

## 学習行動のモデル論 III

——動物の認知過程についての包括的理解の方略——

石 井 澄

筆者はこれまでに動物の学習行動、とくに基礎的な連合学習に関する認知論的モデルの進展の過程とその問題点について議論してきた(石井, 1992, 1993)。そこでは、1970年代以降において、連合学習の進展の基本的条件としての「認知的」過程を、1試行の条件づけ操作に対応した形の定式によって表すモデルが隆盛となったこと、しかしそれらのモデルは、単純な定式のなかに複雑な学習過程のすべてを詰め込もうとしたこと、非経験的要因を無視したこと、そして道具的学習の事態を実質的にモデルの視野から排除したことなどが原因で、学習行動の予測という、それらのモデルが掲げた目的を達成していないことをみてきた。

学習理論の進展はすべて漸進的過程であり、これらの問題点もまた、連合学習についてのこれまでの基本的観点に沿って、モデルを具体的に改良することによって克服されていくと考えられるべきものであるかも知れない。しかし筆者は、少なくともこの領域に関しては、むしろ個々の具体的な理論が形成されてきたより大きな歴史的背景、すなわち動物の学習行動の研究目的についてのメタ理論的な制約に、これまでの問題の大きな原因があると考えざるを得ない。本稿ではその点に関して、これまでに分析してきたモデルが提出された以降に得られた最近の実験的事実を参照しながら検討し、続いてこの問題についての新たな視点と、そこから導かれる当面の作業課題について述べる。

### IV. 従来 of 認知論的モデルの基盤に関する問題

#### 1. 学習理論の目的——行動の予測・統制は可能か

直接観察できない主観的なものであるという理由で、行動主義が心理学の対象から排除してきた「こころ」を、認知心理学は客観的に分析し得ると主張した。この「こころの復権」という意味で、その台頭は心理学史における一つの大きな転換点であるとも言われる。しかし、このような認知心理学の主張の最大の論拠は、コンピューターのような最新の機械が、ある入力変数に対して、ヒトの主観的意識の側面である思考とか記憶とかいった複雑な過程の結果生じられると思われる行動と、外見上はまったく同じ出力パターンを示し得るといふ、科学技術上の進歩にあった。つまりヒトの認知過程も、脳内の小人の存在を仮定することなしに、機械、とく

にコンピューターにおける2進法的論理演算に基づく情報処理に還元できるから、客観科学の対象と成り得るというものであった。実際、現代における認知心理学者のほとんどは情報処理的アプローチをとっており、人間の脳をコンピューター・アナロジーによって理解しようとしている（Bower & Hilgard, 1981）。

しかし、このコンピューター・アナロジーによる認知過程の理解という視点は、突き詰めればそれまでの行動主義的S—R理論の主張と非常に近いものとなる。なぜなら、これら2つの観点はいずれも、刺激（入力）と反応（出力）との関係を、その間に介在する過程に何らの主体性を要求せずに機械論的に説明・予測することを目的とするからである。換言すれば、両者の違いは、認知心理学が刺激と反応との間の仲介変数を記述する際の使用言語という、後述の議論にも関係する問題を別にすれば、それらがどの程度複雑なものかという問題に還元できる（注:Amsel, 1989）。この機械論的決定論という視点を継承した点で、（少なくとも主流の）認知心理学は、結局は心理学における根本的な変革をもたらすことができなかった（Leahey, 1980）。それはただ、心理学の対象を行動主義の時代より拡大したに過ぎなかった。

このような記述に対しては、認知心理学は必ずしもコンピューター・アナロジーの立場を採るものではなく、またその関心はあくまで心的な機構にあり、行動はその指標に過ぎない（たとえば、Neisser, 1967）といった反論が予想される。しかし、すべての認知心理学の観点についてとはともかく、少なくとも前稿までに検討してきた動物の連合学習に関する「認知的」モデルに対しては、このような指摘は的を得たものだと考えられる。

条件づけ手続きに対応した、1試行ごとあるいは瞬間ごとの連合強度の変化に関する定式化されたモデルの必要性の主張（たとえば、Wagner, 1969）は、明確に認識されているかどうかにかかわらず、概念の操作的定義の意図を内包する。つまり、連合強度の変化という内的な学習過程の特質を、それを規定する外的環境条件によって記述しようとする試みとなる。したがってこの意味において、これらの認知論的モデルが将来その目標を達成したとすれば、それは皮肉にも内的概念の必要性を否定する実験的行動分析の主張（たとえば、Skinner, 1950）を裏づけるものとなり、また（S—S連合を仲介過程として含むとしても）刺激によって反応が機械的に決定されるというS—R連合理論の信念を実証することになる。そして論理的には、このような観点に立つ実証的研究の蓄積と、それらに基づく適切なモデル化によって、学習理論が行動の予測と統制というその目標を達成することは可能であろう。

しかし、現実の有機体の行動を規定する要因の多様性とその不安定性を考慮した場合には、このような学習理論が現実にもその目的を達成できる可能性はあり得ないと断言してよい。なぜなら一つには、人間科学においては、適切な理論化のための必要条件である、実験室的環境の条件統制が、物理科学の場合と同等になされることは期待できないからである。物理科学においては、対象とする物質の状態に影響する要因がほぼ特定されており、厳密な実験室統制の達成が可能である。（もちろん、それでも誤差要因は完全には排除できないが）。しかし行動科学

においては、遺伝的特性や過去経験などのいわゆる個体差、あるいは刺激呈示といった操作の影響における誤差は、可能な場合でも緩やかにしか統制できない。また、たとえ温度や明暗周期といった「統制された環境下」で飼育された、遺伝的に均一な近交系動物を被験体とした場合でさえ、すべての個体に対する実験以前の経験、あるいは実験の特定の時点における被験体の内的状態の統制（ということは実験操作の斉一的影響の確保）には非常な困難を伴う。そしてとくに学習心理学においては、特定の経験の影響を条件統制的な実験において分析することを目的とするために、これらの「剰余的変数」は独立変数としての操作と直接的に交絡する可能性が高い。そして現実には、この領域における研究者の多くは、その結果としての指標行動の大きな個体差と格闘せざるを得ない。

