の機能的な側面と現象的な側面が不可分である場合である。つまり、機能的な意識は持つが現象的な意識は持たない、というような状況が不可能な場合である。いわゆる同一説というのがこれにあたる。

もうひとつは、機能的な意識と、現象的な意識が、別々に進化的に生じた場合である。その場合には、現象的な意識が環境適応性の向上のために有利である局面があることを説明できなければならない。

私は直感的には前者であると思うものの、同一説に関する議論は決着しておらず、どちらかという同一説の方が悪そうな雲行きである。また、私自身も残念ながら明確な説明法を用意してはいない。ただし、チャーマーズと、チャーマーズ以来の今はやりのゾンビ論法には反論があるので、そこから話を始めよう。

チャーマーズは、意識の現象的な側面こそが難しい問題であり、これの存在により心身一元論は矛盾があるのだという(3)。

心身一元論の矛盾をつつとつとするチャーマーズが例に挙げるのが、ゾンビの話である。

本来、哲学で言つゾンビとは、外見が人間と全く同じで、表情も話す内容も行動もみな人間そっくりなのであるが、実は現象的な「意識」を持たないような存在のことであった。このような存在を想像できるであろう。

私には想像できる。

将来のロボットがそのような感じになる可能性が高いと考えられる。情報処理やニューラルネットワークの技術が向上した未来社会では、鉄腕アトムのような心を持ったロボットが作られるかもしれない。そのとき、意識の機能的な側面、すなわち、前述のように、自分の無意識の情報処理結果の一部に注目し、体験し、エピソード記憶するための機能を有するシステムは、容易に作れるようになっていいると考えてもよいであろう。しかし、その時点で、意識の現象的な側面を作り出す仕組みがまだわかっていないとしたら、そのロボットは、人間と同じ意識の機能を持つが、人間とは違って現象としての意識のクオリアは持たない存在となっているであろう。

そのようなロボットに、リンゴはおいしい？ と聞くとき、あなたが切ってくれたリンゴは今までのどんなリンゴよりも甘くておいしい、ということができるし、昨年のあなたとの思い出を振り返ってあの日の夕日はとてもロマンチックだった、ということもできるだろう。しかし、ロボットが、いくらリンゴの甘さが舌いっばいに広がって幸せな気分だと言ったとしても、昔体験した夕日が脳裏によみがえって切ない気分だといったとしても、そのように(いわばウンを)言つたための機能を満たすようなプログラムが動作している(あるいはニューラ

ルネットワーク群が発火して意味のあるパターンを生成している)だけであつて、甘さや幸せや切なさのクオリアは、一年前にも現在もロボットの脳裏や感覚器官に広がっているわけではないのである。

一方、チャーマーズのゾンビは、このゾンビの話を一歩進めている。外見が人間と全く同じであるのみならず、脳内のニューラルネットワークの発火分布も人間の場合と全く同じで、表情も話す内容も行動もみな人間のようなのであるが、実は現象的な「意識」を持たないような存在のことなのである。そして、チャーマーズは、そのようなゾンビを想像することが可能で、人間は皆、一人残らずチャーマーズ同様にそのようなゾンビを想像可能であるに違いないというのである。皆さんは想像できるであろうか。

もちろん想像できる、という哲学者が多いようなのであるが、私にはどうしてもできない。

脳内のニューラルネットワークの発火分布が人間とすべて同じであるにもかかわらず、意識の現象的な側面を持たないゾンビを想像できるということは、意識の現象的な側面はニューラルネットワークによって表現されているものではないと考えるということと等価である。つまり、脳内のニューラルネットワークの構造が進化することによって、意識の現象的な側面も生じたに違いない、という目的論的機能主義と相反する。私には、そのようなゾンビは到底想像できない。

では、前者が後者を、あるいは、後者が前者を論破することは可能であろうか。

それは、不可能であると思われる。前者は一種の心身二元論、後者は心身一元論である。これらのいずれが正しいかという論争は形而上学に属するものであつて、いずれかが他方を論破できるような論理構造にはなっていない。したがって、この議論は論理的にはお手上げである。

チャーマーズは、チャーマーズ流のゾンビを想像できると言つた瞬間に心身二元論に立脚しているのであるから、その視点から見ると心身一元論に矛盾があるように見えるのは当然であるし、そこから心身二元論が演繹されるのも当然である。一元論と二元論は前提が異なるのであるから、一元論の問題点は指摘できても論破はできない。

