

# 概念獲得研究の現状と課題

東京大学教育心理学研究室 河 原 哲 雄

## Advances and Issues in Concept Acquisition Research

Tetsuo KAWAHARA

Concept acquisition is one of the major themes in cognitive science. The purpose of this paper is to review the advances and issues of concept acquisition research after mid-80s. We review the prominent studies from human cognitive psychology, machine learning, and conceptual development research, which seem to be important for the explicit modeling of human concept acquisition process.

As for cognitive psychology, impact of theory-based concept view and significance of normative reconstruction of similarity-based concept models are compared. As for machine learning research, explanation-based learning models and connectionist models are introduced, and their applications to human concept acquisition research are explored. As for conceptual development research, studies which hypothesize various constraints are criticized for their lack of computational considerations.

We argue that it is important to make a clear distinction between internal conceptual computation and external concept use, and that modular approaches to concept acquisition modeling are promising.

### 目 次

- I. 概観
- II. 実験心理学と概念獲得モデル
  - A. 「理論ベースの概念観」の反響
  - B. 「類似性ベースの概念モデル」の新展開
- III. 機械学習研究の成果の導入
  - A. 「説明ベースの学習」と「理論ベースの概念」
  - B. コネクショニスト・モデルの展開
- IV. 概念発達研究の現状と問題点
- V. まとめ

### I. 概観

認知心理学の研究論文の冒頭では、そこで取り上げる機能が、人間の認知情報処理全体の中で、いかに重要で中心的かを力説するのがなかば習慣になっている。概念とそれを支えるカテゴリー化の機能に関する研究も、その例外ではない。概念研究の重要性を主張する言明の中で、最も有名でよく引用されるのは、Smith and Medin (1981)<sup>1)</sup>の“概念がなかったとしたら、精神生活は混沌と化してしまうであろう”というものである。しかし概念研究の現状は、ある種の混沌の中にあるように思われ

る。

認知科学の中で、概念の獲得過程を扱ってきた主な研究領域には、1) 実験とシミュレーションを中心とした認知心理学、2) 人工知能研究の1領域である機械学習研究、3) 幼児の概念発達研究、の3つがある。本稿の目標は、これら3つの研究領域における、主に80年代以降の概念獲得研究の動向を概観し、整理することである。しかし現状では、これらの研究全体を包含し得るような統一的な概念獲得モデルは、想像することすら困難である。例えば、理論ベースの概念理論から人工刺激を用いた実験研究への批判に見るように、それぞれの研究領域内部ですら、一見してほとんど両立不可能と思われる概念観を抱く研究者達の間で、激しい論争が進行している。

このような混沌が生じてしまった大きな原因の1つは、概念が中心的な役割を果たすと考えられる認知プロセスが、きわめて多岐にわたることである。Thagard (1992)<sup>2)</sup>は以下の10種のプロセスを挙げている。1) カテゴリー化、2) 学習、3) 記憶、4) 演繹的推論、5) 説明、6) 問題解決、7) 一般化、8) 類推的推論、9) 言語理解、10) 言語産出。これらの中には、意識的な推論が必須であるもの（説明）もあれば、知覚レベルのみで

働き得るもの（記憶）もあり，言語ラベルの存在を前提とするもの（言語理解）もあれば，言語化が困難な場合が多いもの（一般化）もある。現在までのところ，これらすべてのプロセスを扱い得るような概念のモデルは存在していない。本稿で試みる概観では，それぞれの研究パラダイムが，人間の概念というきわめて多面的な問題のどの側面に焦点をあてようとしているのかを出来る限り明確にしつつ，それぞれの準拠枠内での評価と，概念研究の全体像の中での位置付けをきちんと分けて行いたい。

以下の各章では，実験心理学的研究，機械学習研究，概念発達研究の3つの研究領域における近年の研究成果を検討する。ただし，機械学習研究に関しては，認知心理学における理論やモデルとの影響関係に焦点を絞って記述する。いずれの領域についても，論文の選択と評価は，概念獲得プロセスの明示的なモデル化にどれだけ貢献するかを基準とした。

## II. 実験心理学と概念獲得モデル

### A. “理論ベースの概念観”の反響

80年代後半から現在に至る認知心理学における概念研究に，良かれ悪しかれ最大のインパクトを与えたのが，Murphy and Medin (1985)<sup>3)</sup>が提起した“理論ベースの概念 (theory-based concept)”理論であるという主張への異論はほとんどないであろう。この論文以前の概念獲得の研究史については，簡単な背景説明にとどめる。

70年代から80年代前半にかけての実験的な概念獲得研究は，Roschに由来する特徴的特性理論に基づいて行われてきたといっていよい (Rosch, 1973<sup>4)</sup>; Rosch, 1975<sup>5)</sup>; Rosch and Mervis, 1975<sup>6)</sup> など)。この理論は，多くの自然概念には必要十分条件をなす定義的特性などというものはなく，カテゴリー成員の多くに確率的に共有されている特徴的特性があるだけだと考える。この理論は，カテゴリーに含まれる成員の間にカテゴリーへの所属の度合に差があることなど，定義的特性理論では説明が難しい多くの現象を説明することが出来たため，この理論に沿った多くの計算モデルが作られた。代表的な計算モデルに，プロトタイプモデル，範例モデル，特性頻度モデルなどがある。これらのモデルには，カテゴリー毎の要約的表象が作られるか否か，類似性をどのように定義するか，類似性の算出が行われるのは学習時か転移時かなどの，計算アルゴリズムの詳細に相違がある。しかし，概念表象の形成や新事例への一般化が，提示された多数の事例の間で計算された類似性にもとづいて行われると

