学術・技術論文

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―

前　野　隆　司*

How to Make a Conscious Robot
—Fundamental Idea based on Passive Consciousness Model—

Takashi Maeno*

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: Mind, Conscious Robot, Consciousness and Unconsciousness, Cognition Model

1. は　じ　め　に

ロボットまたはコンピュータに心を持たせることは可能か，という命題は，近年，ロボット工学，認知科学，情報科学など，様々な分野で盛んに議論されている．一方，ヒトの心の構成原理を明らかにできるか，という議論も，脳科学，認知科学および哲学における主要な課題の一つである．一般に，心とは，「知」「情」「意」「記憶と学習」「意識」†という五つの働きを総合した情報処理機能であるといわれる [1]．人工知能や認知科学の分野では，いわゆるコネクショニストまたは「強い AI 論者」といわれる人々が，ロボットまたはコンピュータの神経回路を適切に接続することにより心を作れると主張している [2]．ヒト脳の神経系は，任意の非線形関数を写像可能な神経回路が直並列に接続されたものにほかならず，人工の神経回路を適切に接続すれば同様な心を作れることは自明であるという主張である．しかし，これまでに提案されたヒトの心の認知モデル [3] の多くは，「意識」を含んだ心の状態を明確に表現しているとはいえない．その理由の一つは，ヒトの心における「意識」の座の問題（「意識」の座がどこにあるのか，あるいは局在せず分布しているのか），バインディング問題（多数の自律分散計算を「意識」がどのように結びつけるのか），フレーム問題（心は枠外の未知の問題を解けるのにコンピュータはなぜ解けないのか），自己言及性の問題（心は自分の心についても考えられるという再帰的な機能はどのように説明され，どうすればコンピュータで実現できるのか）など，「意識」にかかわる問題が未解決であると考えられていることによる．これに対し，哲学者の Dennett [4] やコネクショニストの Churchland [2] は，「意識」に着目したヒトの心の認知モデルを提案している．すなわち，Dennett は，「意識」は無意識下の多元的草稿を束ねる仮想直列機械であると仮定し，Churchland は，脳の各部に接続した再帰神経回路であると仮定している．しかし，いずれも概念的な提案に留まっており，彼らのモデルを利用して直ちに人工の心を作り出せるものではない．ただし，ヒトの心の解明と，人工の心の構築は難易度が異なる．2 足歩行ロボットの制御則はヒトのそれとは異なっているものの，2 足歩行という機能においてはヒトと同じであるという事象と同様に，構成原理は異なっていても，機能はヒトの心と同じであるようなロボットの心を作ることは可能で

原稿受付　2003 年 2 月 4 日
*慶應義塾大学理工学部機械工学科
*Faculty of Science and Engineering, Keio University

†「知」（intellect）は環境の認識や知識の表象を行う機能，「情」（feelings）は感情や情動を発現する機能，「意」（willpower）は意図や意思決定を行う機能と定義する．また，「意識」は覚醒（arousal，起きている）という意味ではなく，モノやコトに注意を向ける働き（awareness）と，自分は私であることを認識できる再帰的な機能（self consciousness）の総体 [5] を指すものとする．一般に「意識」は「知」「情」「意」「記憶と学習」全体を主体的に統合するメカニズムであると考えられている [1]（本研究ではそうではないと考える）．「記憶」（memory）には宣言的記憶（declarative memory）と非宣言的記憶（nondeclarative memory）があり，宣言的記憶にはエピソード記憶（episodic memory，いつ何をしたかの記憶）と意味記憶（semantic memory，モノやコトの意味の記憶）がある．非宣言的記憶は，運動学習のように記号（言語）で表せないような記憶である．

ある．すなわち，矛盾なく上記の「意識」にかかわる問題を解決する心構築法の仮説を提唱できれば，その方法がヒトの心の構成法として証明されていなくても，それに基づいて既存のロボット・コンピュータに人工の心を持たせることは可能である．

本研究では，まず，これまでの脳科学や認知科学の知見に基づくヒトの心の認知モデルについて述べる．次に，受動意識仮説に基づく心のモデルを提案する．また，仮説に基づく具体的なロボットの心のアルゴリズムを示す．さらに，本モデルの実現可能性と課題を思考実験により示す．最後に，心を持ったロボットを実現することの意義と，心を持ったロボットが実現された際の社会的問題について考察する．

2. ヒトの心に関する仮説

2.1 従来の心のモデル

近年の脳科学の進展に伴い，脳の局在機能は明らかにされつつある．すなわち，大脳新皮質の感覚野・運動野が感覚と運動の処理を，脳の他の部位と神経結合した連合野が司令部として思考の処理を，大脳基底核が競合する大脳の指令信号の選択を，大脳辺縁系が動機付けや情動を，それぞれ担うといった具合である [1]．それぞれの部位が情報を自律分散処理している [2] [6] と考えれば，心の五つの働きのうち，初めの四つ，すなわち，「知」「情」「意」「記憶と学習」がどこで行われているかは大雑把にいえば明らかにされているといえる．また，これら四つの機能を生成する脳の神経回路構造とアルゴリズムについては未知の部分が少なくないものの，これらの機能自体は認知科学の進展とともに明確化されつつある．しかし，これら自律分散処理をバインディングする「意識」の座がどこにあるのか，あるいは「意識」は局在せず脳全体のコヒーレントな場として存在しているのか，といった点は未解決であるというのが現代脳科学の合意事項である．

「意識」の座が局在するという考え方は，古く Descartes の松果体にさかのぼる．また，象徴的な表現として，いわゆるサーチライトモデル [5] あるいはホムンクルス仮説 [7] がある．すなわち，「意識」を統合する一つのシステムが多くの「無意識」†システムの一つにサーチライトを当てたときにのみその「無意識」システムが「意識」に登る，あるいはホムンクルスという小人が「無意識」システムの間を飛び回り，必要な「無意識」システムにスポットライトを当てる，という比喩である．同様な考え方は近年も提唱されており，Churchland [2] は髄板内核に，Damasio [8] は大脳皮質右頭頂葉に，それぞれ「意識」の座が存在すると推定している．

一方，「意識」の座は存在せず，脳全体のコヒーレントな場が「意識」そのものであるという考え方も盛んに主張されている [9]～[12]．例えば，Crick と Koch [9] は，脳の神経細胞群における周波数 40 [Hz] の同期振動現象が「意識」の基盤であると考えている．

「意識」が局在するか，分布するか，という論点は大きな分岐点であるように思えるものの，いずれも，人は「注意（attention）」を払ったものに「意識」を向ける，という性質を当然視しているという点においては類似している．**Fig. 1** に，心の構造についての従来の考え方を包含する模式図を示す．中央に描かれた「無意識」下の情報処理の中に，破線で示された「意識」され得る部分が存在している．「意識」に登る領域の境界は破線の範囲内で変化可能である．例えば，「知」（外界の認識や記憶の想起）に「注意」が向けられているときには「知」が，悲しいときには「情」が，自分が何かしようと意図するときには「意」が，それぞれ意識される．

「情」および「意」の結果は，「運動情報処理（motion data processing）」された後に運動・行動・言語等の形で外界に対し出力され，外界からのセンサ入力は「感覚情報処理（sensory data processing）」部で処理される．同様に，「情」および「意」の結果は，「記憶情報アクセス（memory data access）」部を経てエピソード記憶されたり意味記憶の学習に用いられ，記憶は「想起（recollect）」部により思い出される．

Fig. 1 では，便宜上，心の機能の一部を「知」「情」「意」「記憶と学習」のように分類しているが，これらは他の形で分類されていてもかまわない．Fig. 1 で筆者が主張したい点は，「意識」が局在するか分布するかにかかわらず，一般に，『脳内の自律分散処理のうち，人が能動的に「注意」を払った部分が「意識」される』と考えられてきたという点である．また，一般に，「意」は「意識」に従属すると考えられる．すなわち，自分が何を意図しどのように意思決定するかには，常に「注意」が向けられ「意識」されると考えられる．このため，「意識」と「意」を同一視する研究者が少なくない．ただし，Dennett [4] の主張のように，多元的草稿（無意識）の中に多くの「意」の候補が分散的に励起されており，そのうち「注意」が向けられたもののみが「意識」に登る，という考え方もある．いずれにせよ，「意識」は「注意」を向ける主体であると考える点ではいずれの考え方も一致している．

