2005 年 3 月 11 日(社)精密工学会画像応用技術専門委員会研究会にて講演

ロボットの心のつくりかた
自律分散的無意識と受動的意識から成るロボットの心の構成法

前野隆司（慶應義塾大学）

Mind of Robot Consists of Autonomous Distributed Unconsciousness and Passive Consciousness

Takashi Maeno (Keio University)

Abstract: A fundamental idea for constructing a conscious robot is presented. First, hypotheses of the human mind are presented. The following ideas are shown: 1) The unconscious system is a recurrent network system made of various distributed subsystems. 2) Information of the mind such as intellect, feelings and willpower, is presumably processed in the unconscious system instead of the conscious system. 3) The conscious system just monitors, experiences afterward, models and memorizes the results of the unconscious system. 4) Realistic experiences of quality that the conscious system feels by itself are just illusions that are defined in the brain. Then an algorithm of a robot mind is constructed based on the hypotheses mentioned above. It is shown that a conscious mind of robots can be made using the proposed algorithm. Finally, purposes and issues of the robots with a mind are also discussed.
Key Words: Consciousness, Mind, Autonomous Distributed System

１．はじめに

ロボットに心を持たせることは可能か，という問いは，ロボット学から哲学，SF に至る重要課題である．もちろん，アミューズメントロボットにとっても重要な課題である．心は「知」「情」「意」「記憶と学習」「意識」という 5 つの要素から成るといわれる[1]．このうち，最も未解明と考えられている要素が「意識」である．このため，筆者はこれまでに，ヒトの「意識」は受動的である，とする受動意識仮説を提唱[2]し，これをロボットに適用することの可能性について議論してきた[3][4]．本報では，(1) 自律分散的かつボトムアップ的な「無意識」下の情報処理により，トップダウン的であるように思える「意」のような心の活動を生成できること，および，(2)「意識」は無意識的情報処理の結果を受動的に受け入れるシステムに過ぎないと考えうることについて述べる．最後に，これら 2 つの考え方によればロボットの心をある程度容易に構築可能であることを述べる．

２．従来の「心」のモデル

従来，心を構成する五つの要素，「知」「情」「意」「記憶と学習」「意識」のうち，「意識」は，「知」「情」「意」や「記憶」に対し「注意」を向けるトップダウンの存在であると考えられてきた．従来の心のイメージ図を図 1 に示す．図 1 に示した「意識」は，「注意」を向けた対象に応じてその有効範囲を破線の範囲内で変幻自在に拡大・縮小することのできる存在である．このような「意識」は，脳内の情報処理のうち，意識にのぼる可能性のあるあらゆる情報処理に対し，トップダウンに注意を向け理解することのできる存在でなければならない．つまり，「意識」はあらゆる情報をバインディングできる（結び付けられる）万能な存在でなければならないことになる．このような「バインディング問題」を解くことのできる「意識」が，どのようにヒトの心に宿ったのか，また，どうすれば人工の「意識」を設計できるか，という点は，何千年もの間，謎だと言われ続けてきた．