いう点では共通している。その意味で，これらのモデルをまとめて，類似性ベースの (similarity-based) 概念獲得モデルと呼ぶことが出来る。

理論ベースの概念理論は，こうした類似性ベースの概念モデルが，人間の概念のモデルとして不十分であるという批判から出てきたものである。批判の論点の中で最大のものは，類似性ベースの観点では，概念の“凝集性 (coherence)”を説明できないというものである。概念の凝集性とは，人間が自然に使用しているカテゴリーに見られる，成員の自然なまとまりのことである。

理論ベースの概念理論では，概念獲得を以下のようなプロセスとして考える (Murphy and Medin, 1985)。まず，理論によって重要な特性がピックアップされ，それらが理論によって与えられた説明により組織的に結合しあう。すなわち，多数の特性の中から有意な特性間相関が選択され，そのリンクに理論的説明が付与される。概念の内的凝集性を支えているのは，こうした領域理論に由来する説明的リンクの緊密なネットワークであると考える。

ここでは，Murphy らの論文が類似性ベースの概念モデルへの批判を目的としたものであって，“理論ベースの概念モデル”を明示的に定式化したものではない，ということに注意を促しておく。実際のところ，90年代に入って機械学習の領域から“説明ベースの学習”モデルが導入されるまで，領域理論の内容を具体的に明示して，“理論ベースの概念モデル”に基づく概念獲得プロセスを実験的に検証するような研究は1つも現れなかった。“理論ベースの概念観”は，これからの概念研究が解明していかなければならない問題の所在を明らかにした功績は大きい，人間の概念獲得過程の明示的なモデル構成という点では，“領域理論”の曖昧さゆえの魅力に寄り掛かった不完全な理論であるといわざるをえない。以下では，理論ベースの概念モデルの優位性を支持するとされたきた，実験研究を紹介する。

Wattenmaker et al. (1986)<sup>7)</sup>は，カテゴリー構造の違いによる学習の難しさが，被験者が持ち込む理論とカテゴリー構造の相互作用で決まることを示す実験を行った。彼らの行った実験の1つでは，例えば“金属製—ゴム製”“表面が平ら—でこぼこ”といった2値的属性を用い，これらの属性の組み合わせにより線形分離可能なカテゴリー構造と線形分離不可能なカテゴリー構造を作った。さらに，被験者が持ち込む理論を操作するために，“金槌の代用として使えるものとそうでないもの”というヒントを与える主題あり条件と，与えない主題なし条件を設けた。

結果は、主題あり条件では線形分離可能なカテゴリー構造の方が線形分離不可能なカテゴリー構造よりも学習が容易であったのに対して、主題なし条件ではその逆であった。Wattenmakerらの実験は、線形分離可能性のように重要な構造的属性が学習の容易さにおよぼす効果が、被験者が適用する領域知識と独立には決まらないことを示した点で重要である。先行知識の影響を出来る限り捨象してきたオーソドックスな類似性ベースの概念モデルの枠組みでは、このような相互作用の重要性が理解しにくいのは事実である。しかし、“金槌の代用として使えるかどうか”というヒントを与えるというような、その基準に合う属性値を加算的に考えればよいような場面は、領域理論の適用としてはかなり単純な部類に入るようにも思われる。もっと構造依存的な推論を要する場面を定式化出来ない限り、理論ベースの概念モデル自体を検証したとはいえないであろう。

## B. “類似性ベースの概念モデル”の新展開

80年代の後半から90年代初頭にかけては、それまでに提案された多数の類似性ベースの概念モデルの統合と、理論的観点からの全面的な再定式化が進展した時期でもある。前者の代表的な成果は、Estes (1986)<sup>8)</sup>の配列モデル (array model) である。

配列モデルは、概念獲得の1つの計算モデルというよりも、これまでに提起されてきた異なる計算モデルをその特殊な場合として含むような一種の母型モデルである。すなわち、共通の記憶配列に対して異なる記憶アクセスと決定プロセスを適用することにより、プロトタイプ・モデル、特性頻度モデル、範例モデルといった代表的な計算モデルを統一的に扱うことができることを示したものである。このモデルの出現により、どのモデルとどのモデルの間の比較が、どのような状況で経験的に検証可能であるかといった推論が著しく容易になった。またこれとは逆に、カテゴリー学習研究における係争中の問題を検証するために、どのような計算モデルを用いるべきかといった推論も可能になった。

カテゴリー学習の計算モデルが乱立してしまった原因の1つは、計算理論レベルの問題と、具体的なアーキテクチャへの実装のレベルの問題とがきちんと分離されないままで、いろいろな種類の中途半端なモデル化が行われたためであると思われる。そこで、具体的な実装の問題をひとまず置いておいて、カテゴリー化というものが、そもそもどのような計算目標を実現すべき過程なのかということから考え直す必要が生じてくる。

