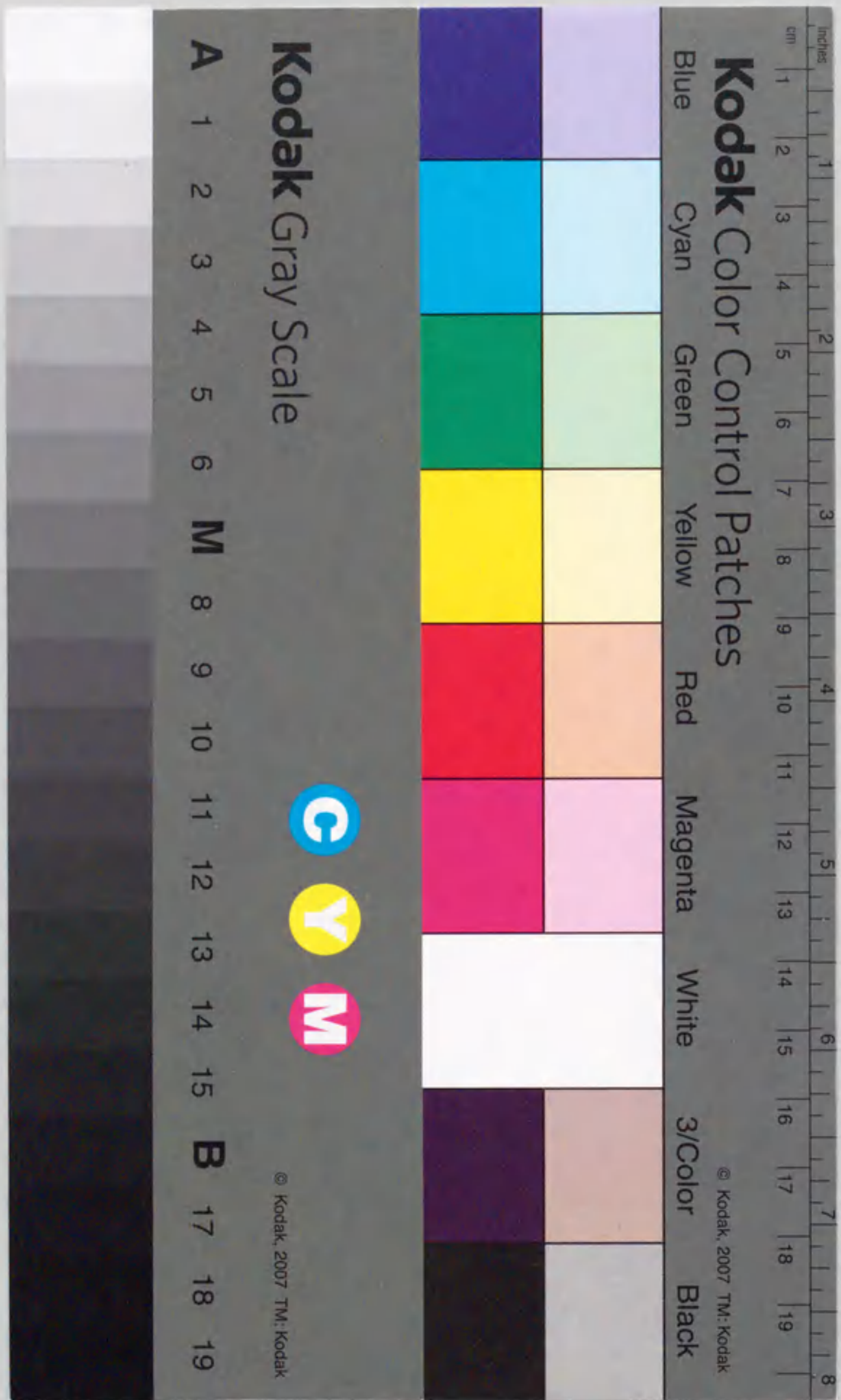


報告年次乙第、4563 号

知識形成と自然言語処理における

文脈の影響の認知科学的検討

水野りか





知識形成と自然言語処理における

文脈の影響の認知科学的検討

水野りか



# < 目次 >

## 序論

第1章 認知科学のアプローチ	1
第2章 本論文の構成	4

## 第1部 知識形成

第1章 はじめに	7
1.1 知識形成の心理学及び人工知能の研究史	7
1.2 第1部の構成	10
第2章 CORESの基本的仮定	11
2.1 記憶の負荷と典型例	11
2.1.1 典型例	11
2.1.2 適合度	13
2.2 文脈と属性	13
2.2.1 属性の相対性と重要度：重み・共有度	13
2.2.2 属性値差異自体の重要度：注目度	14
2.2.3 属性の相関	15
2.3 入力方法	15
第3章 数理モデル	17
3.1 用語の操作的定義	17
3.2 対象の表現方法	17
3.3 対象の入力方法	19
3.4 与えられた対象の認識：適合度	19

3.5 対象が正事例の場合の典型例の属性の強化：共有度	23
3.6 対象が負事例の場合の典型例の強化：重み	24
3.7 文脈にあわせた属性値の差異の認識のされ方の修正：注目度	25
3.8 基本的処理手順	27
第4章 実験1・シミュレーション1	29
4.1 実験1	29
4.1.1 方法	29
4.1.2 結果と考察	33
4.2 1回目のシミュレーション	35
4.2.1 方法	35
4.2.2 結果と考察	37
4.3 2回目のシミュレーション	39
4.3.1 方法	39
4.3.2 結果と考察	40
4.4 総合的考察	40
第5章 実験2・シミュレーション2	44
5.1 実験2	44
5.1.1 方法	45
5.1.2 結果と考察	47
5.2 シミュレーション2	54
5.2.1 教師ありのシミュレーション	55
5.2.2 教師なしのシミュレーション	62
5.3 総合的考察	65
5.3.1 被験者自身の分類上の矛盾のための不一致	65
5.3.2 CORESの仮定との矛盾のための不一致	65
第6章 まとめ	68
6.1 注目度，文脈，及び，知識構造	68
6.2 典型例	69
6.3 ヒューリスティックスの必要性	70



## 第2部 自然言語処理

第1章 はじめに	72
1.1 様々な領域の言語理解研究の融合	72
1.2 第2部の目的と構成	74
第2章 関係情報	75
2.1 関係情報の独立した活性化	75
2.1.1 意味ネットワークモデルの仮定上の問題点	75
2.1.2 関係リンクの独立性を支持する証拠	75
2.1.3 道具推論の研究	77
2.1.4 関係情報の潜在的活性化仮説の提起	79
2.1.5 方法の検討	80
2.2 実験1	81
2.2.1 方法	81
2.2.2 結果と考察	85
2.3 実験2	87
2.3.1 方法	87
2.3.2 結果と考察	88
2.4 総合的考察	89
2.4.1 プライミングの検索理論	89
2.4.2 構成-統合モデル	90
2.4.3 代名詞処理からの知見	91
2.4.4 記憶の負荷	91
第3章 CORESによる代名詞と省略語の処理	93
3.1 心理学的知見	93
3.1.1 代名詞	93
3.1.2 省略語	94
3.2 心理実験	95
3.2.1 方法	95

3.2.2 結果と考察	100
3.3 実験に基づくCORESの拡張	103
3.4 拡張された基本的処理手順	104
3.4.1 与えられた対象の認識	107
3.4.2 与えられた対象を認識できなかった場合	110
3.4.3 与えられた対象を認識できた場合	110
3.4.4 類似関係ノードの生成	110
3.4.5 時間関係ノードの生成	111
3.5 代名詞と省略語の処理手順	114
3.5.1 代名詞の解決	114
3.5.2 省略情報の推論	116
3.6 シミュレーション	119
3.6.1 シミュレーション過程の解説	119
3.6.2 考察	122
3.7 総合的考察	122
第4章 CORESによる次情報の予測	124
4.1 予測の重要性と先行研究	124
4.2 実験	126
4.2.1 実験1	126
4.2.2 実験2	130
4.2.3 全体的考察	132
4.3 予測処理の実現	134
4.4 シミュレーション	136
4.4.1 方法	136
4.4.2 結果と考察	139
4.5 総合的考察	142
第5章 CORESによる新旧情報の影響の新たな解釈	144
5.1 研究の背景	144
5.2 旧情報と新情報の出現順序が説明文の処理に及ぼす影響	144
5.2.1 読み手の知識から見た文の情報構造	144



5.2.2 文章の理解のしやすさと新旧方略	146
5.2.3 実験1	147
5.2.4 実験2	151
5.2.5 実験3	154
5.3 新旧方略による全体的考察とCORESによる新たな解釈の比較	155
第6章 まとめ	159
6.1 動詞の概念属性	159
6.2 助詞に関する統語と意味の問題	159
6.3 文脈の影響の時期	161

## 結論

第1章 研究の要約	163
第2章 今後の課題	165
謝辞	167
参考文献	168

## 序論

### 第1章 認知科学のアプローチ

近年、工学・心理学・言語学・神経生理学など、ともに人間の情報処理過程の解明を目指す様々な分野での知識や方法を結集した、認知科学の重要性は益々その度合いを高めてきている。これは、様々な学問領域が、実際は同一の問題を解決しようとしているにもかかわらず、専門性からくる壁のために互いに無縁で興味をそそらないものになりつつあることへの批判が高まり[戸田, 85], かつ、個々の学問領域の持つ問題が、相補的に解決される可能性と必要性が強く認識されてきたためであろう。