つまり、チャーマーズは、ニューラルネットワークによってクオリアを説明できないという理由によつて一元論を棄却する。しかし、一元論に立脚する者は、一元論で世界のすべての現象を既に説明できるから一元論に立脚しているのではなく、まず、前提としてのその枠組みを決めてから中身の吟味をしているのである。したがって、チャーマーズがいくら一元論の問題点を指摘しても、それでは、一元論自体が間違っている場合と、一元論の一部に未知の部分がある場合とを分離できないのである。

逆の議論も同様であり、一元論の枠内にある目的論的機能主義によつて、チャーマーズ流の二元論に反論はできても論破はできないのであるが、ここでは、反論のしかたを述べておく。当然ながら、目的論的機能主義に立脚する者は以下の議論に納得できるであろうが、チャーマーズや彼に合意する者は納得しないであろう。

意識の機能的側面も現象的側面も進化的に獲得されたのであるが、チャーマーズの言つように、現象的側面に限つてはニューラルネットワークの仕業ではない場合について考えてみよう。それはどのようなケースかと

いうと、あるとき、たぶん、人間という種が生じたときに、突然、ニューラルネットワークの発火などといった物質の振舞いからは説明できないような、新たなメカニズムである現象的な意識というものが進化的に生じたということになるだろう。

前述のように、進化とは、金づちでトントンたたいて無理やり形を変えたり、無理やりたこ足配線することによって突貫工事する準最適化である。そのような場当たりの泥臭いやり方の中から、どうすれば、物質の振舞いからは説明できないような、新たなメカニズムである現象的な意識というものが出現しようとするのだろうか。全くもって謎である。また、そのような、神秘的とも言うべき現象的な意識が、生物の環境適応性という意味でどのような利点をもたらすというのであろうか。やはり、全くの謎である。チャーマーズの言うように、解決の糸口が全く見出せないような謎である。このため、チャーマーズのゾンビ論法を受けて、難しい問題は今後当分解けない謎である、とまで主張する哲学者や脳神経科学者も少なくない。

一方、目的論的機能主義からみると、物質の振舞いと独立な現象的意識が進化的に生じるとは極めて考えにくい。

私は心身一元論に立脚し、脳のニューラルネットワークによって、意識の現象的な側面が作られていると考えられる。心身一元論に立脚するか、二元論に立脚するかは一種の信念である。神を信じるか否か、あるいは、自由主義と共産主義はいずれが本質的と考えるか、と問うのと似て、個人個人の過去の深く長い経験から帰納して、どちらが直感的に妥当だと考えるかを選択するしか、残念ながら形而上の議論にヒリオドを打つ方法はない。絶対的価値基準を持たない構造主義以降の世界において、正解が存在しないことは自明なのである。そのような枠組みの中で、私は、心身二元論に陥り、物質とは別の全く未解明な意識のクオリアが存在すると考えるより、心身一元論という現代日本の科学技術者の常識的かつ素朴な考え方の範囲内で、ニューラルネットワークが意識の現象的な側面を生成するやり方は不明だが、しかし、得体の知れない別のものではなく、ニューラルネットワークが意識のクオリアを作り出しているはずだ、という方が妥当であろうと、信念として思っている。

直感と言った方が妥当かもしれない。つまり、私は、脳のニューラルネットワークによって、(どのよう)にして生成できるのか、詳細は現状ではわからないけれども(意識の現象的な側面が作られると直感的に思うのである。もっといえば、ニューラルネットワークにより現象的な意識が作り出せる様を想像できるのである。コネクシオニストの言い分も同様であろう。なぜなら、ニューラルネットワークは身体や外界の運動・行動といった文字(記号)では表せないようなパターンを順モデルとして生成することができるシステムなのであるから、脳の他の部分の順モデルであるところの現象的な意識をも場として作られる様を、漠然とはあるが直感的にイメージできるのである(説明不足であることは承知している。将来もっと明確に説明したいが、納得できない方はまずは人工のニューラルネットワークを使ってみられることをお勧めする。そうすれば、私の「イメージできる」様子をイメージできるのではないかと思う)。

もちろん、逆の立場の者も信念または直感に基づいているに過ぎない。つまり、神経活動が同じで意識のクオリアを持たないゾンビを想像できる、あるいは、ニュー

ラルネットワークが意識の現象的な側面を生成することとはありえないはずだ、という信念に。実際、チャーマーズ自身、自分がゾンビを想像できる根本的な理由は、直感によるとしか説明できないと言いつつ切っている。

