

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation



松香 敏彦 (Toshihiko MATSUKA, Ph. D.)

千葉大学 文学部 准教授

(Chiba University, Faculty of Letters, Associate Professor)

Psychonomics Society, 日本認知科学会、日本人工知能学会、日本認知心理学会 会員

研究専門分野：認知科学、計量心理学、認知モデル

**あらまし** 人間はあらゆる種の情報をカテゴリ化している。カテゴリ化は推論・予測・理由付けなど高度な思考を可能としている。このことから、カテゴリがどのように我々の頭の中で表象・形成され、どのようにカテゴリ化が行われているかは、認知科学・心理学の分野において重要な研究対象とされ様々なモデルが提唱されてきた。既存のモデルは大きく、規則モデル、典型モデル、事例モデルの3つに分類でき、ほとんどの全てのモデルは排他的な見解を示してきた。しかし、最近の研究結果の知見から、これら3つの主要のモデルの妥当性はカテゴリの構造・目的によって変化するという考えが広まってきた。このことから本研究では、3つの主要モデルを包含する SUPERSET モデルを提唱した。既存のモデルにない SUPERSET の特徴は、単に異なるカテゴリの表象を可能にするだけでなく、多様な複数の表象を同時に保持することができることである。3つのシミュレーションの結果、SUPERSET は柔軟で多様な表象が可能であることが示された。

## 1. はじめに

我々人間は自然にそしてほぼ自動的にあらゆる種の情報をカテゴリ化している。例えば、通常我々が「トリ」を見た時に、「クチバシを持ち、羽と足がある飛ぶことのできる」動物など、その物体を個々の特徴の集合として認識するのではなく、多くの場合単に「トリ」と認識する。

このようなカテゴリ化によって抽象化された情報は高次認知における重要な知識ユニットであり、よってカテゴリ化は推論・予測・理由付けなど人間の高度な思考を可能にしている主要なメカニズムと考えられている (e.g. Murphy, 2002 [17])。また、カテゴリ化とは無数の次元からなる刺激の抽象化であることから、一種のデータ圧縮とも考えられる。しかも、単なるデータ圧縮ではなく、高速かつ汎用・予測力の高いデータ圧縮であると同時に、主成分分析などとは違い、全ての圧縮された情報 (e.g. カテゴリ) は解釈可能な出力・分析結果を導く点においても特徴的な性質をもっている。

このように、カテゴリ化人間の認知処理に重要な役割を担っており、これまでに数多くの研究者・科学者の興味を引きつけてきた。特に、カテゴリがどのように我々の頭の中で表象・形成され、どのようにカテゴリ化が行われているかは、古代ギリシャ時代から、知的欲求の対象とされてきた。本論文では、一連のカテゴリの内部表象および知識構造に関するモデル研究を紹介する。

まず、カテゴリの内部表象 (人間の頭の中でのカテゴリの表象・知識) についての3つ主なモデルとそれらの問題の部分的な克服を試みた拡張モデルを説明する。次に既存の主要モデルを包含する統合モデルを提案する。

内部表象の説明にあたり、用語を整理しておく。これまでの認知科学の慣例にならい、あるまとまりのある個々の集合を「カテゴリ」とよび、カテゴリの内部表象およびカテゴリに対する知識を「概念」と呼ぶ。

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

## 2. 既存のカテゴリ表象のモデル

### 2.1 規則モデル

概念のモデルで最も早くに提唱されたのが規則モデルである。規則モデルでは、カテゴリは必要十分条件で定義されていて、事象がカテゴリに属するか否か一意的に決定されると考えられている。この考えはアリストテレスの時代から1960年代後半まで、一般的に受け入れられていた (Murphy, 2002 [17])。カテゴリが必要十分条件で定義され決定論的な表象となり、演繹的推論を可能にする概念である。決定論的な表象は曖昧な表現や複数の解釈を抑制でき、厳密な表現を必要とする科学的な思想に適合している。このことが、このモデルを妥当なものとして長らく受け入れられていた理由の1つであろう。事実、科学の世界においては、カテゴリや集合を人工的に定義することによって発展してきた。

しかし、規則モデルの記述妥当性 (人間の自然なカテゴリの認知プロセスを正しく記述できているか否か) が1960年代後半から疑問視されてきた。まず、挙げられたのが、カテゴリ C の全ての事象が条件 R を満たし、かつ、その条件 R を満たすことのないカテゴリ C に属さない事象が1つもない、といった必要十分条件を全てのカテゴリにおいて定義が可能であるのかといった点である。例えば、全ての犬がその条件を満たすが、同時にその条件を満たす犬以外のものが存在しない、といった条件を定義することはできるだろうか? 非常に難しいであろう。犬1つのカテゴリの規則を定義するだけでも十分困難であるのに、全てカテゴリにおいて重複しない必要十分条件を定義するのは、現実的には不可能であろう。実際に、様々な行動実験の結果から、多くの自然カテゴリにおいて、必要十分条件を定義することが不可能もしくは非常に困難であることが示された (e.g. McCloskey & Glucksberg, 1978)。

規則モデルは必要十分条件が基礎となり決定論的であるため、同一のカテゴリ内では全てのメンバーは均一に適切なメンバーである。よって、より適切なメンバーは存在しない。しかし、行動実験の結果、同一のカテゴリに属するメンバーであっても、認識される速さの違いや、カテゴリの例として挙げられる頻度に差

があるなど、カテゴリ内の典型性効果が示され (e.g. スズメはペンギンより早くトリとして認識され、トリの具体例としてスズメがペンギンより高い頻度で挙げられた)、規則モデルの問題点が指摘された (Batting & Mongatue, 1969, Rips, Shoben & Smith 1973)。

最後に、必要十分条件が簡単に定義可能なカテゴリがあったとしても、実生活における我々のカテゴリ化は必ずしも決定論的ではない例を示しておく。例えば、概ね三角形であるが2つの辺が隣接してはいるものの接していないような幾何学図形や、概ね三角形であるがある頂点が丸みを帯びているものは、厳密には三角形ではない。しかし、多くの人間は、三角形の幾何学的厳密性が問われることの少ない実生活においては、それらの図形を三角形と認識するのではないだろうか。このように、規則モデルは決定論的で演繹的推論を可能し高度な思考を可能にするものの、決定論的表象で人間の自然なカテゴリに関する認知処理を説明するのは不可能である。

### 2.2 典型モデル

先述のカテゴリの典型性効果から、カテゴリは典型例によって表象されているといった説が提唱された (Rosch & Mervis, 1975 [20])。典型モデルでは、入力刺激は、様々なカテゴリの典型例との心理的類似性によって確率的にカテゴリの属性が判断される。つまり、典型例に近いメンバーほど、高い確率でその典型に対応するカテゴリのメンバーとして判断され、心理的類似性の減少に伴い確率は減少する。よって、典型モデルは規則モデルでは説明不可能であった典型性効果を説明することが可能である。なお、典型モデルと後述の事例モデルは、類似性による確率的な表象であるため帰納的推論を可能とするものの、規則モデルとは異なり演繹的推論の妥当性を保証するものではない。

