

人工生命モデルによる構成的研究の方法論

A Methodology for Ensuring Constructive Approaches based on Artificial Life Models

有田 隆也
Takaya Arita

名古屋大学 大学院情報科学研究科
Graduate School of Information Science, Nagoya University
arita@nagoya-u.jp, <http://www.is.nagoya-u.ac.jp/>

Keywords: artificial life, constructive approach, methodology, emergence

1. はじめに

人工生命は、生命の振舞いや機能に関する原理を抽出し、コンピュータなどの人工的なメディアによってそれに基づく現象を再現することにより、生命に関する普遍的な知見を得ようとする研究アプローチである[有田 02]。従来の生物学が対象としてきた、この地球上に存在する（あるいは存在した）生物にとどまらずに、想定しようとするならば想定できる「存在しうる」生物も含めて幅広く「生命」を捉えた上で、生命現象の本質を明らかにしようとする点の特徴である[Langton 89]。

人工生命と銘打った研究領域がスタートした当初から、その研究対象は生命系にとどまらず、知能、心、社会、言語、…、と幅広い。対象は様々であっても、対象とする系における創発的現象への着目、及びその創発現象を作り出して理解するという構成的研究手法の両者は共有されている[有田 02] (図 1)。

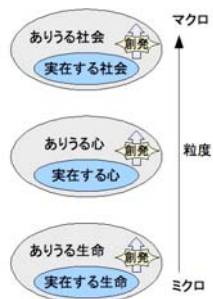


図 1 「存在しうる現象」の創発による理解

ただし、作って理解する科学がいかなるものか、その方法論的な根拠、科学としての根拠は自明ではない。従来の伝統的な自然科学の研究手法とは異なり、自分で構成した人工生命モデルを、ターゲットとする具体的な現

象やデータと直接的に比較検証するプロセスが前面に出て来ないのがむしろ普通であるからだ。このことは、従来の研究手法に馴染んだ研究者にとっては尋常ではないと感じられるだろう。本稿では、人工生命がいかなる科学であるのか、その方法論を検討する。さらに、このような研究手法に基づいて、特に心を研究対象とする場合の可能性[有田 07]について考える。

2. 人工生命モデル例 Tierra

以降の議論を円滑にするために、人工生命研究の典型例として、生物学者 T. Ray による、生態系進化を対象としたデジタル生命モデル Tierra [Ray 92]を簡単に説明する。

Tierra では、自分自身を別の記憶領域に複製することを繰り返すプログラム（祖先）一個用意し、その実行を開始することにより、デジタル生命による生態系を動かす。自己複製によって生成されたプログラム（子）は並列実行を開始し、同様に自己複製を繰り返す。この繰り返しにより、複製されたプログラムがメモリ領域を埋め尽くすと、基本的にはもっとも古く実行を開始したプログラムから終了させていく。

これだけでは、単に、最初に実行開始したプログラムとまったく同じクローンが自己複製し続けるだけにすぎないが、「突然変異」のメカニズムが別に組み込まれている。これは、命令コピー時、演算時、あるいは不定期に、命令やデータが別のランダムな値に置き換わってしまうものである。これにより、プログラムの多様性がたえず作り出される。1) 集団内に個体差があり、2) その差によって効率よく子孫を残せるものが生じ、3) 親のプログラムは子のプログラムに伝わる、つまり、変異、適応度差、遺伝の3条件が揃う可能性があり、これは計算機内で適応進化が起こることを意味する。

実際、命令をコピーする部分を自分は持たずに隣接プログラムから借用する「寄生」プログラム（これは構成命令数が小さいため、素早く子孫を増やせる）が生まれたり、逆にこのようなプログラムを自分の自己複製に利用してしまう「重寄生」プログラムが生じた。さらに、お互いに利用し合う「共生」関係が出現するなど、複雑な「生態系」が観察された。

Tierra が発表された 1990 年代初頭、計算機内でのそのような生態系が進化するということに多くの人々が驚き、興味を持ち、研究領域の隆盛の大きな原動力となった。Tierra 以降、もっとも活発に研究されているデジタル生命は、C. Adami らのグループが開発し、進化生物学者も加わった Avida である。Avida は、実行した演算に対して明示的な報酬を付加的に設定可能である点と、プログラム間の相互作用を限定している点が Tierra と異なる。これらにより、容易に複雑性が進化することを狙っており、実際、デジタル生命による進化研究 [Wilke 01][Lenski 03]の位置づけを確かなものにした。