人工知能研究では、結果として人間の知と等価な、機械の知を「作る」ことを目的としてきた[辻井 88]。したがって、規則的で顕著な現象に焦点を当て、限られた条件でそのシミュレーションを行い、主に主観的・常識的判断に基づいて結果を評価するため、実証過程に弱点を持つという批判がある[Ahn 92, 辻井 88]。

一方、認知心理学は、人間の知の過程を「知る」ことを目的としてきた。したがって、人間行動に関する様々な仮説が立てられ実験が行われるが、パラダイムの限界から、得られた知見を体系的にまとめあげて応用することは非常に困難であり、孤立した知見が集積される傾向にある。そこで、それらの知見を統合し、包括的な人間モデルを作成する必要性の認識が高まり、それを可能にする計算モデルやシミュレーションの重要性が見直されてきている[Pylyshyn 84]。

このような各領域の問題を解決すべく生まれた認知科学は、「作ることによって知る」ことを目的とする[安西 86, 佐伯 86]。認知科学では、モデルを実験によって吟味し、得られた知見すべてを包括的なモデルとして体系化する。したがって、このアプローチによって作られたモデルは、機械の知のモデルとしても人間の知のモデルとしても高く評価されているものが多く[e. g., Anderson 91a, Gluck 88a, 88b, McClelland 81, 86, Rumelhart 82, 86], その幅広い適用可能性と包括的な説明力は広く認められている。



しかし、現実には、特に日本では、依然として学問領域の壁は厚く、認知科学研究とは言っても、いずれかのパラダイムに大きく依存した研究が多く、お互いの理解や融合は、現在、必ずしも達成されているとは言えない。

本研究は、本来の認知科学の理念にのっとり、知識形成と自然言語処理における、極めて人間の処理過程に近い柔軟で適応性の高い学習モデル、CORES(Contextual and Relative Existence System)を作成し、これを心理実験とシミュレーションを利用した学際的アプローチによって、吟味・発展させることを目的とする。そのアプローチとは、具体的には、1) 心理学・工学のみならず、言語学等で得られた知見と実際の人間の処理、言い換えれば、理論と実際との比較考察、2) 見いだされた矛盾や問題を解決できる、実行可能で、なおかつ、現実に即したモデルの作成(修正)、3) そのモデルの心理実験による部分的、あるいは、全体的な妥当性の検証、4) モデルの吟味(修正)、5) 吟味(修正)されたモデルのシミュレーションによる妥当性の確認、そして 2)へ戻り、最終的に5)の段階で問題がなければその時点でモデルを完成とみなす、という何段階もの手順を経る。

このようにCORESは、人間の情報処理過程の解明という1つの問題を、学際的アプローチによって、様々な角度から検討して作成された。よってCORESは、人間の基本的な処理過程に忠実であるがゆえに、往々にして別々に扱われる知識形成と自然言語処理という2つの問題に対処できる、適応性に富む人工知能モデルとなった。そして、その結果、CORESは、人間の学習と極めて類似した形で、言語的に与えられる情報を学習し、概念や知識を形成し、そして、その知識を応用して新たに与えられる言語情報を処理する能力を持つに至った。

これまでの機械学習や自然言語モデルの対象は、極めて領域特殊なものに限られていた。そして、その限られた適用範囲の中ですら、十分な成果を上げたと言言できるものは数少ない。しかし、人間は、恐らく基本的にはさほど異なるないメカニズムで、様々な領域の、しかも、文脈によって異なる情報をも、効率的に処理することができる。このような機械と人間の学習能力の違いは、我々に、現時点で必要なのは、原点に立ち戻ってもう一度人間がどのように情報を処理するのかを十分検討し参考にするのであることを、示唆しているのではなかろうか。そして、そうすることが、工学における機械学習や自然言語処理にとって多大な意義を与えることになると確信する。

本研究が、CORESによって文脈特殊な知識形成から自然言語処理に至るまでの広範囲で柔軟な機械学習を実現しえたのは、人間がどのような学習過程を経て、どのような知識を形成するかについての様々な実証的な知見を集積し、人間の認知過程についての理解を深めつつ研究が進められたからに他ならない。そしてその結果、CORESは、高度で柔軟性に富む機械学習を実現した工学モデルであると同時に、包括的な人間のモデルとしても極めて妥当であり、有効なモデルとなったと考える。



## 第2章 本論文の構成

第1部では、CORESを、人間の知識の基本的構成単位である概念（カテゴリー）形成過程に適用した研究[水野 92a, 92b, 92c, 93a, 93b, 93c]について、第2部では、CORESを、より高度な知識形成の手段としての自然言語処理に適用すべく拡張・発展させた研究[水野 93d, 93e, 93f, 93g, 93h, 93i, 94]とCORESによる言語心理学での研究結果[水野 86a, 86b, 88]の新たな解釈について述べ、結論部では、CORESの説明力と適用可能性について、その限界を含めて総括し、今後の研究課題を述べるものとする。

各部のはじめは、心理学、工学の双方からの問題提起が行われる。特に、第1部冒頭では、人間の処理過程の現実的な考察から生まれた多くの基本的仮定について述べるとともに、これらを実現するための基本的手順を詳述する。したがって、これらの部分には多くの紙面が割かれており、第1部の知識形成のみならず第2部の自然言語処理にも共通するため、本論文全体の礎であり、CORESを理解する上で極めて重要な部分となっている。また、第2部、第3章の代名詞と省略語の研究の冒頭は、第1部の基本的手順と、その段階までの知見を基に、CORESを自然言語理解システムとして発展させるための拡張された手順を述べた部分であり、現時点でのCORESの全容が記述されている。Table 1.1には、本論文で用いた主な変数の出現ページ、数式番号、及び、定義を記すので、随時参照されたい。

Table 1.1

### Variable List

---

Pages (Corresponding Equation Numbers) and Definitions of Variables

---

#### PART i

p.17 (1.1), p.21 (1.4), p.103 (2.2)

$doc_{Ab}$ : degree of correspondence of Object  $b$  toward Concept  $A$

p.17 (1.1), p.20 (1.3), p.21 (1.4), p.23 (1.5)

$res_i$ : resolution of Attribute  $i$

p.18 (1.2), p. 17 (1.1)

$ex$ : executing value of unmatched nodes

p.20 (1.3)

$c_{Ai}$ : communality of Attribute  $i$  in Concept  $A$

p.21 (1.4), p.17 (1.1), p.103 (2.2)

$w_{Ci}$ : weight of Attribute  $i$  in Concept  $A$

p.23 (1.5) & p.23 (1.6), p.17 (1.1), p.18 (1.2), p.20 (1.3), p.21 (1.5), p.103 (2.2)

$attn_i$ : degree of attention of Attribute  $i$

---



---

## PART 2

p.101 (2.1), p.103 (2.2)

$act_A$ : activation level of Concept A

p.103 (2.2)

$next$ : performance index determining next search path

p.106 (2.3)

$td$ : time delay

p.106 (2.4), p.113 (2.6), p.130 (2.7)

$act_{AB}$ : activation level of temporal proximity node between Concept A and  
Concept B

p.112 (2.5)

$rang_{A,B}$ : psychological distance or range between Concept A and Concept B

p.113 (2.6), p.112 (2.5)

$root_r$ : aggregated activation level of all roots between Concept A and Concept B

p.130 (2.7)

$nc$ : determinant of prediction of next coming information

---

Note: Each first equation number indicates the defining equation of each variable.

# 第1部 知識形成

## 第1章 はじめに

### 1.1 知識形成の心理学及び人工知能の研究史

心理学の分野での概念,あるいは,カテゴリー形成モデルは大きく3つに分けられる[Homa 84a]. 1つ目は,典型モデルで(prototype models)[e. g., Fried 84, Homa 76, Newmann 77, Posner 68, 70, Reed 72, Rosch 75a, 75b],過去の学習から形成された,算術平均[Reed 72],分散[Fried 84],出現確率[Medin 89]などの少数の要約的表象,すなわち,典型例に基づいてカテゴリーが形成されるとする. 2つ目は,事例モデルで(exemplar models)[e. g., Brooks 78, Hintzman 86, Medin 78, Whittlesea 87],過去に学習されたすべての事例との類似性に基づいてカテゴリー化が行われるとする. 3つ目は,特徴モデルで(feature models)[e. g., Hayes-Roth 77, Newmann 77, Reitman 73],属性値や属性値の組み合わせが出現する頻度を累積的に記憶し,その頻度に基づいて分類が行われるとする.