しかし、典型モデルも提唱後比較的早くから問題点が指摘されてきた (e.g. Medin & Schafer, 1978)。まず、挙げられるのがこのモデルでは、概念が「典型例」といった心理的スペース内でのある「点」で表象されていて、カテゴリのバリエーション (分散) やカテゴリ内での特徴間の相関に対する知識、つまりカテゴリの統計的特性 (分散や相関) に関する情報を持ち得な

# 柔軟で多様な概念形成モデルの構築

## A cognitive model for flexible and diverse concept formation

いことである (Medin, Ross & Markman, 2005 [15])。典型例のみによる表象では、「大きなトリはさえずらない傾向があるが、小さなトリはさえずる傾向がある」といったカテゴリ内での特徴間の相関に関する推論は不可能である。また、同様に点による表象なため、(心理的空間で) 線形分離が可能であるカテゴリのみしか表象できない。つまり、排他的論理和 (exclusive OR) などのような比較的シンプルだが線形分離が不可能なロジックを表象することができない。

Medin と Schaffer (1978)の行動実験の結果は典型モデルの反証として頻繁にあげられている。この実験の結果では、典型例に近いはずの事例が典型例に遠いはずの事例より正しくカテゴリ化される確率が有為に低かった。この現象の説明は諸説あるが、典型例から遠い事例は、実はカテゴリ化に有用な特徴をもっており、その特徴に選択的注意が向けられたため、典型例に近いがその特徴を有していない事例よりも正しくカテゴリ化がされた可能性が定量的に示唆された (Matsuka, Corter & Markman, 2010 [7])。つまり、より正しくカテゴリ化された事例は「典型的」ではなかったが、カテゴリの統計的特性を踏まえると、より「特徴的」な事例であり、カテゴリの統計的特性に関する情報を持ちえない典型モデルではこの実験結果を再現することは不可能である。

### 2. 3 事例モデル

典型モデルの問題点を踏まえ提唱されたのが事例モデルである。このモデルでは、入力刺激は記憶に残る今まで経験してきた様々な事例との心理的類似性によってカテゴリ判断される。規則モデルは必要十分条件、典型モデルでは平均値などの代表値の探索・算出が必要であるのに対し、事例モデルでは、記憶の処理が複雑にはなるものの「単に事例を記憶する」といったもので表象に必要な認知処理は最もシンプルである。

記憶に残る全ての事例を参照することによって、概念の整理をすることなく、各カテゴリの統計的特性を推論することを可能とした。また、その統計的特性から典型例を理解し、よって典型モデル同様に、しかし平均値の計算をおこなうことなく、カテゴリの典型性効果を説明することを可能とした。また、記憶された

事例を基準とするため、線形分離できない複数のカテゴリの表象が可能である。

事例モデルに対する問題点は、前述の2つのモデルよりも致命的では無いが、少なからずある。1つは、事例モデル自体の問題点ではないが、事例モデルの妥当性を示した研究の多くは、規則モデルや典型モデルの反証の実験であり、内的妥当性 (実験自体で操作されている因果関係は妥当ではある。つまり、事例モデルの表象は起こりうる) はあるものの、外的妥当性 (実験で想定している状況が実際に起こりうるか) に疑問が残ることである (Blair & Homa, 2003 [2])。また、事例の記憶についての問題点として、非例外的事例に比べ例外的事例 (ほ乳類におけるコウモリなど) はより正確に記憶されることが行動実験で示されたが、事例モデルでは非例外的事例も例外的事例も同等のレベルで記憶されるれ行動実験の結果を説明できないことが挙げられる (Palmeri & Nosofsky, 1995)。

#### 2. 3. 1 事例モデルの数理モデル

概念に関する主要3モデルのうち現時点でもっとも広く受けられているのが、前述の事例モデルである。ここで、事例モデルで最も広く用いられている計算モデルの1つである ALCOVE (Kruschke, 1993) を簡単に説明する。

ALCOVE では、カテゴリ化される  $I$  次元からなる入力刺激  $x$  と記憶されている  $j$  番目事例  $R_j$  の心理的距離  $d_j$  が以下の式で計算される：

$$d_j(x) = \sum_i a_i (R_{ji} - x_i)^2 \quad (1)$$

ここで  $a_i$  は特徴次元  $i$  に向けられた選択的注意で、多くの選択的注意が向けられた場合 (値が高い場合) その次元の特徴は強調され、僅かな違いであっても認識することが可能であり、逆に選択的注意が向けられない場合、その次元にどのような違いがあっても認識されない。

この入力刺激と事例の心理的距離をもとに、その2つの心理的類似性  $s$  が以下の式で決まる。

$$s_j(x) = \exp(-\beta \cdot d_j) \quad (2)$$

# 柔軟で多様な概念形成モデルの構築

## A cognitive model for flexible and diverse concept formation

ここで  $\beta$  は刺激スペース全体の類似性勾配で実験者が定義するパラメータである。

次に心理的類似性をもとにカテゴリ  $k$  の活性度が決まる。

$$O_k(x) = \sum_j w_{kj} s_j \quad (3)$$

ここで、 $w_{kj}$  は事例  $j$  とカテゴリ  $k$  の関連性の重みであり、カテゴリ  $k$  と事例  $j$  の関連性が高いほど正の大きな値をとる。例えば、ある入力刺激が過去に見たカテゴリ  $Z$  に属する多くの事例に心理的に近い場合、つまり、その入力刺激とそれらの事例の心理的距離が短い場合、 $Z$  の事例群と心理的類似性が高いということになり、類似性の高い事例はカテゴリ  $Z$  を集合的に活性化化する。

なお、ALCOVE においては、選択的注意の  $a$  および事例・カテゴリ間の関連重みの  $w$  が（教師あり）学習の対象となっており、勾配法によって間違ったカテゴリ化を最小化するように、重要な特徴次元に選択的注意を向け、また事例と適切なカテゴリを紐づけるように設計されている。

入力刺激がカテゴリ  $Z$  と判断される確率は、カテゴリ  $Z$  の相対的活性度で決まる。

$$P(Z|x) = \frac{\exp(\phi_{O_z}(x))}{\sum_k \exp(\phi_{O_k}(x))} \quad (4)$$

ここで  $\phi$  は決定率であり、その数値が大きければ、より極端な決定を下す。

なお、典型モデルでは、式 1 の  $R_j$  が典型で置換えられている。一般的な典型モデルでは各カテゴリに 1 つの典型しかないため（つまり、1 つの  $R_j$  しか存在しない）、非常にコンパクトな表象となる。

### 2. 4 規則モデルと典型モデルの発展形

規則モデルと典型モデルに対する事例モデルの優位性は行動実験や計算機シミュレーションによって示されてきたが、規則モデルと典型モデルを発展しそれぞれの問題点を部分的にはあるが克服しているモデルも提唱されている。