さらに、操作変数の行動に対する影響は直線性をもたないことも多く（たとえば、Yerkes & Dodson, 1908）、また刺激あるいは反応の閾値についても複雑な過程を考慮する必要（たとえば、Frey & Sears, 1978）があるが、行動の予測にとってのこれらの種々の課題をどう克服するかについては、現状では見通しすらたない。事実、過去におけるこのような試み（たとえば、Hull, 1952）が成功した例はまずない。そのような制約のなかで、直線的に行動の予測・統制を目指す理論化の試みがどれほど有効性をもつかは疑問である。

このような主張に対しては、石と紙を屋上から投げれば、かならず前者が先に落下するからといって、物理学の理論が予測性をもたないことにはならない、という反論が生じるかも知れない。しかし、これは論旨を取り違えている。上述の議論の要点は、物理学に対応する実験室的統制でさえ、人間科学の場合には現時点では事実上不可能であり、したがって、そのような状況下で得られた実験の事実は偶然誤差を超えた系統誤差が含むことが多く、それが行動の量的な予測のために不可欠な法則性の発見にとって大きな障害となっているということである。モデルを構築あるいは検証するための実験場面における条件統制の困難さから生じる問題は、理論の現実場面への適用における困難とは別の問題である。

また比較的近年になって生物学的な観点からなされた、普遍的学習過程の存在に対する批判（たとえば、Hinde, 1973; Rozin & Kalat, 1971）も、学習行動についての予測性をもったモデルの構築がいかに困難かを示している。たしかに、学習における生物学的制約の事実は、種を超えた普遍的な学習過程の存在を否定することにはならない（たとえば、Dickinson, 1980; Revusky, 1985）。だが、少なくとも行動の予測という目的を維持する限り、学習過程に関するモデルと言えども、種に特殊な制約の存在を「モデルの視野の外にある」として無視することはできないが、現在にいたるまでこの問題についての予測的モデルは皆無である（注: Domjan & Galef, 1983）。

## 2. 連合主義の限界

これまでに検討してきたモデルは、出発点としての認知的着想においては互いに異なってい

た。たとえば, Rescorla と Wagner (1972) は「強化の意外性」を, Mackintosh (1975) は「冗長な情報の無視」を, そして Pearce と Hall (1980) は「統制された情報処理から自動化された処理への移行」を, それぞれ連合学習の進展を規定する基本的な要因と考えた。しかしそれにもかかわらず, 最終的な定式化されたこれらのモデルは, すべて連合強度という共通の変数を中心として表現されている。それは一つには, 連合という用語が経験主義哲学の時代から考えられてきた, 経験の効果についての代表的な仮説的機構を表すことによる。しかし, より大きな理由は, 連合という機構をすべてのモデルが共通に仮定し, それを1試行の条件づけ操作との対応関係を表す定式に組み込むことによって, 実験に基づくモデル間の妥当性の検証が相対的に容易になるという点にあったと考えられる。また, 2者間の連合を要素として仮定することは, たとえそれを強化刺激以外の刺激間の連合にまで拡大しても, Hull (1943) の  $rG$ - $sG$  結合の連鎖の着想のように, 最終的には強化刺激という反応を喚起する力をもつ刺激との連合を想定することで, 入力と出力との関係を規定することができるので, 実験的操作による反応の予測についての記述が可能となる。つまり, 連合強度という変数に集約されたこれらのモデルの形式上の統一も, 行動主義以来の学習理論が維持してきた心理学の目的観, すなわち, 心理学は有機体の行動の予測・統制を目的とする科学である, という Watson (1925) 以来の信念の反映であると考えてよい。そしてたしかに, 実証的事実のレベルにおいても, ある種の経験によって大脳の2つの神経細胞間に新たな回路が形成され(たとえば, Fanselow, 1993), しかも神経間の伝達を媒介する生化学的過程を妨害すると, パヴロフ型条件づけも阻害される(たとえば, Kim, DeCola, Landeira-Fernandez, & Fanselow, 1991) といった, 2者間連合が条件づけ的学習における一つの過程であるという主張を支持する証拠が存在する。

しかし他方では, すべての学習行動を2者間連合という機制によって理解しようとする試みは, 近年までに示されてきた様々な実験的事実に対しては, 非常に大きな困難を伴わざるを得ない。最近になって進展してきた, 条件づけにおける文脈の多様な役割に関する知見を一例として検討すると, その点が具体的に明らかになるであろう。

条件づけに関係した諸現象を連合という機構によって説明する場合, 個々の現象に関してはかならず, 「何と何が連合するのか」を想定しなければならない。そこで, S—S 連合を学習の基本的な機構と考えるモデルは, 潜在制止や消去といった単一の事象の経験の効果についても, その事象と連合するもう一つの事象を想定する必要に迫られる。これまでに問題としてきた連合主義的モデルでは, それは文脈(背景刺激)であった。実験状況においては, それ以外には何ら存在する刺激はないのであるから, そうなるのは必然だったとも言える。そしてとくに Wagner (1981) の SOP モデルが, 多くの現象の説明のために, 文脈に決定的とさえ言える役割を与えたところからは, 文脈の機能に関して多くの研究がなされるようになった。

しかし最近になって, 文脈には連合の一要素としての役割のみでなく, その他の多くの特殊な機能を持つことが判明してきた。一例として Balsam (1985) は, 過去の研究を分析した結果