しかし,これらのいずれもが,現実に反したいくつかの問題を有する. まず,典型モデルは,1)学習の進行とともに事例数が際限なく増大する可能性がある以上,全属性値に基づく典型例の利用は非現実的であり,2)多峰性の分布など,非一般的分布をする属性の平均,分散,あるいは,出現確率等の要約的値には意味がなく,さらに,構造的に異なるためにそのような要約的値を算出できない場合もあり,3)相関のある属性は扱えない. 特徴モデルは,1)何を特徴(feature),すなわち概念の構成素(constituent)とするのかが明確でなく,2)扱われる属性値は離散的であるがゆえに,組み合わせの数の爆発的増大が起こりうるだけでなく,3)カテゴリーを形成するのに重要でない特徴の出現頻度まで計算されるため,必然的に,矛盾が生じる場合がある. そして事例モデルは,1)典型モデルと同様,事例数は学習とともに増大するため,全事例の記憶は非現実的であり,2)カテゴ



り自体の一般的特徴を測りえない。

一方、人工知能の分野での概念学習モデルは、目的によって大きく2つに分けられる。1つは、帰納推論システムに特定の概念の例を与えてその定義を決めさせるものであり[e. g., Dietterich 83, Mitchell 82], もう1つは、階層的な技法を用いて入力进行分类させるものである[e. g., Feigenbaum 84, Fisher 87, Gennari 89, Lebowitz 87]。そして、これらにも心理学のモデル同様、いくつかの問題がある。前者は、1) 頻度モデルと同様、概念は離散的で順序性のない属性値で表されており、2) その上で、原則としてすべての属性および属性値が記憶されている、などの点で非現実的である。そして後者は、1) 離散値でなく連続値を使用する場合も区分表現を使用しており、2) 厳密に階層的な概念構造において固定的な機能的、演繹的規則を利用している、などのために、一定の分類方法しか行うことができない。しかし人間の場合は、後に実験によって示すように、個人間のみならず個人内であっても、状況や目的が異なれば異なった分類を行うはずである。

心理学のモデルにせよ、人工知能のモデルにせよ、これらのモデルに共通する第1の問題は、記憶の過負荷である。したがって、いずれのモデルでも、課題が多く、複雑になるほど事例や属性の組み合わせの数が増大することから、極めて限られた条件や事例数・属性数の課題にしか適用されていない。ただし、最近提唱された、適合ネットワークモデル[e. g., Estes 89, Gluck 88a, 88b, Kruschke 92, Nosofsky 84, 92]では、Rumelhart, Hinton, and Williams[Rumelhart 86]の提唱したバックプロパゲーション法と同じく、得られた出力と正解を比較し、正しい分類に最も適合するように誤差を用いてパラメータを修正しつつカテゴリー化が行われるので、すべての事例や属性値を保持する必要はなく、記憶の過負荷の問題は解消された。しかし、これが解消されたのは、このモデルが、正解と事例を照合するためのフィードバックを利用しているからであって、現実の世界では、あらかじめ正解の明らかなカテゴリー形成はむしろ限られていると考えられる。

さらに、最も決定的な問題は、いずれのモデルも概念やカテゴリーの形成に対する文脈の影響を実現しえないことである。Labov[Labov 73]は、異なる文脈下ではコップかボールかの判断が異なることをことを実験的に示した。これは、概

念には本質的には明確な境界がなく、文脈に依存していることを意味する。また、概念の認識のされ方は固定的ではなく、同じ「鳥」でも、時には生きている鳥を、時には食べ物を、そして時には飛行機を意味するというように[Bergadano 92], 文脈によって変化すると考えられる。

また、文脈にはこの他にも、目的、文化、そして個人差など、様々なものが考えられる。例えば、1950年代までのアラスカエスキモーは、「飛行機」も「つばめ」も同じ「飛ぶもの」という1語で表した(現在はアメリカ文化の浸透から英語の借用語が利用される)。無論、彼らの目にも飛行機とつばめは明らかに異なって見えたはずである。しかし、彼らには、飛行機とつばめを識別する必要がなかった。識別して呼ぶことで果たす目的もなかった。したがってこの文脈では、1つのカテゴリーで十分なのである。

これが一種の文脈である証拠は、我々でもその必要のない文脈では、彼らと同じカテゴリーを利用することにある。例えば、何か動くものが上空を飛んでいる場合、「飛ぶもの」というカテゴリーを利用し、検索する。好奇心を満たす必要から「鳥か、飛行機か、スーパーマンか?」と考えるのはその後であり、最初から区別して考えてはいない。

概念は、Saussure[Saussure 60]の言うように、特定の文脈下での概念間の相対的關係によってのみ定義可能だと考えられる。なぜなら、前もって存在する概念があるなら、概念を表す言語間での相違があるはずがないからである。

文脈の問題は、このように実験的あるいは理論的には明らかにされても、特定の文脈でのカテゴリー化に焦点を絞れば、実行上の問題として必ずしも表面化しないために、モデルに十分組み込まれてはこなかった。しかし、本来、概念の意味やカテゴリーが文脈で異なりうるならば、個々の概念は柔軟な属性や属性値を持つように、そして文脈は、そのような属性の重要度や属性値、さらには、概念間の類似性など、すべての認識の仕方とそれに基づくカテゴリー形成に影響するように、どのような状況も再現できる上位の変数として具体的なモデルにも組み込まなければならないはずである。

CORESでは、記憶の過負荷と文脈の問題に最大の焦点を置いた上で、先述したいくつかの問題点のすべてを解決すべく、数々の基本的仮定が設定された。



## 1.2 第1部の構成

2章ではまず、その基本的仮定の各々について詳述し、3章では、基本的な数理モデルを解説し、4章及び5章では、それらの妥当性を検証するための2種類の実験とシミュレーションの紹介を行い、そして6章をその結びとする。

## 第2章 CORESの基本的仮定 [水野 92c, 93c]

現実の世界の分類では、分類する対象の数や属性は極めて多いが、記憶の過負荷は生じない。また、属性には様々な種類があり、その中には相関のあるものも、非一般的な分布をするものも存在する[e. g., Malt 84]。そして、属性値の認識上の大きさ、属性自体の重要度、さらには人間の認識する概念間の類似性は、すべて固定的ではなく、置かれた文脈や目的、そして、概念の入力順序によっても異なってくる。以下には、これらの1) 記憶の負荷、2) 文脈と属性、3) 入力方法、に関して、現実的考察に導かれた基本的仮定を順に詳述する。

### 2.1 記憶の負荷と典型例

#### 2.1.1 典型例

典型モデルで用いられた要約的表象は、記憶の負荷の問題や、相関のある属性に適用できないなどの問題を有していることは既に述べた。一方事例モデルでは、そのような要約的表象を用いない代わりに、全事例を記憶しておく必要が生じるため、これを軽減すべく、予測性の最大化[Anderson 91a]やカテゴリ有用性の最大化[Fisher 87]などの制約を用いてカテゴリ数を制限しているが、その計算の基となるのは事例間に共通の特徴の数だけであり、根本的な記憶の過負荷の問題が大き過ぎるための苦肉の策でしかない。適合ネットワークモデルでは、分類の正誤[e. g., Kruschke 92, Nosofsky 84, 92]を用いて記憶の過負荷を解決はしたが、先述したようにあらかじめ正誤のわからない分類過程には適用できない。したがって、残された可能性は、系列的に入力される事例の中から、限られた適切な事例を複数選択して典型例として記憶し、それらに基づいて連続的な属性値を有する事例を分類する方法をとることだと考えた。

分類と同様に情報が系時的に入力される文章処理においては、逐次入力される命題が、1) 初頭で処理された命題、2) 最近処理された命題、とリンクを形成することが明らかにされている[Kintsch 78]。ただし分類のための情報は、文章のように必ずしも順序だてて入力されるわけではないこと、系列的分類は常に先に



処理されたカテゴリとの関係から分類されていくことから、新しい情報は、1) のようなカテゴリ初頭の事例を基準に処理されていくと考えられる。

そこで、CORESでは、系列的に入力される事例のうち、各カテゴリの最初的事例を典型例として記憶するものとした。そして、フィードバックのない場合は、系列的に入力される個々の事例と既存のすべての典型例との間の、近接性の指標である適合度(2.1.2)を特定の閾値と比較する(3.4で詳述する)。そして、もし適合度がその閾値以上であれば、CORESはその事例がその典型例の属するカテゴリの成員であると認識する。もし、その閾値以下であれば、属すべきカテゴリを決定するまで他の典型例との適合度を算出する。最終的にもし見つからなければ、その事例を新たなカテゴリの典型例として記憶する(フィードバックのある場合は、各カテゴリの最初の正事例をそのまま典型例とすればよい)。このように仮定するのは、初めて系列的に入力される事例から概念が形成される場合には、例えば初めて宇宙人を見た時はそれが宇宙人の典型例となるように、最初的事例が典型例となる可能性が高いと考えたからである。ただしこの典型例は、情報、特にその正事例が増えても不変であるとは想定していない。例えば、8本足の宇宙人しか知らない時に、ETのような宇宙人と遭遇してしまった場合のように、既に存在する典型例で新しい正事例を識別不可能な場合は、その正事例を異なる典型例として追加する。

こうして、CORESは、限られた典型例だけを保持することで記憶の負荷を軽減することができる。さらに、このような典型例とともに、後に述べる適合度、重み、注目度等を使用することによって、非常に多くの事例の学習を必要とする適合ネットワークモデルとは対照的に、少数の事例だけに基づいてカテゴリを形成することも可能であり、また、これまでは無視されてきたカテゴリ間の近接性をも扱うことができる。