#### 2. 4. 1 RULEX モデル

RULEX とは RULe plus EXception の略で、その名前が示すようにこのモデルではカテゴリは規則 (rule) と例外 (exception) によって表象される (Palmeri & Nosofsky, 1995 [18])。例外を「例外」という独自の規則として表象することによって、カテゴリ全体に当てはまる必要十分条件の定義を不要とし、おおよそ当てはまる規則とその規則には当てはまらない例外で、決定論的なカテゴリ化を可能にした。一般的な規則モデルより、簡素な表象を可能としたが、決定論的であるため、依然カテゴリの典型性効果の説明は不可能である。

#### 2. 4. 2 OEDIPUS モデル

OEDIPUS とは Matsuka et al. (Matsuka, 2006 [5]; Matsuka, Nickerson, Jian, 2006 [11]; 松香・本田・吉川, 2009 [10]) が提唱した拡張型典型モデルである。OEDIPUS では各カテゴリに独自の選択的注意の演算子があるとし、各カテゴリの統計的特性を学習する機能を付け加えたものである (詳細は松香・本田・吉川, 2009 [10])。具体的には、カテゴリ  $k$  の典型例と入力刺激の心理的距離に多変数間の相関に基づくマハラノビス距離を用いている。

$$d_k(x) = \sum_i \sum_m a_{jim} (R_{ki} - x_i)(R_{km} - x_m) \quad (5)$$

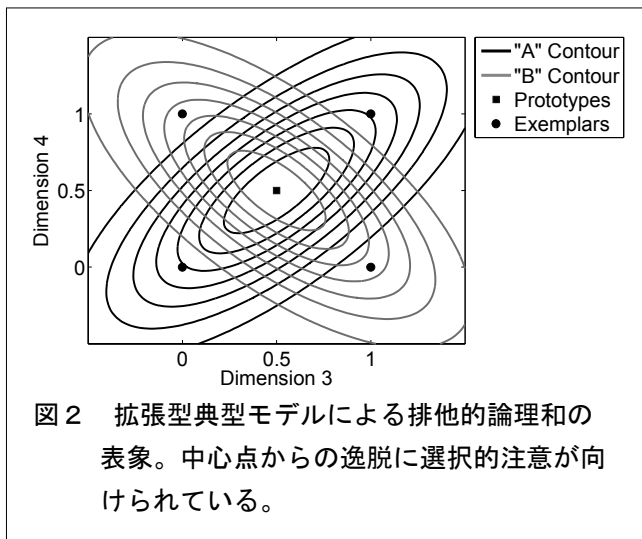
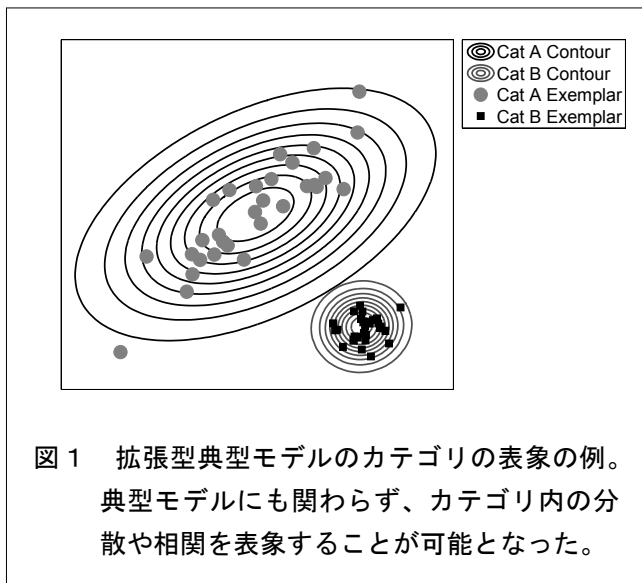
ここで  $a$  は選択的注意であるが、添字が示すようにカテゴリ毎に独自の値をとり、また、2 つの次元  $i$  と  $m$  の相関にも選択的注意が向けられているとした。 $a$  の絶対値は注意の強さを表し、符号は注意の方向性 (相関のみ) を表している。

このような拡張を適用することによって、OEDIPUS は、図 1 にあるように、典型モデルであるにも関わらず、カテゴリ内の分散や相関を表象することを可能にした。また、図 2 にあるように、線形分離不可能な排他的論理和の構造をもつ 2 つのカテゴリ、つまりロジック (「真」カテゴリと「偽」カテゴリ) を学ぶことを可能にした。OEDIPUS は中心点からの逸脱に選択的注意を向け、各カテゴリの独自の広がりを実算し排他的論理和を学習する。また、Medin と Schaffer が典型モデルの反証として示した「典型的で

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

はないが特徴的である事例は、典型的ではあるが特徴的でない事例より正しくカテゴリ化される」といった現象も、OEDIPUS はカテゴリの統計的特性を参照することによって再現することを可能とした (Matsuka, 2006 [5], 松香・本田・吉川, 2009 [10])。しかし、OEDIPUS は 2 変数の関係までにはしか選択的注意が払われないため、3 変数の交互作用をもつようなカテゴリの表象は現状では不可能である (松香・本田・吉川, 2009 [10])。



## 2. 5 モデルの比較・妥当性

ここまでに紹介した 5 つのモデルを比較したものが表 1 である。規則モデルと典型モデルに対する事例モデルの有意性は行動実験や計算機シミュレーションによって示されてきたが、先ほど述べた通り、外的妥当性には疑問が残る。また、拡張型規則モデルの RULEX は事例モデルでは説明不可能であった、記憶における例外事例の優位性が説明可能とした。同様に、拡張型典型モデルの OEDIPUS は事例モデルの有意性を示してきた現象 (典型的ではないが特徴的な事例の正しいカテゴリ化、非線形分離カテゴリの学習、カテゴリの統計的特性の理解など) を再現可能とし典型モデルに対する事例モデルの有意性を非常に弱めるものとした。では、どのモデルが最も適切なのであろうか。1960 年代からごく最近まで、どのモデル・学説が正しいか、様々な研究がおこなわれてきたが、カテゴリの構造や目的などを超越した普遍的な単一の表象形式はなく、状況に応じて適切な表象が適用・獲得されているといった考えが徐々に広まっている (e.g. Medin, Ross & Markman, 2005 [15]; Murphy, 2002 [17])。

そこで、本研究課題で、ここで紹介した 5 つのモデルを包含する SUPERSET モデルを構築した。SUPERSET は、状況に応じて 5 つのモデルに対応する表象や、それらの混合的表象を獲得するモデルである。

## 3. SUPERSET モデルの提案

規則モデル、典型モデル、事例モデル、拡張型規則モデル (RULEX)、拡張型典型モデル (OEDIPUS) の全てを包含するためには、これら 5 つのモデルで最も複雑な認知処理を組み込む必要がある。最も複雑な表象の参照点 (R) は記憶されている事例で、最も複雑な注意の処理は OEDIPUS の統計的特性へ向けられた選択的注意である。つまり、5 つのモデルで最も複雑な認知処理を組み込んだ結果、参照事例  $j$  と入力刺激  $x$  の心理的距離は以下の通りとなる：

$$d_j^n(x) = \sum_i \alpha_{ji}^n (R_j^n - x_i)^2 + \sum_{i=1}^{I-1} \sum_{m=i+1}^I 2C_{jim}^n (R_{ji}^n - x_i)(R_{jm}^n - x_m) \quad (6)$$