として、以下の7つを文脈の機能として挙げている。すなわち、①一つの手がかりとして、刺激と連合強度あるいは注意の獲得を巡って競合する；②行動の決定のために、他の手がかりのもつ連合強度と文脈のそれとが比較される；③文脈と連合した動機づけがCSの連合強度と加重して反応を増強する；④文脈が、その存在下でCSが獲得した連合の検索手がかりとなる；⑤刺激の意味を決定する弁別手がかりとして機能する；⑥刺激に対して生じる反応の形態を決定する；⑦刺激の知覚過程に影響する。そして、これら7つの機能は相互に排他的ではない。

しかし Balsam は、文脈にこのような多様な機能があることを指摘している一方で、学習心理学における文脈の研究は、「連合主義的心理学の新たな境地への進展のために、既存の理論に挑戦することを目指す」と述べている。すなわち、これらの機能はすべて基本的には連合という観点から説明されるべきだという、伝統的な連合主義の観点を明確に示しているのである。しかし、このような多様で相互に関連する機能の各々が、具体的に文脈と他の事象とのどのような連合構造によって理解できるのか、というもっとも重要な点については、Balsam ほとんど触れていない。その限りでは、この問題に対する連合主義的アプローチという主張は、Watson の行動主義の教義的主張と同様に具体性をもたないものである。また、道具的学習の連合構造を分析している Rescorla (たとえば、Colwill & Rescorla, 1986; Rescorla, 1991) は、道具的学習行動が弁別刺激、手段反応、そして強化子のいかなる2者間の連合、あるいはそれらの連鎖によっては説明できないことを認めた上で、なお「連合の階層構造」という連合論的発想を維持した説明を主張している。だが、このような階層構造の強調は、逆に説明における2者間連合という概念の重要性を弱めるという矛盾を内包している。

もちろん、これらの文脈の機能についての個々の理論的解釈は、Balsam の主張とは別により詳細になされている。たとえば Miller と Schactman (1985) は、文脈の比較機能について、連合強度の獲得においては刺激と文脈との間に競合はなく、反応の決定の際に両者が比較されるという仮説を提出している。また Bouton (1993) は、消去は条件づけによって獲得された連合強度の減少ではなく、刺激を単独で呈示することによる刺激一事象なしという別個の連合の獲得過程であり、ただ条件づけ期と消去期の文脈が同一であるために、それによって両方の連合が同時に検索されるため、たがいに反応を巡って干渉することが原因であると主張している。

これらはいずれも、連合強度という概念を維持しながら文脈の機能を説明しようとしている点において、連合説の枠組みのなかにあると考えることができる。しかし他方では、これらにおいて実験的事実を説明するためにもっとも重要な概念は、実際にはむしろ記憶表象の検索とか反応決定の段階といった非連合的な過程に置かれている。そのことは、これらの仮説が連合強度の変化の過程については、何らの記述も行っていないことから明らかである。事実、この問題についての最近の実験的事実を概観した Bouton (1994) は、文脈は非連合的な形で刺激に対する学習行動に影響する側面が大きいことを認めている。

このような新たな視点は、従来の注意説の主張などとは異なる理論的意味をもっていると考

えられる。たしかに、前稿までに検討してきたモデルのなかでも、たとえば Mackintosh (1975) あるいは Pearce と Hall (1980) などのモデルは、連合強度の変化そのものではなく、むしろ刺激に対する注意の大きさ（明瞭度）の変化の過程によって、実験的事実を説明しようとしていた。しかしこれらのモデルにおいては、そのような連合以外の過程の変化は、連合強度の変化によって2次的に規定されるものであり、その意味では他のモデルと同様に、連合強度の変化という過程に最終的には依存したモデルであったと言える。しかし、文脈の機能に関する上述の仮説は、それらの機能を連合強度の変化と少なくとも直接的には関係させておらず、その点において、学習理論の新たな展開を促す可能性をもっている。

この他にも、たとえば Pearce (1987a) が提出した、連合強度の獲得過程ではなく刺激間の知覚的般化の過程によって、条件づけにおける諸現象の説明を試みたモデルなども、連合という仮説的機構への過度の依存からの脱却を目指すものだとみなすことができる。これらの新たな理論化の方向は、連合強度の変化は学習過程に含まれる一つの段階ではあるかも知れないが、それ以外にもいくつかの同等に重要な機構が存在し、学習行動の全体的な理解のためには、それらを総合的に考慮する必要があることを示唆している。

このような主張に対しては、多数の機構を仮定するモデルはその複雑さゆえに、予測をもたない単なる解釈になってしまうという批判が生じるであろう。しかし前述のように、現状で予測性を追求することがどれほどの意味をもつのかを考えた場合、連合という単一概念に依存することの適切さも、それとの関連においてあらためて問いなおす必要があると考えられる。

## V. 学習行動の理解——適切な理論化のための作業課題

連合学習の進展が強化の意外性によって規定されるという認知的着想を、連合強度という仮説的概念に置き換え、その変化の過程を定式によって表そうとした Rescorla と Wagner (1972) のモデルは、その予測性の拡大によって脚光を浴び、1試行の操作に対応する形の定式によるモデル化という方向が、いずれは連合学習の過程についての十全な説明を可能にするという期待を抱かせた。しかし現時点で分析するかぎり、これらモデルは、これまでにみてきたようないくつかの難点を抱えていた。そしてそれらの根底にある問題は、行動の予測と統制という心理学理論の目的観と、連合という単一概念によって多様な事実を包括しようとする姿勢という、メタ理論的な次元にある可能性が指摘された。筆者は、したがってそのような現状の打開はモデルの漸進的な改良の試みという対処ではもはや困難であり、動物の学習過程についての根本的な視点と目標の変更が必要であると考え。すなわち、学習過程を複数の認知的機構から構成されるシステムとして捉え、少なくとも当座は関与する複数の機構の理解を目指し、それに基づいて、条件づけや馴化のパラダイムに基づく従来の実験的知見の各々に主とし