これまでこのような可能性が実験的に確認されなかったのは、人間が内的に用いている典型例は直接外部から観察できないからである。しかし、どのような典型例を想定した場合に人間の分類過程を最もよくシミュレートできるかという認知科学的アプローチならば、これを間接的に明らかにすることは可能だと考えた。よって、本研究ではその可能性も検討することになる。

### 2.1.2 適合度

CORESは、事例と典型例の近接性を、それらの属性値の差異に基づいて算出される、適合度という指標で表す。この適合度は、従来のモデルで扱われたような概念間の固定的な類似性とは異なり、様々な文脈の影響が反映される。なぜなら、例えば動物園の動物全体からすれば、チンパンジーとオランウータンは非常によく類似したものと認識されるかもしれないが、猿だけがいるモンキーセンターでは両者は極めて異なって認識されるように、概念間の近接性は一様ではなく、文脈によって変化すると考えたからである。そして、その文脈の影響は次の2.2に述べる属性値の差異、属性の重み、そして、注目度によって実現される(算出の仕方の詳細は第3章 数理モデルにおいて述べる)。

## 2.2 文脈と属性

### 2.2.1 属性の相対性と重要度：重み・共有度

Ripsら[Rips 80]は属性を相対属性と絶対属性に分けた。彼らによれば、前者は、物理的な測定に関する比率尺度や順序判断に依存する属性であって、大きさ、高さ、広さに関したものであり、後者は、質的な属性であって、色(赤、青、緑、etc.)、形(四角、円、三角、etc.)、あるいは、物理的構成(金属、木、プラスチック、etc.)などに関したものとされる。しかし、色や形や物理的構成は本当に絶対的に存在するであろうか。

菅沼[菅沼 92]は、絶対属性は存在しないと仮定した。そしてその根拠は、先に述べたSaussure[Saussure 60]の概念の認識についての考え方と一致しており、以下のような深い洞察に基づく。すなわち、任意の属性を絶対属性であると定義するためには、それ自体で認識可能でなければならないはずである。しかし、Ripsらが絶対属性と定義した四角にしても、それを認識するためには、それとは異なる多くの対象(円や三角形)が存在し、かつ四角をそれらと識別する必要がなければならないはずである。そしてそれらとの相対的な関係から初めて長方形という概念を記述するために必要な属性(特徴)が浮かび上がってくる。例えば円と四角だけで成立している世界を考えた場合、四角という概念を記述しようとするれば、直線であるか否かの属性で十分である。より単純な例で言えば、もし、この



世に男性しかいなければ、男性、という概念は存在せず、人間であるか否かの属性だけで十分であろうのと同じである。

このように概念が相対的存在だと仮定した場合に問題となるのは、対象の有する属性値の相対的な差異と、差異のあるその属性が、特定の文脈で、識別上どの程度重要であるか、という属性自体の重要度だと考えられる。極端な例で言えば、ここに対象aと対象bの2つの対象だけしか存在せず、対象aが $(a_1, a_2, a_3, a_4, a_5, \dots)$ 、対象bが $(b_1, a_2, a_3, a_4, a_5, \dots)$ 以下も対象aと同様)という属性値の集合からなり、さらに、対象aと対象bは識別される必要があると仮定する。すると、その異なる属性、すなわち、属性値 $a_1, b_1$ を持つ属性1は、他の属性に比べて、識別上で極めて重要でなければならないはずである。先述した頻度モデルは、このような属性の重要度を考慮せず、カテゴリー化に重要でない属性を含めて、単純に共通属性の頻度だけを問題としたために、最終的に矛盾を招く結果となったのだと考えられる。

そこでCORESでは、属性は基本的にすべて連続値を有すると仮定し、分類すべき対象の属性値の差異に基づく学習を基本とし[菅沼 92]、これに基づいて、特定の文脈における属性の識別上の重要度である、重みを算出するよう設計された。ただし、適合ネットワークモデルは、これと類似した属性の重要度を設定はしている[e. g., Gluck 88a, 88b, Kruschke 92, Nosofsky 92]。しかし、先に述べたように、このモデルではフィードバックのあることが前提となっており、あらかじめ決められた分類に即した属性の重要度に過ぎず、相対性を重視するCORESの重みの考え方や設定の仕方とは異なる。さらに、CORESの仮定する重みは、カテゴリーの成員が共有し、なおかつ、属性値の差異が小さいほど増加する、共有度という指標によって強化される(詳細は後述の第3章 数理モデル参照)。

### 2.2.2 属性値差異自体の重要度：注目度

冒頭にも述べたように、従来の分類モデルは、属性値に連続値を仮定した場合でも、一定の区分表現を使用していた[e. g., Feigenbaum 84, Fisher 87, Kruschke 92, Lebowitz 87, Michalski 83b]。しかし、属性の識別上問題となるのは、属性の重みの掛かった属性値の差異だけとは限らない。例えば100点満点の試験の得点が標準正規分布した時に優、良、可、をつける場合と、得点の分散が極めて

小さい時に優、良、可、をつける場合のように、比較的大きな差異だけを問題にすればよい文脈と小さな差異も問題となる文脈では差異の持つ重要度が異なってくるはずである。したがって、一定の区分表現による差異では不十分なばかりか、特定の文脈下での差異の大きさの認識のされ方、言い換えれば、特定の文脈下での属性値の差異自体の重みも考慮されねばならない。

文脈によるこのような認識のされ方の違いを扱ったモデルはこれまでになく、連続値を利用した適合ネットワークモデルでも、属性自体の重要度は設定しているが、属性値の差異の重要度までは考慮に入れていない。

CORESでは、特定の文脈での属性値の差異の大きさの認識のされ方、つまり属性値差異自体の重みを、注目度と命名・設定し、特定の文脈でその差異がカテゴリーの形成にどの程度重要か、という側面に初めて焦点を当てた。したがってこの注目度は、文脈の影響と相対的関係を重視するCORESの、中核を成すものである。

### 2.2.3 属性の相関

典型モデルを中心とする心理学モデルや人工知能のほとんどのモデルでは、相関のある属性は問題だとされ、独立した属性を仮定する必要があった[e. g., Elio 81, Feigenbaum 84, Fisher 87, Hintzman 86, Homa 76, 84b, Medin 78, Posner 68, Ross 90]。しかしいくつかの実験研究は少なからず属性間に相関があることを明らかにしている[e. g., Malt 84, Medin 82]。さらに、現実的にみても、完全に独立した基本的属性[Rosch 78]のようなものが特定できない以上、属性間相関の有無が問われること自体が問題なのではないかと考えられる。

相関の有無を問題にし、相関のある属性を意図的にできるだけ避け、その結果、限られた属性だけを用いた単純で人為的な実験しか行わないのは、本末転倒であろう。CORESは現実に即し、相関の有無によらずすべての属性を区別なく扱えるように作成された。そして最終的な属性間相関は、重み、注目度などの属性変数に反映され、処理に反映されるだけでなく、逆にこれらの数値から属性の性質を客観的に吟味することも可能となる。

## 2.3 入力方法



これまでのモデルは、分類の結果に重きを置くために、同時入力(simultaneous)形式をとるものが多かった[e. g., Michalski 83a, 83b, Quinlan 86]. しかし、研究者の関心がカテゴリーの形成結果ではなく過程そのものに移行し、新たな学習では情報は常に系列的に入力されるという事実が重要な意味を持つようになったため、系時的入力(sequential)の方式をとるモデルが近年増加してきた。そして、その中には、系列入力だと入力系列のゆがみに敏感過ぎてバイアスのかかった分類を生じるとして[Ahn 92], それを修正するための対策を講じたモデルも出現した[Anderson 91b, Fisher 87].

しかし、このバイアスは、不自然な独立したカテゴリーや属性を仮定し、それに基づいた理想的分類を想定しているからこそ不都合なのであって、本来はバイアスではないのではなかろうか。なぜなら、人間の場合は、例えば、時間的に接近して共通の特性を持つ事例を提示すると、帰納的推理が促進されるという実験結果も示すように[Elio 81], 入力系列によって分類の仕方に違いが出てくるのは当然だと考えられるからである。

CORESでは、系列入力が本来あるべき入力方法であると想定している。その上で、系列順序の影響は、一種の文脈の影響として実現される。

次の3章では、これらの基本的仮定を実現する数理モデルを解説する。

### 3.1 用語の操作的定義

以下の記述では、説明の便宜上、長期記憶に記憶される前の、短期記憶にある事例を対象(ただし、正事例・負事例という用語はそのまま用いる。添え字小文字)、記憶された事例を概念(以後、長期記憶と呼ぶ。添え字大文字)、類似性に基づく概念の有意味なまとまりをカテゴリーと呼ぶものとする。ただし、これらはいくまでも操作的な定義であって、本質的な定義ではない。

### 3.2 対象の表現方法

与えられた対象および概念の内部表現(以後、短期記憶と呼ぶ)は、Fig. 1.1のように表現される。ここで、○は対象の各構成要素(components)に対応し、構成要素ノードと呼ばれる。点線は、上位のノードが下位のノードを構成要素としていることを表している。また、ノード間を結ぶ実線は、それらの構成要素ノード間に何らかの関係(componential relationship)が存在することを表している。CORESでは、これらの関係も1つのノードに対応づけ、それを構成要素間関係ノードと呼んでいる。また、これらのノード及び構成要素間関係ノードは複数の属性を有し、各々が連続値を有すると仮定する。例えば、対象である数字の「4」は、「4」を構成する各線分を1つの構成要素ノードとみなせば、Fig. 1.2のような形で表現可能となる。この図において、○及びそれらを結ぶ実線が、構成要素ノード及び構成要素間関係ノードであり、その横に書かれているのが、それらを記述する属性の例である。