Fig. 1 の破線部分のサイズを自在に変化させる「意識」を仮定する上記のモデルは，一般の心のイメージにはよく合致している．しかし，脳内の「知」「情」「意」の膨大な自律分散演算に「注意」というサーチライトを当てるためには，「意識」は膨大な演算を理解し観測できる必要があり，そのために「意識」はやはり膨大なシステムにならざるを得ないため，局在するとは考えにくい．一方，「意識」とは脳全体のコヒーレントな場，例えば，神経回路の連成振動（共鳴現象）[9] [12] であると考えるとすると，では，その振動の「観測者」はどこにいるのか [11] という疑問が生じ，結局，ホムンクルス探しの無限縮退 [7] に陥ってしまう．このため，「意識」は物理現象に還元することの不可能な複雑システムであると考えざるを得ない [13]，あるいは，「意識」とは現状では解明されていない「創発」現象なのだ [14]，といった回答の先送りに陥る．結局，未解明の疑問点が残ることになるのである．

2.2 受動意識仮説に基づく心のモデル

苧阪 [5] は，環境への適応の最適化をはかるために，ヒトは脳の中に「意識」という認識と行動の「シミュレータ装置」をもっているのではないかと述べている．川人 [15] は，「脳の計算理論」の最終章で，「意識」とは「無意識」下に生じる脳内の

† 「無意識」（uncousciousness）は，精神分析学でいうところの生活不適応的な意味ではなく，「意識」に登らない多様な脳内処理全般を指すものとする．

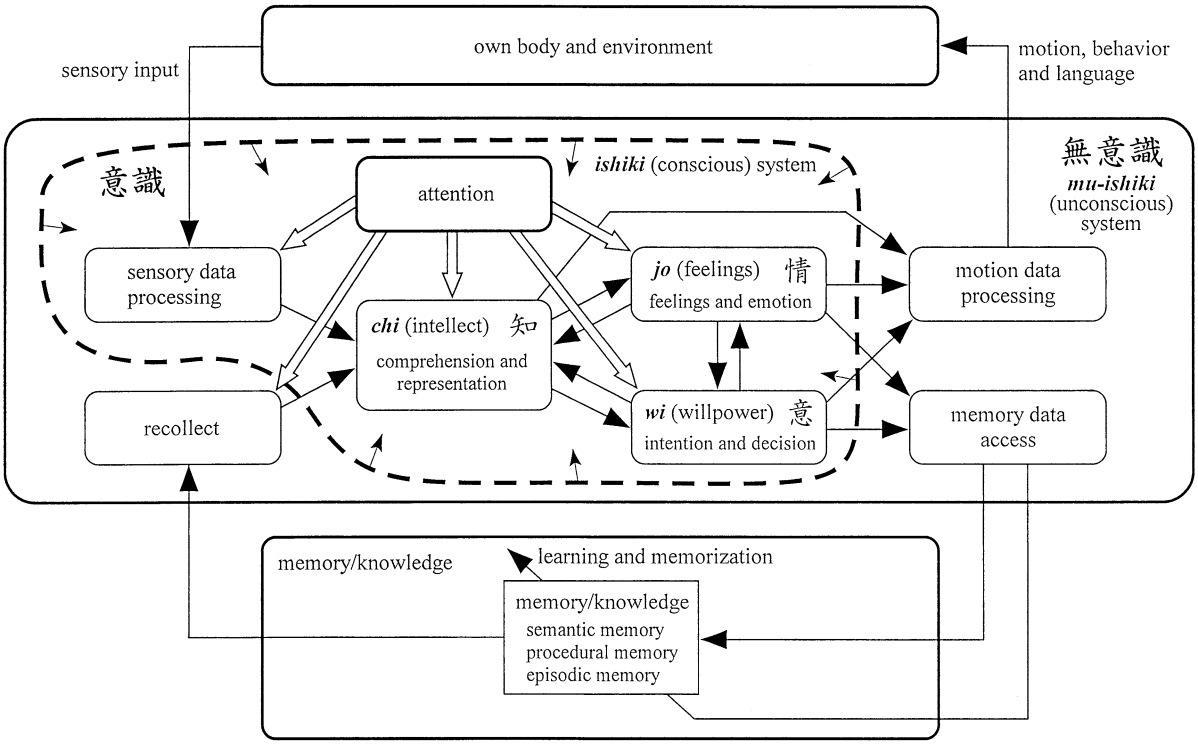

Fig. 1 One of the formarly believed *ishiki* (conscious) and *mu-ishiki* (unconscious) systems

膨大かつ並列な計算を非常に単純化された「うその」直列演算で近似する「モデル」であり，それぞれのモジュールの結合を加速するために存在する機能なのではないかと述べている．いずれも，「意識」は「無意識」の単純化されたモデルに過ぎないという考え方である．これらの考え方を拡張すると，以下のような認知モデルを考えることができる．すなわち，『意識システムは，無意識システムの膨大な自律分散的処理の一部を，あたかも自分が行っていることであるかのように錯覚しながら，単純化し追体験している受動的なシステムに過ぎない』という仮説に基づく認知モデル [16] である．

Fig. 2 に，提案する心のモデルを示す．本モデルでは，「想起」「知」「情」「意」は「無意識」に従属する機能であると考える．すなわち，「無意識」システムは，「意識」に登らない多様な処理を自律分散的かつ再帰的に行うシステム（**Fig. 3** に概念図を示す）であると考える．この考え方は，概念的に Dennett [4] による脳内の多元的な草稿という考え方と類似している．また，Minsky [6] のいう脳内のマルチエージェントと同様な機能であるといってもよい．ただし，後で述べるように，多元的草稿あるいはマルチエージェントを束ねる「意識」の主体性についての考え方は，Dennett や Minsky とは異なる．また，本「無意識」は，Brooks [17] のサブサンプションアーキテクチャの考え方を，反射による行動生成から，脳内における連想や推論といった認知情報処理全般のパターン生成に拡張したものということもできる．

以上に例示したように，本研究では，「無意識」は自律分散的再帰計算を延々と続けるシステムであると考える．また，「無意識」へのトップダウン指令は存在しない，自動的かつ無目的的なシステムであると考える．無目的であるにもかかわらず，何らかの秩序がボトムアップ的に自己組織化される理由は，連想神経回路 [18] のように，情報処理の再帰的連鎖が埋め込まれているためである．すなわち，「無意識」とは連想神経回路が階層的・重層的に結合されたようなシステムであると考える．このような分散的「無意識」の考え方は別段特異なものではなく，ヒトの日常体験と一致している．例えば，定型的な運動の制御は一般に「無意識」下で反射経路や内部モデル（非宣言的記憶）を用いて行われる．また，島田紳助 [19] は，優れた芸人は瞬時におもしろいネタを三つくらい思いつき，その中から最良のものを選ぶという．立花隆 [20] は，本を読むときに流し読みをしていれば，重要な点に自然に注目できるという．「無意識」が「意識」に登らない膨大な処理をこなしていると考えることは，以上の例からも自然と考えられる．

また，Fig. 2 のモデルでは，Fig. 1 とは異なり，「意識」は「無意識」とは独立な，境界が明確なシステムであると考える．また，「注意」は「意識」が能動的に発するものではなく，「無意識」下の自律分散的処理のうち，発火頻度の高いものを受動的に選択する機能であると考える．「意識」は認知の “原因” ではなく “結果” に過ぎないと考えるのである．言い換えれば，「意識」とはワーキングメモリの特殊な状態の一つであり，「無意識」下の処理を必要最小限に単純モデル化し，エピソード記憶として保存するために存在していると考えるのである．そうであるならば，多様な「無意識」システムからの多数の入力を単純化しシミュレーションするモデルであるヒトの「意識」は，自律