チャーマーズへの不毛で形而上の反論はこれくらいにして、機能的な意識を持つが現象的な意識は持たないというような状況の困難さを述べよう。つまり、機能的な意識が進化的に生じた際に、現象的な意識も同時に生じたのではないかと私は思うのだが、なぜそのように思うのかについて述べよう。もちろん、心身一元論の範囲内です。

さきほど、機能的な意識は持つが現象的な意識は持たないロボットは想像できると述べた。このロボットが「切ない」というとき、本当は胸がキュンとする、「切ない感じ」を持たないのである。また、「痛い」というとき、「痛み」を持たないのである。では、感じていないクオリアを感じているというウソをついているのであろうか？ 未来のロボットが人間同様に切ないとか痛いというとき、そのロボットの表情も(存在するならば)脈拍も顔のほてりも、あらゆる反応はヒトが本当に切ないまたは痛いと感じるときと同様であるはずである。また、「それは本当はウソなのではないですか、本当は痛くないのではないですか。」と聞いたら、「何を言っていますか、本当に痛くてしょうがないんですよ、早く薬を飲ませてください。いたたたた……。」といったのた打ち回るはずである。ロボットは、本当は痛くないとか、実はウソだなどとはみじんも思わないのである。なにしろ、意識の機能は人間と同じなのであるから、本当に痛いかどうかのように振舞うのである。また、「一時間前の痛みはどうでしたか」と聞くと、「一時間前は今よりももっと痛くて失神しそうでした。」などと言っているのである。本当は一時間前にも今と同様、痛みをクオリアを感じていなかったにもかかわらず、一時間前の痛みを痛烈に感じていたと真顔で言うのである。しかも、本当にそう思っている場合の人間と全く同じように振舞うのである。

そのようなロボットは当然チューリングテストにパスする。なにしろ、百パーセント何の隙もなく、本当に痛いというのだ。自らウソをついているという自覚さえないのだ。もはやロボットを外側から見ても、本当は痛いといっているのはウソだ、と見破るすべはどこにもない。そこまで臨場感のあるロボットは、もはや痛みをクオリアを持っていないと言いつつ切れるのかどうか疑問であるといつていいほどである。なにしろ、一時間前に痛かった様子を、ことごとくに真顔で説明できるのだ。本当にここが痛かったと顔をしかめて痛々しく痛かった箇所を指し示すのである。そこまで臨場感のあるロボットをつくるなら、いつそのこと本当にその場所を痛がらせてやればいいのではないかという気がする。しかし、現象的な意識のクオリアの作り方はわからないから残念ながら今のところ作れないのである。

さて、ロボットを進化的に生じた人間そっくりの生物に置き換えてみよう。

つまり、機能的な意識はエピソード記憶のために進化的に生じたのだが、現象的な意識は持たない生物がたまたま地球上が他の惑星上に生じたとしてしよう。

そのような生物は進化的に生じそうであらうか。ただし、もちろん、ここでは心身一元論に立脚するものとする。もっといえば、物理主義の中の、コネクシオニズムと目的論的機能主義の和集合の部分に立脚する。したがって、前提として、この世界は、ニューラルネッ

トワークによって現象的な意識を進化的に作り出せるような場であることとしよう。

さて、現在の痛い様子または一時間前に痛かった様子を、痛かった箇所を指し示しながら、ことまかに顔をしかめて痛々しく真顔で説明するにもかかわらず、本当は痛みのクオリアを感じていないような生物が、進化的に生じるであろうか。

ロボットの場合には、クオリアの作り方を人間が知らないから、そのような奇妙なロボットを作らざるを得なかったといえる。しかし、ニューラルネットワークによって（現代の人間はまだやり方を解明していないとはいえず）現象的な意識を作り出せるのであったら、巧妙に痛いふりをする生物を作るよりも、本当に痛みのクオリアを持った生物を作った方が、わかりやすい。何しろ、本当に痛くて、それを痛がるのだ。本当は痛くないのにあたかも痛かった場合と同じように振舞うといったようなまどろっこしい生物を作るよりも、クオリアを持つ私たちから見ると直感的でわかりやすい（わかりやすいだけであって、必然性の証明にはなっていないが）。