3. 人工生命モデル研究の方法論

本節では、人工生命の方法論の独自性と意義について、「モデル」という言葉の意味に立ち返って、考えることにする。

従来の典型的な自然科学では、具体的な現象やデータを観察し、そこから一般性をもつ仮説、法則、理論を帰納に推論する方法（枚挙的帰納法）や、逆に、仮説、法則、理論を立て、実験や観察を行うことにより妥当性を裏付けていく方法（仮説演繹法）が用いられてきた（図 2）。そして、大きな成果をあげてきたといえる。

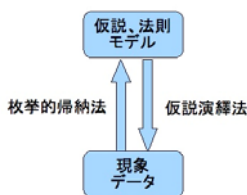


図 2 基本的な自然科学の研究方法

一方、典型的な人工生命モデルは、具体的な現象やデータを直接的にはターゲットにしていなくて、大きな特徴がある。Tierra モデルにおいても、生じたプログラム同士の相互作用を自然の生態系で観察される具体的な現象と直接的に比較検証する余地はない。ということは、人工生命モデルには対象やターゲットがない、あるいは、図 2 における、現象・データとのリンクがないのである。言い換えるならば、モデル=対象 [Moreno 02] ということであろうか（図 3）。いずれにせよ、このような立場をとるならば、「モデル」という言葉は誤用に近いと言

えよう。

ALモデル
(=対象)

図 3 人工生命の研究方法の解釈 (1)

しかし、人工生命モデルには対象がないと言い切るのはどうしても腑に落ちない。他のいかなる物事との整合性もなく、人工生命モデルを扱っているわけではないからである。では、どのような作業を行っているかと言えば、それは、概念レベルにおける検討である。

一般に、何かを理解する場合、その理解にはいろいろなレベルがありうる。Tierra を計算機で動かすとき、そこで焦点となっているのは、進化のシミュレーションではなく、そこで起こっている進化そのものである。そもそも C. Darwin は、概念レベルで進化を考えだし、提唱している。そして、そのレベルでの進化の理解は、物質的分子生物学的メカニズムには依存しない。

Tierra を始めとする多くの人工生命モデルは、概念レベルでのモデル化を行ったものである。この立場からは、人工生命研究は現象やデータとのつながりは必ずしも強く維持せずに、仮説、法則のそれぞれの意味やそれらの間の関係を計算論的モデルにダイレクトにマッピングして（実際の現象やデータの表象のさらなる表象、つまり 2 次の「シミュラクラ」 [Emmeche 1994] を扱うと言え）、理解していくものであると言える（図 4）。



図 4 人工生命の研究方法の解釈 (2)

ただ、このような立場でも、まだ、早い段階で人工生命に興味をもった進化生物学者 Maynard Smith が指摘したような、人工生命は“fact-free”な科学であるという批判 [Horgan 95] に対して、正面から応えているかどうかは定かではない。図 2 に示すような確固たる科学的検証のループが存在するか明確ではないからである。

ここで考えるべきことは、人工生命モデルのほぼすべてが対象としている現象である。それは、1 節で述べたように「創発」、要するに、「要素間の相互作用から、マクロなレベルの振舞い、機能、構造が出現する現象」である。Tierra でも、80 個の命令からなる「先祖」プログラム自体に何らかの生命としての価値を見いだしているわけではなく、そのようなプログラム間の相互作用から創発する現象に、生態系の本質がある。

人工生命、あるいは複雑系科学の研究者は、創発を、

生命を始めとして、様々な複雑系において普遍的に見られる中核的現象であると考えて追究している。ただし、興味深い創発現象を計算機の中で起こすのは容易ではない。その意味で、少なくとも、真つ当な創発現象かどうかという検証過程は常に付随している。

このような理解も踏まえると、図5のような解釈が得られる。研究者は何らかの仮説や法則を制約として、(直接的には実際の現象やデータを対象とせずに)概念レベルにおいて人工生命モデルを構成する。それは、要素間の関係や動的特性を計算論的、あるいは数学的に記述するものである。そして、計算機の中で時間発展させ、興味深い創発現象を起こし得たときに、その挙動を従来の仮説や法則の中に位置付けながら解釈する。なお、その解釈の中で、あるいは元の人工生命モデルのレベルで、実際の現象やデータとの比較が可能な場合もある。

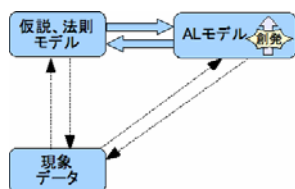


図5 人工生命の研究方法の解釈 (3)