カテゴリー学習への規範的アプローチとして、近年最

も徹底したものは、Anderson (1990)<sup>9)</sup>の合理論的 (rational) 分析である。合理論的分析では、認知システムの行動が、環境の構造に対して最適化されたものであるという仮定に立つ。そして、システムの目標、環境の統計的構造、計算論的な制約を順に明らかにして、問題となっている認知領域の理論を特定する。カテゴリー化の場合、認知システムの目標は、過去に経験した事例に基づいて、新たな対象の属性に関してできるだけ正確な予測を行うことである。

Andersonの合理論的分析では、まず経験した $n$ 個の対象のあらゆる可能な分割に基づいて、 $n+1$ 番目の対象の特性を推定するベイズ統計学的な意味での最適解を求めている。その上で、あらゆる可能な分割が指数的に増大することを考慮に入れて、既に構築した以外のカテゴリー構造を後から再考することはできないという制約を加えた繰り返しアルゴリズムを定式化している。

Andersonは、この具体的アーキテクチャへの実装の問題を捨象した、純粋に計算論的な分析から生み出された繰り返しアルゴリズムが、プロトタイプや特定事例の効果、基礎レベルカテゴリーの抽出、基準比率の効果、確率マッチングなど、カテゴリー学習にまつわる数多くの実験事実を再現できることを示している。

これらの、あくまで類似性ベースの概念理論の枠内での統合や再定式化の成果に対して、理論ベースの概念理論を奉じる立場からは、どのような反応が返ってくるであろうか。

Andersonの合理論的分析における、人間のカテゴリー構造への制約に関する主張は、世界内の対象が離散集合へと分割可能であり、集合のメンバーが様々な特性を示す共有確率によって定義可能であるというものである。この意味で、Andersonの合理論的分析は、世界に存在している特性相関構造が凝集性の原理であるという、理論ベースの概念観から批判されている発想を、そのまま踏襲しているように見える。

ここで注意しなくてはならないのは、Andersonのアルゴリズムが構築するカテゴリー構造が、通常使用されている語彙ラベルとは必ずしも一致しないことである。合理論的分析では、語彙ラベルはなんら特権的な地位にない。ラベルは予測されるべき多数の特性の中の1つにすぎない。合理論的モデルでは、予測が改善されるならば、システムは実験者の設定したものや自然のカテゴリー・ラベルと一致しないカテゴリーを形成する。

Andersonは、Murphy and Medin (1985)<sup>10)</sup>が挙げたユダヤ教の食習慣に関する例 (ラクダ、ダチョウ、ワニ等が不浄な動物、ガゼル、蛙、バッタが清浄な動物) を、

カテゴリー化とラベルづけの混同であるとして批判している。合理論的分析の観点からは、これは異なるカテゴリーに属する動物が清浄、不浄といったラベルを共有しているに過ぎないのであって、これらの種を包含するカテゴリーが実際に存在するわけではない。

### Ⅲ. 機械学習研究の成果の導入

#### A. “説明ベースの学習”と“理論ベースの概念”

人工知能における概念学習研究では、80年代後半頃から“説明ベースの学習 (explanation-based learning, 以下では EBL)”のパラダイムが注目を集めてきた。それ以前の概念学習モデルの主流は、多数の事例から類似性を抽出することによって概念を形成することを前提としている意味で、“類似性ベースの学習 (similarity-based learning, 以下では SBL)”としてまとめることが出来る。EBL理論では、概念学習において先行知識が果たす役割を重視し、知識が豊富な領域では、1事例からでも新概念の獲得が可能であると考ええる。SBLとEBLの関係は、心理学における類似性ベースの概念と理論ベースの概念の関係によく似ているが、これらを単純に同一視することは出来ない。

EBLの中核をなすのは、説明ベースの一般化とよばれるプロセスである。説明ベースの一般化問題は、2つのステップに分けて解かれる。

第1の“説明”ステップでは、領域理論を用いて訓練事例が目標概念の定義を満足していることを証明する。ある事例がカテゴリーに所属するかどうかは、その所属を領域理論から演繹的に証明できるかどうかで決まると考えているわけである。

第2の“一般化”ステップでは、その下で説明構造が有効であるような、操作性基準を満たす十分条件の集合を決定する。

EBLによる概念獲得モデルは、SBLによるモデルと比べてどのような利点を持っているのであろうか。SBLアプローチには、いくつかの問題点があるとされている。これらの問題点にEBLがどのように対処しているかを、順にみてゆこう。

第1にSBLでは、領域知識の影響を考慮することが難しい。獲得される概念は学習事例セットの形式的構造のみに依存すると考えられており、領域が異なれば学習内容や学習の困難さが異なるという現象を説明できない。EBLでは、領域知識の具体的な内容と量こそが獲得概念の最大の決定因である。第2にSBLでは、説明的情報と非説明的情報の区別をしていない。事例を構成してい