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

$$\alpha_{ji}^n = \frac{1}{1 + \exp(-D_{ji}^n)} \quad (7)$$

ここで  $D_{ji}$  は事例  $j$  と入力刺激の比較時に特徴次元  $i$  に、 $C_{jim}$  は特徴次元  $i$  と  $m$  との相関に向けられる選択的注意である。なお、 $C_{jim} = C_{jmi}$ 、 $|C_{jim}| \leq (\alpha_{ji} \alpha_{jm})^{1/2}$  といった制約がある。特徴次元に向けられる選択的注意は  $D_{ji}$  は式 7 を介して、0~1 の値 ( $D$  が  $+\infty$  の場合 1、 $-\infty$  の場合 0 に) をとるものとした。SUPERSET では学習の対象となる選択的注意の重み  $D$  と  $C$  は確率的最適化法を用いているため、 $D$  の値では振幅が起りやすい。実際に向けられる注意量  $\alpha$  は、式 7 を用いることよっての振幅が抑制されている。

詳細は後述するが、SUPERSET では同一のカテゴリに対し同時に複数の表象をもつと仮定しており、各表象は添字  $n$  によって表されている。心理的距離の算出以降の認知処理は、ALCOVE と同様である。しかし、SUPERSET は複数の表象を持つため、式 2、3 は以下のように変化する。

$$O_k^n(x) = \sum_j w_{kj}^n \exp(-\beta d_j^n) \quad (8)$$

次節では、どのような条件に SUPERSET が既存のモデルのように振る舞うか説明する。ここでは、事例モデル、拡張型規則モデル、拡張型典型モデルについてのみ説明する。

### 3. 1 事例モデルとしての SUPERSET

次元間の相関に選択的注意を払わず ( $C_{jim}=0$ )、また全ての事例において同一の選択的注意の分配パターンを持つ時 ( $D_{ji} = D_i, \forall j \& I$ )、SUPERSET と一般的な事例モデル (ALCOVE など) の表象は同質となる。「次元間の相関に選択的注意を払わず、全ての事例において同一の選択的注意の分配パターンを持つ」ということは、カテゴリの統計的特性に選択的注意を払わないことと同じである。事例モデルは、記憶された全ての事例を参照することによって、統計的特性に選択的注意を払わなくとも、統計的特性を推論することが可能である。

事例とカテゴリの関連性の重み ( $w$ ) に関する制約はない。

### 3. 2 典型モデルとしての SUPERSET

カテゴリとそのカテゴリの典型的でない事例と関連性が十分低い場合 ( $w_{kj}^2 \leq \epsilon \forall j \notin$  典型事例)、つまり、カテゴリが典型事例とのみ意味のある繋がりを持つ場合、SUPERSET と拡張型典型モデル (OEDIPUS) の表象は同質となる。つまり、入力刺激と典型的事例のみとの類似性を基にカテゴリ化する場合 SUPERSET は OEDIPUS と同質となる。

選択的注意に事例モデルと同様の制約を課すと SUPERSET は各カテゴリを「典型的な事例」といった心理的スペース上の点で表象することとなり、一般的な典型モデルと同質となる。

### 3. 3 規則モデルとしての SUPERSET

SUPERSET が規則モデルと同様の表象となるには複数の制約が必要である。まず、典型モデルと同じように、カテゴリとそのカテゴリの規則と対応していない事例と関連性が十分低いこと ( $w_{kj}^2 \leq \epsilon \forall j \notin$  規則)。次元間の相関に向けられる選択的注意がないこと。必要十分条件に対応する事例とその特徴次元に十分な選択的注意が向けられること ( $D_{ji} > \delta, \forall j \& I \in$  規則)。

また、上記の条件に例外を付け加えると、拡張型規則モデル (RULEX) と同質の表象となる。

## 4. 概念の多様性

SUPERSET では既存のモデルに比べ多様な表象形態を可能とした。しかし、概念の多様性とは表象形態だけに留まらない。例えば、三角形を考えてみよう。幾何学的厳密性を必要としない場合は、事例や典型などを基とした視覚的類似性によってカテゴリ化されるのではないか。一方で幾何学的厳密性を必要とする場合、必要十分条件である規則によってカテゴリ化されるであろう。つまり、同じ「三角形」といったカテゴリであっても同時に複数の表象が保持されていることもある。複数の表象の同時保持は直感的には妥当な考察だが、行動実験などをもちいた検証はされていない。そこで、本研究課題で複数の表象の同時保持の可能性について、行動実験をおこなった。

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

## 4. 1 行動実験

本実験では複数の表象の同時保持の可能性について、Matsuka & Corter (2008)[6]の Experiment 2 で使用された課題を用いて以下のような実験を遂行した。表 1 は Matsuka & Corter (2008)[6]で用いたカテゴリ構造である。表 1 において、カテゴリ列は各事例が属するカテゴリを示している。D(Dominant), R(Recessive), U1(Unrelated 1), U2(Unrelated 2) は各事例の持つ特徴次元を示している。このカテゴリ構造の特徴は、次元 D、または次元 R を見ることによってどのカテゴリであるかを完全に識別できることである。Matsuka & Corter はこのような構造を用いた場合、多くの実験参加者は学習の途中から次元 D にほとんどの注意を向けるようになることを示した。なお、本質的には、次元 D と R は完全に重複しており、識別力から考えた時に構造上、全く同等の性質を持つことになるので区別ができない。よって次元 D および R は実験結果から定義した。完全なカテゴリの識別力を有する 2 つの次元のうち、最終ブロックにおいてより注視時間が長かった次元を D、もう一方の次元を R と定義した。しかし、注視時間差が 5%未満の場合は、汎化課題での次元が明白であった場合により成績を良いものを D とした。

カテゴリ	D	R	U1	U2
A	1	1	3	4
A	1	1	4	1
A	1	1	1	2
B	2	2	2	1
B	2	2	3	2
B	2	2	4	3
C	3	3	1	3
C	3	3	2	4
C	3	3	3	1
D	4	4	4	2
D	4	4	1	3
D	4	4	2	4

表 1 本研究の行動実験で用いたカテゴリ構造

本実験では、学習したカテゴリ知識に基づいて、学習課題時には呈示されていない新規事例の呈示を行い、

カテゴリ判断を行う汎化について分析を行った。特に、本実験では完全なカテゴリ識別力がある次元 D、R のいずれかの情報が得られない、不完全情報に基づくカテゴリ判断の内容から検討を行う。不完全情報に基づくカテゴリ判断とは、例えば、事例 A1 の各次元 (D, R, U1, U2) の特徴は [1, 1, 3, 4] と表現されるが、D の特徴を不明とし、[?, 1, 3, 4] という事例が与えられた場合のカテゴリ判断である。