では、人間はどうなっているのだろうか。指先に大きな痛みが生じ、痛覚受容器が発火すると、その情報が大腦皮質に伝達される。大腦皮質の感覚野には全身の地図があり、指が痛いときには指の感覚に対応する箇所のニューロンが発火する。対応した大腦皮質の感覚野で痛みという質的情報が生成されるように思えるのに、実際の痛みのクオリアは指先で感じるのである。指先に痛みのクオリアという現象を作り出すような装置はなく、単に痛覚受容器があるだけであるのに、である。ラマチャンドランの幻肢(4)はさらに不思議に思えるかもしれない。右腕を失った人が、右の指を握り締めていて爪が手のひらに食い込む痛みのクオリアを生き生きと感ずるといふ。右手はないにもかかわらず、である。

これらはどういうことであろうか。本当は痛みというのは脳の神経発火パターンによって生み出されているはずなのに、何らかの方法によって、ある身体部位に痛みという意識の現象的側面が生じるようにできているとしか考えようがないように、私には思える。拙著(1)にも書いたように、人は、そのように幻想または錯覚するように作られているから、そう感じるのであると考えるしかないように思える。どのようにすればそのような幻想または錯覚を作り出せるのかはわからないが、前述のまどろっこしい生物よりも、痛みのクオリアをその場所で感じる生物の方が直感的でわかりやすいからそうなっているに違いない。

なお、「痛み」といふ脳の神経発火パターン」と「身体部位の痛みのクオリア」は不可分である。クリプキはそのような物理主義はおかしいというが、私は、不可分であると考え。そうであれば、死んだ人の脳に痛みの神経発火パターンを再現したとき、痛みは生じないではないか、とクリプキはいうが、それは条件を見誤っている。意識のスイッチがオンになっているときに、「痛み」といふ脳の神経発火パターン」と「身体部位の痛みのクオリア」は条件付き不可分なのである。そのような条件付き不可分な状態が、別のニューラルネットワークの演算によって定義されているのである。眠っているときや死んだ後はその条件を満たさないから、神経発火パターンがあっても痛くないのである。

以上、この章で述べたことをまとめると、痛みというクオリアの存在の進化的な妥当性の説明には至らなかったものの、「現象的な意識はニューラルネットワーク

によって作られる（に違いない）」という前提（信念）のもとでは、旧来ゾンビ型の生物よりも、幻想または錯覚であるところのクオリアを感じることでできる生物の方がありえそうである、という程度の主張を行ったわけである。

いずれにせよ、現時点で言えることは、現象的な意識は、私が立脚する二元論の枠内ではニューラルネットワークによってどのように作られるのかがまだ明確にはわからず、チャーマース流の二元論では解決の糸口すらつかめない謎となる、ということである。そしていずれを選ぶかは、個人の信念にゆだねる他ない。

7. 死は進化のためにある

最後に、死生観について述べたい。ここからは工学者としてのコメントではなく、意識を持ったある一人の人間の情緒的なコメントである。

もともと私が意識の問題に興味を持ったのは、自分自身の意識はどこから来てどこへ行くのかを第一人称的に知りたかったからであった。今は、私という身体が生まれた後、たぶん二歳くらいときに無から生まれ、将来肉体の死とともに失われるのだと考えている。

進化という文脈で述べると、少なくとも機能的な意識は個体のエピソード記憶作成のための前処理なのであって、個体の役割が終われば意識の役割も終わると考えるのが自然である。もちろん、生物に死があるのは進化のためであろう。死なない生物が生き延びていたのでは、新しい個体の生きる場がない。新しい個体が生まれなければ、環境に適応した新たな種が生まれる進化というメカニズムも働きようがない。したがって、ドーキンス(5)が述べたように、個体は遺伝子の乗り物に過ぎない。意識も、使い捨ての乗り物の部品に過ぎない。

以上のように、第三人称的に機能的意識の由来を理屈としてわかった気にはなれたとしても、第一人称的な私自身の現象的意識にとつては容易には納得できない。死にたかない。むなし。もっとほっとする結論はないのであるうが。

先日、ある老婦人と心について話したときの、彼女のコメントが印象的であった。彼女は言った。人間は風になりたいんだと思います。風は、何も考えずに吹いていきます。昆虫も、何も考えずに生きていて、何の悩みも持たない。悩みを持つのは人間だけなんです。人間が、古い、記憶を失い、意識が薄れて恍惚の人になるのは、風や虫たちのように、幸せになつていくためなんだと思います、と。