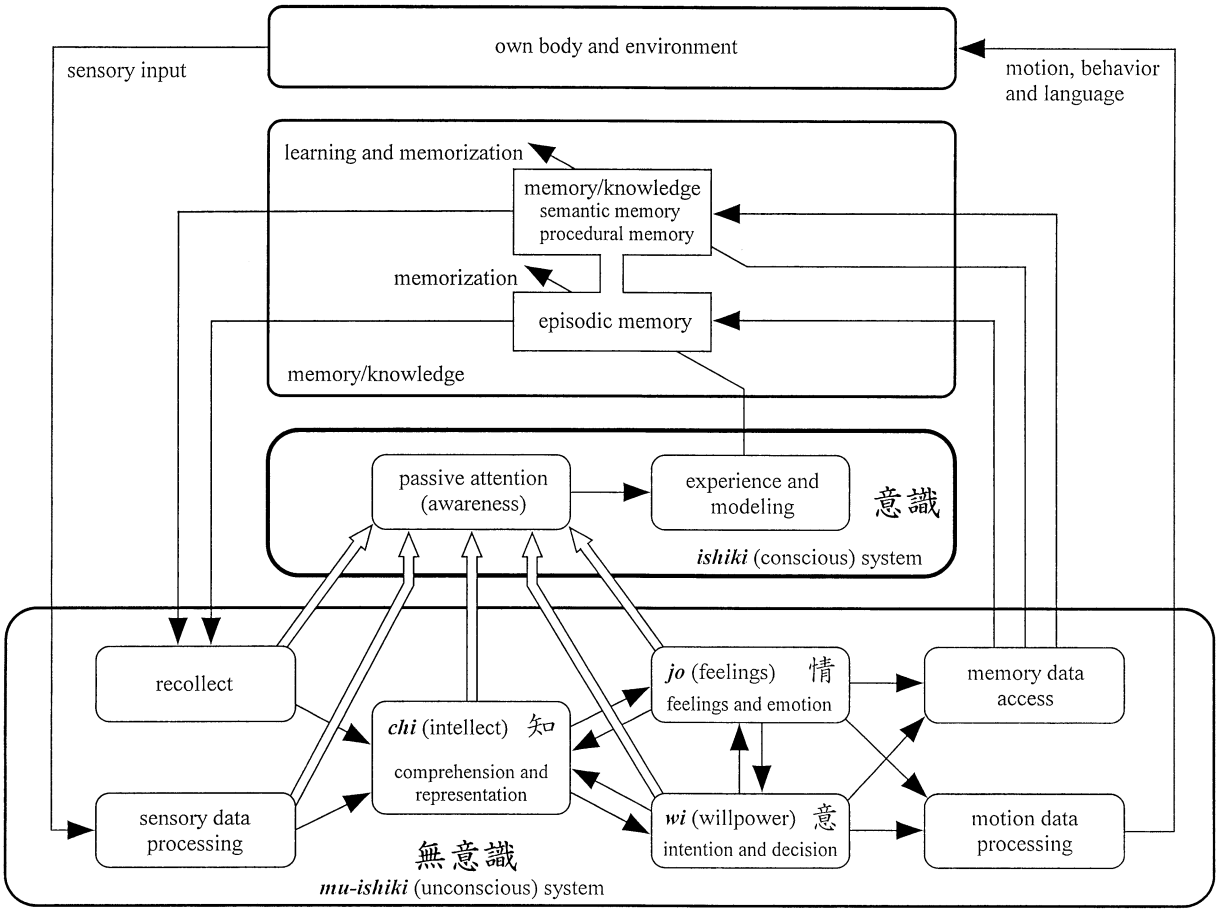

Fig. 2 Proposed *ishiki* (conscious) and *mu-ishiki* (unconscious) systems

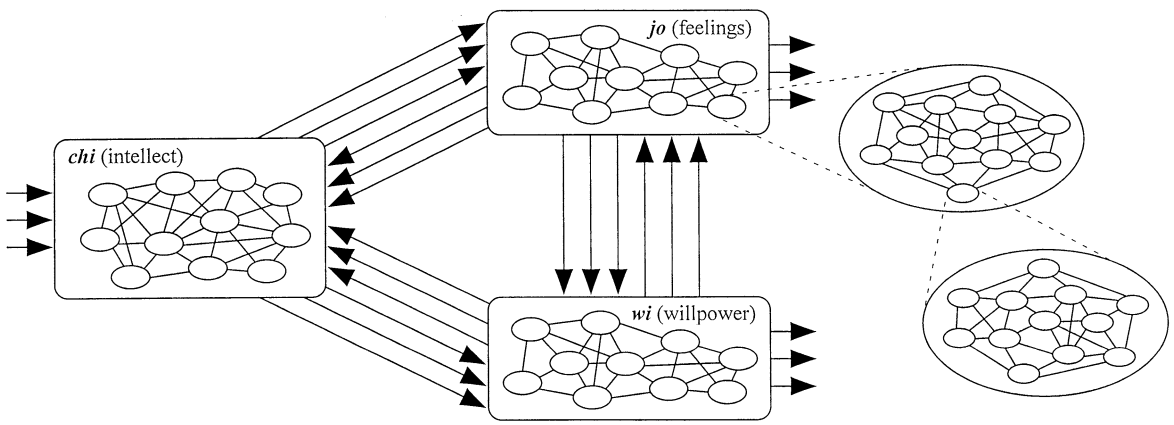

Fig. 3 Schematic view of subsystems of intellect, feelings and willpower system

分散処理を行う大脳内システム群よりも単純な情報処理により近似を行えればよく，その神経回路は大脳全体に比べ極めて小さくてよい．

「意識」が「無意識」システムを単純モデル化し保存するシステムであると考えれば，心に関する前述の疑問点に対する解答が用意できる．すなわち，「意識」が「無意識」よりも単純なモデルであるならば，思考の無限縮退 [7] には陥らない．なぜなら，分散しない一つの小さな「意識」は単純な意識下の処理を行うだけだからである．情報を受け取り「注意」する「意識」の座の境界は明確なので，コヒーレントな場の「観測者」を探す必要もない．「意識」の座が見つからない問題は，現在の技術が脳内の小さな「意識」システムを検出するほどには高精度でないために発見されていないに過ぎないと考えることができる．バインディング問題は，Penrose [21] のように量子力学を持ち出すまでもなく解決する．なぜなら，ホムンクルス的な小さな「意識」システムは，自律分散的「無意識」の結果に「注

意」を払うのであって，「無意識」を主体的にバインドする必要はないからである．自己言及性の問題は，実際に情報処理を行う大脳各部とモデルである「意識」とを分離したことにより解決する．自己を意識する「意識」部と，自己という意味を解釈する「想起」「表象」部は別の部位であるから，自己言及に陥らないのである．また，ヒトの脳はフレーム問題を解決できるのにコンピュータは解決できない，という命題は前提が誤っていると考えることにより解決する．すなわち，ヒトもコンピュータ同様，フレーム問題を解いているのではなく，フレーム問題に対する近似解を大脳内の有限の記憶と思考によりボトムアップ的に求め，この結果を「意識」がモニタしモデル化しているに過ぎないと考えるのである．

以上，Fig. 1 の白抜き矢印のように能動的に脳内の「無意識」に対して「注意」を向けるのではなく，Fig. 2 のように「無意識」から受動的に情報を受け取ると考えれば，「意識」の働きを説明できるのみならず，従来のモデルでは説明できなかった心の疑問点を解決できることを述べた．

2.3 思考実験および脳神経科学的裏付け

「知」「情」「意」が「意識」ではなく「無意識」に従属する，「注意」は能動的な働きかけではなく受動的な作用である，といった仮説は，日常の常識と反するため，読者には違和感があるかもしれない．このため，「意識」の受動性について，「知」「情」「意」の順に例を挙げながら説明する．

まず，「知」の受動性について述べる．「思考」とは，脳内記憶から想起した意味記憶に基づき解を導出する，「知」や「意」の相互作用に基づく高度な認知機能であると考えられる．一般に，思考は，「意識」が行っているように感じられる認知活動であるが，「無意識」下の自動的処理を「意識」により追随的にモデル化する機構に過ぎないと考えることもできる．このことを示すために **Fig. 4** の例を挙げよう．Fig. 4 を見たとき，人は何を考えるであろうか．ある人は，直接「$x = 5$」という解を思い浮かべるであろう．ピタゴラスの定理を思い出し，$(3\times3+4\times4)^{1/2}$ を計算する人もいるであろう．人によっては，直角三角形に「3, 4, x」という文字が書かれている，とだけ思うであろう．