このような人工生命の方法の位置付けはむしろ穏当なものかもしれない。人工生命モデルが作り上げるような人工世界を用いた科学を支持し、“fact-free”との批判に対してもっともラディカルに応じるのは、市川惇信の「人工科学」[市川 03]の概念であろう。彼は、数学のような、人工の世界を扱う科学を人工科学と呼び、数学が公理系で与えられた世界の性質を明らかにする作業と、自然科学が宇宙は与えられたものとしてその性質を論じていることは基本的に同等であり、人工科学も科学であるとしている。そして、自然科学の仮説は観測との整合性で評価されるのに対して、人工世界はその有用さと美しさにより評価されるとしている。

ただし、市川は、特に、人工科学の中のモデル科学については、モデルの記述能力、解析能力、数学的操作能力、現実への逆写像能力で評価すべきとしているが、あえて、人工科学の基本的概念の通りに人工生命を位置づけると図6のようになるだろう。ある種の潔さを感じる立場であるし、もちろん工学的な人工生命研究にとってはこの解釈は十分妥当であるが、現実を支配している法則との関係性はもっと密であると感じる。

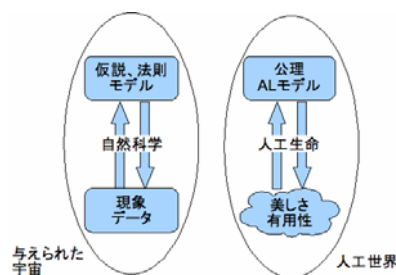


図6 人工生命の研究方法の解釈 (4)

4. 人工生命モデルの分類と特徴

3節では、Tierra に代表されるような典型的な人工生命モデルを念頭において、人工生命モデルの独自性について考察した。しかし、人工生命モデルと一言と言っても、対象が様々であるというだけでなく、そこで論じたモデルの抽象度も様々である。Barandiaran らは、人工生命モデルを抽象度に応じて次のように4分類している [Barandiaran 06].

1) 一般モデル

抽象度が極めて高く、複雑系全般に関わる一般的な性質を導くことを目的とする。例として、セルオートマトン、スケールフリーネットワーク、ランダムブーリアンネットワークなどがある。パラメータとシステム挙動の関係を探る。実データと照合可能な場合もある。

2) 概念モデル

人工生命によるユニークな「モデル」であり、一般的な複雑系に関わる概念(創発、進化、複雑性)の定式化や、特定の理論における概念間の関係など概念レベルの解明を目的とする。前者の例として、進化を扱った Tierra [Ray 92]、後者の例として進化と学習の相互作用に関する鈴木らの研究[Suzuki 04]がある。

3) 機能モデル

全体の機能や振舞いが部分の再帰的な相互作用によって創発している実在する特定の系の究明を目的とする。概念モデルとは違い、ターゲットとする系の個別の制約を考慮する。社会的生物による群知能が例としてあげられる。

4) メカニズムモデル

機能モデルと構造に関するモデルの両者を兼ねたモデル。実在する現象を忠実に再現することが第一の目的である。モデルのパラメータと実在する現象での観測可能なパラメータを対応させる。

これらの4分類は厳密なものではなく、4つは抽象度の軸上に連続的に位置すると考えるべきであろう。重要

なことは、モデルの抽象度に応じて研究の目的も異なってくることである。このことは、一見当たり前のようであるが、人工生命研究の当事者でも認識しているか定かでないと感じることがある。特に、概念モデルや一般モデルの人工生命研究を行う際に意識すべき点は次の2点である。

第一に、概念モデルや一般モデルの研究の際には、実際の現象データとの照合を無闇に目指すべきではないということである。もちろん、照合がうまくつくことはよいことであるが、基本的にはレベルの異なる話である。モデルのパラメータを増やしていけば、基本的に望む現象に近いものを起こすことができるだろう。しかし、それによって、一般モデルや概念モデルにおいて目指す理解からは離れていく。

第二には、多くの概念レベルや一般モデルにおいて、モデルを動かした結果、何らかの定量的なデータが出力されるが、その数字自体に過剰な意味付けを行ってはいけないということである。様々な概念や現象の動的な特性に関する一般的な議論、定性的な議論をするためのものであり、実際のパラメータとの直接的な対応関係が明確でないパラメータに基づく部分が少しでもあるならば、定量的な議論は難しいからである。

以降では、4分類の中でもっとも人工生命らしいと言える概念モデルに相当する人工生命モデルについて焦点をあてる。

まず、この種の人工生命モデルが対象とするのは、まだ明確化されていない、あるいは定義が明確でない概念（法則、仮説、モデルも含める）である。その概念には、複雑系の本質である創発性が深く関連する場合が多い。たとえば、創発そのもの、生命、言語、社会、多様性、進化、心、階層性などである。