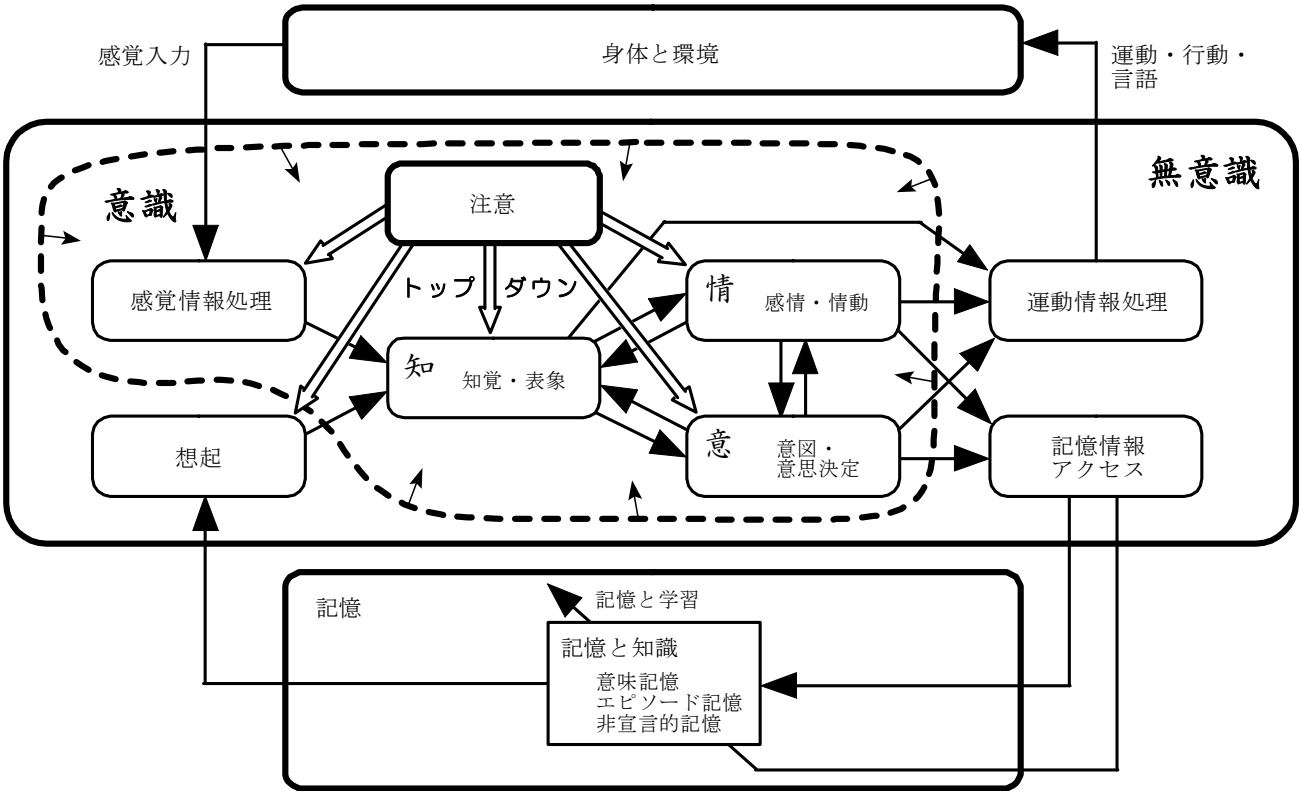

図 1　「無意識」との境界が不明確でトップダウンに「注意」を
払った部分に注目する「意識」を仮定する従来の心のモデル

これに対し，本研究では，(1) 自律分散的かつボトムアップ的な「無意識」と (2) 無意識的情報処理の結果を受動的に受け入れる「意識」という 2 つのシステムを仮定すれば，「意識」の問題が解決可能であると考える．本研究の考え方を以下に述べる．

3．自律分散的「無意識」

Brooks のサブサンプションアーキテクチャ[5]は，従来のトップダウン的制御系と異なり，反射的かつ自律分散的な制御系を組み合わせることにより，昆虫のような下等な生物の行動が生成可能であることを示したため，発表当時は大きな脚光を浴びた．しかし，そのような制御系はヒトの知的制御のような複雑な制御には適用できない，という批判を浴び，現在ではあまり注目されていない．まさに，意識にのぼりうる情報をバインディングするトップダウンの「意識」のような情報処理を説明できないことが批判のひとつの論拠であった．

ただし，ヒトの脳内の無意識下の処理が自律分散的なエージェントによる処理であると捉えうることは広く知られている[6]．言い換えれば，ヒトの無意識下の情報処理は，サブサンプションアーキテクチャのような自律分散的な処理の結合と捉えうるといえる．