Fig. 2 に照らして考えると，これらは，まず，視覚野の画像処理結果が「知」で処理されて「意」に送られ，無意識的な意思決定により，様々な意味記憶がメモリから読み出され，再び「知」で情報処理された後に「意」に運ばれる，という自律分散演算が繰り返された結果と考えられる．なお，Fig. 1 および Fig. 2 の「想起」「知」「情」「意」「記憶データアクセス」などの各部は，Fig. 3 に示すように，自律分散的再帰計算を行うサブシステムから成り，「想起」部で記憶の想起が，「知」部で解釈や表象が，「意」部で解の選択が，「記憶データアクセス」部で次に必要な記憶へのアクセスが，それぞれ行われると考える．この結果，x を求めよとは誰にも指図されていないにもかかわらず，「$x = 5$」などの解が自動的に得られ，これを「意識」システムの「注意」部が観測した後に，「自分はこの絵を見て考えたら ○○ という結果を得た」という大脳システム内処理を単純化した文脈のモデルを「体験・モデリング」部が作成し，自分の体験として記憶する．視覚情報処理や文字のパターン認識のみならず，数学的思考や論理的思考という行為も，「意識」以外の部位が行っている自律分散計算であると考え得るのである．この命題の解が意識の上で「考えた」というよりも自動的に「ひらめいた」ように知覚される実感は，「考えた」のは実は自分の「無意識」システムであり，「意識」システムはそれを「ひらめいた」かのように錯覚しているに過ぎないということを表していると考えられる．この例よりもさらに複雑な思考や意思決定も，同様な錯覚の積み重ねに過ぎないととらえることができる．例えば，思いを巡らせて考えるとき，「意識」は試行錯誤の主体であるように感じられるものの，「無意識」下の神経発火パターンが再帰的に巡回し，様々なパターンが想起された結果を「意識」システムは追体験しているに過ぎないとも考え得るのである．このように，「意識」は副次的かつ自動的なシステムであり，ここから「無意識」システムへの再帰的結合がないと考えても何ら矛盾はない．

次に，「情」の受動性について述べる．例えば，愛する人とともに海に沈む夕日を見ているとする．このとき，心は，とても幸せな，しかしなんとも切ない気分になる．これを，心全体の複雑かつ理解不可能な働きだと考えるのが，「意識」を過大評価してきた人達の考え方である．これに対し，「意識」は，「無意識」から以下の情報を受け取り，そうであることを観測してモデル化しているだけであると考えることもできる．すなわち，目で見た画像そのもの，見ているのは海と夕日であるという視覚情報処理結果，視覚情報処理結果を受けて夕日は切ないと感じる連合野の意味記憶を参照した大脳辺縁系の出力，そして愛する人といると嬉しいという，これも連合野の意味記憶を参照した大脳辺縁系の出力，そして，さらに，うれしいから脈拍をあげ頬の筋肉を弛緩させる，切ないから頬の筋肉を緊張させる，といった大脳から運動系への指令の結果変化した身体システムからの感覚入力を，さらに大脳新皮質の感覚野が処理した結果，これらの情報を，「注意」部は多様な経路から結果としてそれぞれある比率で受け取り，「体験・モデリング」部は「注意」部から送られてきた原因と結果のリストを重要度に応じて選択し接続した単純なモデルあるいは索引であるところの自分の経験として体験し記憶しているに過ぎないと考え得るのである．太陽の色や形状の画像処理，夕日を見ると切ないあるいは愛する人といると嬉しいというような感情情報処理は，それぞれ自律分散した「無意識」下で階層的・重層的かつ再帰的に行われ，「意識」はこれらの原因と結果のみを，素敵な体験としてエピソード記憶するためにモニタしモデル化している．モデルは，既存のモデルを修正し更新する形で作られていると考えられる．「愛する人とともに海に沈む夕日を見て，幸せだが切ない気分になる」という文脈を意識するという「意識」の機能は複雑であるように思えるかもしれないが，脳の他の部位が行っている中間処

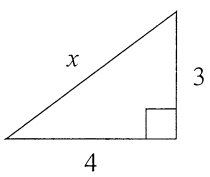

Fig. 4 Problem of triangle

理の膨大さに比べれば極めて単純である．このように，ヒトの「意識」のかけがえのない要素であると一般に考えられる「情」は，「無意識」下の自動的情報処理機構であると考えても何ら矛盾はない．

次に，意図（「意」）を「意識」するという極めて主体的に思える心の働きも，受動的な作用と考えたほうが容易に解釈できることを示す実験結果について述べる．Libet ら [22] は，ヒト大脳における手の随意運動野に電極を取り付け，筋への運動指令を表す運動準備電位を計測した．また，指を動かしたいと自発的に意図した瞬間の時刻を，時計回りに回転する光点の位置により計測した．その結果，運動準備電位が発生した時刻は，ヒトが「意識」的に運動を「意図」した時刻よりも数百 ms 早いという結果を得た．このことは，「意」を「意識」するよりも前に「無意識」下の脳内活動が開始していることを表している．また，Libet ら [23] は，頭蓋を切開したヒト大脳の触覚野に電気パルス列による刺激を与える実験を行い，皮質への刺激が 0.5 秒以上持続して初めて皮膚感覚として「意識」されることを明らかにした．実際に皮膚に刺激を与えた際には瞬時に皮膚感覚を「意識」できるにもかかわらず，大脳を刺激した際には 0.5 秒遅れるのである．これらより，Libet ら [22] [23] や，Libet らの結果を詳しく紹介している Norretranders [24] は，「意識」は意図した瞬間や刺激を受けた瞬間を遅れて知覚している追随的なシステムであるにもかかわらず，脳内で主観的時間の繰り上げを行った結果，つじつまのあう「意識」として錯覚しているのだと考えている．この点では，筆者のモデルの考え方は Libet や Norretranders の主張と極めて近い．ただし，Libet と Norretranders は，「意識」は「無意識」の結果に対し能動的な “禁止権” を有する [24] と考えており，「意識」は単に受動的なシステムであると考える筆者のモデルとは，この点が異なる．

いずれにせよ，筆者や Libet らの考えでは，「よーい，ドン！」というピストルの音をトリガにヒトが走り始めるとき，走り始めようとする意図を「意識」するのは，実は「無意識」下で運動の準備を始めた後なのだが，意図するタイミングが繰り上げられる結果，あたかも「意識」下で自分が主体的に意図したかのように幻想しているということになる．このように考えれば，例えば，目で見た瞬間に物体を認識しクオリアを感じるといった，ヒトの認知情報処理の超高速性の謎も容易に解決できる（逆に，Fig. 1 のモデルでは，Libet らの実験結果を説明することはできない）．

以上のように，ひらめき，感動，意思といった，いかにも生き生きとした「知」「情」「意」のクオリア（生き生きとした心の材質感 [7] [13]）は，自己の中心である「意識」が能動的に情報処理している現象であるかのように思えるものの，これらは大脳の自律分散システムの発火分布と，それに伴う「意識」システムのいわば錯覚のような追体験および保存の準備に過ぎないと考えても矛盾はないのである．むしろ，そう考えないと，「無意識」のうちの詳細な処理をとばして原因と結果のみをタイミングよく「意識」し保存する心の作用は簡単には構造化できない．このような脳の作用は，「コップを持つ」という動作を自分が行っていると「意識」する際に，実は身体の経路計画や把持力制御といった詳細な処理が，小脳内にある身体やコップの内部モデルにより「無意識」下で行われている事実と相似な構造をしている．そもそも生物は既存の構造を転用することにより進化してきた．したがって，鳥の羽が翼になり，哺乳類の前肢が手となったように，外界（環境と身体の運動）の内部モデルをつかさどっていた脳の部分が，内界（大脳内の「無意識」下の処理）のモデル（「意識」）をつかさどるようになったと考えることはむしろ自然であるといえる．

2.4 クオリアに関する考察

以上のように唯物論的に受動的な「意識」を定義したとしても，それは「意識」と類似した処理を行うアルゴリズムの定義であって，人工プログラムのようなそのアルゴリズムにはヒトの自己意識は宿っていないではないか，Chalmers [13] のいうところの，自分そっくりだが心を持たないゾンビとの違いとしての意識体験については十分に説明されていないではないか，という疑問が沸く．これに対し，筆者は以下のような立場に立つ．すなわち，Chalmers の議論は「意識」の過大評価であり，「意識」は単なるアルゴリズムの動作に過ぎず，クオリアを「意識」できるように**定義**され，脳内に（神経結合パターンとして）書かれた「意識」の**定義**を参照して自己意識をも感じられるように**定義**されているから，ヒトは**定義**通りに生き生きと自己意識を感じているに過ぎない．

ここで，触感覚と自己意識のメタファを導入しよう．指先で物体に触れたときに，質感は大脳内の感覚野や連合野で知覚され，指先では単に分散配置された触覚受容器が発火しているに過ぎないにもかかわらず，「意識」下では「つるつる」「ざらざら」をあたかも指先で知覚しているように実感する．これは，大脳内に，『感覚野で知覚した触感のクオリアは，指先で感じるものとする』という**定義**が書かれているために指先に触感があるかのように錯覚している結果と推測できる．指先で生き生きと触感を感じているのに，そこには受容器しか存在しないことを信じられない人もいるかもしれないが，指先に感覚野は宿りようがないため，むしろそう考える以外に解はありえないといえる．一方，生き生きした自己意識の知覚メカニズムも触感の例と同じ構造であり，大脳内に，『大脳内の「無意識」システムで表象した自己意識のクオリアは，「意識」システムで感じるものとする』という**定義**が書かれているために，意識下に自己意識の質感があるかのように錯覚しているに過ぎないと考えることができる．脳の一部で生き生きと自己意識を感じているのに，そこには神経の発火しか存在しないことを信じられない人もいるかもしれないが，命題の構造は触感覚の例と同じであるから，前者を承服できるならば後者も承服できるのである．