属性は主として実験で得られた被験者のプロトコルを基に決定する。本質的な属性、あるいは、基本的属性[Rosch 75a]が明らかでない以上、この方法以外に適切な方法は現時点ではないと考えたからである。

客観的な属性の選択方法としては、多次元尺度構成法が考えられる。しかし、たとえ多次元尺度構成法を用いたとしても、基本的属性としてはじめに何を設定



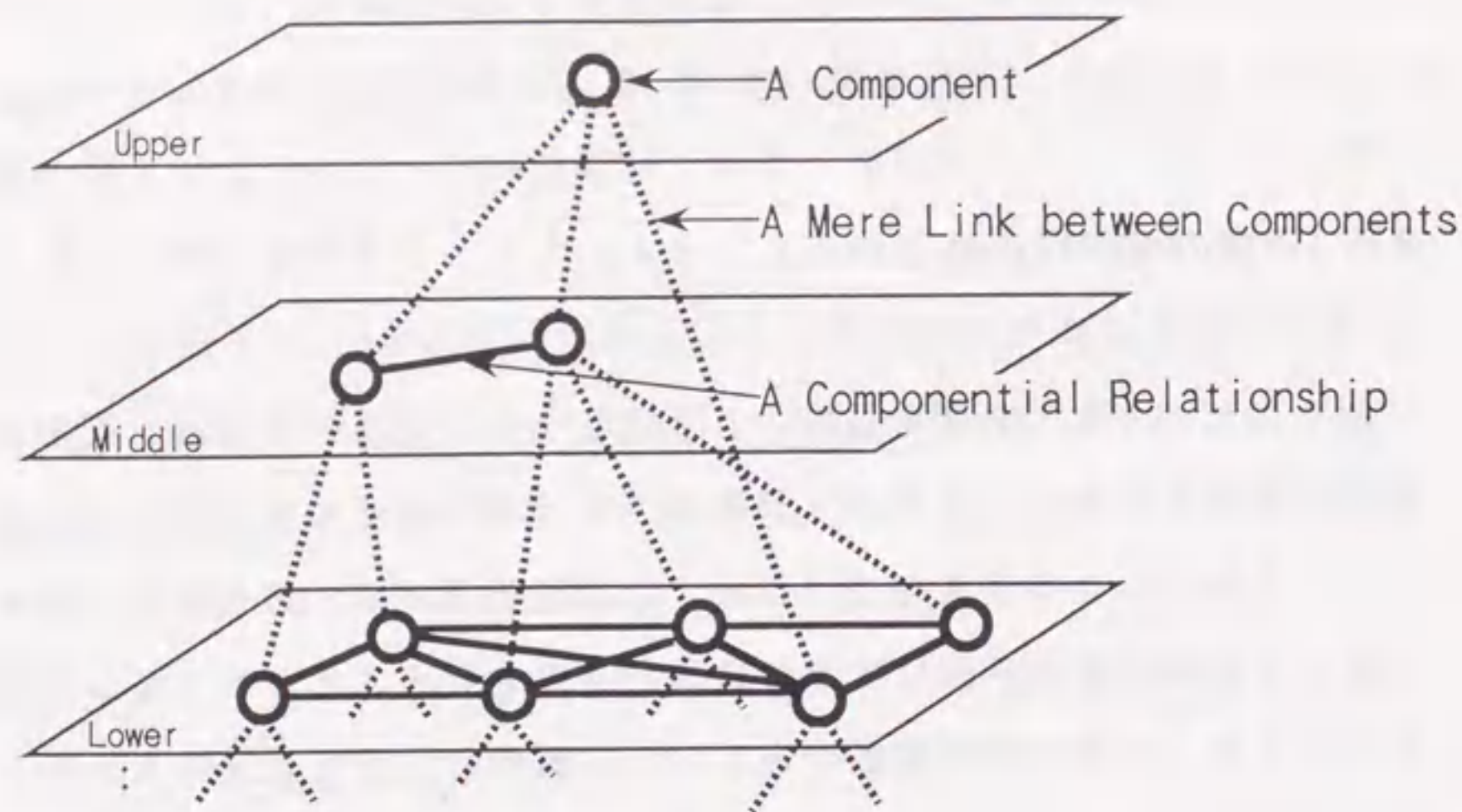


Fig.1.1. Representation of an object.

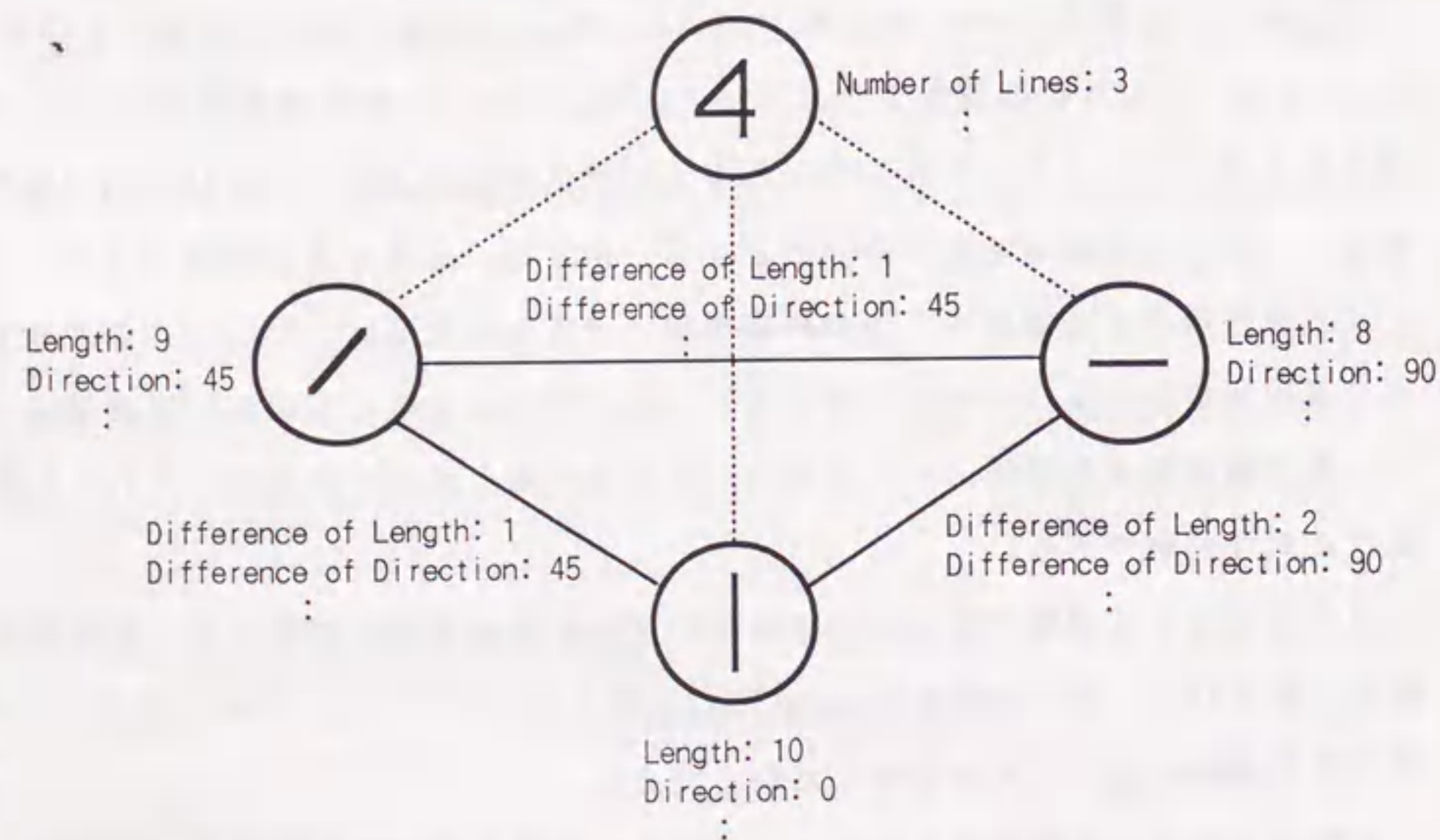


Fig.1.2. Representation of an object "4".