る特性のうち、説明的な関連づけが可能である特性とそうでない特性では、異なる処理が適用される可能性がある。SBLでは、あらゆる特性を非説明的なものとみなして一律に処理してしまう。EBLでは、説明の構築に必要とされる特性のみが処理の対象となる。SBLの第3の問題点は、学習事例セットにたまたま含まれていた無意味な相関に影響されやすいことである。EBLの概念学習モデルでは、目標定義と関係のない特性は除外されるため、このような問題は起こらない。

このように多くの利点を持つEBLのアプローチを、理論ベースの概念理論の基盤に据えようとするのは、ある意味で当然のことである。両者は先行知識の役割を重視し、事例の分類が“説明”を介して行われると考える点で、共通の発想に基づいている。しかし両者の間には重要な相違点もある。

まずEBLでは、“説明”を定理証明に匹敵するような演繹的な論理演算と見なし、“理論”を良く定義されたルールの集合と見なしている点が挙げられる。Murphyらが想定していた“説明”“理論”が、このような完全に整合的なものに限られるとは考えにくい<sup>11)</sup>。人間の、特に素人の推論が形式論理的な合理性に必ずしも従わない事実に関しては、既に多くの実証的研究が蓄積されている。人々の日常的理論は、事実と虚構、科学的原理、ステレオタイプ、インフォーマルな観察体験などの混合物であり、こうした理論を用いて人々が構築する説明は、演繹的というよりは蓋然的な性格を持つものである。

また、EBLでは(SBLの問題点とは逆に)非説明的な情報が単に無視されてしまうという点も問題になる。多くの概念は、本性上必要であるが説明的ではないような重要な特性を含んでいる。例えば“白色”“球形”といった特性は、電球がなぜ発光するかの説明には現れないが、これらは電球を同定する上で重要な特性である(Wisniewsky and Medin, 1991<sup>12)</sup>)。以下では、EBLをいわば不完全な理論ベースの概念獲得モデルとみなしつつ、SBL的な発想をも取り込むことによって、柔軟性の高い概念獲得モデルを構想しようと試みる最近の認知心理学研究を紹介する。

Wisniewsky and Medin (1991)<sup>13)</sup>は、複雑な学習刺激に対するルール帰納課題を用いて、概念獲得におけるEBL的なプロセスとSBL的なプロセスの関係を検討している。彼らの実験の1つでは、子供の描いた人物画のセットを素材として、“創造性の高い子供が描いた絵—非創造的な子供が描いた絵”といったラベルを付加することによって、被験者が帰納する分類ルールがどのよう

に変化するかが調べられた。その結果、それぞれのラベルによって活性化された理論がルール帰納のデータ処理過程に3つの形で影響することが明らかになった。

第1に、異なる理論は訓練項目の異なる特性に選択的に注目させる。第2に、異なる理論は同一データに対する異なる解釈へと導く、第3に、理論は異なる特性を高次の抽象水準では等価と見なさせる。これらの結果は、理論的期待と経験的証拠が密接に相互作用しながら帰納が作用する特性を決定していることを示している。また、特性解釈が学習の重要な一部であり、その過程がトップダウン的知識とボトムアップ的知識の複雑な相互作用を通じて生じていることも示唆している。

Wisniewskyらの研究は、EBL的なプロセスとSBL的なプロセスを概念獲得の不可欠な要素と認めた上で、それぞれの適切な位置付けと相互作用の在り方を明らかにしようとしたものである。しかし、EBLの導入によって、曖昧さを批判されてきた“理論”の内容と作用が十分に明示的になったかという点、必ずしもそういうわけではない。この研究で対象とした子供の人物画のカテゴリーを学習するというような場合に、被験者が直接的に適用可能な領域理論をあらかじめ持っているとは考え難い。Wisniewskyらは、創造性、子供、典型的な人の外観といったスキーマないしフレームに相当する一般的知識構造から創発的に構成されるのであろうと想定しているが、このように柔軟でダイナミックな知識の利用は、少なくとも現存のEBLモデルの範囲外にある。

EBL対SBLという枠組みを人間の概念獲得研究に導入したもう1つの実験研究として、Ahn, Brewer and Mooney (1992)<sup>10)</sup>を紹介する。EBLでは、領域知識が十分にあれば、1事例からでも新概念の獲得が可能であると予想する。Ahnらの実験では、アメリカインディアンのポトラッチ儀式を具体的に記述したテキストを提示して、1事例からのスキーマ獲得がどのような条件の下で行われるかが調べられた。

その結果、スキーマ獲得には、具体的な1事例の記述と背景知識の提供だけでは不十分であり、儀式の目的を説明させるなどの課題を追加することによって、背景知識を積極的に使用するよう誘導する必要があることが分かった。また、事例を記述したテキストを複数提示することも、スキーマの変数に相当する部分を抽出するのに役立ち、スキーマ獲得をある程度促進することが分かった。

これらの実験を通じて、人間が実際にEBL、すなわち1事例からのスキーマ獲得を行う場合があることが確かめられた。しかし、それは純粋なEBLモデルが予測する

ように、常に自動的に起こるものではなかった。また、複数事例の提示の効果も純粋なEBLモデルでは説明できない。

結論として、この研究は人間の概念獲得モデルとして、適用できる知識が少ない時はSBL、十分な知識があり、またその知識を積極的に使用しなくてはならないような状況ではEBLというような複合的なモデルが必要であることを示唆しているといえよう。