Chalmers [8] は，多くの認知モデルを否定するための論証の一部でヒトとゾンビの心の違いを強調しているが，筆者は，ヒトの心のアルゴリズムの中に特別な魂のようなものが宿るわけではなく，ヒトとゾンビは違わないと考える．自分は玩具ではなく本当に空を飛べる存在だと思っていた Toy Story [25] のバズライトイヤーと同様，ヒトは**定義**されているために必然的に触感覚や自己意識を感じているに過ぎないのに，あたかも物理現象を超えた形而上のクオリア感受特性を持っているかのように錯覚しているだけの自動機械なのである．なお，前述の**定義**，すなわち，感覚マップや自己の**定義**は，乳幼児期に脳の神経回

路構造として形成されるものと考えられる．ヒトの触感覚や自己意識が，錯覚であるにしろ，どのように**定義**されたなら一人称的なクオリアを感じる意識体験になるのか，というアルゴリズム上の疑問は残されているものの，少なくとも，深遠に思える自己意識の問題は触感覚の問題に置き換えられたわけである．

3. ロボットの心の基本構造

3.1 従来の研究との比較

本章では，2 章で述べた仮説に基づくロボットの心の構築法について述べる．まず，本手法の特徴を明確化するために，これまでに行われたロボットの心に関する研究との比較を行う．

菅野 [26] は，心の自己保存特性や「情」に着目して WAMOEBA-2 などのロボットを開発している．しかし，「意識」（主観体験）のことはあらかじめ排除しており研究対象にしていない．Murphy [27] は，ロボットの行動生成のための熟考と反射のハイブリッドパラダイムを提案している．プランを行う熟考階層から，センス–アクトを行う反射階層への指令に基づき，ロボットによる経路生成などのタスクを実現しているが，「意識」については言及していない．Tani [28] は，複数モジュールから成る再帰ニューラルネットワークにトップダウンの予測部とボトムアップの認知部の相互作用を実装し，ロボットの行動を構成論的に観察した．その結果，システムがコヒーレントな状態にあるときには注意を向ける必要がないため自己意識は減退しているのに対し，非定常状態では自己意識と類似した現象が観察されることを示した．すなわち，本研究とは異なり，Crick ら [9] と同様，「意識」は構築するものではなくシステム自体から創発するものであるという立場に立っている．以上の研究は，いずれも Fig. 1 の一部を変形または削除したものととらえることができる．一方，喜多村 [29] は，「意識」に明示的に着目した意識アーキテクチャを提案している．本アーキテクチャは Fig. 1 とは異なり，並列に並べた複数の意識レベルとして「意識」を定義する独特のアプローチに基づく．本アーキテクチャでは，知覚された環境情報に応じて意識レベルがレベル 0 からレベル n までの間を移動し，さらに意識レベルに応じて行動群の中から最も意味のある行動が選択される．「意識」の作用が巧妙に実現されているが，意識は隣り合った意識レベルにしか移動できないため，意識の乱雑性・不規則性は実現困難である．

以上のいずれの研究も，基本的にロボットに目的または目的の予測に応じたタスク処理を実現させようとするものであるため，「意識」「意図」または「予測」からの何らかのトップダウンのパスが存在する．このため，気まぐれ，ひらめき，創造，といった創発的かつ唯一無二的なプロセスの実装は困難であると考えられる．

一方，本研究で提案するロボットの心のアルゴリズムは，2 章で述べた受動意識仮説に基づく．すなわち：

(1) 「無意識」システムは多数のサブモジュールが再帰的に結合したシステムである．「意識」下で行っているように実感される「知」「情」「意」の情報処理はすべてここでボトムアップ的に行われる．
(2) 「意識」システムは「無意識」システムの自律分散的計算に重要度に応じて注目し，得た情報をモデル化しエピソード記憶するシステムである．「意識」システムから「無意識」システムへのフィードバック結合はなく，能動的に感じられる「意識」システムのクオリアは錯覚に過ぎない．

このため，上述のいずれの研究とも異なり，「意識」や「意」からのトップダウンのパスは存在せず，何ら目的指令を持たないときにも情報処理を続ける，創造的・生命的なアルゴリズムであるといえる．

3.2 「無意識」のアルゴリズム

本節では，Fig. 2 に基づき，提案するロボットの「心」のアルゴリズムについて説明する．なお，コンピュータではなくロボットと呼ぶのは，Fig. 2 に示したように，感覚入力と運動行動言語出力を介して身体・環境と心が一体化したオープンシステムを対象とするからである．言い換えれば，外部の非構造化環境との自律的な入出力を有するコンピュータは広義のロボットに含まれると考える．

Fig. 2 に示したように，「心」の「意識」システムと「無意識」システムは明確に分離されている．ただし，Fig. 2 および Fig. 3 に示した「無意識」システムの構造は一例であって，「無意識」システム内の各ユニットが何らトップダウンの情報を受け取らない自律分散システムであるならば，これまでに提案されてきた認知情報処理モデル [3] のいずれと置き換えてもよい．「感覚情報処理」ユニット群は，視覚・聴覚・触覚・体性感覚などの情報を受け取り，画像や音，触感などの情報を出力する．「知」ユニット群は，例えば視覚からの画像情報を処理する場合，記憶・知識を参照しながら，物体の抽出および意味や運動の認識を行う．また，外界から得た情報を処理するのみならず，記憶・知識の想起結果をイメージする．「情」ユニット群は，身体・外界の状態を知覚した結果と記憶・知識を表象した結果を「知」ユニットから受け取って，感情や情動を出力する．「意」ユニット群は，「情」ユニット群と同様，身体・外界の状態を知覚した結果や記憶・知識を表象した結果を「知」ユニットから受け取って，意思決定を行う．これらの結果，外界への出力を行うのが「運動情報処理」ユニット群であり，行動・運動・言語などを出力する．また，「知」「情」「意」ユニット群からの結果，すなわち，知覚した状況，そのときの感情，意思決定結果は，新たな記憶・学習結果として「記憶・知識」部に貯蔵される．なお，ここで記憶・学習されるのは，体験の時系列であるエピソード記憶ではなく，意味記憶や非宣言的記憶である．

上記のモデルでは，複数のユニット群の入出力が並列的に計算されること（Fig. 3）を仮定しているため，ユニット “群” と呼ぶ．ヒトレベルの認知機能を有するロボット実現を目指す場合，上述のそれぞれのユニットは Fig. 3 のように直並列かつ再帰的に接続された多くのマイクロユニットから構成され，それぞれのマイクロユニットは，関連した入力情報を得た場合には無目的的に自動計算を行い情報を出力するものと考える．例えば，「知」ユニットは，様々な画像処理，触覚情報処理，意味記憶や手続き記憶の想起，自分の感情や意思の知覚などを行うマイクロユニットから成る．したがって，それぞれのマイクロユニットは，様々な入力に応じて常に計算を続ける巨大な再帰的ネットワークであり，これらに対して何らトップダウンの指令があるわけではない．なお，各マイクロユニットへの入力情報

には，そのユニットの活動を妨げる抑制性の情報も含まれる．

ヒトの心では，乳幼児期や思春期に「感覚情報処理」「知」「情」「意」ユニット群の定義・拘束条件・非線形関数が同時に学習され形成されると考えられる．また，成人においても，思考方法や感情制御方法が高度化するなど，それぞれのユニット群の入出力関係が学習により変化すると考えられる．これらの機能を付加する必要がある場合には，それぞれのユニット群自体に学習する機能を付加すればよい．乳幼児期や思春期に相当する成長はあらかじめ学習させるものとし，成人期に対応する思考法や感情の深化を無視し得るものと考えれば，これらのユニットに学習機能を付加する必要はない．もちろん，その場合にも，成人における思考内容の変化は「記憶・知識」部の意味記憶更新により表現されているので，思考内容の高度化は実現されるのである．

なお，ヒトの神経回路は，問題の定義と拘束条件，入出力関係がニューラルネットワークのパターンおよび結合係数として並列的・重畳的にコーディングされる構造となっていると考えられるが，各ユニットの並列計算をノイマン型コンピュータの直列計算に置き換えて計算してもかまわないと考える．