学習を通じて実験参加者が獲得している知識の性質からは、不完全情報に基づくカテゴリ判断に関して以下のような予測が成り立つ。1 つは、単一の知識 (e.g. 次元 D を見てカテゴリを判断する、という知識) を獲得しており、次元 D の情報が得られなければ実験参加者はカテゴリ判断ができなくなってしまうことが考えられる。よって、例えば [?, 1, 3, 4] のような事例が呈示された場合、この事例が属するカテゴリに関する選択率は A, B, C, D それぞれへ 25%、つまりランダムな選択になることが予測される。もう 1 つは、実験参加者は次元に対する注視時間に反映されない、学習ブロックでは顕在化されなかった潜在的知識も獲得しており、次元 D の情報が得られないような場合でも、残りの次元から正確なカテゴリ判断ができることが考えられる。つまり、[?, 1, 3, 4] のような事例が与えられた場合でも、実験参加者は有意にカテゴリ A を選択することが考えられる。

方法の詳細は文字数の問題のため省略するが (松香・本田を参照 [10])、"Mouselab" と呼ばれるインタフェイスを用いて実験参加者の各特徴の注視時間を記録し、注視時間をその次元へ向けられた選択的注意とした。実験は、学習課題 (12 事例の学習を 1 ブロックとし、計 8 ブロック、合計 96 試行) とカテゴリ判断課題で構成された。カテゴリ判断課題では、12 事例 (以下、完全事例とする) と、次元 D、または次元 R を不明とした 24 事例 (以下、不完全事例とする) に対するカテゴリ判断、計 36 事例に対するカテゴリ判断が求められた。

複数の表象の同時保持の可能性を示唆する結果を図 3 に示す。図 3 では、学習課題の最終ブロック (LB)、完全事例判別課題 (CI)、次元 R が欠如した不完全事例判別課題 (MR)、次元 D が欠如した不完全事例判別

# 柔軟で多様な概念形成モデルの構築

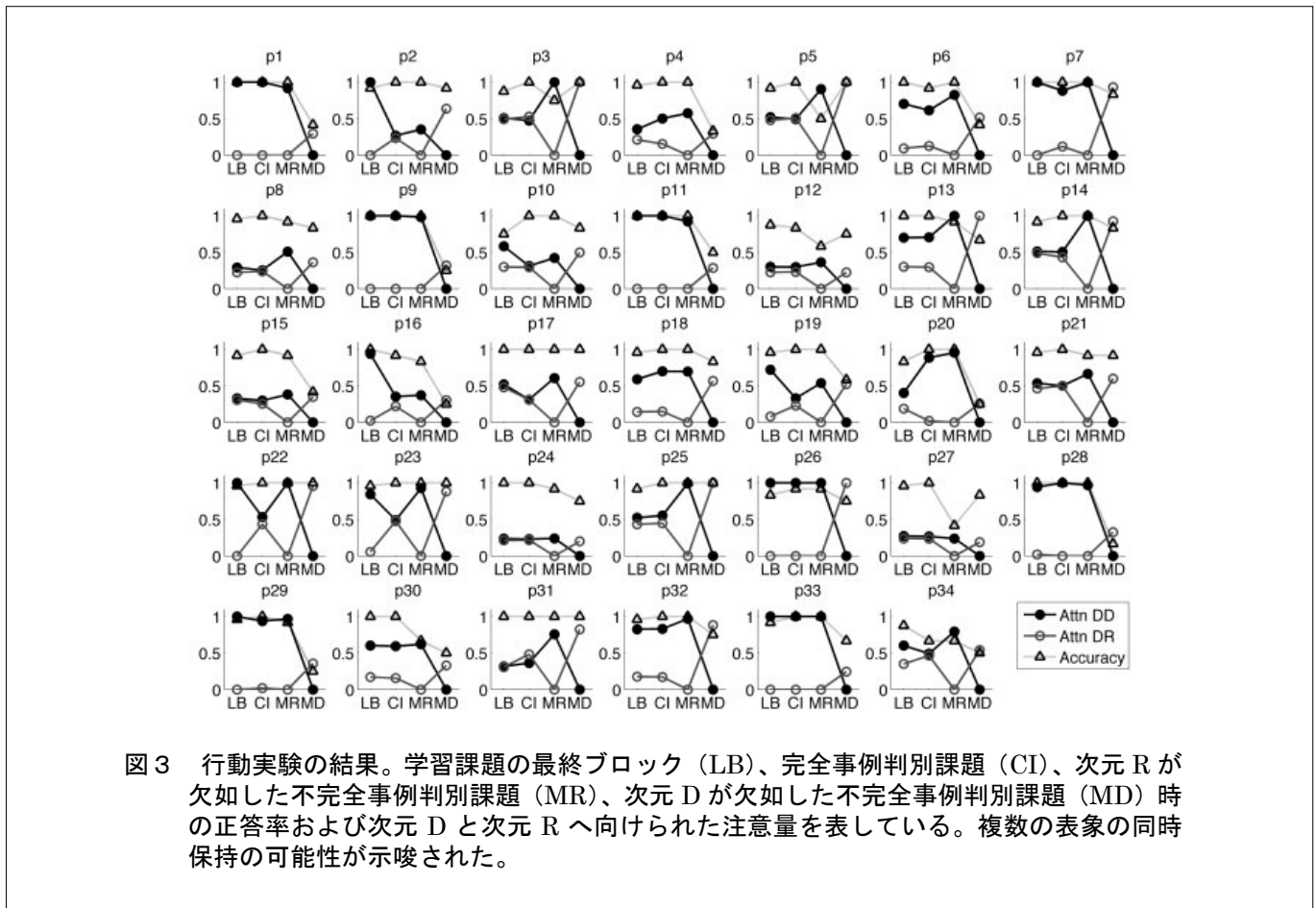
A cognitive model for flexible and diverse concept formation

課題 (MD) 時の正答率および次元 D と次元 R へ向けられた注意量が示されている。

まず、実験参加者 p1 の行動パターンを見てみる。p1 は学習最終ブロック、完全事例判別課題、次元 R が不完全事例判別課題 (MR)、において次元 D のみに選択的注意を向け、次元 D の情報がなり不完全事例判別課題 (MD) において、チャンスレベルの正答率となり、よって単一のカテゴリ表象を獲得したと解釈できる。一方で、実験参加者 p29 のように、学習ブロックと次元 R の情報のない不完全事例判別課題では最低限必要な次元 D のみに注意を向け、また、次元 D の情報の無い不完全事例判別課題でも最低限必要とされる異なった次元 R に注意を向け、完全事例判別課題では次元 D、R を含む全ての次元に注意を向けるといった、状況に応じて異なったカテゴリ方略をとった実験参加者も複数名いた。

今回用いたカテゴリ刺激に対する選択的注意やカテゴリ化のパターンのみでは、本実験で獲得された概念が、三角形の例のように複数の表象モデルに対応するものであるかは不明だが、部分的には重複するかもしれないが複数の表象を獲得していることが示唆された。