するかという問題は変わらない。また、結果として得られるのは、与えられた刺激を表現する最小の次元数の属性であるが、CORESでは必ずしも属性をこのように限定する必要はないし、また、限定することで多様な文脈に適用できるという柔軟性を損なう恐れがある。

本来、CORES自体には、特定の文脈や対象において重要な属性、及び、属性値の差異を算出する能力がある。したがって、重みや注目度の減少によって不適切な属性は自ずと分類に関与しなくなるばかりか、相関のある属性の場合はそれらが適切な比率で配分されると期待される。したがって、多次元尺度構成法などによってあらかじめ限られた数の属性を設定するよりもむしろ、被験者のプロトコルに基づく柔軟な方法が、実験者にとっては最も客観的かつ合理的であると考えられた。

### 3.3 対象の入力方法

あらかじめ入力されるデータには、上記の属性データに加え、重み・注目度の初期値、重み・注目度・共有度の変化量を制御するパラメータをはじめとする、各種パラメータがある。これらの詳細に関しては、個々の式の説明で逐次解説する。

なお、CORESが認識した結果が正しいか否かを判断する方法としては、教師ありの場合(supervised simulation)と教師なしの場合(unsupervised simulation)がある。教師ありの場合は正しい解があらかじめ入力データとして与えられる。教師なしの場合は、適当な適合度の閾値が利用される。この閾値は、いくつかのシミュレーションを行った後に、適切なカテゴリ数となるように、任意に値が設定される。この閾値は、例えば、1.0(適合度の最大値)に設定されればすべての対象を異なるカテゴリに、0(最小値)設定されればすべての対象を1つのカテゴリに分類する、というように、どの程度厳密にCORESが対象を分類するかを決定する値であって、どのように分類するかを決定する値ではない。したがって、任意に設定されても、CORESの分類方法自体は影響されない。

### 3.4 与えられた対象の認識：適合度



CORESは、先に述べた内部表現を有する対象を、系列的に認識していく。与えられた対象は典型例と比較マッチングされ（系列初頭はまだ典型例が存在しないため、それが典型例となる）、それらの間の各々の属性値の差異が求められる。次に認識すべき対象を対象*b*、典型例を概念*A*とすると、両者の近接性は、以下の適合度の指標で求められる。

$$doc_{A,b} = \sum_i \sum_j (w_{A,i} * \exp(-x_{A,b,i}^2)) + ex \quad (1.1)$$

$$x_{A,b,i} = attn_i * (x_{A,i} - x_{b,i}) / res_i$$

ただし、 $doc_{A,b}$ ：対象*b*の概念*A*に対する適合度(degree of correspondence)

$w_{A,i}$ ：概念*A*の構成要素ノードまたは構成要素間関係ノードの属性*i*の重み(weight: 3.6で詳述する)

$attn_i$ ：属性*i*に対する注目度(degree of attention: 3.7で詳述)

(以上の注目度と重みは、後に述べるように学習によって自動的に修正され、決定される。)

$x_{A,i}$ ：概念*A*の構成要素ノードまたは構成要素間関係ノードの属性*i*の値

$x_{b,i}$ ：対象*b*の構成要素ノードまたは構成要素間関係ノードの属性*i*の値

(なお、対応する構成要素ノードまたは構成要素間関係ノードが、属性*i*を持たない場合は、 $x_{A,b,i}$ が無限大であるものとみなす。)

$res_i$ ：属性*i*の分解能、すなわち、どこまでの属性値の差を識別できるかを決定する値(resolution)

(なお、以下の記述では、簡略化のために、概念*P*、対象*p*の後の「構成要素ノードまたは構成要素間関係ノードの」という記述を、単に「ノードの」と記述するものとする(Fig. 1.1参照)。

ここで、 $n$ に関する総和は、対象*b*の各ノードに対応する概念*A*のノード毎の値の総和を計算することを意味している。なお、この対応関係は1種類とは限らないが、それらの中で(1.1)式の値が最も大きくなるものを採用する。そして、 $ex$ (executing value)は、概念*A*と対象*b*に対応するノードがない場合に、これら処理し

数値化したもので、次のようにして計算される。

$$ex = sum_{Ab} * \exp(-(attn_{s,i} * sum_{Ab})^2) - sum_{bA} * (1.0 - \exp(-(attn_{s,i} * sum_{bA})^2)) \quad (1.2)$$

ただし、 $sum_{Ab}$ ：概念*A*のノードのうち、対応するものなかったノードに付加された重みの和。ここで、ノードに付加された重みとは、そのノードを記述する各属性に付加された重みの和である。

$sum_{bA}$ ：対象*b*のノードの内、対応するものなかったノードに付加された重みの和

$attn_{s,i}$ ：長期記憶の構造に対する注目度

$attn_{ss}$ ：短期記憶の構造に対する注目度

$ex$ の意味は次の通りである。例えば、Fig. 1.3の左を長期記憶の概念*A*、右を短期記憶の対象*b*とする。ただし、ここでは便宜上、すべてのノードの属性は一致するものとし、各構成要素ノード( $co_1, co_2, co_3, co_4$ )及び構成要素間関係ノード( $cr_{12}, cr_{13}, cr_{23}, cr_{24}$ )の持つ重みの和がその下に示す値だとする。ここで、概念*A*と対象*b*は、構成要素ノード  $co_1, co_2, co_3$ 、及び構成要素間関係ノード  $cr_{12}, cr_{23}$ によって部分的に一致する。そして、概念*A*の構成要素  $co_4$ 及び構成要素間関係ノード  $cr_{24}$ は対象*b*に対応するノードが存在せず、逆に、対象*b*の構成要素間関係ノード  $cr_{13}$ は、概念*A*の方に対応するノードが存在しない。

長期記憶の構造に対する注目度は、対象*b*が概念*A*の持つノードを部分的に有するだけでも適合度を高いとみなすか否か、言い換えれば、両者の部分的一致をどの程度認めるかを表し、一方、短期記憶の構造に対する注目度は、対象*b*が概念*A*の持たないノードを有する場合に、どの程度適合度を低めるか、言い換えれば、それをどの程度ノイズだとみなすかを表す。したがって、これらの注目度がともに0の場合は、対象*b*が概念*A*の有する  $co_4, cr_{24}$ を持たなくても、概念*A*に対する対象*b*の適合度は十分高いとみなし、対応しないこれらのノードの重みの和( $sum_{Ab}$ )、すなわち、 $co_4$ と  $cr_{24}$ の  $0.05 + 0.05 = 0.1$ の分だけ適合度を高める。また、与えられた対象*b*の方に長期記憶の概念*A*の持たないノード、ここでは  $cr_{13}$ 、が存在し



でも、これをノイズだとみなさないため、その重みの和( $sum_{bA}$ )の項を0として適合度を下げない。一方、これらの注目度がともに1.0だとすれば、 $sum_{bA}$ は適合度には加えられず、しかも、 $sum_{bA}$ 、ここでは $cr_{13}$ の重みの和、0.05の分が適合度から引かれることになる。

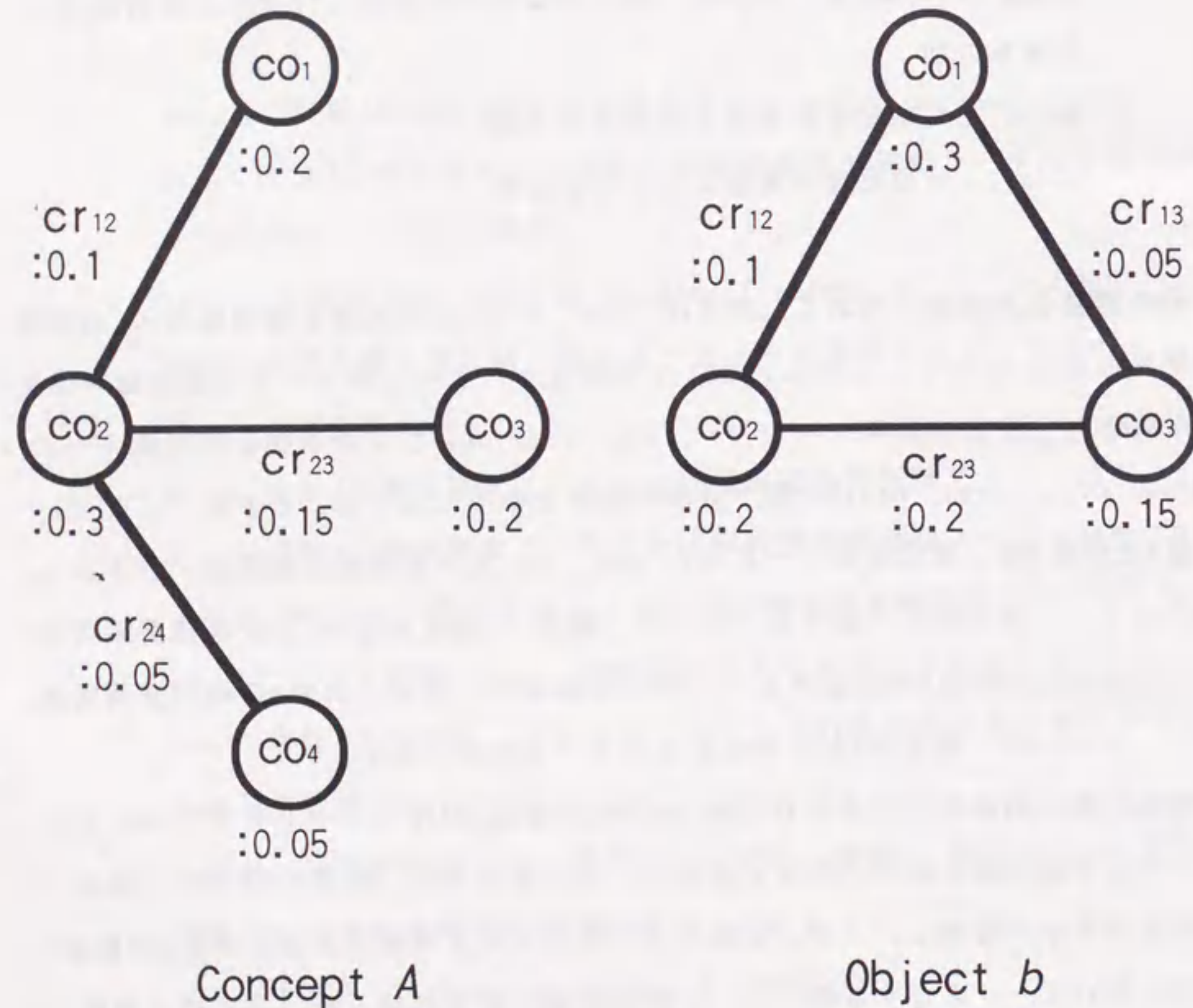


Fig.1.3. Examples of unmatched nodes.