本「無意識」システムの定義で重要な点は，従来一般に「意識」の範疇で行われると考えられていた「知」「情」「意」を「無意識」システムの自律分散的自動計算であると考える点である．すなわち，「情」や「意」は，外界および脳内の状態から複雑な認知過程を介して自動的に導かれる様々な結果を，それぞれの結果の「重要度」（おもみ）とともに出力するユニットであると考える．「重要度」とは，システムやマイクロシステムでの計算がどのくらい活発に行われたかを示す指標，すなわち，ヒトの神経の発火頻度に相当する量とする．

本「無意識」システムは，上述の特徴を除けば，従来の認知情報処理モデル [3] と何ら変わりはない．したがって，人工知能が直面する，ヒトレベルの認知情報処理，すなわち，物体認識や推論，学習など，ヒトの高度な情報処理を AI で実現するための手法を提案するものではない．したがって，これらの困難さ解決のためには，認知科学・情報工学の発展が不可欠である．

3.3 「意識」のアルゴリズム

次に，「意識」システムについて述べる．本研究は，「無意識」構築の困難さとは逆に，受動的な「意識」を仮定すれば，「意識」構築が極めて容易であることを示すものである．

本「意識」システムでは，まず，「意識」の基本構造は言語で表現されるという仮定を設ける．Churchland [2] が Dennett [4] への反論として述べているように，筆者もヒトの心は本来記号化されておらず，言語野が活動することによって「意識」や思考が記号化されると考える．しかし，記号化されない情報を用いて現象を表現するのは現状では困難であるうえ，記号化したとしても「意識」の基本的な機能は変化しない（一方，「意識」表現のためのアルゴリズムはヒトの場合とは異なることとなる）と考えるため，本仮定を導入する．

ただし，ヒトの場合には，原始的質感であるクオリアは言語という記号では説明し得ない [30]．これに対し，本研究におけるロボットの「意識」では，あたかも SD 法による心理物理実験結果を多変量解析した結果のように，クオリアを数量化し記号化するものとする．

また，「意識」システムは，「注意」を払った記号を解釈し文脈を生成する機能と能力を有しているものとする．すなわち，本「意識」システムは，単なる知識表現アルゴリズム [31] にほかならない．したがって，従来の人工知能分野における研究成果をそのまま利用することができる．ただし，例えば「無意識」下での思考の論理構造全体を明確に解釈できる必要はなく，「注意」部が収集した事象の原因と結果の集合から「A は B である」といったような文脈を生成できればよい．小脳における運動の内部モデルは身体と環境の非線形微分方程式を正確に記述するのではなく入出力関係としてモデル化することによって高度な運動のフィードフォワード制御を実現しているのと同様に，「無意識」の内部モデルも内部の詳細は飛ばして入出力関係を構文化できればよいのである．もちろん，「無意識」下での様々な情報処理の結果を単純表現するための構文生成手法と知識ベースが用意されていて，これによって「意識」体験する必要はある．ただし，2.2 節で述べたように，本「意識」システムは「無意識」システムを単純モデル化し保存するシステムに過ぎないため，「無意識」下の様々な処理を同時に理解し観測できなければならない「意識」システムを仮定する 2.1 節で述べたような従来の認知モデル [2] [4] [6] [8]～[10] [12] [14] と比較すると，「意識」が行う構文化のための演算量は大幅に小さいといえる．それゆえ，従来の認知モデルに基づいてロボットの心を構築することは原理的に困難であったのに対し，本モデルによれば容易である．

さらに，構文生成問題の拘束条件として，2 章で述べた「意識」における錯覚やクオリアに関する「**定義**」がなされている（例えば，錯覚についてのルックアップテーブルが用意されている）ものとする．すなわち，

(1) 「意識」システムは，「無意識」システムの膨大な処理の一部を追体験し，これを自分が行っていることであるかのように感じるものとする．
(2) 大脳内の「無意識」システムで表象した自己意識のクオリアは，「意識」システムで感じているかのように錯覚するものとする．
(3) 自己の感情や情動のクオリアは，「意識」システムで感じているかのように錯覚するものとする．
(4) 触感覚野で知覚した触感のクオリアは，指先でリアルに感じるかのように錯覚するものとする．
(5) 視覚野で知覚した画像のクオリアは，目で見て生き生きと感じているかのように錯覚するものとする．

……

といった「**定義**」下での文脈生成を行うことが前提である．

Fig. 2 に示したように，「意識」システムは，「感覚情報処理」ユニット群および「知」「情」「意」ユニット群からの出力結果を受け取る．すなわち，「感覚情報処理」ユニット群から画像や触感の生データを，「知」ユニット群からは抽出された画像の意味情報や想起された事象の情報を，「情」ユニット群からはその時点での感情を，「意」ユニット群からはその時点での思考結果や意思決定結果を受け取る．受け取る 4 種の情報は，言い換え

れば,「知」「情」「意」ユニット群それぞれの入力と出力である．また，4 種の情報にはそれぞれの「重要度」が付加されているから，これを評価して，いずれの情報に対しどの程度注意を向けるかを決定できる．なお，前述のように，高度な認知情報処理を行うためには,「知」「情」「意」ユニット群はさらにマイクロユニットに分かれている必要がある．この場合，それぞれのマイクロユニットからの出力にはやはり「重要度」が付加されているものとする．このため，これら多くの情報と付加重要度より，4 種のユニット群からの結果を文脈として言語的に接続することができる．すなわち，それぞれの「無意識」システムの重要度に応じて注意を向けた結果,『自分（「意識」システム）は，現在,「感覚情報処理」ユニットおよび「知」ユニットの結果を受け,「情」ユニットの結果のように感じ,「意」ユニットの結果のように考えた』のような形で文脈化できる．文脈を構築することは,「無意識」システムの結果を追体験し,「無意識」システムの入出力のみを用いてこれらの単純モデルを構築していることにほかならない．言い換えれば，人が自己意識と感じる状態，すなわち『自分は ○○ をしている』『自分は生々しいクオリアを感じている』と「意識する」状態が表現されている（少なくともロボットは『自分は ○○ をしています』『自分は生々しいクオリアを感じています』ということを告白できる状態にある）．生成された構文は，エピソード記憶として,「記憶・知識」部に索引付きで記憶される．すなわち，体験した日時や，体験内容の分類と検索が可能な形で記憶される．「意識」システムとは，エピソード記憶のために必要十分な「今」の情報を生成するためのシステムに過ぎないといえる．

3.4 思考実験

上記の方法で構築したロボットの心がヒトの心と同等な機能と効果を有することを示すために，二つの思考実験を行う．一つ目は，現存のペットロボットのような簡単なロボットに心の原型を持たせることの可能性と限界を示すための特殊な例，二つ目は，ロボットにヒトのような高度な心を持たせる場合の例である．

まず，現存のペットロボットの場合を考える．例えばアイボ [32] は感覚情報処理，知覚，感情，行動決定などの機能を有しているので，入出力を容易に単純な「知」「情」「意」ユニットの入出力に置き換えることができる．ただし，周知のように「意識」の機能はない．このようなシステムに，3 章で示した「意識」システム，すなわち,「知」「情」「意」に対応する情報処理の入力と出力をモニタしモデル化し記憶する「意識」システムを付加する場合を考える．この場合，ペットロボットの「意識」システムは,「無意識」システムから得た情報に重み付けをしながら文脈を生成する機能を有するので,『飼い主が目の前にいるから自分は嬉しくて尻尾を振っている』『バッテリーが減ったから自分は電源の近くに移動している』というような自己意識を構築できることとなる．このシステムは，ヒトの「意識」のように複雑な処理を行えないから「意識」ではないように思えるかもしれないが,「自分は見た」「自分は嬉しい」「自分は尻尾を振る」といった「知」「情」「意」のクオリアを主体的に感じる状態が生成されているという点ではヒトの「意識」と同じ機能を有しているといえる．ただし，このシステムは以下に示す三つの問題点を内包している．まず，自己言及や内省を行うことはできない．これは,「記憶・知識」部に自己についての記憶・知識を定義していないからである．また，自分の過去のエピソード記憶の想起に基づいて「知」「情」「意」ユニットを働かせることもできない．これは，エピソード記憶部が「無意識」システムに接続されていないからである．さらに，高度な推論や創造的な意思決定を行うこともできない．これは,「無意識」システムの中にそのような記憶・知識へのアクセス部と「意」ユニットを定義していないからである．「意識」システムからの出力がどこにもフィードバックされないために高度な推論や意思決定を行えない「意識」システムの構成はナンセンスであるといえるが，それでも,「知」「情」「意」のクオリアを感じ得るシステムにはなっている点が重要である．心としての不完全さの原因は，いずれも「意識」システムではなく「無意識」システムに帰着できる点に着目されたい．