この辺りで、EBLと理論ベースの概念モデルの関係や、両者の概念獲得研究における位置付けについて、もう少し整理してみる必要があるだろう。まず、EBLを叩き台とした最近の研究が、理論ベースの概念観の元々の構想のうち、何を実証したのかについて考察する。理論ベースの概念観には、自然概念の凝集性を成り立たせているのは、類似性ではなくて素朴理論の原理や説明モデルであるという含みがある。しかし、ここで紹介した2つの研究は、いずれも類似性に基づく一般化を、概念獲得における重要なプロセスとして認めている。

EBLによる概念獲得を人間が確かに行うことは分かったが、それは十分な背景知識のある領域において、一般化の必要に迫られる状況で、知覚的な類似性の情報と相互作用しながら行われるものだったのである。説明モデルが類似性にとって代わるという、理論ベースの概念理論のより強い方の主張に関しては、何1つとして実証されたわけではない。

もちろん、EBLが人間の概念獲得のきわめて重要なモードの1つであることには疑いを容れない。必要なのは、EBLを、序章でも触れた人間の概念的行動の多彩なスペクトルの中に適切に位置付けることである。そこで考えなくてはならないのは、EBLが想定している学習事態が、どれだけ典型的なのかということである。これまでの研究で見る限り、EBLが起動されるのは、おおむね意識的で、目標志向的な推論が必要とされる場面であると考えられる。そして、EBLプロセス（説明と一般化）が適用されるのは、既に言語化された特性に対してである。言語や意識が関与しなくても概念形成が可能であるという実験的証拠は数多くある。日常的な概念獲得の場面としても、1事例だけの経験から目標志向的な推論をへて新概念を獲得するというのが、それほど典型的な場面であるとは思えない。人間が所有している概念の膨大さを考えれば、その大部分はさしたる推論の負荷もなく、自動的・受動的に形成されると考えるのが自然であろう。その大部分の概念の形成においては、類似性が重要な役割を果たしている可能性が否定されたわけではない。

こうした議論は、EBLや理論ベースの概念獲得がとるに足らないプロセスであることを主張しようとしているわけでは決してない。筆者の主張は、“人々が持つ素朴な領域理論というものは、概念的行動を成り立たせる計算過程の原理としてはたらくものではなく、概念を用いた推論やコミュニケーションの原理としてはたらくものであると考えるべきだ”というものである。そもそも理論ベースの概念理論の出発点となった、概念の凝集性というものの自体が、言語レベルでは、認知主体にとって（推論）も他者にとって（コミュニケーション）も理解しやすい概念でなくてはならないという制約から派生した現象であると考えられる。

### B. コネクショニスト・モデルの展開

この節では、EBLと並んで概念獲得モデル実装の有力なパラダイムであると見なされている、コネクショニスト・モデルの最近の成果を紹介する。とはいえ、コネクショニスト・モデルを用いた概念研究は、草創期の McClelland and Rumelhart (1985)<sup>15)</sup> 以来、最近では概念階層や変数の使用を可能にした Hinton (1988)<sup>16)</sup> まで、数限りなくある。ここでは人工刺激を用いたカテゴリー学習実験の再現という意味で重要な Gluck and Bower (1988)<sup>17)</sup> と、概念獲得とモジュール性の問題を考える上で重要な Schyns (1991)<sup>18)</sup> を紹介するにとどめる。

Gluck and Bower (1988)<sup>19)</sup> は、症候のリストがコネクショニスト・ネットへの入力、病名の診断が出力となるような、医療診断を模したカテゴリー学習実験のシミュレーションを行った。彼らは、隠れユニットがなく、入力層と出力ユニットが直接リンクで結ばれている最も単純なネットワーク構造を採用した。そのため、学習アルゴリズムも誤差逆伝播ではなく、より単純な LMS (Least Mean Square ; 最小平均平方)<sup>20)</sup> 学習規則を用いた。実験では、被験者に4つの症候からなるリストを提示して、その患者が可能性のある2つの病気のうち、どちらに罹っているかを学習させた。この2つの病気は、生起確率をそれぞれ75%（高頻度病）と25%（低頻度病）とした。また、4つの症候のうちの1つは、その症候を所与として一方の病気に罹っている条件付き確率が、高頻度病と低頻度病で等しくなるように設定した。訓練ステージが終了した後で、被験者にそれぞれの症候を単独で持つ患者が、一方の病気に罹っている確率を評定させたところ、問題の症候に関する被験者の推定値は、ベイズの定理で計算される規準値とは異なり、低頻度病に罹っている確率の方が高頻度病に罹っている確率よりも大きくなる傾向があった。GluckらのLMSネットワーク

モデルは、この人間の確率判断のバイアスをうまく再現した。

この結果が重要であるのは、従来のカテゴリー学習モデルでは、ベイズ確率に準拠した確率評定値を予測したり（範例モデル）、基準比率の効果を全く考慮に入れられなかったり（プロトタイプモデル）と、いずれもこの実験の結果を再現できないからである。