行動実験結果から複数の表象の同時保持の可能性が示唆された。既存のモデルでは概念が1つベクトルによって表記され、よって行動実験で示されたような複数の表象の同時保持は不可能である。対する SUPERSET では式 6~8 にあるように、概念が複数の独立したベクトルで表記されているため、複数の表象の同時保持を可能としている。次節では、複数の独立したベクトルで表されている概念がどのように形成・発達していくか説明する。





# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

## 5. 概念形成モデル

SUPERSET では、学習可能な係数 (i.e.,  $D, C, w$ ) は確率的最適化法によって学習されるとする。具体的には進化的アルゴリズムを応用した、当研究課題の成果の 1 つである CLEAR フレームワーク (Matsuka, Sakamoto, Chouchourelou & Nickerson, 2008 [14]) を用いる。CLEAR がモデル化している学習プロセスは至ってシンプルである。まず、複数の表象を組み合わせ新たな仮説を生成する。具体的には、学習可能な係数を  $\theta$  とした場合、新たな仮説  $\theta^{\circ}$  はランダムで選出された 2 つの表象を組み合わせることによって生成される：

$$\theta_i^n = \begin{cases} \theta_i^{p1} & \text{if } UNI < 0.5 \\ \theta_i^{p2} & \text{otherwise} \end{cases} \quad (9)$$

この仮説生成にはなんらかの誤差が発生すると仮定している。具体的には次式のように正規分布に従う誤差が生じる

$$\theta_i^n(t+1) = \theta_i^n(t) + N(0, \sigma_i^n(t+1)) \quad (10)$$

$$\sigma_i^n(t+1) = \sigma_i^n(t) \exp(N(0, \gamma_G) + N(0, \gamma_I)) \quad (11)$$

ここで  $t$  は時間を表し、 $N$  は正規分布に従う乱数を発生させる関数、 $\sigma$  は各係数に対応する誤差の振れ幅であり、 $\gamma$  はその振れ幅をコントロールするパラメータである。

生成された仮説に対し学習目的に応じた効用が算出され、良い仮説は保持され悪い仮説は忘却される。本研究課題においては、学習目的は間違っカテゴリ化の最小化と、表象の簡素化の 2 つから成り立っていると仮定した。CLEAR は確率的であり、online 学習 (1 入力刺激毎に係数を最適化する) では振幅が起こりうるため、Matsuka と Chouchourelou (2006)[5] が提唱した retrospective verification を応用して過去の複数の事例に現在の仮説を当てはめ、その妥当性を検証するといった学習法を組み込んでいる。つまり、カテゴリ化の誤差は、

$$E(\theta^n) = \sum_{g=1}^G \sum_{k=1}^K \Xi(d_k^g, x^g) (d_k^g - O_k^n(x^g)) \quad (12)$$

となる。ここで  $d$  は正しいカテゴリラベルであり、式 12 は Anderson と Schooler (1991)[1] の知見を基にした記憶の保持率を表す関数で、過去の入力刺激とその入力刺激に対する正しいカテゴリを想起できる割合を出力する。

表象の簡素の度合いは以下の 3 つの式から成り立っている。

$$S_D(\mathbf{w}^n) = \sum_{K=1}^K \sum_{j=1}^J \left[ 1 + (w_{kj}^n)^{-2} \sum_{l=1}^J (w_{kl}^n)^2 \right]^{-1} \quad (13)$$

$$S_D(\mathbf{D}^n) = \sum_{j=1}^J \sum_{i=1}^I \left[ 1 + (\alpha_{ji}^n)^{-2} \sum_{l=1}^I (\alpha_{jl}^n)^2 \right]^{-1} \quad (14)$$

$$S_D(\mathbf{C}^n) = \sum_{j=1}^J \sum_{i=1}^I \sum_{m=i+1}^I (C_{jim}^n)^2 \quad (15)$$

式 13 は事例とカテゴリの関連重みに対する簡素化で、不必要な関連を減らす働きがある。式 14 は特徴次元に向けられた選択的注意に対する簡素化で、式 13 と同様に、不必要な特徴次元への注意量を減らす働きがある。式 15 は特徴次元間の相関に向けられた選択的注意に対する簡素化で、注意の絶対量を減らす働きがある。SUPERSET では、式 12~15 を学習の目的関数として、一連の仮定生成と検証を繰り返し、獲得する概念を最適化することによって学習する。

## 6. シミュレーション

SUPERSET モデルの記述的妥当性を検証するため 3 つのシミュレーションをおこなった。

### 6. 1 シミュレーション 1

シミュレーション 1 では 2 次元の特徴からなる 2 つカテゴリを SUPERSET に学習させた。具体的には、図 4 にある異なる表象の複雑度をもつ 3 つの刺激セットを学習させた。先行研究の知見から、図 4A の刺激は規則的表象 (Estes, 1996 [3])、図 4B の刺激は (拡張型) 典型的表象 (松香・本田・吉川 2010 [10]; Matsuka & Sakamoto, 2007 [12]; Minda & Smith 2002 [16])、図 4C の刺激は事例的表象 (Blair & Homa, 2003 [2]) に落ち着くであろうと予測される。

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

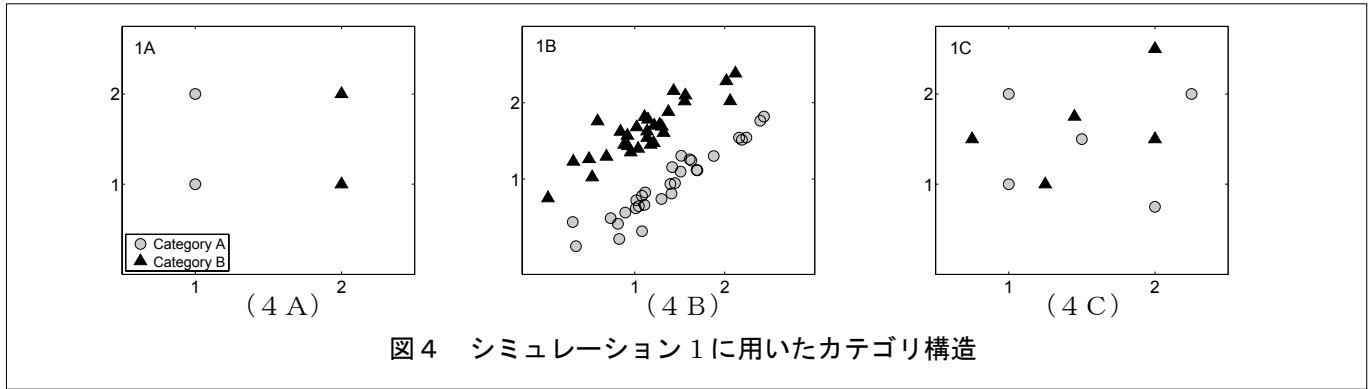


図4 シミュレーション1に用いたカテゴリ構造

シミュレーションの設定については字数の問題で省略する（詳細は Matsuka, Sakamoto, & Chouchourelou, 2008 [13]を参照）。図4Aの刺激を学習した結果、SUPERSETはカテゴリ化に必要な一方の次元のみを注目し、また、各カテゴリの2つの事例のうち一方のみを基準とするといった、規則モデルに適合する表象を獲得した。図4Bの刺激に対して、SUPERSETは二次元とそれらの次元間の相関に注目し、また各カテゴリの多くの事例のうちほぼ1つの事例を基準とする拡張型典型モデル的表象を獲得した。また、図4Cの刺激に対しては、ほぼ全ての事例とカテゴリの関連性を見いだすといった事例的モデルに適合する表象を獲得した。シミュレーション1の結果、SUPERSETは先行研究の知見から予測される3つの表象を異なった刺激において獲得することが示された。