このため，図 1 に示した従来の心のモデルにおいても「感覚情報処理」「想起」「知」「情」「意」等のうちの一部は自律分散的な情報処理であると考えられてきた．

たとえば「知」は感覚の知覚や記憶の表象を行う知的な情報処理であり，ヒトはあたかも自

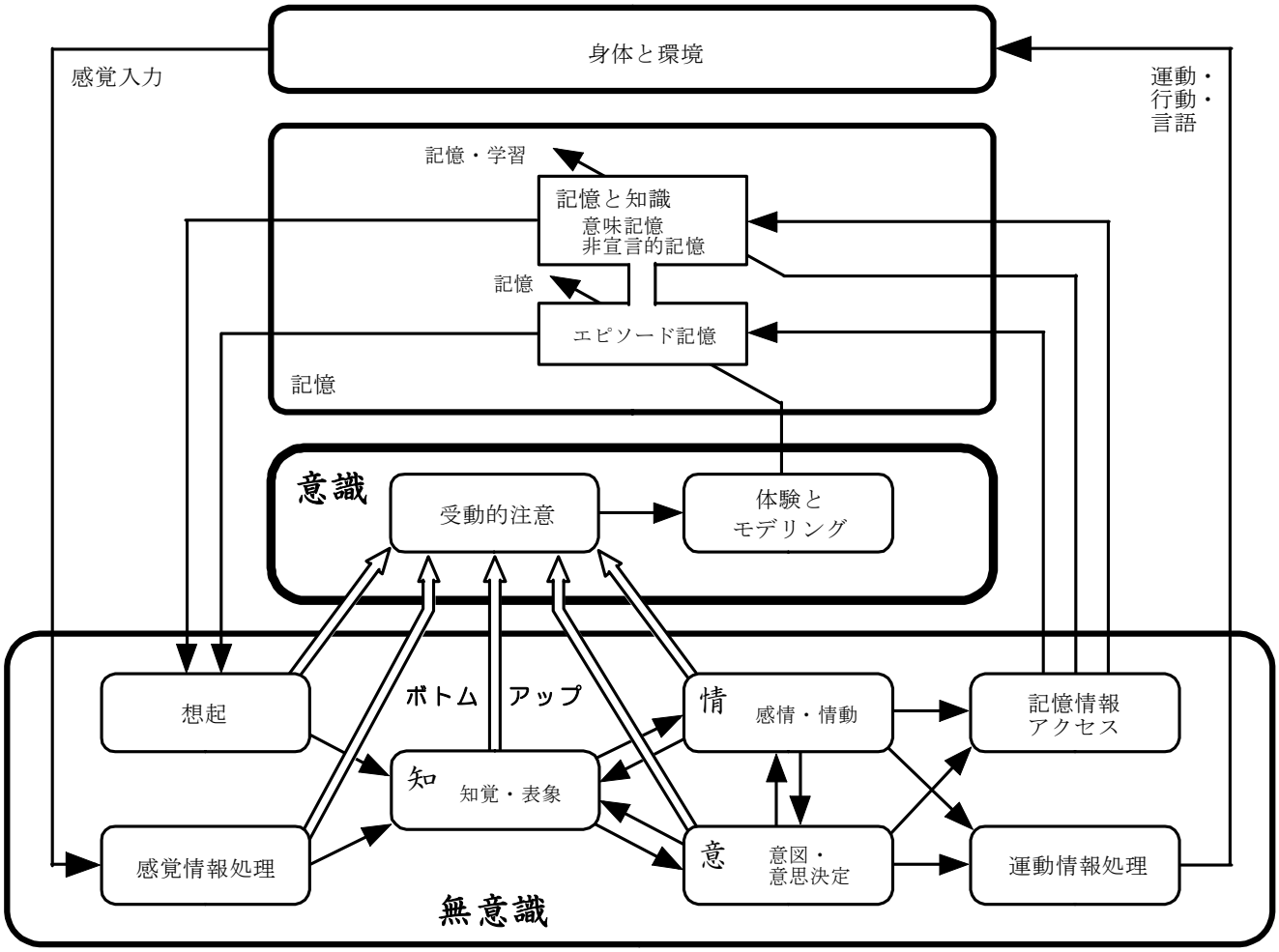

図 2　「意識」の外にある「無意識」の自律分散計算結果にボトムアップに「注意」を払う受動的な「意識」を仮定する心のモデル

動的に物事が思い浮かぶがままに,「知」の情報処理を行っているといえる. ヒトは決して,「赤いリンゴがここにある」というような情報処理をトップダウンには行わない.「知」の情報処理は, 昆虫のそれよりもはるかに複雑であるとはいえ, 自律分散的な処理により行いうる.