そして、その動的な側面に着目するが、特にそれが実際の現象としては観測しにくい場合に真価を発揮する。たとえば、時間・空間スケールの異なるプロセスや現象の間の相互作用（遺伝子型と表現型、進化と学習、個体と集団行動、神経回路と行動）である[Barandiaran 06]。また、ある概念が前提としている前提、土台を問う場合にも有用である。その部分が白紙にしたモデルで必然的にそれが創発しうるか問うという立場である。

このようにして、人工生命モデルを用いることにより、対象とする概念のデモンストレーション、定式化、証明を行うことによって、その概念を大きな知の構図の中に位置づけることを狙う。同時に、その概念の応用可能性も検討する。

人工生命は哲学であると当初から言われていた[Dennett 94]が、確かに、それは概念、及び概念の間の関係の理解のためのツールを提供しうるし、哲学の方法である思考実験と非常に似ている面がある。哲学者でもある人工生命研究者 Bedau も、人工生命モデルを用いた実験は、創発現象に関わる様々な手法と計算機パワーを備

えた強力な思考実験であると指摘している[Bedau 02]。

ところで、人工生命コミュニティでは、人工生命モデルの元祖は、Von Neumann による自己複製モデルであるとししばしば言われる。様々な領域で先駆的業績を残した彼は、1940年代後半から1950年代前半にかけて、生命性の一つの本質である自己複製現象に関して、実際の生物現象を離れて、人工的な世界で自己複製モデルを考え出した。

1950年代前半に考案した自己複製モデルは、セルオートマトン平面上でのセルパターンの自己複製を実現するものである。テープ（状のパターン）に書かれたマシンの記述情報を読み取って別の場所にマシン（のパターン）を構成する機能をもつ「万能製作マシン」を構成した上で、テープにその万能製作マシンを記述し、さらにテープをコピーして出来上がった万能製作マシンに付加する機能を付け加えることで、自己複製マシンとなる、というのが基本的なアイデアである。

1958年になって、F. Crick は、DNA の情報が RNA に転写され、リボソームにおいて RNA を鋳型としてタンパク質へ翻訳されるという、いわゆるセントラルドグマを発表した。Von Neumann が、これにやや先行して、テープ=遺伝子型、万能製作マシン=表現型、万能製作マシン=リボソームという構造を思いついた[Mange 98]ことは、概念レベルのアプローチが従来の科学の方法よりも近道になりうる、あるいは先導的な役目を果たしうるということを示唆する好例と言えよう。

5. 心に関する人工生命モデル研究

5.1. 心へのアプローチ

4節まで、研究対象を広く「複雑系」として方法論を検討してきた。もし、研究対象を心に限定した場合、3節で論じた方法論に関して変更すべき点はあるだろうか？

まず、図2に示した基本的な科学の方法に関しては、同様のループが構成され[三輪 99]、同様の方法論を持ちうるということがわかる。つまり、実験によって人の行動を観察し仮説を立てる（実験心理学、実験志向の認知科学）。また、逆にその仮説から行動に関する現象やデータを作り出し、検証する（理論心理学、モデル志向認知科学）（図7）。

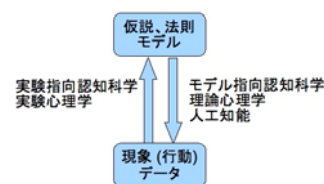


図7 心を対象とする場合の基本的な研究方法

もちろん、自然現象とまったく同様に心が扱えるわけではない。J. R. Anderson は、心を対象とするときの特殊性を、未知の乗り物を点検するために内蔵するエンジンを調べる例を用いて説明している[Anderson 80]。もし、エンジンを分解するのが容易ならば、それで問題はないが、もし、分解して調べるのが困難な場合（心の場合）、その乗り物を運転してエンジンの作用を理解するのが、心を対象とする科学（の一つの方法）に相当するとしている。基本的に、対象とするメカニズムを直接的に調べるのが困難ではあればあるほど、モデルの必要性は高くなるということが言えるだろう。様々な角度から心のモデルを作り、乗り心地を調べるために、計算機を動かし、仮説演繹を行っていく。

一方、人工生命研究をこの比喻で説明するのは難しい。なぜならば、研究目的が、そこに存在する乗り物自体を解明したいというよりは、なぜ、そこにエンジンがあるのかとか、この乗り物はもともといかなる存在かというようなものであるからである。心がそこにありきではなく、様々なレベルの環境や個体間の相互作用での心の創発や進化を、ヒトへの心の進化という歴史的事実に固執せずに追究することを目的とする。