## 6.2 シミュレーション2

SUPERSETの特徴は既存の主要なカテゴリの表象モデルを包含することと多様な表象を同時に複数保持できることである。

前者はシミュレーション1で確認した。シミュレーション2では、後者の複数の表象の同時保持について検証する。

シミュレーション2では、表象が複雑ネットワーク、具体的にクラスター性の高いsmall-world network的 (Watts & Strogatz, 1998[22]) 構造となっていて、概念の効用（どのような概念が適切か）に対する見解の違いがクラスターを生み出している場合、つまりクラスター内では学習目的は共有されるが、クラスター間では共有されない場合、どのような概念を形成するか検証した。シミュレーション2では、各表象にはカテゴリ化の誤差と表象の複雑性を同時に最適化した、それぞれの重みに多様性を持たせた（概念の効用の多様性）。

シミュレーションの結果を図5に示す。学習の結果SUPERSETはpareto最適的な表象群を獲得した。ここでいうpareto最適的な表象群とは、ある表象は至ってシンプルであるが代償としてカテゴリ誤差が多く、またある表象は至って誤差は少ないものの代償とし、

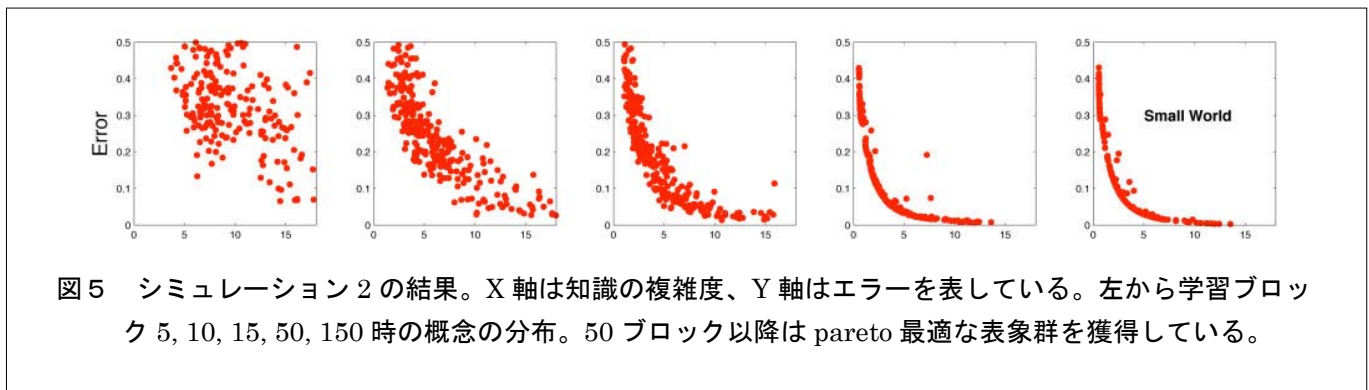


図5 シミュレーション2の結果。X軸は知識の複雑度、Y軸はエラーを表している。左から学習ブロック5, 10, 15, 50, 150時の概念の分布。50ブロック以降はpareto最適的な表象群を獲得している。

# 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

複雑性の高い表象となり、他の表象はそれらの間に分布するといったもので、殆どの全ての表象は他の表象に対し2つの学習目的において同時に劣ることのない、バランスのとれた表象群ということである。

## 6.3 シミュレーション3

シミュレーション1では、先行研究の知見から予測される3つの表象を異なった刺激において獲得することが示された。シミュレーション2では、概念が多様な表象群として獲得される可能性を示唆した。しかし、これら2つ結果は先行研究からの推論的予測であり、人間の行動データによる検証はおこなっていない。シミュレーション3で実際の人間のデータを比較してSUPERSETを検証する。

シミュレーション3では、SUPERSETによるNosofskyら(1994)[19]がおこなった行動実験の結果の再現を試みる。Nosofskyらは実験参加者にShepardらが(1961)提案した6種類の3次元からなる2つのカテゴリ学習を課した(図6)。行動実験の結果、1次元の情報のみでカテゴリ化が可能なType1が最も速く学べ、次元1と2で構成される排他的論理和であるType2が次に速く学ばれた(図6左)。3次元の情報が必要となるType3~5は同様の速さ・正確さで学ばれた。3変数の交互作用をもつType6では学習が最も遅かった。