LMS ネットワークモデルには、線形分離不可能なカテゴリー構造を原理的に学習できないという大きな限界がある。この限界は、入力層と出力層の間に隠れユニットの層を導入し、LMS 学習規則に代えて、誤差逆伝播（一般化デルタルール）を適用することで克服される。そこで Gluck らは、誤差逆伝播を用いた多層のコネクショニスト・ネットワークでも、LMS ネットワークの場合と同様の結果が得られるだろうとの希望的観測を述べている。しかし、Shanks (1990)<sup>21)</sup> はこの期待が裏切られることを確認した。3層のネットワークに誤差逆伝播を適用した結果は、ベイズの規準的確率を正確に再現し、したがって人間の確率判断に特有のバイアスを再現することには失敗してしまったのである。人間のカテゴリー学習の実験データを再現する上では、能力の低い LMS モデルの方が、能力の高い誤差逆伝播モデルよりも有能であったことになる。

誤差逆伝播にはこれ以外にも、学習速度がきわめて遅く、実験における1事例の提示とシミュレーションの1サイクルを同一視することが出来ない、局所最小に陥る危険があるなどの問題点がある。

Gluck らが用いた LMS ネットワークモデルは、カテゴリー学習実験の結果をシミュレートするという点では優秀であったが、人間の概念獲得過程全体のモデルとしては、しよせん限界のはっきりしているものであった。これと比較して、Schyns (1992)<sup>22)</sup> のモデルは、より広範囲の現象を説明し得る強力なモデルである。

Schyns のモデル（以下では、モジュールモデルと呼ぶ）の特徴は、カテゴリー化と命名（naming）のプロセスを、それぞれ別々の機能モジュールに分離している点である。カテゴリー化モジュールは、教師なし学習アルゴリズムで動作する。具体的には、Kononen (1984)<sup>23)</sup> のトポロジカルマップを若干変形した自己組織化アルゴリズムにより、入力刺激を2次元マップ上の活性化パターンへと写像する。お互いに類似した入力刺激は、マップ上の近接した領域を活性化する。こうして、入力刺激の構造に対応して、マップ上で活性化される領域が次第に分割されてゆくのがカテゴリー化の過程である。

命名モジュールは、教師付き学習アルゴリズムによ

り、カテゴリー化モジュールが生成した内的な概念表現と、カテゴリー名（言語ラベル）の連合を行う。

こうした、機能モジュールごと分割した上で接続するアーキテクチャには、2つの計算論上の利点がある。第1は、誤差逆伝播で深刻な問題になるスケールアップ問題（Rumelhart, 1990）<sup>24)</sup>が大幅に軽減されることである。第2は、複数のモジュールから共通に利用可能な、課題独立な概念表現が生成される可能性があることである。もちろん、仮にモジュール的なアーキテクチャの採用が正しいとしても、このモデルが採用したようなモジュール分割の仕方が最適であるかどうかには、大いに議論の余地がある。Schynsはカテゴリー化と命名を分離する根拠の1つとして、親による正しい語彙ラベルの使用が、カテゴリーを区別する特性を子供が自発的に発見した後になる場合もあるという、語彙発達上の事実を挙げている。

Schynsのモジュールモデルは、プロトタイプ効果、相互排他性制約（Markman and Watchel, 1988）<sup>25)</sup>、語彙ラベルの効果、カテゴリー階層の構築など、広範囲にわたる現象の説明に成功している。このように、モジュール的なコネクショニストモデルは、これからの研究が期待されるきわめて有望なアプローチであると思われる。Schynsのモデルの現状での限界は、入力刺激の構成に心理学的な根拠がなく、心理学実験や概念発達データの再現ではない、単なる計算機実験に過ぎない点である。

最後に、概念獲得のコネクショニストモデルの、現状での限界について検討する。Thagard (1992)<sup>26)</sup>は、コネクショニストモデルは以下の現象を説明できなければ完全な概念理論とは言えないと指摘している。1) 事例からでなく既存概念の組み合わせによる新概念の構成、2) 説明、問題解決、演繹のような系列のプロセスにおける概念の使用、3) 複雑な一般化や類推を含む広範な推論課題において分散表象が果たす役割、4) 言語生成・理解の一般理論において分散表象が果たす役割。Thagardはまた、“パーティでプールに飛び込んだ男を酔っぱらいとして分類する”のは、将来のコネクショニストモデルが直面すべき難問であると述べている。分散表象を用いたコネクショニストモデルでこのような説明的推論が可能かどうかを検討するのは、工学的観点からも心理学的観点からも確かに興味深い作業である。

#### IV. 概念発達研究の現状と問題点

幼児の概念発達研究は、それだけに関して長大なレビューを必要とする、きわめて多くの研究が蓄積されて

いる領域である。しかし、今までのところ概念発達における“理論”の研究では、EBLの導入などによる形式化の努力がなされていないこともあり、以下では、概念獲得の計算モデルを目指す上で参照すべきであると思われる、語彙獲得における制約の問題を検討するととどめる。