「情」も同様である. 感情や情動は, 基本的に, ヒトが自らの「意思」によりコントロールするものではなく, 身体と環境の状態に応じて湧き上がってくる自律分散的な作用であると考えられる.

一方,「意」は, 意図や意思であり, 一般には, 自律分散的かつボトムアップ的な情報処理結果ではなく,「意識」が行うトップダウンの情報処理であると考えられている.

しかし, たとえば, これから「仕事をする」か「遊ぶ」かという意思決定を行う際に, 最後に「意識」下の「意」がトップダウンに判断を下した, と考えるのではなく, サブサンプションアーキテクチャのように, 意思決定のための様々な行動モジュールの反射結合の結果としてたまたま「仕事をする」に関連するモジュールの活動が活発になり, それが「注目」され「意識」されたものと考えても特に問題はない. つ

まり，ヒトの意思決定は，昆虫が様々な反射の重ね合わせの結果として右に行くか左に行くかを意思決定するメカニズムと本質的に大差ないと考えうるのである．トップダウンに選択されたのではなく，いずれかの情報処理結果がボトムアップに生き残った様子がトップダウンの選択のように見えているだけであると考えうるのである．

４．受動的「意識」

従来「意識」の範疇で行われる情報処理であると思われてきた「知」「情」「意」は，無意識的な情報処理とも考えうることを述べてきた．しかし，そのように考えると，ヒトにとって能動的かつトップダウン的であるように感じられる「意識」と矛盾するように思える．このため，「無意識」の問題は「意識」の問題とセットで考えなければならない．

筆者の心のモデルでは，図 2 に示したように，「意識」は「無意識」下の自律分散的・ボトムアップ的・無目的的情報処理結果を受け取り，それをあたかも自分が行ったことであるかのように錯覚し，単一の自己の経験として体験した後にエピソード記憶するための受動的・追従的なシステムであると考える．

すなわち，「意識」は「無意識」の内部で変幻自在に領域のサイズを変更できるシステム（図 1）ではなく，「無意識」とは独立なシステム（図 2）であると考える．また，「注意」は「意識」が無意識下の情報処理結果にサーチライトを当てるトップダウン処理の中心的役割を担っているように実感されるものの，そうではなく，自律分散的情報処理結果としてたまたま選択されたものについての情報を受動的に受け取る存在であると考える．

なお，「意識」が進化という淘汰圧のもとで獲得された理由は，自己の唯一の経験をエピソードとして記憶するためと考えられる．なぜなら，エピソード記憶ができることは，その種の生存のために有利だからである．逆に言えば，エピソード記憶をしない動物は，「意識」が存在する必要がないので，「意識」を持たないものと考えられる．

つまり，ヒトが「仕事をする」か「遊ぶ」かを無意識的に決定するメカニズムは，昆虫が右へ行くか左へ行くかを決定する反射的メカニズムと同様な自律分散的情報処理であり，唯一の違いは，ヒトの方には，自分が選択した結果を体験しエピソード記憶するための「意識」というシステムが存在する，という一点のみであるといえる．

言い換えれば，筆者が提唱する受動的「意識」とは，心の因果関係についての見方を反転させることであると言える．すなわち，「意識」が行う主体的な意思決定の結果として，行動や脳内イメージが開始されるのではなく，自律分散的な情報処理の結果としてボトムアップ的に開始された無意識的な行動や脳内イメージを，あたかも自らが主体的に開始した行為であるかのように錯覚するシステムが「意識」であると考えるのである．