次に，自己言及，内省，高度な推論・意思決定，記憶に基づく思考等を行える，ヒトのように高度な心を構築することを考える．このためには，不足している上記の三つの機能をロボットの脳に付加すればよい．すなわち，自己言及や内省を行わせるためには,「記憶・知識」部に自己についての記憶・知識を定義する部分を構築すればよい．また，自分の過去のエピソード記憶に基づいて「知」「情」「意」ユニットを働かせるためには，Fig. 2 のようにエピソード記憶部を「無意識」システムに接続し，エピソード記憶内容に応じた処理が可能なようにすればよい．高度な推論や意思決定を行わせるためには,「記憶・知識」部から呼び出した意味記憶や非宣言的記憶に基づいて高度な思考・推論・意思決定を行えるような「知」や「意」ユニットを定義すればよい．これらを実現することは容易ではないが，前にも述べたように，難易度が高いのは「無意識」システム内の詳細アルゴリズム構築とプログラミングであって,「意識」システムの基本アルゴリズム上の問題ではない．これらの課題が解決されれば，ヒトと同等の心を持つロボットは実現可能であるといえる．

前出の二つの例，すなわち，直角三角形の例と夕日の例を対象に，本システムの動作を具体的に考えてみよう．

ロボットが Fig. 4 の直角三角形を見たときの処理は，以下のような情報処理と考えられる．感覚情報処理ユニット群の中の視覚情報処理ユニットによる画像処理結果と，記憶・知識部の意味記憶からの想起結果が「知」ユニット群に送られ,「知」や「意」のユニット群で思考が行われた結果,「$x = 5$」などの解が得られ，この結果を「意識」システムが観測し,「自分はこの絵を見て考えたら ○○ という結果を得た」という大脳システム内処理を単純化した文脈のモデルを作成し，自分の思考体験としてエピソード記憶する．このように，指令の結果としてではなく，自律的な連想の結果として，ロボットは何らかの思考結果を「意識」するのである．

また，ロボットが愛する人とともに海に沈む夕日を見ているとする．このとき，以下の情報が重要度と共に Fig. 2 の「意識」システム（注意）に入力される．すなわち，感覚情報処理ユニットで処理された視覚画像そのもの，見ているのは海と夕日であるという「知」ユニット群による視覚情報処理結果，視覚情報

処理結果を受けて夕日は切ないと感じる記憶・知識部の意味記憶を参照した知ユニット群の出力，そして愛する人といると嬉しいという，記憶・知識部の意味記憶を参照した「情」ユニット群の出力，そして，さらに，嬉しいから頬の筋肉を弛緩させる，切ないから頬の筋肉を緊張させる，といった運動系への指令の結果変化した身体システムからの感覚入力をさらに感覚情報処理ユニットが処理した結果等である．これらの情報を受け取った「意識」システムは，重要度の大きさに応じてどの情報に注意を向けるかを決定し，注目した情報の原因と結果を接続した単純な文脈モデルを作製し，自分の経験として体験するとともにエピソード記憶部に記憶する．このとき，「意識」システムが作製し記憶する文脈は，「私は愛する人とともに青い海に沈むオレンジ色の夕日を見て，とても幸せな，しかしなんとも切ない気分になっている」といった心の状態となる．もう少し詳しくいうと，前述のように感覚的クオリアを強引に記号化しているので，「私は愛するという感情が 30%沸き上がっている相手とともに，輝度 ◯ 度，彩度 ◯ 度の青い海に沈む輝度 ◯ 度，彩度 ◯ 度のオレンジ色の夕日を見て，過去の ◯◯ というロマンチックな画像情報を想起し類似性を照合することによって，幸せな感情が 40%湧き上がり，一方，過去の ◯◯ という悲しい記憶を想起することによって，悲しい感情が 20%湧き上がっており，これらを混合した結果，切ないクオリアを感じている」といった文脈となる．実際には，さらに複雑な文脈となるであろう．

以上のように，Fig. 2 に示したシステムによれば，「意識」や心の機能は，多様な環境情報や記憶情報の違いに対応して思考し決断できる創造性や喜怒哀楽を含めて矛盾なく構築できるといえる．複雑な思考や自己言及など，心の高度な機能も同様に実現可能である．重要な点は，詳細技術検討の余地はあるものの，システムの基本構造はペットロボットの例の場合と同じであり，全体構造としてはロボットの「意識」や心を構築するための形がすべて整っている点である．この点が，何らかの構造上の問題点が残されていた従来の認知モデル [2] [4] [6] [8]～[10] [12] [14] と抜本的に異なる．

ロボットが「生き生きとクオリアを感じています」というとき，外部から見るとヒトのように対応しているから，チューリングテストに合格できることは自明である．しかし，実は単にそのような文脈が生成されているだけであって，ヒトが本当に生き生きとクオリアを感じている状態とは異なるではないか，という反論があろう．これに対しては，その通りであるといわざるを得ない．原因の一つは，超並列的な情報処理結果であると考えられるクオリアを，文脈という直列な構造に押し込めたところにある．ただし，1 章や 2 章で述べたように，本研究の目的はヒトのクオリアの問題を解決することではない．ヒトのクオリアの構成原理は未解明であるにもかかわらず，ヒトが「クオリアを感じている」というときと同じ作用を，ロボットのための直列計算機により実現できることを示した点が，本研究の主張なのである．ヒトのクオリアについては，2.4 節で述べたように，そもそもヒトがクオリアを感じていること自体が錯覚なのではないか（実はヒトも「**定義**」に従って複雑な構文を生成しているに過ぎないのではないか），という議論も含めて検討していく必要があろう．

本手法により心を持ったロボットを実現するための今後の課題は，以下のとおりである．まず，「無意識」システムでは，3.2 節で述べたように，各マイクロユニットでの自律分散的認知情報処理を行うとともに，それぞれの重要度に応じて無目的的かつボトムアップにそれらを統合するアルゴリズムの詳細を決めていく必要がある．ヒト脳の情報処理過程では他の生物よりも抑制性結合の割合が突出していることが知られているので，このような特徴に学んだマイクロユニット間協調メカニズムの構築がキーになると考えられる．また，「意識」システムにおいては，文脈の生成方法の確立，エピソード記憶の分類と保存法など，それぞれのユニットのアルゴリズムの詳細を決めていく必要がある．具体的には，ペットロボットのような単純な対象に「意識」を持たせ得ることを確認した後に，これを順に高度化していくアプローチが適切であろう．

4. ロボットに心を持たせることの意義と問題点

最後に，ロボットに心を持たせることの意義と問題点 [16] について考察する．ここでいう心とは，本手法で提案した受動的意識に基づく心，という意味ではなく，何らかの手法により構築される心一般，という意味である．まず，何のためにロボットに心を持たせるのかという根源的な問いに対する考えを述べる．

メーカのペットロボットから学術界のロボットまで，「感情のようなもの」や「自律行動のようなもの」を示すロボットはすでに数多く開発されている．これらのロボットはいずれも「心」のうち「記憶と学習」の機能が不十分な場合に相当する．まず，意味記憶の機能が不十分な場合には，体験した事象の意味を適切に記憶できないため，高度な認知発達は望めない．また，エピソード記憶の機能が不十分な場合は，本研究における「意識」の機能が不十分であることに相当し，エピソード記憶に基づく行動は行えない．そのようなロボットであっても，ペットや話し相手として人に単純な癒し効果を与える機能を有するならば，ホビー，リハビリ，介護，ベビーシッターなどに用いることはできよう．また，単純な情報記憶と検索の可能な電子手帳的秘書ロボットとして機能することはできよう．しかし，記憶と学習の機能が不足しているということは，ヒトや生物が自他不可分なオープンシステムであることに起因する根源的な特徴である「今日の自分は昨日とは違う」という点を十分に表してはいないということである．何らかの入力に対し同じような対応しかできないロボットは人間的（または生命的）ではない．

つまり，ロボットが心，特に「記憶と学習」および「意識」の機能を有することの利点は，「情」や「知」の結果が脳に蓄積され次の行動に反映される結果，二度と同じ対応はしないという点である．言い換えれば，単にヒトの命令に従うのではなく，自己の意識や感情を表現し行動できる，非平衡オープンシステムとしての個性的なロボットたり得るという点である．