SUPERSETにこの6種類のカテゴリ学習課題を課したところ、図7右にあるように、行動実験と同質の学習パターンが得られた。

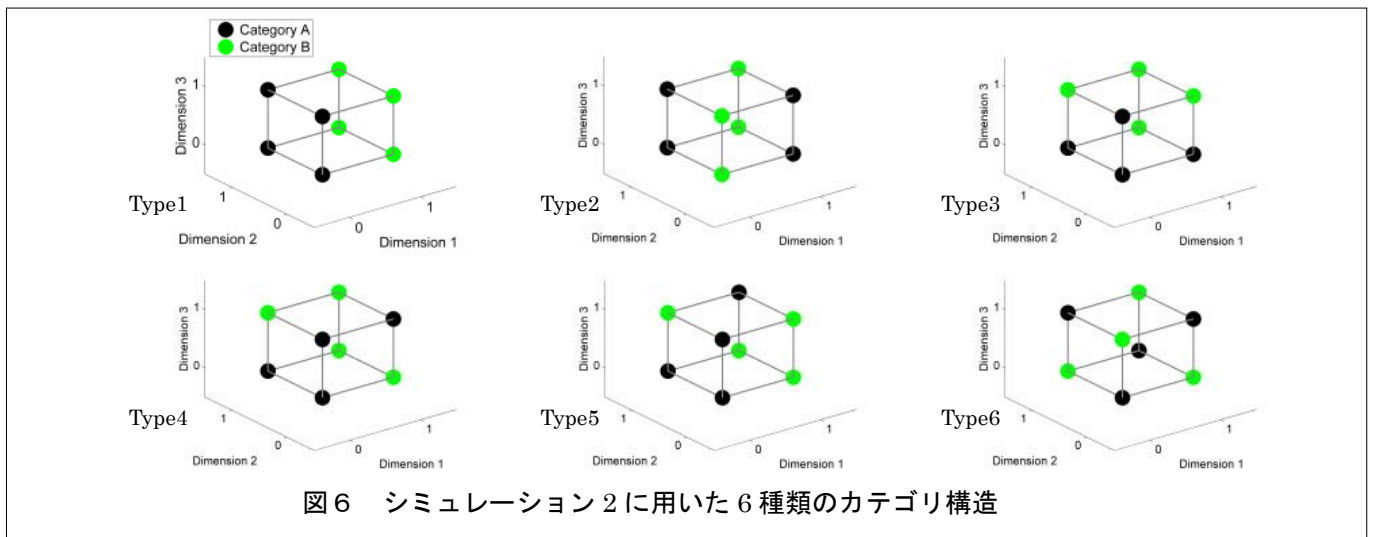


図6 シミュレーション2に用いた6種類のカテゴリ構造

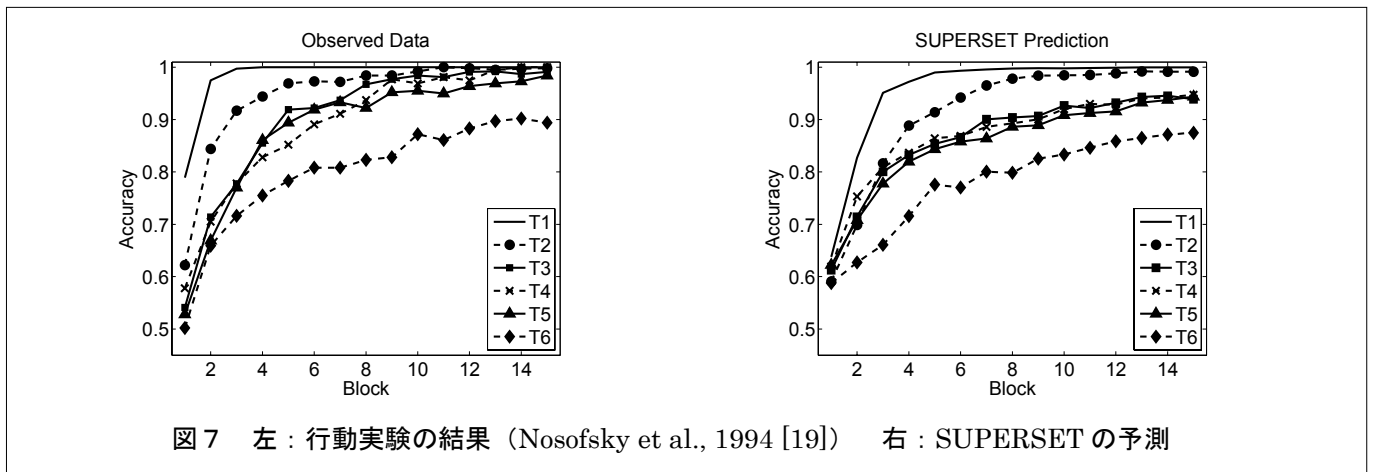


図7 左：行動実験の結果 (Nosofsky et al., 1994 [19]) 右：SUPERSETの予測

## 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

### 7. まとめと今後の展望

人間はあらゆる種の情報をカテゴリ化している。カテゴリ化は推論・予測・理由付けなど高度な思考を可能としている。このことから、カテゴリがどのように我々の頭の中で表象・形成され、どのようにカテゴリ化が行われているかは、認知科学・心理学の分野において重要な研究対象とされ様々なモデルが提唱されてきた。既存のモデルは大きく、規則モデル、典型モデル、事例モデルの3つに分類でき、ほとんどの全てのモデルは排他的な見解を示してきた。しかし、最近の研究結果の知見から、これら3つの主要のモデルの妥当性はカテゴリの構造・目的によって変化するという考えが広まってきた。このことから本研究では、3つの主要モデルを包含する SUPERSET モデルを提唱した。既存のモデルにない SUPERSET の特徴は、単に異なるカテゴリの表象を可能にするだけでなく、多様な複数の表象を同時に保持することができることである。3つのシミュレーションの結果、SUPERSET は柔軟で多様な表象が可能であることが示された。本研究ではモデルの構築に重点がおかれ、人間の行動実験を用いた検証が豊富ではない。今後の展望として、様々なカテゴリ構造や学習状況を操作した行動実験を行い、モデルの更なる記述的妥当性の検証を試みたい。

### 参考文献

- [1] Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- [2] Blair, M.R. & Homa, D. (2003). As easy to memorize as they are to classify: The 5-4 categories and the category advantage. *Memory & Cognition*, 31, 1293–1301.
- [3] Estes, W. (1996). *Classification and cognition*. New York: Oxford.
- [4] Kruschke, J. E. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- [5] Matsuka, T., & Chouhourelou, A. (2006). A model of human category learning with dynamic multi-objective hypotheses testing with retrospective verifications. In *Proc. IJCNN 2006* (pp. 3648–3656).
- [6] Matsuka, T & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology*, 61, 1067-1097.
- [7] Matsuka, T., Corter, J.E. & Markman, A. B. (2010). Some Attention Learning “Biases” in Adaptive Network Models of Categorization In *The Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1762 – 1767). Austin, TX: Cognitive Science Society.
- [8] 松香敏彦・本田秀仁. (2010). 概念学習における知識の多様性. 日本認知科学会 27回大会発表予稿集, 196-198.
- [9] Matsuka, T. & Honda, H. (2012). Effects of Individual Differences on Knowledge and Wisdom of Society: A Social Modeling Approach. In *The Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems*. (pp 401-406).
- [10] 松香敏彦・本田秀仁・吉川詩乃 (2010). プロトタイプ理論再考. 『認知科学』 17, 95 - 109.
- [11] Matsuka, T., Nickerson, J. V., & Jian, J. (2006). A prototype model that learns and generalize Medin, Alton, Edelson, & Frecko (1982) XOR category structure like humans do. In *the Proceedings of the 28th annual meeting of the cognitive science society*. (pp 1795 – 1800).
- [12] Matsuka, T. & Sakamoto, Y. (2007). A Model of Concept Formation with a Flexible Representation System. In *Advances in Neural Networks, Lecture Notes on Computer Science (LNCS) Vol. 4491 Part I.*, (pp. 1139 -1147). Berlin: Springer-Verlag.

## 柔軟で多様な概念形成モデルの構築

A cognitive model for flexible and diverse concept formation

- [13] Matsuka, T. Sakamoto, Y. & Chouhourelou, A. (2008). Modeling a flexible representation machinery of human concept learning. *Neural Networks*, 21, 289-302.
- [14] Matsuka, T. Sakamoto, Y. Chouhourelou, A. & Nickerson, J. V. (2008). Toward a descriptive cognitive model of human learning. *Neurocomputing*, 71, 2446-2455.
- [15] Medin, D. L., Ross, B. H., & Markman, A. B. (2005). *Cognitive psychology* (4th ed.). Hoboken, NJ: Wiley.
- [16] Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 775-799.
- [17] Murphy, G. L. (2002). *Big book on Concept*. Cambridge, MA: MIT Press.
- [18] Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition*, 22, 352-369.
- [19] Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53-79.
- [20] Rosch, E. & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- [21] Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75.
- [22] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440-442.

この研究は、平成19年度SCAT研究助成の対象として採用され、平成20～22年度に実施されたものです。