て関与する機構を推測し整理するという作業を行なうということである。以下では、それらに関する一つの試論を述べるが、そこでは非実証的な議論も含まれるために、そのままでは明快な理解を得るのが困難になる恐れがある。そこで、最初に筆者の主張の要点を、ジグソーパズルという作業課題に比喻させて説明してみたい。

「学習過程」というタイトルの難解なジグソーパズルを完成させるために、いろいろ推理しながら個々の実験的事実という小片（ピース）を組み合わせる時に、最初は連合という一つの大きな箱が描かれていると思って完成を目指したが、うまくいかない。そのうちに、これは認知機構という、異なる特徴をもつ複数の小さな箱の絵から全体が成り立っているらしいと気づいたら、多くのピースを（自然言語という）色や線の手がかりに基づいて、各々の箱の絵に対応すると推測されるいくつかの群に分類するという作業が、最初の段階における一つの方略である。分類してみても、それはジグソーの完成を意味するものではなく、また各ピースが確実にその箱の部分の構成要素であるという保証もないが、すべてのピースを漫然と眺めているよりは有効であろう。そこで、各々の箱の絵のもっとも特徴的な部分を構成するピースから組み合わせ、その特徴を理解し、つぎに別の箱の絵との境界部分へと作業を進める。これらの段階において、特定のピースがじつは他の箱の絵の構成要素であったということが判ることもあるが、これも仮の群分けという段階を経ている方が気づきやすい。したがってとりあえずは、実験的事実という多くのピースを、自然言語という手がかりを利用して、いくつかの仮説的な認知機構に対応するいくつかの群に分類してみるべきだというのが、ここでの提案である。

### 1. 動物の学習過程についてのシステムの観点

Fig. 1には、ヒトの認知過程（知覚および記憶システムの主要素）についての一つのモデルが例として示されている。この図には種々の情報処理機構の各々と、その間のネットワークがいわゆる「箱型モデル」の形で示されている。このような図をここで示した理由は、モデルの妥当性を議論するためではなく、それがこれまで検討してきた動物の連合学習のモデルとは対照的に、刺激事象の受容という経験によって開始される内的機構は決して単一ではなく、いくつもの異なる機構が存在するという前提の上に形成されていることを再認識するためである。付言すれば、このモデルには連合という機構はどこにも記述されていない点にも注意すべきである。このことは、動物の学習理論において中心的な位置を占めてきた連合という概念は、ヒトの情報処理に関するモデルでは、これら複数の機構の組み合わせによって表される総体を意味する、ということを示唆していると考えられる。もちろん、両者がモデルの対象としている現象の範囲が異なることも、そのような観点の違いに影響しているであろうが、ヒトと他の動物の認知過程が質的にまったく異なるという仮定に立たない限り、このような認知機構の複雑さを動物の学習行動の場合に限って無視することは困難であり、したがって複数機構のシステムとして理解するという観点が必要である。

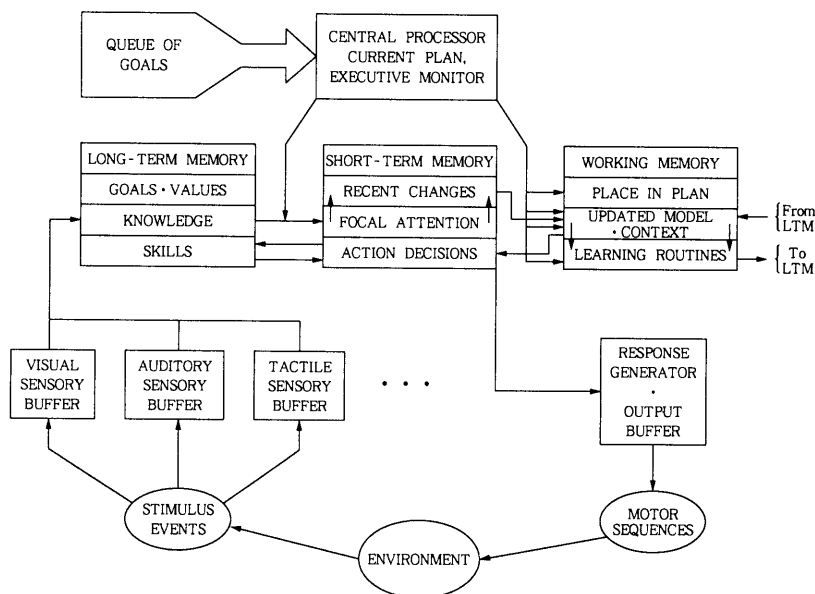


Fig. 1. An example of "box model" for perception and memory system.  
(After Bower 1975).

Hall (1991) は、動物の馴化と潜在制止に関する従来の実験的証拠を詳細に分析し、刺激を繰り返し単独で提示するという単純な手続きでさえ、そのことの効果として少なくとも、①刺激表象の精緻化; ②刺激の連合力(注意)の減少; ③刺激が重要な事象を何も伴わないという知識の獲得、の3点が並列的に生じると考える必要性を主張した。このような分析結果も、多様な経験効果のすべてを連合という仮説的機構によって説明することが、現実にはきわめて困難であることを示唆していると考えべきであろう。

## 2. 情報処理モデルの動物の認知過程への適用の問題

では、動物の学習のための認知機構としてどのようなシステムを想定すればよいのであろうか。最初に生じる着想は、すでにそのような観点からなされた実証的研究の基盤をもつヒトの認知過程に関するモデルを、そのまま借用するということである。ヒトと動物が基本的に同じ認知機構を有するという仮定そのものは、現在ではかつてのような強い拒否反応を引き起こさなくなってきた。それは一つには、脳の解剖学的・生理学的所見から、脳の機能の種差は、(少なくとも哺乳類間では) おもに量的なものであり、質的な差異は非常に少ないことが理解されてきたからであり、他方ではヒト以外の動物にも、計数や推論、はては言語理解までの、従来はヒトに特有なものとされた高次の能力が備わっていることが、行動学的に示されたからである(これらの点については、Pearce, 1987bなどを参照されたい)。それなら、この「動物のヒト視」という観点を情報処理システムにも適用することは可能であろう。そして実際には、