また、分解能  $res_i$  は、任意の属性において、どの程度の属性値の差を識別できるかを決定する値で、例えば、注目度が1.0で概念Aと対象bの属性iの属性値の差異が1.0の場合に、0.4に設定したとすれば、(1.1)式の  $\exp(-\chi_{A,b,i}^2)$  はほぼ0となり、属性iの属性値の差1.0を、厳密に識別することができる。

以上、適合度は、各属性値の差異が大きいほど近接性が小さくなることを示す。正規関数を利用しているのは属性値の小さな差異の影響を少なくしたりある程度以上属性値が異なった場合は飽和的性質を持たせるためである。そしてこの程度が、後に述べる注目度によって制御されることになり、2.2.2で述べた成績の例のような文脈による違いが実現されることになる。

### 3.5 対象が正事例の場合の典型例の属性の強化：共有度

ここからは便宜上、教師なしで、かつ、適当な閾値が設定されている場合を例にとって解説する（教師ありの場合は評価のプロセスが含まれないだけである）。

与えられた対象bの、ある典型例の概念Aに対する適合度が閾値以上の場合、その対象bは正事例と認識される。そして、次式で算出する共有度(communality)によって、両者に共通の属性の強化を図る。

$$c_{A,i} = \Delta c_{A,i} + c_{(old)A,i} \quad (1.3)$$

$$\Delta c_{A,i} = par_{13} * \exp(-\chi_{A,b,i}^2)$$

$$\chi_{A,b,i} = attn_i * (\chi_{A,i} - \chi_{b,i}) / res_i$$

ただし、 $c_{A,i}$ ：概念Aのノードの属性iに付加された共有度の修正後の値

$c_{(old)A,i}$ ：概念Aのノードの属性iに付加された共有度の修正前の値

$par_{13}$ ：共有度基準増加量

(これ以下のパラメータのsuffixはプログラム内部で使用しているものをそのまま用いた。)

$\chi_{A,i}$ ：概念Aのノードの属性iの値

$\chi_{b,i}$ ：概念Aの正事例となる対象bのノードの属性iの値

(なお、対応するノードが、属性iを持たない場合は、 $\chi_{A,b,i}$ が無限度で



あるものとみなす.)

$res_i$ : 属性  $i$  の分解能

この共有度は、ある属性値の差異が小さいほど、また、その属性がその概念の事例の間で共有されているほど増加することになる。

### 3.6 対象が負事例の場合の典型例の強化：重み

上とは逆に、与えられた対象  $b$  の、ある典型例への適合度が閾値未満だった場合、CORESはその対象を負事例と認識し、次式で算出する重みによって両者を識別する属性の強化を図る。ここで、上と区別するために、その典型例を概念  $C$  とおく。

$$\begin{aligned}w_{c,i} &= \Delta w_{c,i} * w(oid)c,i & (1.4) \\ \Delta w_{c,i} &= par_{10} * docc_b * (1.0 - \exp(-cc_i)) * \exp(1.0 - \chi_{c,b,i}^2) \\ \chi_{c,b,i} &= attn_i * (\chi_{c,i} - \chi_{b,i}) / res_i\end{aligned}$$

ただし、 $w_{c,i}$ : 概念  $C$  のノードの属性  $i$  に付加された重みの修正後の値

$w(oid)c,i$ : 概念  $C$  のノードの属性  $i$  に付加された重みの修正前の値

$par_{10}$ : 重み基準増加量

$docc_b$ : 対象  $b$  の概念  $C$  に対する適合度

$cc_i$ : 概念  $C$  のノードの属性  $i$  に付加された共有度

$attn_i$ : 属性  $i$  に対する注目度

$\chi_{c,i}$ : 概念  $C$  のノードの属性  $i$  の値

$\chi_{b,i}$ : 概念  $C$  の負事例となる対象  $b$  のノードの属性  $i$  の値

(なお、対応するノードが、属性  $i$  を持たない場合は、 $\chi_{c,b,i}$  が無限大であるものとみなす.)

$res_i$ : 属性  $i$  の分解能

なお、上の式を使用して重みを修正した後、概念  $C$  のノードの各属性に付加された重みの和は1.0になるように正規化される。

冒頭で、概念とは相対的な存在に過ぎず、他の概念の存在により始めてその本質が明らかになると述べた。この強化は、正にこの仮定に従っており、概念  $C$  と負事例の対象  $b$  によって、概念  $C$  の記述を明確にしようとするものであり、対象  $b$  と概念  $C$  の対応するノードの属性間において、その値が異なれば異なるほど、その属性に付加された重みを増加させる。さらに、両者の適合度が1.0に近いほど重みの修正幅は大きくなる。なぜなら概念間の近接性は、この2つだけでなく、すべての概念間の相対的關係から決まるのであり、全体の適合度をバランスよく配分するには、概念  $A$  と事例  $b$  の適合度が初めから1.0に近くなるような場合には、反発する力を持たせて識別力を高める必要があるからである。

こうして重みを修正した後、CORESは対象  $b$  の属するカテゴリーを探すために、別の典型例を検索し、再び(1.1)式の適合度の算出へ戻る。これを繰り返しても最終的に適合度が閾値以上となる典型例がない場合、すなわち、属するカテゴリーが見つからない場合には、CORESはこの対象  $b$  を、新しいカテゴリーの典型例、すなわち、概念  $B$  として長期記憶に保存する。その際には、概念  $B$  のすべての属性に対する重みは同じ値に設定される。

### 3.7 文脈にあわせた属性値の差異の認識のされ方の修正：注目度

以上から、ある典型例に対する正事例と負事例が認識されると、CORESはその文脈に合わせて、全属性の注目度を修正する。

この注目度とは、先にも述べたように、特定の文脈下では、どの程度属性値が異なれば異なるカテゴリーに分類するか、を決定する、いわば属性値の差異自体の重みである((1.1), (1.2), (1.3), (1.4)式,  $attn_i$ 参照)。

仮に、与えられた対象と典型例が、識別上重要な、すなわち、重みの大きい1つの属性の属性値によって識別されたとする。先の成績の例のように、ある文脈では、両者のその属性値の差異がかなり大きくても同じカテゴリーに分類されるが、別の文脈では、ほんの少しの差異でも同じカテゴリーには分類されない、という場合が起こりうる。

このような判断を適合度への影響として実現するためには、該当する属性の、(1.1)式における注目度を変化させてやれば良いと考えた。つまり、その注目度を



適度に小さくしてやれば、対象と典型例の属性値の大きな差も無視されて適合度は1.0に近づき、適度に大きくすれば、小さな差異でも考慮され適合度は0に近づく。ただし、仮にその注目度が不適當に小さければ、対象と異なるカテゴリーの典型例に対する適合度も1.0に近づいて識別が不可能になってしまうし、不適當に大きければ同じカテゴリーの典型例との適合度まで0に近づいて識別されてしまう。

そこで、CORESでは、対象と同カテゴリーの概念に対する適合度と異カテゴリーの概念に対する適合度の差ができるだけ大きくなるように注目度を修正していく。具体的には、次式における $f(\text{attn}_i)$ を最大にするよう修正する（極値をとる値を計算する）。

$$f(\text{attn}_i) = \exp(-X_{A, b, i}^2) - \exp(-X_{C, b, i}^2) \quad (1.5)$$

$$X_{A, b, i} = \text{attn}_i * (x_{A, i} - x_{b, i}) / \text{res}_i$$

$$X_{C, b, i} = \text{attn}_i * (x_{C, i} - x_{b, i}) / \text{res}_i$$

ただし、 $\text{attn}_i$ ：属性  $i$  に対する注目度

$x_{b, i}$ ：対象  $b$  のノードの属性  $i$  の値

$x_{A, i}$ ：対象  $b$  と同カテゴリーの概念  $A$  のノードの属性  $i$  の値

$x_{C, i}$ ：対象  $b$  と異カテゴリーの概念  $C$  のノードの属性  $i$  の値

$\text{res}_i$ ：属性  $i$  の分解能

なお、通常、属性  $i$  を使用している概念やノードは複数個存在するため、(1.5)式の $X_{A, b, i}$ 及び $X_{C, b, i}$ における差は、それらの重み付き平均値が使用される。