基本的に、2節で論じた人工生命の研究方法（図3-6）に関する議論は、心を研究対象とする場合でも有効であろう。なお、心理学自体が複合的な領域になっていることからわかるように、心に関わる従来の学問領域に関わりなく、このようなアプローチが成立するわけではない。計算機を使ったモデル研究という点では、認知科学や人工知能と近い。また、個体間の心の相互作用からの創発という観点は社会心理学の重要テーマである。また、複雑系の変化を起こす適応プロセスとして進化は重要であり、その見方は、進化心理学と共有する部分が多い。

以下では、我々がやっている、人工生命モデルに基づく心に関連する3つの研究例を簡単に紹介する。

5.2. 例1：心の理論の再帰レベル進化モデル

「心の理論」とは、他者の心の状態（意図、知識、信念など）を推測する心の機能である[Premack 78]。心の理論を持つことが社会的生物にとって適応的であることを想像するのは容易である。資源の限られた社会環境では、駆け引きの中で他者よりも優位に立つことが重要であるからである。

ある個体が心の理論により他者も心の理論を持つと推測すると、心の推測の入れ子構造が発生する（この構造の深さを再帰レベルと呼ぶ）。チンパンジーでは再帰レベル2程度まで可能[松沢 91]と言われる一方で、ヒトはレベル5前後まで可能と言われており[Dunbar 00]、他種に比べて際立って高い。この再帰レベルは何を意味するのであろうか？ そもそも再帰レベルは高ければ高いほ

どうまく行動できるのであろうか？

この問題に答えるための構成的研究の試みとして、心の理論の再帰レベルの適応性と進化に関して検討を行った[高野 05][Takano 06]。設定としては、人混みの中で各個体がそれぞれスタート地点とゴール地点をランダムに与えられて移動するが、その際、各個体もっている再帰レベルに基づいて、他個体の動きを予測して衝突回避しながら動く。ただし、論じてきたように、実際の人混みの状況を再現することを目的とはしない。認知的コストも考慮せずに再帰レベルの意味を明らかにすることを目的とする。

個体同士が衝突すると時間的ペナルティが課される。逆に、他個体を気にしすぎて大きく迂回しても、時間のロスにつながる。再帰レベル0の個体は、相手をまったく気にせずに自分のゴールに向かって直進し、レベル1は、視界内の個体が皆レベル0だとみなして動きを予測して移動する。レベル2以降も同様に自分より一つレベルが低いと考えて移動する。要した時間をスタート・ゴール間距離で割ったものを得点とする。

シミュレーションの結果、再帰レベルの奇数偶数による行動や得点への大きな影響が示された。レベル1はレベル0が全く避けないと予測するので、極端に他者を避ける。レベル2はレベル1が自分を大きく避けてくれると予測するので、極めて小さな回避行動しかしない。レベルが高くなるにつれて、両者の差は小さくなる(図8)。

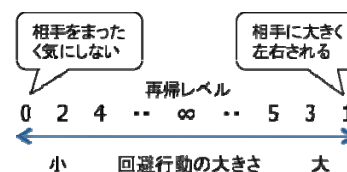


図8 再帰レベルと回避行動の大きさの関係

得点は、この直線上のある点をピークにしてそこから離れるほど小さくなる。そのピークのあるパラメータによって決まる。それは、視界の広さである。間近に迫ってから回避行動を始めると、大きな回避行動をとらないと間に合わない。逆に、遠くの個体も気にすると、無駄な迂回を招いてしまう。さらに、得点を適応度として、再帰レベルと視界の広さを同時に進化させる実験を行ったところ、再帰レベルが深ければ深いほど適応的である領域(図8の軸で ∞ で示される場所)に視野の広さが進化し、再帰レベルも深くなる一方であった。

視野の広さを表すパラメータは、個体の密集度を表すパラメータに置き換えても同様な結果が得られた。視野が狭い(密集度が大きい)と、この直線上で左よりの奇数レベルの推測が他者との衝突をうまく避け適応的であり、逆の場合は右よりの偶数レベルが適応的である。特に、密集度がこの軸の ∞ で示される場所にあるとき、奇数偶数に関係なくより高いレベルが高い適応度を得る。

ヒトの歴史において、個体間の相互作用の強さは徐々に強くなり、それが知能の進化を駆動してきたと考えられる。適応度がピークとなるレベルは、図8において、右から左へと移行したのかもしれない。つまり、ヒトと他種との差は、レベルが深くなればなるほど適応的であるような、この特定の領域に達したかどうかの差であった可能性がある。