これまでに検討してきたモデルのなかにも、その出発点としての着想を、ヒトの情報処理論的モデルのそれから借用した例がいくつか存在するのである。ただそれらが、認知過程が複数の機構からなるシステムであるという前提を無視し、単一の着想によって、しかもそれを連合という「方法主義的 (methodologic) な理論言語 (Bolles, 1985)」に置き換えて、すべてを説明しようとしたところに問題があったとも言える。

しかし、情報処理的モデルの借用にはいくつかの問題もある。まず、動物を対象とする実験では、ヒトの注意過程に関する両耳分離聴の実験の場合のように、言語的教示が特定の認知機構のみを操作できるという前提 (このことのもつ意味に関しては後に議論する) に対応するものがない。刺激の呈示と行動の測定という限定された方法のみによって、関与し得る機構を操作的に特定することは容易ではない。したがって、特定の仮説的機構の特質を演繹的な実験に基づいて決定するという方法は相対的に困難となり、既存の実験的事実について、その基礎にある認知機構を帰納的に推理するという方式に頼らざるを得ない。

また、これも行動を指標とすることに関係するが、言語的な情報を用いることができない動物実験では、行動の変化によって認知機構を推測するという手段をとらざるを得ない。もっとも単純な手続きは、単に刺激を繰り返し呈示し、反復に伴う定位反応の減少、いわゆる馴化を指標とする方法である。しかし、より複雑な条件下で認知機構を分析する場合には、いわゆる条件づけの手法を用いることが一般的となるが、そこでの強化刺激、あるいは無条件刺激の呈示は、必然的に動物に動機づけ (情動) 的变化をもたらす。したがってこの過程の行動への影響は、そのような刺激を用いる必要が小さいヒトの認知機構の研究と比較して、相対的に大きな問題となるが、ヒトの場合にはこのような要因は、まったく考慮されていないわけではない (たとえば、戸田, 1992) にせよ、十分に検討されていない。

さらにこの動機づけ的要因は、すでに述べたような生物学的制約と呼ばれる事実にも強く関係している可能性があり、したがってそれをどのような形で理解するべきかについては、慎重な検討を要すると思われる。その点を無視して実験的事実を説明しようとする、かつて知覚心理学において、理論から説明できない現象をすべて「注意」に起因するものとして処理したのと同様に、都合の悪い事実をすべて、これらの「種に特有な過程」に押しつけて済ませてしまう危険性がある。

しかし、情報処理論的なモデルへの依存を回避する最大の理由は、先に述べたように、そのようなモデルは最終的にはS-R行動論に帰着するが、筆者にはそのような一貫した機械論的観点への固執こそが、心理学の進展を却って妨げていると思われるからである。それについての全面的な議論は本稿の目的を超えているので割愛するが、以上のような理由から、ここでは他の観点からの一つの方略の可能性について考察する。

### 3. 自然言語に基づく認知機構の分類

どのような視点を採用するにせよ、動物の学習過程を理解するためのシステムの機構を考えていくためには、何らかの分類基準に基づいた、仮説的な認知機構のリストの作成から始めるべきであると考えられる。そして、ここではあえて、そのリストをヒトの自然言語に依存して作成することを提案したい。その論拠を三段論法的に述べると以下ようになる。①ヒトの個人間で同一の認知機構が存在することを認めるなら、論理的にはヒトと動物（少なくとも哺乳類）は基本的に同じ認知機構を有すると考えることができる。②ヒトの「意識（あるいは内観的）内容を記述した」自然言語の各々が内包する意味は、基本的にヒトの脳内の特定の認知機構（の特質）を反映している。だからこそ、ある程度まではそれらの自然言語によって、同じ機構を共有するヒトの脳の間では、お互いの理解が可能なのである。③したがって、それらの自然言語によって、動物の認知機構についてもおおまかに分類できる。

①の論拠についてはすでに述べた。したがって筆者の提案に係わる問題の焦点は、②の主張の妥当性にある。①と②の叙述が容認されれば、③は論理的に帰結されるからである。

②のような着想は、純粹に自然科学的な観点からすれば、実証性をもたない感情移入的なものとして退けられるであろう。実際、ヒトの自然言語が脳内の過程の反映であるということは実証することができない。しかし、それはたとえば数学における公理が、実証が不可能なものであっても、万人に正しいと認められ、数学における基盤的な位置を占めるのと同様に、だれもがそれによって表出される内容を、(完全ではないにせよ)たがいに理解できるという意味で、ヒトの脳内の機構についての基本的な手がかりであると考えられる(養老, 1989)。そのような公理的側面が認められるからこそ、同じく脳の機構の反映と考えられる数学的公理や科学的認識が自然言語によって表現され、そして認知心理学者はヒトの自然言語のもつ意味内容に基づく実験を行なうのであろう。この点についてもう少し詳細に検討してみよう。

ヒトを対象とした認知心理学が、種々の批判を受けながらも、認知行動に関する諸機構の解明に関して一定の成果を挙げってきた一つの大きな理由は、ヒトの場合は言語的教示による特定の認知機構の操作、あるいは言語的反応に基づく特定の機構の分析が可能だという前提を、研究者が基本的に受け入れてきたからであると考えられる(戸田, 1984)。たとえばある数字列を呈示して、「最大公約数を計算するように」と教示すれば、被験者は演算という機構に基づいた反応を示し、同じ刺激について「3分後に思い出すように」と指示すれば、記憶という機構に基づいて反応するとみなされる。だが、反応が本当にそのような機構に基づいたものかどうかは、「計算する」とか「思い出す」という自然言語が意味する仮説的認知機構を個人間で理解できるという、科学的にはきわめて曖昧な前提に依存している。教示を除けば2つの刺激事態の間には何の差もないからである。また、教示に反して誤反応が続いた場合などには、これも言語による質問とそれに対する回答(あるいは内省報告)を分析することで、たしかに演算あるいは記憶の間違いなのか、あるいは単に「ぼんやりしていた」ためなのかを判定するが、そ