このように「心」の中の原因と結果を反転させると，従来「心」の謎であるといわれてきたバインディング問題を解決する方法がわかる．すなわち，トップダウン的な「意識」が無意識下のあらゆる情報処理結果のうちの一部に能動的に「注意」を向ける（バインディングを行う）のではなく，ボトムアップ的な情報処理結果に受動的に「注意」を向ける「意識」はもともとバインディングを行っていないと考えればよいのである．

また，以上の受動意識仮説によれば，他の「心」の謎，すなわち，フレーム問題や，意識の自己言及性の問題，「なぜなんのために哺乳類の意識は生じたのか」という疑問，独我論が問題にしてきた〈私〉（自己意識のクオリア）の問題など，あらゆる「心」の謎に対する回答を用意することができる[2]．また，ロボットの脳であるコンピュータに意識を持った生命的・無目的的な「心」を作り出すことも困難ではないと考え

られる[3].

5. 自律分散的「無意識」と受動的「意識」を持つロボット

自律分散的「無意識」と受動的「意識」を仮定すれば，心を持ったロボットを作ることは容易であることを以下に述べる.

自律分散的「無意識」と受動的「意識」を持つロボットとは，たとえていうならば，要するに，アイボ[7]のようなペットロボットに意識とエピソード記憶を追加するようなものであるといえる.

アイボのようなペットロボットは，プログラムされた結果とはいえ，「知情意」のようなものを示す.すなわち,飼い主の顔を認識する「知」，感情のようなものを提示する「情」，タスクレベルの目的は与えられていないにもかかわらず何らかの行動を開始する「意」の機能を，単純とはいえ持っている．そして，外部にいる観察者にとっては，並列処理のように感じられるそれらの処理から,いずれかの行動を選択している．このような行動生成法はまさに自律分散的「無意識」と等価であるといえる.

ただし，既存のペットロボットは「意識」や「エピソード記憶」の機能は持たない．このため，ペットロボットに「意識」と「エピソード記憶」の機能を付加することを考えてみる.

筆者の定義によれば，「意識」は自分が無意識に行った様々な行動のうちのひとつに「注意」を払い，それをあたかも自分の意図により行ったものであるかのように感じるシステムである.このため,まず,様々な自律分散的行動のうち，その時点でのそのロボットの行動を象徴するような行動を選択すればよい．たとえば，筆者が提案した[3]ように，無意識レベルの自律分散的演算に「重要度」という情報を付加しておき，この大きさを比較するという方法が考えられる.また，たとえば，ニューラルネットワークによって自律分散的無意識を実現している場合には，ニューラルネットワークの発火頻度をモニタし，発火頻度の大きい情報処理に着目するようにしてもよい.いずれにせよ,何らかの方法により，自分の行動のうちのひとつに着目するようなシステムを作成すればよいのである.

つぎに，注目した結果を，自らのみずみずしい実体験としてロボットが意識下で体験する必要がある．このようなみずみずしい体験の質感をクオリアという[2][8].ヒトがなぜどのようにクオリアを感じているのか，という謎の解明は容易ではない．同様に，クオリアをロボットに感じさせるような手法を確立するのは容易ではないと考えられる．しかし，ロボットに，「みずみずしいクオリアをまさに今感じています」と「言わせる」ことは難しくない．なぜなら，実際にクオリアを感じていなくても，「いえ,まさに，感じています．うそじゃないですよ.」と言わせるだけであれば，既存の文脈生成法により十分実現可能と考えられるからである.

したがって，ペットロボットに，自らのある行動に着目させ,その結果を,「私はまさに○○をしています.」「私はしみじみと嬉しさをかみ締めています.」「自分で考えて○○という意思決定をしました.」といわせることのできるシステムを作れば，それはまさに「意識」の機能を持ったシステムであるといえる.