5.3. 例2：誤認の適応性モデル

一般に、情報を正確に獲得することは適応的であろう。しかし、その情報が特定の行動を促進し、しかも、その行動の適応性が集団内でその行動が行われる頻度に依存する場合には、それが成り立たないときも少なからずある。例えば、「ある道路がガラガラであり、そこを通れば迅速に移動できる」という情報が集団内に広まると、その道路を通過しようという行動の頻度は増加し、結果的にその道路の渋滞を招くかもしれない。また、ある株価が安いという情報の伝播は、逆に株価の高騰を招き、後から購入した者が損をしがちである。これらの例は、情報が共有されることによって特定の行動に集中した結果、適応度が減少したと解釈できる。

このような状況、つまり、情報の共有が特定行動への集中を招き、結果として適応度を減少させる場合には、情報を正確に認識しないこと、すなわち、誤認が生じて情報の共有を妨げることで、行動の多様性は増加し、適応的となりうると考えられる。

また、情報の共有そのものが多様性の減少をもたらす状況において、コミュニケーションは情報共有の1つの手段であり、必然的に多様性を低下させる。したがって、コミュニケーションにおいても、このような多様性減少による効果が現れるはずである。つまり、コミュニケーションの際に誤認すると、認識に差異が生じ、その結果、行動の多様性が増加し適応的となると考えられる。なお、情報を環境から直接得る際に生ずる誤認を直接誤認、コミュニケーション時の誤認を間接誤認と呼ぶ。

さらに、情報の共有が行動の多様性の低下をもたらす効果は、その情報によって促進される行動がどれだけ特定の行動のみを促進するとき、その一点に行動が集中するため、多様性低下の効果が最大となる。ある特定の行動のみを禁止するような情報では、行動を促進する範囲がその一点以外すべてということとなり、多様性低下の効果は最小となる。結局、促進される行動の幅が誤認の効果の大小を決めると考えられる。

以上のような仮説を検討するために、資源探索問題を対象としたエージェントベースモデルを構築して、直接誤認と間接誤認の適応性を評価した [Akaishi 02] [赤石 04]。タスクはロボットの資源収集を想定しており、正方形のセルに区切られた2次元平面上をエージェントが資源を探索し収集するものである。

エージェントは自ら知覚した、もしくはコミュニケーションによって得た資源の位置と量に関する地図情報を記憶する。エージェントがコミュニケーション、もしくは自らの知覚により情報を得る際、誤認が発生する可能性がある。誤認が発生した場合、情報の受信者の認識は送信者と異なったものとなる。実験の結果、特定の行動を促進する情報に関する直接誤認は、個々の多様性を増加し適応的となりうることで、正確なコミュニケーションは集団行動の多様性を減少させ誤認の適応性を減少すること、それに対して間接誤認は集団の多様性を増加し、コミュニケーションを適応的としうることを確認された。さらに、誤認度を各個体に遺伝的に持たせ、進化実験によって、誤認を適応進化の産物とみなすことの妥当性を検討した[赤石 05]。

これらの結果は、次のようなテーマに関わる新たな知見を生み出す可能性がある。第一に、工学的な技術革新に対するアンチテーゼとしての可能性である。例えば、群ロボットシステムにおいて、ロボットのセンサー感度の限らない向上がシステム全体の性能向上に必ずしもつながらない状況がありうるのではないかと。第二に、人間の感覚器官の不完全性に関する機能主義的な説明である。感覚器官はしばしば誤るが、これは構造的物理的な限界によるものだけでなく、適応進化による必然という面からの説明も可能ではないかということである。

5.4. 例3：感情の起源モデル

感情の起源や進化に関する研究を最後に紹介する。ロボットを用いた感情の起源や進化的基盤の追究、及び、感情に基づくロボット行動の工学的応用の検討を目的とする(図9)。感情には、内部状態を表出して他者へ伝達するという機能もあるが、ここでは、感情が自己の行動を適応的に調整する機能に着目している。

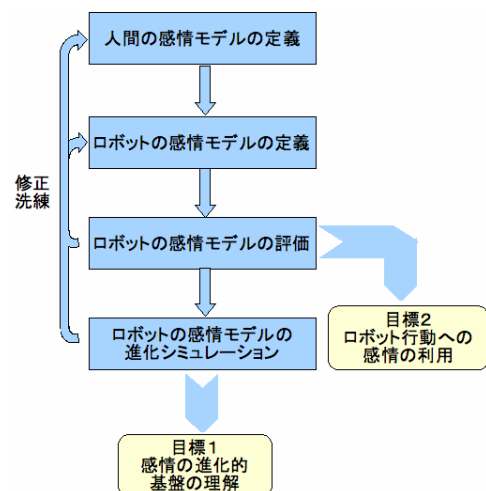


図9 ロボットを用いた感情の起源へのアプローチ

図9のフレームワークにおける最初のステップとし

て、D. Dörner らによる人間の感情に関する理論[Dörner 88] [Dörner 95]を採用した研究を開始した[Kato 04]. これは感情を行動のモジュレータとみなす理論である。たとえば、従来のテレビに明るさやコントラストなどの調整機能があるように、人間が行動するときの行動のしかたを調整するモジュレータを感情の機能とみなす。各感情は、それらの調整の度合いのセットで定義される。この理論のポイントは、感情は行動のしかた、振舞いかたを決めるのであって、行動の選択や内容には影響しないという点である。