こでも自然言語が意味する内的機構についての普遍的理解が前提として存在する。

これに対して動物実験の場合は、言語的コミュニケーションが不可能なために、シェーピングや刺激・反応閾値に関する予備的測定などによって、実験者が操作に対応する行動を測定できるような準備を行なう。しかしそれでも、そのようにして決定された個々の操作は、言語的な意味に対応する前提をもたない、単なる刺激事象の呈示であり、したがって、操作と概念との対応関係の記述については、神経質なほど慎重な態度が要求されてきたのである。

純粹に科学的な論理からみれば、ヒトの実験の場合にも同じ慎重さが求められなければならない。つまり、日常的な意味をもつ自然言語ではない刺激とそれらの配置によって実験的操作を行ない、かつ反応を測定することが求められる。しかし現実には、自然言語による教示が特定の認知機構を作動させ、また被験者の言語的反応の「自然言語的意味内容」を分析することに、心理学者の多くは疑問をもたない。そのことは、自然言語が実験場面に限定された単なる刺激ではなく、むしろヒトの認知機能の特質をある程度反映した「意味」をもつものであるという前提を、実証性がないままに研究者が暗黙理に認めていることを示唆している。

しかし筆者は、そのような実験的研究は厳密な科学性をもたないから放棄せよ、と主張しているのではない。むしろ、自然言語はヒトを含む有機体の認知機構の検討のために、ある程度までは有効な分類的概念を提供しているからこそ、このような前提に立った認知機構に関する研究が進化したのだと考えたいのである。別の表現をすれば、自然言語についてのこのような理解は、ヒトの認知機構についての「メタ認知」的基盤を反映しているとみなせる。

#### 4. 出発点としての分類作業の方法

以上の議論が認められるなら、つぎにはヒトの自然言語に基づく動物の認知機構の分類の方法を検討しなければならない。その具体的な進め方について、確信ある方略を現段階で示すことはできないが、それでも敢えて筆者の着想の一つの例を示そう。

見慣れぬ音を最初に聞いた場合、それによってわれわれの内部でどのような意識が生じるかについて、「一人称的記述」(ナイーブなレベルの内観)を最初の手がかりとしてみる。するとたとえば、ぼんやりしていた状態から「目が覚め (arousal)」、何事がおきたのだろうか、「不安になり (emotion)」, 続いてその事象に対して「注意を向ける (attention)」。さらにそれが何であるかを「確かめ (identify)」, そしてそれがその状況では無害であると「判断する (judgement)」。しかし、同じ音を同じ状況で繰り返し経験すると、「安心し」、「退屈になり」、それを「無視するようになる」。一般に「馴化」と呼ばれるこの現象の背後にある機構を自然言語でこのように表現すれば、以上のようなものになるであろう。馴化は手続き的にはある刺激を繰り返し呈示するという単純なものであるが、このように表現すると、先に述べた Hall (1991) の示唆のように、それらは複数で、かつ経験によりある機構の機能は変化し、他の機構は変わらないと考える必要が生じる。さらに別の状況、たとえば条件づけの事態では、馴化

の場合には関与しない機構が記述されることが考えられ、また各機構の機能変化の様態も異なることが考えられる。

しかし、上に述べたような様々な自然言語が表す（可能性がある）認知機構が、すべて別個のものであるとは限らない。それらのなかには、同一機構についての反対語の関係にあるものや、別の表現に過ぎないものもあろう。そこで、Allport (1937) が性格の特性論的記述を行なうために、性格に関係した幾千もの単語を順次統合して「心誌 (psychogram)」を作成した過程のように、種々の実験的事実に対する上述のような試行錯誤的な記述を種々の現象について行ない、重ね合わせることで、共通の機構を自然言語のレベルで抽出するという作業を行なう方法が考えられる。だが、これまでの膨大な知見についてこのような思考実験を行なうことは多大の労力を要する作業であり、またその過程においては多くの問題が生じることが当然予測される。しかも、複雑な条件操作の結果としての行動の群間比較によって理解される現象が多いために、性格特性の場合のような因子分析的手法による分類も困難である。したがってこのような作業は、今後の課題として、これまでの実験的研究によって成立条件がある程度確定した、より単純な現象から一步一步進めていくしかないであろう。

#### 5. 予想される反論への回答

以上のような提案に対しては、2つの反論が考えられる。一つは、自然言語が認知機構を反映しているとしても、それは実証性をもたないがゆえに科学的には不完全な解釈に過ぎないという主張である。ヒトが他人の「こころ」を不十分にしか理解できない、というありふれた事実はまさに、自然言語の脳内過程の投射としての限界を示している（養老, 1989）。したがってそのような批判はたしかに的確であり、このような作業課題の達成をもって行動の説明とすることはできない。それは、動物の学習過程に関与する可能性のある認知機構についての、概念的な分類を示すための一つの方略として位置づけられべきものである。しかし他方では、その不完全さを含む段階を最初から排除したモデル化の試みは、結局は従来の定式化されたモデルについてすでに分析してきたような問題を蒸し返すだけのことになる恐れが強い。不十分で曖昧な概念でも、出発点においてあえてそれを用いることが、従来繰り返されてきたような、モデルが提出されるとすぐにその適用範囲の限界ばかりが目立つようになるという事態を避けるためには必要なのではないか。