語彙獲得メカニズムに関する80年代以降の研究は、制約（constraint）の概念を中心に動いてきたといえよう（Carey, 1988<sup>27)</sup>; Markman, 1990<sup>28)</sup>）。幼児の語彙獲得は、1才の半ば頃から6才頃までに、1日に約8から10語というおそるべき速さで進む（Carey, 1978<sup>29)</sup>）。おそるべきというのは、特にその初期段階では、語彙獲得というのがとうてい解決不可能な真正の帰納問題であると思われるからである。子供が新たな語を聞いた時、その語が指す可能性のあるもの（事物、様態、属性）は原理的には無限にあるはずなのに、子供はなぜ正しい意味を推測することができるのだろうか。生得的な制約があるために、子供はこれらすべての可能性を考慮するのではなく、ほんのわずかの仮説しか考慮する必要がないのだ、というのが制約説の回答である。

具体的に提案された制約の中でおそらく最も有名であり、論争的にもなってきたのが、Markmanの相互排他性（mutual exclusivity）<sup>30)</sup>制約である。この制約は、簡単にいうと、“子供は最初、事物がそれぞれ唯一のラベルを持つと仮定することによって語の意味を制約する”というものである。この制約を厳密に適用すると、1つの語彙領域の中の2つの語の外延は重なり合わないというきわめて強い主張になる。それでは概念の階層や別名のようなものがまったく習得できなくなってしまうので、実際にはこの制約は悉皆的に働くものではなく、3才頃までには廃棄されると想定されている。

語彙獲得においてはたらく制約として提案されてきたものには、これ以外にも、相互排他性原理の一般化版と考えられるClark (1986)<sup>31)</sup>の対比原理や、“子供の最初の仮説は、語は事物の分類カテゴリーを指すというものである”という分類仮説（taxonomic assumption）（Markman and Hutchinson, 1984）<sup>32)</sup>などがあるが、概念獲得プロセスの明示的なモデル化という本稿の目的に資するような研究はというと、ほとんど見当たらない。あるのは、個別の制約を提案し、それを擁護する実験を行うといった記述的な研究ばかりである。“子供が直面する帰納問題を回避するためには、何らかの制約が必要である”という設定は一見して説得力があるのだが、そもそも語彙獲得というのがいかなる計算プロセスであり、どれだけの計算論上の制約を導入すれば解決可能になる

のかという計算論レベルでの考察が欠如しているため、いったいいくつの制約をどのように組み合わせれば目的が達せられるのかが一向に明らかにならないのである。

既に触れたように、Schyns (1992)<sup>33)</sup>は、彼のモジュール的なコネクショニストモデルが、相互排他性制約の現象を再現することができたと主張している。重要なのは、このモデルでは相互排他性と階層概念の獲得という一見して対立する現象が、経験の蓄積に伴う一連の経過として再現可能であり、相互排他性制約が一定の年齢で放棄されるというようなアドホックな仮定を導入する必要がないことである。計算モデルの設定の仕方によっては、制約の必要性そのものが雲散霧消してしまう可能性があるということは、真剣に受け止めておく必要がある。

## V. まとめ

本稿では、80年代後半以降の概念獲得研究の中から、特に概念獲得のモデル化のこれからの動向を考える上で重要であると思われるものを中心に検討した。概念の問題をめぐる最近のレビュー論文と比較すると、理論ベースの概念理論に対して批判的な内容が多かったと思われる。だが、それはこの理論がとるに足らない劣ったものであるということを主張しているのではなく、人々の素朴な理論という原理が、これまでに現れてきた他の原理と同様に、概念獲得研究の混沌を永遠に封じ込める黄帝の鏡ではないことを確認しているに過ぎない。

本稿で扱った3つの研究領域（実験心理学、機械学習、概念獲得研究）間の対話は始まったばかりで、まだまだ不十分である。したがって本稿の全体としての印象が、混沌に混沌を重ねるものであるのは、ある意味では当然である。にもかかわらず、ここでの検討作業を通じて、考慮を要すべき共通の問題も浮かび上がってきたように思われる。

第1は、“概念を支える内部計算”と“概念の使用”の混同の問題である。一方で、概念的行動を成り立たせるのに必要な計算プロセスを考える計算論的見地からは、言語ラベルは他の数多くの特性とまったく同じように処理される、1つの特性にすぎない。他方で、人間の概念的行動を外から観察したり内観したりした場合、言語ラベルは推論や伝達の単位として特権的な位置にある。両レベルの混同が、概念モデルをめぐる議論を著しく見通しの悪いものにしてきたのは否めない。

Andersonの合理論的分析は、今のところ“概念を支える内部計算”のレベルだけを対象にしている。しか

し、本質的に“概念の使用”のレベルに属する理論ベースの概念理論の問題にこれを適用してみるのも興味深い。合理論的分析では、理論の使用によって新事例の特性に関する予測力が高まるならば、理論が使用されると予測する。ここで問題は、“概念の使用”レベルでは、特性の予測以外に、コミュニケーションの容易さや注意資源の制約といった要因が関わってくることである。したがって、理論の使用は必ずしも合理論的分析の予測には従わないものと予測される。