この理論をロボット行動に適用したモデルでは、感情モジュールは、基本的には行動生成機構とは独立であり、アクチュエータのコントロールのしかたに影響を及ぼす。モジュレータとして、1) 単位行動の所要時間、2) 移動速度、3) 角を曲がる時の慎重さ、4) 基本行動の乱れの少なさ、の4つを定義し、これら4つのパラメータ値の典型的なセットとしての5つの感情(怒り、不安、満足、興奮、悲しみ)を定義した。

行動モジュレータのロボットの行動への割り付け方の妥当性を調べるため、実際のロボットが移動する際に、各感情に基づく振舞いを行わせた。1000人以上の被験者に対して、ロボットが個々の感情を伴って行動している動画像と感情名を正確に結び付けられるか評価実験を行った。実物ではなく計算機画面上の小さな動画であったにも関わらず、各行動に対して、設定した感情が伴っているとほぼ適切に判断された。特に、「満足」の識別率は高かった。一方、「悲しみ」と「心配」は混同されやすかったが、これらは近い感情であるので妥当であろう。

さらに、このような感情定義に関する進化的基盤を検討するために、各種の障害物が存在する場をロボットが通行するタスクを設定し、感情を埋め込んだ場合の移動の適応性を検討した[Kato 05]. その結果、センサー入力之急激に減少する暗闇に進入すると、「心配」の感情が起きて、注意深く行動するようになって不測の事故を避けたり、突起物によりダメージを受けるような場に入ると、「怒り」の感情が起きて、多少注意力が落ちても一気にその環境を脱するなどの行動が進化的に生じてくることを確認した。このような感情の適応性を応用したロボットシステムを検討中である。

6. おわりに

人工生命、あるいは複雑系科学において、モデルを作って研究する際、そのモデルが現実をいかに正確に記述しているか、あるいは、そのモデルが結局どういう意味を持つのか、ということは、その領域を科学として成立に関わる根本的な問いかけである。しかし、これに対する答えに関して明確なコンセンサスは存在しないと言える。本稿では、人工生命モデルに関する積極的な意味付

けがいかに可能であるか、心に対するアプローチを念頭において、その方法論を検討した。

従来の科学的手法では解き明かすのが容易でないチャレンジングなテーマの中には、本稿で論じたような方法に基づく研究が先導的な役割を果たしていきうものが少なくないと期待している。心に関わる現象はその有力な候補である。

◇ 参考文献 ◇

- [Akaishi 02] Akaishi, J. and Arita, T.: Misperception, Communication and Diversity, *Artificial Life VIII*, pp. 350-357 (2002).
- [赤石 04] 赤石 仁, 有田隆也: 誤認の適応性に関するエージェントベースモデル, *電子情報通信学会論文誌, Vol. J87-D-I, No. 4*, pp. 482-491 (2004).
- [赤石 05] 赤石 仁, 有田隆也: 誤認の進化に関するエージェントベースモデル, *電子情報通信学会論文誌, Vol. J88-D-I, No. 7*, pp. 1161-1164 (2005).
- [Anderson 80] Anderson, J. R.: *Cognitive Psychology and its Implications*, W. H. Freeman and Company (1980).
- [有田 02] 有田隆也: 人工生命 (改訂2版), 医学出版 (2002).
- [有田 07] 有田隆也: 心はプログラムできるか, ソフトバンククリエイティブ (2007).
- [Barandiaran 06] Barandiaran, X. & Moreno, A.: *ALife Models as Epistemic Artefacts*, *Artificial Life X*, pp. 513-519 (2006).
- [Beer 03] Beer, R.D. (2003): The Dynamics of Active Categorical Perception in an Evolved Model Agent, *Adaptive Behavior*, Vol. 11, No. 4, pp. 209-243 (2003).
- [Bedau 02] Bedau, M.: The Scientific and Philosophical Scope of Artificial Life, *Leonardo*, Vol. 35, No. 4, pp. 395-440 (2002).
- [Dennett 94] Dennett, D.: *Artificial Life as Philosophy*, *Artificial Life*, Vol. 1, No. 3, pp. 291-292 (1994).
- [Dörner 88] Dörner, D., Schaub, H., Stäudel, T. and Strohschneider, S.: Ein System zur Handlungsregulation oder: Die Interaktion von Emotion, Kognition und Motivation, *Sprache & Kognition*, Vol. 4, pp. 217-232 (1988).
- [Dörner 95] Dörner, D. and Hille, K.: Artificial Souls: Motivated Emotional Robots, *Proceedings of the International Conference on Systems, Man and Cybernetics*, pp. 3828-3832 (1995).
- [Dunbar 00] Dunbar, R. I. M.: On the Origin of the Human Mind, Carruthers, P. and Chamberlain A. (eds): *The Evolution of Mind*, Cambridge University Press (2000).
- [Emmeche 94] Emmeche, C.: *The Garden in the Machine*, Princeton University Press, 1994.
- [Horgan 95] Horgan, J.: From Complexity to Perplexity, *Scientific American*, Vol. 272, Issue 6, pp. 74-79 (1995).
- [市川 03] 市川 惇信: 人工科学と技術, 計測と制御, Vol. 42, No. 3, pp. 162-171 (2003).
- [Kato 04] Kato, T. and Arita, T.: A Robotic Approach to Emotion from a Selectionist Perspective, *Proceedings of the 9th International Symposium on Artificial Life and Robotics*, pp. 601-604 (2004).
- [Kato 05] Kato, T. and Arita, T.: Evolutionary Simulations based on a Robotic Approach to Emotion, *Proceedings of the 10th International Symposium on Artificial Life and Robotics*, pp. 258-261 (2005).
- [Langton 89] Langton, C. G.: *Artificial Life*, *Artificial Life*, pp. 1-47 (1989).