予想される他の一つの反論は、自然言語の使用はあくまで便宜的な隠喩であって、それらが表す「意識される」過程は学習の原因としての機構ではなく、むしろそれらこそが実証的な研究に基づいて、（おそらくは刺激と反応を記述する）客観的な用語によって説明されるべき対象である、というものである（たとえば、Mandler, 1985）。筆者も、最終的にそのような客観性が追求できる可能性を否定しない。しかしこれに対しても、真に科学的な説明とは何かという問題を棚上げしてでも、当座は関与する可能性のある諸機構の分類を行なうことが必要であり、

そのためにはヒトの自然言語を用いることが一つの方略として考えられると答えたい。繰り返すが、それは学習の機構の説明ではなく、あくまでそのための予備的だが必要な作業課題なのである。そして、これまで述べてきたように、それによって高等動物が共有する認知機構のシステムをある程度までは理解できるという仮定が根拠をもつなら、作業の遂行過程における具体的な相互批判によって、より洗練された分類基準へと漸進的に改良され、次の段階への展望が開けることが期待できるであろう。

動物心理学においては、Romanes などの逸話主義的な解釈に対する反動としての機械論的観点が、そのまま行動主義の流れに繋がり (Boaks, 1984), すでにみてきたように、それが「認知論的」と呼ばれるモデルにも受け継がれてきたという歴史がある。したがって Bolles (1985) が分析したように、この領域の研究者は、真の意味での行動の「過程 (processes)」に関心を向け言及することを巧妙に避けてきた。連合という用語は動物の内的過程を意味するものではなく、実験の方法論と心的過程の所産 (product) としての行動との関係についての、もっとも形式的なレベルの記述に過ぎず、それを超えた動物の内的過程に関する思索を拓けるべきだと Bolles は主張する。もちろん他方では、たとえば Crowder (1985) が、短期記憶課題においてヒトの心のなかで進行する過程について、一時は多くの着想の「沸騰状況」が生じたが、それらの多くは内的過程というよりも単なる隠喩に過ぎなかったと批判したように、内的過程についての方法主義的分析を超えた思索には種々の問題も伴う。だが、真剣に動物の認知「過程」に関心を向けようとするならば、それは少なくとも当初は隠喩によってしか表現できないほど複雑なものだと考え、そのようなレベルでの説明を (他の種々のレベルの説明と同時に) 試みる必要も生じるのではないかと思われる。

## 6. 自然言語にもとづく従来のモデルの再解釈

一方でこのような作業努力は、従来のモデル間の対立を異なる観点から解釈しなおすことによって調和させる可能性をも示す。たとえば、Rescorla-Wagner モデルと Pearce-Hall モデルはある意味で同じ着想から出発しているとみなすことが可能である。Rescorla-Wagner モデルの「強化刺激の意外性」という着想は、認知機構そのものではなく、認知機構の一つ以上に影響を与える要因を記述したものである。認知機構はむしろわれわれが「意外」な事象を体験した時に動物の内部で生じるものであり、したがってその内容をヒトの自然言語に基づいて上述のように表現すれば、「強化刺激の意外性」を強調した Rescorla-Wagner モデルと、「信号刺激への注意の減少」を強調した Pearce-Hall モデルは、(たとえば刺激への注意という) 同じ類の認知機構の関与を、別の刺激事象に関して主張していたと考えることができる。そして現実にも、たとえば Pearce (1987b) が、Wagner (1981) の SOP モデルにおける STM 内の表象の「1次活性化 (A1) 状態」という記述を、いわゆる注意を受けた状態を意味すると解釈することで、選択的注意の機構に関するモデルに分類している、といった例が存在する。

これまでそれらがまったく異なった認知機構を反映すると考えられていたのは、手がかり刺激と強化刺激とを最初から区分して考えるという、前述のような操作主義的かつ連合論的な行動論の枠組みのなかでそれらが議論されてきたからである。

換言すれば、その枠組みを外してナイーブに「内側から」、つまり被験体の側から事態を眺めれば、従来は対立すると考えられてきたモデルの着想が決して相互排他的なものではなく、表現のレベルが異なるだけのこともあり、また事態によって関与の程度が変化する別個の機構の一方のみを、それぞれが強調していたことに起因している場合もあるであろう。さらにそのような観点に立てば、これまでの理論的な論争のように、すべての事態を包括する認知的着想の択一的な是非を議論することは意味がなくなる一方で、同一の機構的システムによって、連合学習に限定されないより広い範囲の事実を、（説明ではないにせよ）「理解」することも可能になるであろう。

## 結 語

まだ筆者が研究者として駆け出しだったころに、教科書の一章として担当した学習行動に関する草稿について、他大学出身の2人の動物心理学者に批評をお願いしたところ、両者から、「このような表現はみたことがない。学習行動は操作と反応の用語によって記述されるべきものだ。」と批判されたことがある。その貴重なアドバイスは結果的に無視されることになったのだが、そのことによって、学習行動の研究の伝統がない大学で教育を受けた筆者の視点が、この領域における多数者のそれといかに大きく異なっているかが認識できた。

したがって、これまでに述べてきた従来のモデルについての分析や、進むべき方向についての試案などは、それらの多数の方々からは空虚で非科学的な妄言だと片づけられるかもしれない。しかし、特定の目的観や信念に固執することなくこれまでの経過を冷静に分析すれば、条件づけのような比較的単純な実験操作によって生じる、連合学習と呼ばれる行動の変化の過程でさえ、すべてを連合の（あるいはそれに付随した）機構によって説明しようとする従来のモデルの視点では理解できなくなっていることは明白である。それは、連合という概念を他の仮説的機構に置き換えたところで同じであり、ましてすでに破綻が明らかな操作主義に回帰しても、何の解決にもならない。操作の単純さは内的過程の単純さを意味するものではない。学習の過程は複数の認知機構のシステムとして理解されなければならない。そして、従来の実験的証拠について、そのような視点から整合的に再解釈することが、最初に必要とされる課題である。筆者がもっとも強調したかった点はそのところにある。