1つの解釈では、理論ベースの概念観というのは、“概念の使用”レベルでの原則が、“概念を支える内部計算”のレベルにも貫徹しているという主張であるといえよう。概念の使用が埋め込まれている活動や文脈とは独立に概念表象や類似性を考えても意味がないという主張などがこれにあたる。SBLプロセスとEBLプロセスが相互に密接に編み込まれているというWisniewskiらの議論は、この主張をやや弱めたものである。

第2の問題は、概念獲得モデルにおけるモジュール性の問題である。ただし第1の問題と決して独立ではない。カテゴリー化モジュールと命名モジュールを分離したSchyns (1991)<sup>34)</sup>のアプローチは、両方の問題を同時に解決する1つの試みである。ただし第1の問題をうまく解決しているように見えるのは、“概念の使用”の部分が、最も単純な命名の処理で終わってしまっているからである。特性間の複雑な（理論的）推論をも可能にするような内部表現を持つコネクショニストモデルは、まだ現れていない。また、機能モジュールの分割の仕方も、これ以外のあらゆる可能性について、これから検討を重ねてゆく必要がある。機能だけでなく、内容領域によっても別個のモジュールを想定しなければならないのかもしれない。もしかすると概念獲得における生得的制約というのは、脳内の機能モジュールの分業と、その配置関係に由来しているのかもしれない。こうした問題はすべて、これからの検討課題である。

## 注・引用文献

- 1) Smith, E.E., and Medin, D.L. 1981 Categories and concepts. Cambridge, MA: Harvard University Press
- 2) Thagard, P. 1992 Conceptual revolutions. Princeton, NJ: Princeton University Press
- 3) Murphy, G.L., and Medin, D.L. 1985 The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- 4) Rosch, E. 1973 On the internal structure of perceptual and semantic categories. In Moore, T.E. (Ed.) *Cognitive development and the acquisition of language*. New York: Academic Press, 111-114.



- 5) Rosch, E. 1975 Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- 6) Rosch, E., and Mervis, C.B. 1975 Family resemblance : studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- 7) Wattenmaker, W.D., Dewey, G.I., Murphy, T.D., and Medin, D.L. 1986 Context, relational properties, and concept naturalness. *Cognitive Psychology*, 18, 158-194.
- 8) Estes, W.K. 1986 Array models for category learning. *Cognitive Psychology*, 18, 500-549.
- 9) Anderson, J.R. 1990 The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 10) Murphy, G.L., and Medin, D.L. 1985 op.cit.
- 11) Murphy and Medin (1985) は、彼らのいう“理論”について、“完全に組織された科学的説明というよりも、あらゆる形での心的な“説明”であり、因果的知識、スクリプト、ルールなどを含む”と述べている。
- 12) Wisniewski, E.J., and Medin, D.L. 1991 Harpoons and long sticks: The interaction of theory and similarity in rule induction. In Fisher, D.H., Pazzani, M.J., and Langley, P. (Eds.), *Concept formation: Knowledge and experience in unsupervised learning*, Palo Alto, CA: Morgan Kaufman
- 13) Wisniewski, E.J., and Medin, D.L. 1991 op.cit.
- 14) Ahn, W., Brewer, W.F., and Mooney, R.J. 1992 Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 391-412.
- 15) McClelland, J.L., and Rumelhart, D.E. 1986 Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-188.
- 16) Hinton, G.E. 1988 Representing part-whole hierarchies in connectionist networks. Tech. Rep. CRG-TR-88-2. University of Toronto, Department of Computer science
- 17) Gluck, M.A. and Bower, G.H. 1988 From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.
- 18) Schyns, P.G. 1991 A modular neural network model of concept acquisition. *Cognitive Science*, 15, 461-508.
- 19) Gluck, M.A. and Bower, G.H. 1988 op.cit.
- 20) この学習規則は、ネットワークの現在の出力と目標出力の差異を漸近的に最小化するという意味で、デルタルール (delta rule) とも呼ばれる。
- 21) Shanks, D. 1990 Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology*, 42A, 209-237.
- 22) Schyns, P.G. 1991 op.cit.
- 23) Kohonen, T. 1984 *Self-organization and associative memory*. Berlin: Springer-Verlag.
- 24) Rumelhart, D.E. 1990 The architecture of mind: A connectionist approach. In M.I. Posner (Ed.) *Foundations of cognitive science*. Cambridge, MA: Bradford Books
- 25) Markman, E.M., and Watchel, G.F. 1988 Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1-27.
- 26) Thagard, P. 1992 op.cit.
- 27) Carey, S. 1988 Lexical Development -- the Rockefeller years. In W. Hirst (Ed.) *The making of cognitive science*. Cambridge, MA: Cambridge University Press.
- 28) Markman, E.M. 1990 Constraints children place on word meaning. *Cognitive Science*, 14, 57-77.
- 29) Carey S. 1978 The child as word learner. in M. Halle, J. Bresnan, and G.A. Miller (Eds.) *Linguistic theory and psychological reality*, Cambridge, MA: MIT Press
- 30) Markman, E.M., and Watchel, G.F. 1988 op.cit.
- 31) Clark, E.V. 1986 The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.) *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- 32) Markman, E.M. and Hutchison, J.E. 1984 Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, 16, 1-27.
- 33) Schyns, P.G. 1991 op.cit.
- 34) Schyns, P.G. 1991 op.cit.