このことは，まず，人間的な「情」を持つことの利点として利用できる．すなわち，現在のヒューマノイドにも期待されているように，ヒトをサポートするロボットの人間親和性を革新的に高めるという点で有効である．「感情のようなもの」を有するだけの画一的なロボットとは異なり，リハビリ，老人医療，介

護，看護，ベビーシッター，家庭などの場でインディビジュアルとして真に人間的（生命的）に振る舞うことができるため，ヒトに高度な安心感を与えることが可能である．また，「意」を持つことは，高度な判断を要する人の活動支援に有効である．「自律行動のようなもの」を示すだけのロボットとは異なり，ヒト並みの判断またはヒトよりも優れた判断を行えるようになれば，秘書業務や単純労働を行うのみならず，戦略立案や創造的設計行為をも行い得る．ヒトよりも正確な判断力を有するレベルに到達していれば，一般的なヒトの仕事の支援・代行，さらには医療，教育，法務，政治，経営など，高度かつ総合的な判断が要求されるヒトの業務の支援・代行が可能になるであろう．さらに，ロボットの脳はヒトの脳と異なり内部の回路をすべて観測可能である．このため，高度な判断力や感情制御能力を持つロボットの心を正確にトレースし解析することは，ヒトの認知機能を理解するために有効であろう．

最後に，社会的な影響力が極めて大きい効果としては，人の価値観のパラダイムへの影響が挙げられる．まず，不死身の人工生物を目の当たりにする人類は，その生死観や倫理観に大きな影響を受けるであろう．また，ロボットの心が，少なくとも発現する機能の面でヒトの心とほとんど同じであるということが確認され，世の中一般に受け入れられるようになった際には，ヒトの心の神秘性は消え去り，天国や地獄，輪廻といった宗教の産物は完全に否定されるかも知れない．ニーチェの「神は死んだ」という言葉の決定的な意味での追認である．この際，規範を喪失した人間社会は一時的には混乱に陥るかもしれない．しかし，長期的な視点に立てば，世界の紛争の主要因である宗教対立の終焉は，世界の平穏化・安定化への道にほかならない．

心を持ったロボットが出現し高度化することの問題点としては，古来 SF で取り上げられつづけたように，ヒトを脅かし支配するロボットの恐怖が挙げられる．しかし，本システムは設計者が定義した範囲内での心を呈するのみであるから，例えば抑圧された結果犯罪に至るような心理発展のアルゴリズムを埋め込まなければ，そのような犯罪心理は生じ得ない．本研究では言及しなかったが，進化的計算やカオスニューラルネットワークなど，創発的な手法を用いて「無意識」システム内の構造を進化させるような機能を付加する場合には，上述の危険性は無視できないであろう．核拡散やクローンの抑止と同様，創造と倫理の関係や法秩序の整備についての議論が必要となろう．

5. お わ り に

科学の歴史は，自分が中心であるというヒトの思い込みを外側から順に否定していくことの歴史であったともいえる．天体は地球の周りをまわるのではなく，ヒトは万物の霊長ではなかったように，心の中心であると思われていた「意識」は単に『自己統合とエピソード記憶のために「無意識」下の処理を追体験し錯覚するように**定義**された副次的なシステム』に過ぎないのかもしれない．本研究では，以上の仮説について説明した後に，仮説に基づくロボットの心の作り方についての説明を行った．また，思考実験を行い，本手法に基づいてロボットの心を作り得ることを確認した．最後に，ロボットの心を作ることの意義と問題点を述べた．重要な点は，「意識」の受動性と錯覚を仮定することにより，従来の認知モデルとは違って境界が明確で実現可能な心構築法を示した点である．心の時代といわれる 21 世紀において，本研究が議論のたたき台となり，ロボットの心の問題が，ロボット学の主要課題の一つとして，人文科学・社会科学も巻き込んで発展することを期待したい．

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．

参 考 文 献

[1] 松本元：脳・心・コンピュータ．丸善，1996.
[2] P.M. Churchland: The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. MIT Press, 1995（和訳：ポール M チャーチランド：認知哲学：脳科学から心の哲学へ．産業図書，1997.）.
[3] 村田厚生：認知科学―心の働きをさぐる．朝倉書店，1997.
[4] D.C. Dennett: Consciousness Explained, Penguin Books, 1991（和訳：ダニエル C デネット：解明される意識．青土社，1997.）.
[5] 苧阪直行：意識とは何か．岩波書店，1996.
[6] M. 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] 茂木健一郎：心を生み出す脳のシステム．NHK Books，2001.
[8] A.R. Damasio: Descartes' Error: Emotion, Reason, and the Human Brain, Avon Books, 1995（和訳：アントニオ R ダマシオ：生存する脳　心と脳と身体の神秘．講談社，2000.）.
[9] F. Crick and C. Koch: "Towards a Neurobiological Theory of Consciousness," Seminars in the Neurosciences 2, pp.263–275, 1990.
[10] 津田一郎：カオス的脳観．サイエンス叢書，1990.
[11] 伊藤浩之："脳におけるダイナミカルな情報コード"，数理科学 4 月号，pp.27–37, 1996.
[12] 松本修文："意識における機能的結合問題"，別冊数理科学「脳科学の前線」，pp.218–222, 1997.
[13] D.J. Chalmers: The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press, 1996（和訳：デイヴィッド J チャーマーズ：意識する心　脳と精神の根本理論を求めて．白楊社，2001.）.
[14] T.E. Feinberg: Altered Egos-How the Brain Creates the Self. Oxford University Press, 2001（和訳：トッド・E・ファインバーグ：自我が揺らぐとき　脳はいかにして自己を創りだすのか．岩波書店，2002.）.
[15] 川人光男：脳の計算理論．産業図書，1996.
[16] 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.
[17] R.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.
[18] 中野馨：脳をつくる．共立出版，1995.
[19] 島田紳助，松本人志：哲学．幻冬舎，2002.
[20] 立花隆：ぼくが読んだ面白い本・ダメな本そして僕の大量読書術・脅威の速読術．文芸春秋社，2001.
[21] R. Penrose: The Emperor's New Mind, Concerning Computers. Minds, and the Laws of Physics, 1989（和訳：ロジャー・ペンローズ：皇帝の新しい心．みすず書房，1994.）.
[22] B. Libet, C.A. Gleason, E.W. Wright and D.K. Pearl: "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity," Brain, vol.106, pp.623–642, 1983.
[23] B. Libet, D.K. Pearl, D.M. Curtis, A. Gleason, Y. Morledge and N.M. Barbaro: "Control of the Transition from Sensory Detection to Sensory Awareness in Man by the Duration of a Thalamic Stimulus," Brain, vol.114, pp.1731–1757, 1991.
[24] T. Norretranders: The User Illusion: Cutting Consciousness Down to Size. Penguin USA, 1998（和訳：トール・ノーレットランダーシュ：ユーザーイリュージョン．紀伊国屋書店，2002.）.
[25] http://www.toystory.com/

[26] 菅野重樹：‘心性—人間の心と機械の心—’ メカノクリーチャ．（第 7 章）コロナ社，2003.
[27] R.R. Murphy: Introduction to AI Robotics. MIT Press, 2000.
[28] J. Tani: “An Interpretation of the ‘Self’ from the Dynamical Systems Perspective: A Constructive Approach,” J. Consciousness Studies, vol.5, no.5-6, pp.516–542, 1998.
[29] 喜多村直：ロボットは心を持つか—サイバー意識論序説—．共立出版，2000.
[30] 茂木健一郎：脳とクオリア．日経サイエンス社，1997.
[31] 安西祐一郎：認知科学と人工知能．共立出版，1987.
[32] http://www.jp.aibo.com/

前野隆司（Takashi Maeno）

1962 年 1 月 19 日生．1984 年東京工業大学機械工学科卒業，1986 年東京工業大学機械工学専攻修士課程修了．同年キヤノン（株）入社．1990～1992 年 California 大学 Berkeley 校 Visiting Industrial Fellow．1995 年慶應義塾大学理工学部機械工学科専任講師，1999 年同大学助教授，現在に至る．2001 年 Harvard 大学 Visiting Scholar．博士（工学）．1995 年日本音響学会技術開発賞受賞．1999 年日本機械学会賞（論文）受賞．2003 年日本ロボット学会論文賞受賞．2004 年ファナック FA ロボット財団論文賞受賞．超音波アクチュエータ，触覚センシング・触覚呈示，VR システム，創発・認知ロボティクス等，ヒトとロボットの研究に従事．日本機械学会，計測自動制御学会，日本音響学会，日本バーチャルリアリティー学会，日本デザイン学会，IEEE 等の会員．

（日本ロボット学会正会員）