考えてみると、子供時代は親に守られていて幸せだった。いつの間にか、知識が増え、記憶が増え、煩惱が増えて今に至った。それを次第に忘れ去って、機能的な意識も現象的な意識も薄れ、次第に自然に戻っていく様が、人生の後半なのであろう。無から人間ができ、再び無へとソフトランディングしていくのが、種の中の個体としての人間の営みなのである。

註

(1) 前野隆司『脳はなぜ「心」を作ったのか 「私」

の謎を解く受動意識仮説』(筑摩書房、二〇〇四)

(2) 川人光男『脳の計算理論』(産業図書、一九九六)

(3) デイヴィッド・J・チャーマース『意識する心 脳

と精神の根本理論を求めて』(白楊社、二〇〇一)

(4) V.S.ラマチャンドラン『脳の中の幽霊』(角川

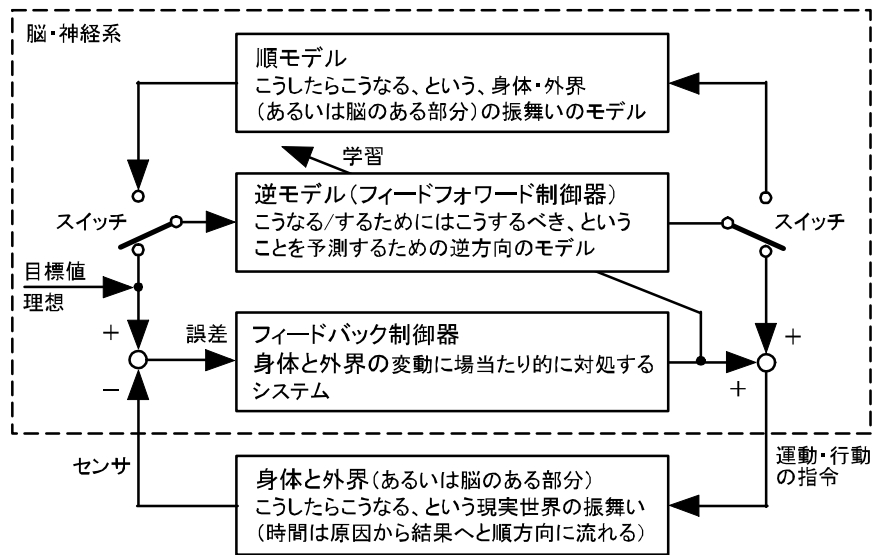
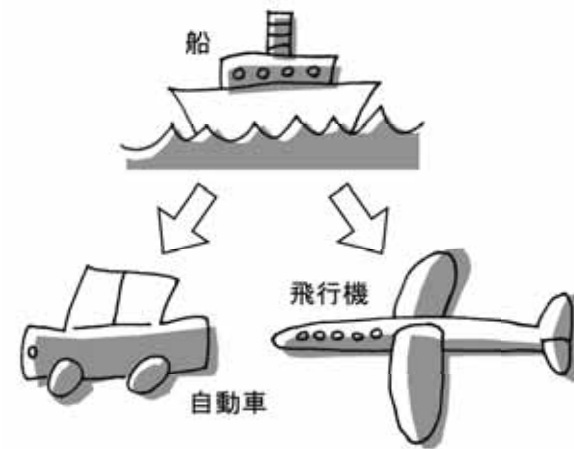
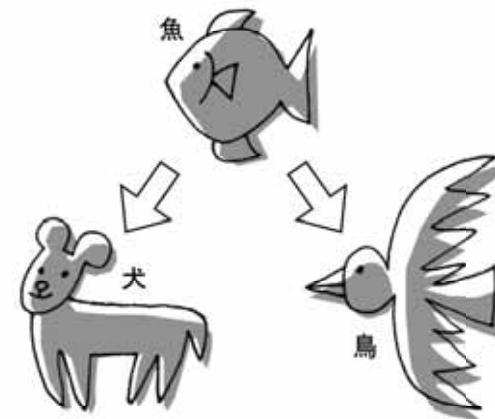


図2 脳の制御系の概念図



(a) 人工物: 構造も機能も不連続に変化



(b) 生物: 遺伝子はわずかに変化/機能は大きく変化

図1 人工物の設計と生命の進化的設計の違い

書店、一九九九)
 (5) リチャード・ドーキンス『利己的な遺伝子』(科学
 選書、一九九一)

(まえの たかし・ロボティクス)