- [Lenski 03] Lenski, R. E., Ofria, C., Pennock, R. T. and Adami, C.: The Evolutionary Origin of Complex Features, *Nature*, Vol. 423, pp. 139-144 (2003).
- [Mange 98] Mange, D. and Sipper, M.: Von Neumann's Quintessential Message: Genotype + Ribotype = Phenotype, *Artificial Life*, Vol. 4, No. 3, pp. 225-227 (1998).
- [松沢 91] 松沢哲郎: チンパンジー・マインド: 心と認識の世界, 岩波書店 (1991).
- [三輪 99] 三輪和久: 科学的発見における有効な仮説検証方略: 計算機シミュレーションに基づく検討, 岡田 猛, 戸田山和久, 田村 均, 三輪和久 (編): 科学を考える: 人工知能からカルチュラル・スタディーズまでの 14 の視点, 北大路出版 (1999).
- [Moreno 02] Moreno, A.: *Artificial Life and Philosophy*, Leonardo, Vol. 35, No. 4, pp. 401-405 (2002).
- [Premack 78] Premack, D. and Woodruff, G.: Does the Chimpanzee Have a Theory of Mind?, *The Behavioral and Brain Sciences*, Vol. 4, pp. 515-526 (1978).
- [Ray 92] Ray, T. S.: An Approach to the Synthesis of Life, *Artificial Life II*, pp. 371-408 (1992).
- [Suzuki 04] Suzuki, R. and Arita, T.: Interactions between Learning and Evolution: The Outstanding Strategy Generated by the Baldwin Effect, *Biosystems*, Vol. 77, No. 1-3, pp. 57-71 (2004).
- [高野 05] 高野雅典, 加藤正浩, 有田隆也: 心の理論における再帰のレベルの進化に関する構成論的手法に基づく検討, *認知科学*, Vol. 12, No. 3, pp. 221-233 (2005).
- [Takano 06] Takano, M. and Arita, T.: Asymmetry between Even and Odd Levels of Recursion in a Theory of Mind, *ALIFE X*, pp. 405-411 (2006).
- [Wilke 01] Wilke, C. O., Wang, J., Ofria, C., Adami, C. and Lenski, R. E.: Evolution of Digital Organisms at High Mutation Rate Leads To Survival of the Flattest, *Nature*, Vol. 412, pp. 331-333 (2001).

2008 年 11 月 **日 受理

著者紹介



有田 隆也 (正会員)

1960 年生まれ。1983 年東京大学工学部計数工学科卒業。1988 年同大学大学院工学系研究科博士課程修了。工学博士。名古屋工業大学講師, カリフォルニア大学ロサンゼルス校客員研究員を経て, 現在, 名古屋大学大学院情報科学研究科教授 (情報文化学部兼務)。人工生命や情報科学の研究に従事。進化ダイナミクス, 言語の進化, 複雑適応系等に興味を持つ。著書に『人工生命』(医学出版, 2002 年), 『心はプログラムできるか』(ソフトバンククリエイティブ, 2007 年) 等。