それは,「意識」ではなく,「意識」を模擬したシステムに過ぎないではないか，という批判が考えられる．確かに，そのような言い方はできる．哲学者のいう「ゾンビ」という存在がこれに近いと考えられる.しかし,これまでは「意識」を模擬したシステムの作り方すら不明であったため，そのようなものさえ実現されていなかったことを考えれば,進歩であるといえよう.また，このロボットをヒトに見せたとき，たとえばチューリングテストにかけたとき，このロボットが「心」を持っているのか,「心を模擬したシステム」を持っているのかを観察者が区別することは不可能である．このため，このロボットを外から観察すると，心を持っているとしか思えないという結果となる．心というには幼

稚すぎる，という批判はあるかもしれないが，それは単純さのレベルの問題であって，心であるか否かの議論ではない．

また，この主張は，一見，システムを複雑にしていけば意識が自然に形成されると考える創発論の立場と類似しているように思えるかも知れない．しかし，両者は明確に異なる．創発論は,意識のようなものが自己組織的に生成され，これをチューリングテストにかければヒトの意識と区別が付かなくなるのではないか，と推測しているに過ぎない．これに対し，筆者のシステムは,「意識」というシステムを明示的に作成しているため，外から見たときに意識の機能を必要十分に有する．

また，ロボットは,「意識」した結果をエピソードとして記憶するものとする．もちろん，ロボットの場合には，エピソード記憶をしないけれども「意識」を持つ，ということはありえるが，その場合の「意識」はヒトの意識とは機能が異なることとなる．

一方，ロボットが生き生きとした意識体験をエピソード記憶できることは，記憶した内容を後に読み出してきて過去の経験として思い出せることを意味する．これは，その結果として，過去の経験を参照した情報処理が可能であるのみならず，経験に基づいて情報処理アルゴリズム（たとえば思考法）を変更することも可能であるということである．言い換えれば，情報処理のしかたと結果が常に同じではないということであり，ヒトの場合と同様，同じことは二度と繰り返さない個性的なロボットたりうることを意味する．このこと，つまり，情報処理の手法と結果が同時に時間発展し，二度と同じ行動を繰り返さないような非平衡システムであるということが，心，中でも「意識」を持つロボットの最大の特徴である．

このようなロボットは，トップダウンに行動の目的を与えられたロボットと異なり，ヒトのように生物的に振舞い自ら目的を生成することができるので，様々な場でのヒトの代替者としての活躍が期待されると考えられる．

6. おわりに

本研究では，(1) 自律分散的かつボトムアップ的な「無意識」下の情報処理により，トップダウン的であるように思える「意」のような心の活動を生成できること，および，(2)「意識」は無意識的情報処理の結果を受動的に受け入れるシステムに過ぎないと考えうることについて述べた．最後に，これら 2 つの考え方によればロボットの心を容易に構築可能であることを述べた．心を持ったロボットを実現することは今後の課題である．

謝辞　本研究の一部は，21 世紀 COE プログラム「知能化から生命化へのシステムデザイン」の援助により行われた．記して謝意を表す．

参考文献

[1] 松本元, 脳・心・コンピュータ, 丸善, 1996.
[2] 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, 2004.
[3] 前野隆司，ロボットの心の作り方 ―受動意識仮説に基づく基本概念の提案―，日本ロボット学会誌 23 巻 1 号，pp.51-62, 2005.
[4] 前野隆司，生命模倣ロボティクス ―生命のボトムアップ的設計原理に学ぶ, 平成 16 年度日本デザイン学会秋季企画大会講演論文集，pp. 21-27, 2004.
[5] Rodney A. Brooks, A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation.*, vol. 2, no. 1, pp. 14-23. 1986.
[6] Marvin Minsky, The Society of Mind, Simon & Schuster, Inc., 1885.（和訳：マーヴィン・ミンスキー，心の社会，産業図書，1990.）
[7] http://www.jp.aibo.com/
[8] 茂木健一郎，心を生み出す脳のシステム，NHK Books，2001.