そのための、「内側からの」視点に基づくヒトの自然言語を手がかりとするという着想の展望については、正直なところ確信はない。本稿の冒頭でも述べたように、科学的心理学にとって

の行動の予測・統制の困難さは現時点での問題であり、漸進的な努力によって少しずつ解決されていくと楽観することも可能である。しかし、Mackenzie (1977) が指摘したように、Kuhn (1962) が定義した意味でのパラダイムの確立を経験していない心理学において、様々な異なる立場から提出された雑多なモデルがそれぞれの方法で自説の正当性を主張している現状では、科学性の枠そのものを問い直すような視点の変革も必要ではないか (佐伯, 1982)。

しかし、だからと言って、Griffin (1984) や Walker (1983) のように、動物における意識過程の存在そのものを事実から推測するという方向で議論を進めることは、共感を呼ぶことはあっても、実証が永久に不可能な問題を直接に問うているがゆえに説得性が薄い。したがってヒトの自然言語を利用するとしても、それがもつ意識主義的傾向が喚起するこれらの問題を回避するためには、あくまで過渡的の分類のための方略として限定的に位置づけるべきであろう。

#### 引用文献

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Amsel, A. (1989). *Behaviorism, neobehaviorism, and cognitivism in learning theory: Historical and contemporary perspectives*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Balsam, P. D. (1985). The functions of context in learning and performance. In P. D. Balsam (Ed.), *Context and learning*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Boaks, R. A. (1984). *From Darwin to Behaviourism: Psychology and the minds of animals*. Cambridge: Cambridge University Press.
- Bolles, R. C. (1985). Short-term memory and attention. In L. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the influence paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80-99.
- Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior processes*, 20, 219-231.
- Bower, G. H. (1975). Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes*. Vol. 1. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning*, 5th edition. Englewood Cliffs, N.J: Prentice-Hall.
- Colwill, R. M., & Rescorla, R. A. (1986). Associative structures in instrumental learning. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol.20, New York: Academic Press.
- Crowder, R. G. (1985). Basic theoretical concepts in human learning and cognition. In L. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.

- Domjan, M., & Galef, B. G. (1983). Biological constraints on instrumental and classical conditioning: Retrospect and prospect. *Animal Learning and Behavior*, 11, 151-161.
- Fanselow, M. S. (1993). Associations and memories: The role of NMDA receptors and long-term potentiation. *Current Directions in Psychological Science*, 2, 152-156.
- Frey, P. W., & Sears, R. J. (1978). Model of conditioning incorporating the Rescorla-Wagner associative axiom, a dynamic attention rule, and a catastrophe rule. *Psychological Review*, 85, 321-340.
- Griffin, D. R. (1984). *Animal thinking*. Cambridge, N.J.: Harvard University Press.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Clarendon.
- Hinde, R. A. (1973). Constraints on learning: An introduction to the problems. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on learning*. London: Academic Press.
- Hull, C. L. (1943). *Principles of behavior*. New-York: Appleton-Century-Crofts.
- Hull, C. L. (1952). *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven: Yale University Press.
- 石井 澄(1992). 動物の学習行動のモデル論 I : 1970年代以降の認知的モデルの発展とその特質. 名古屋大学文学部研究論集, 哲学38, 123-136.
- 石井 澄 (1993). 動物の学習行動のモデル論 II : 従来の認知的モデルにおける問題点. 名古屋大学文学部研究論集, 哲学39, 103-116.
- Kim, J. J., DeCola, J. P., Landeria-Fernandez, J., & Fanselow, M. S. (1991). N-methyl-D-aspartate receptor antagonist APV blocks acquisition but not expression of fear conditioning. *Behavioral Neuroscience*, 105, 126-133.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Leahey, T. H. (1980). *A history of psychology: Main currents in psychological thought*. Englewood Cliffs, N.J.: Prentice-Hall.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Mackenzie, B. D. (1977). *Behaviourism and the limits of scientific method*. London: RKP.
- Mandler, G. (1985). *Cognitive Psychology: An essay in cognitive science*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Miller, R. R., & Schactman, T. R. (1985). The several roles of context at the time of retrieval. In P. D. Balsam (Ed.), *Context and learning*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Pearce, J. M. (1987). A model of stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61-73. (a)
- Pearce, J. M. (1987). *An Introduction to animal cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates. (b)
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Rescorla, R. A. (1991). Associative relations in instrumental learning: The eighteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 43 (B) , 1-23.



- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.
- Revusky, S. (1985). The general process approach to animal learning. In T. D. Johnston & A. T. Pietrewicz (Eds.), *Issues in the ecological study of learning*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Rozin, P., & Kalat, J. W. (1971). Specific hungers and poison avoidance as adaptive specializations of learning. *Psychological Review*, **78**, 459-486.
- 佐伯 胖 (1982). 「わかること」の心理学. 佐伯 胖 (編) 推論と理解 (認知心理学講座 3). 東京大学出版会.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, **57**, 193-216.
- 戸田 正直 (1984). 認知とは何か. 大山 正・東 洋 (編) 認知と心理学 (認知心理学講座 1). 東京大学出版会.
- 戸田 正直 (1992). 感情——人を動かしている適応プログラム. (認知科学選書24). 東京大学出版会.
- Wagner, A. R. (1969). Stimulus-selection and a "modified continuity theory". In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol. 3*, New York: Academic Press.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Walker, S. (1983). *Animal thought*. London: RKP.
- Watson, J. B. (1925). *Behaviorism*. New York: Norton.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, **18**, 459-482.
- 養老 猛司 (1989). 唯脳論. 青土社.