ここで、(1.5)式によって修正された属性  $i$  の注目度を、 $\text{attn}_{(\text{new})i}$ とすると、この値をそのまま採用すると、問題が生じる可能性がある。その問題とは、与えられた対象ごとに、注目度が大きく変動しうることである。特に、与えられた対象の中に特殊な例が混じっていた場合、その例の影響をそのまま受けてしまう。そこで、上の式で計算された注目度 $\text{attn}_{(\text{new})i}$ を以下の式によって修正した実行上の値を使用するものとする。

$$\text{attn}_i = \text{par}_{20} * \text{attn}_{(\text{old})i} + (1 - \text{par}_{20}) * \text{attn}_{(\text{new})i} \quad (1.6)$$

ただし、 $\text{attn}_i$ ：実行上の属性  $i$  の注目度

$\text{attn}_{(\text{old})i}$ ：修正前の属性  $i$  の注目度

$\text{attn}_{(\text{new})i}$ ：(1.5)式による修正後の属性  $i$  の注目度

$\text{par}_{20}$ ：修正前の注目度に対する優先度

### 3.8 基本的処理手順

以上の処理をフィードバックがない場合を例にとって簡単に解説する。

ここに、すべてのノードが対応する3つの事例、事例  $a$ 、事例  $b$ 、事例  $c$  だけが存在すると仮定する。事例  $a$  と事例  $c$  は同じカテゴリーに属し、事例  $b$  だけ異なるカテゴリーに属する。3つの事例は  $a, b, c$  の順序で入力される。そして属性は、属性1と属性2しか存在しない。そして、それぞれの事例の属性1と属性2の値は、事例  $a$  が(1.0, 1.0)、事例  $b$  が(10.0, 1.0)、そして事例  $c$  が(2.0, 1.0)であるとする。

今、適合度の閾値は0.6に、共有度の基準増加量のパラメータ( $\text{par}_{10}$ )は0(説明の簡略化のため)、重みの基準増加量のパラメータ( $\text{par}_{10}$ )は0.2、重みの初期値は、均等かつ和が1.0になるよう正規化されるので、属性1に対しても属性2に対しても0.5、そして、注目度の初期値はともに1.0に設定されているものとする。

CORESに、まず、事例  $a$  が与えられると、短期記憶に表象されて対象  $a$  となる。CORESの長期記憶にはまだ概念がないので、対象  $a$  はまったく新しいカテゴリーの典型例と認識され、長期記憶に保存されて概念  $A$  となる。次に、事例  $b$  が与えられると、CORESはこれを対象  $b$  とし、対象  $b$  と概念  $A$  の属性値の差異を求め、属性1の差異が9.0、属性2が0となる。そしてこれを基に対象  $b$  の概念  $A$  への適合度を(1.1)式によって算出し、その結果、 $\exp(-81)$ は無視して良いほど小さく、適合度は約0.5となる。この適合度を閾値の0.6と比較すると、閾値未満であるので、対象  $b$  は概念  $A$  と異なるカテゴリーに属すると判断し、(1.4)式によって重みを修正する。その結果、概念  $A$  の属性1の重みは、 $0.5 * \text{par}_{10} + 0.5$ で0.6、属性2の重みは、 $0 * \text{par}_{10} + 0.5$ で0.5となり、和が1.0となるよう正規化されるので、それぞれ、0.55, 0.45となる。この時点では、概念  $A$  と同じカテゴリーで、なおかつ、異なる



属性値を持つ対象がないために、注目度はまだ修正されない。そして対象**b**は、別のカテゴリーの典型例、すなわち、概念**B**として長期記憶に記憶される。

次に、事例**c**が入力されると、対象**c**と概念**A**の属性値差異が算出され、属性1の差異は1.0、属性2の差異は0となる。そして、この差異と、先に求められた属性1の重み0.55と属性2の重み0.45を使用して、対象**c**の概念**A**への適合度が算出され、約0.65が得られる。すると、CORESはこの適合度と閾値0.6を比較し、閾値以上であるため、対象**c**は概念**A**と同じカテゴリーに属すると判断し、(1.3)式によって共有度を修正する（この例の場合は、共有度の基準変化量のパラメータ(*paris*)が0のため、変化しない）。

ここで初めて概念**A**と属性値の異なる正事例と負事例の両方が認識された。よって、注目度は(1.5)式及び(1.6)式によって、対象**c**と同カテゴリーの概念**A**の属性1の差1.0、異カテゴリーの概念**B**の属性1の差9.0（属性2については差がないので修正されない）の適合度に及ぼす違いが最も大きくなるように修正される。特にこの場合、初期値が1.0なので、注目度はそれよりも小さくなり、その結果対象**c**の概念**A**への適合度は大きくなり、かつ、対象**c**の概念**B**への適合度は小さくなる。

以上の処理を終えると再び事例**a**が認識され、続く同様の処理のすべてが行われる。そして、これを繰り返すことによって、属性1の重みは徐々に大きく、かつ、属性2の重みは小さくなり、注目度はより適切な値に修正されていき、最終的にCORESはすべての対象をより正確に認識できるようになる。

#### 第4章 実験1・シミュレーション1 [水野 92c, 93a, 93c]

ここでは、2つの問題に焦点を絞り、実験とシミュレーションの比較によってその妥当性を検討する。1つは、属性の差異の認識の仕方に対する文脈の影響を実現する注目度の変数について、もう1つは、人間の記憶容量を考慮して考案された典型例についてである。

これら2つは、従来のモデルの文脈の影響の問題と記憶の過負荷の問題を解決するために考案された、CORES独自の変数と考え方であり、人間のデータとの綿密な比較が必要だと考えたためである。

第1の、注目度の妥当性を吟味するために、まず、実験1では、課題文脈を操作して2つの文脈条件を設け、典型例と事例の任意の属性値の差異に対する認識の仕方が、人間の場合はどのように異なるのかを明らかにする。そして続くシミュレーション1では、実験と同じ2種類の文脈を与えた場合の注目度の違いを調べ、これを実験結果と比較することで、人間の差異の認識の仕方に対する課題文脈の影響を、CORESが注目度によって実現しているか否かを検討する。

第2の、典型例については、先述したように人間が実際に用いているものを直接的に明らかにする方法はない。しかし、実験1で得られた典型例と事例の属性値の差異の認識の仕方が、想定された典型例を用いた場合のシミュレーション1の適合度と一致するかどうかを比較すれば、その妥当性を間接的に検討することができる。もし人間がCORESの想定した典型例と異なった典型例を用いていれば、実験とシミュレーション結果は一致しないはずである。

### 4.1 実験 1

#### 4.1.1 方法

心理測定関数の考え方によれば、同じ頻度で混同される2つの物理的差異は心理学的には等しい[Lindsay 77]。この考え方に従えば、典型例と1つの属性の値だけが異なる事例がその典型例と同じカテゴリーと判断されるか否かの確率は、その属性値の認識上の差異を表すと考えられる。よって、本実験では、特定の属



性の最小値を有する図形を基本となる典型例とし、様々な属性値を持つ他の刺激図形がそれと同じカテゴリーに属するか否かの判断をさせた場合の分類反応確率を、両者の認識上の差異の指標とする。ただし、CORESでは、冒頭に述べたように、基本的なカテゴリー初頭の典型例以外にも、被験者が分類に必要な典型例を内的に追加すると仮定している。しかし、これを確認ために、被験者が内的に利用する典型例をあらかじめ設定したり、あるいは、外部から観察することは不可能である。したがって、実験1ではこの基本的な典型例だけを提示し、得られた分類反応確率と続くシミュレーション1との比較から、カテゴリー初頭の典型例以外にも、内的な典型例が追加されているという仮定を検証するものとする。

なお、分類の判断に要する反応時間は、概念間の近接性の指標であり[Collins 75]、認識上の差異の大きさを反映すると考えられる。よって、反応時間と分類反応確率の相関関係を調べ、指標としての分類反応確率の信頼性をも確認する。

文脈は、冒頭に述べた成績の例のように、まったく同じ対象が、ある時は小さいものだけで分類され、ある時は大きいものまで含めて分類されるように操作されねばならない。そこで、テスト段階の前に学習段階を設け、属性値の差の小さいものだけを正事例とするフィードバックを与えて学習させる条件と、大きいものまで含めて正事例とするフィードバックを与えて学習させる条件の、2種類の条件を設定して文脈を操作するものとする。

**被験者** 大学生40名を、以下の条件(2)×課題(2)からなる4群に、10名ずつランダムに配置した。

**材料** 2本の直線からなる図形(Fig. 1.4)で、用いた属性は、それまでの研究[水野 92a, 92b]を参考に、第1課題では角度(angle)、長さの比率(ratio)、線の太さ(width)、第2課題では水平移動(horizontal move)、垂直移動(vertical move)、線の太さ(width)、とし、それぞれ最初の角度と水平移動を典型例と比較すべき属性(以下、主属性)とした。

各課題の主属性値、ならびに、学習条件の詳細をFig. 1.5に示す。ここに示すように、角度課題では、典型例(prototype)の主属性値が58°で、刺激図形(stimulus)はそれとの差異が0°から135°まで9°間隔で16個、水平移動課題では、典型例の縦の直線のX座標を0とすると、刺激図形は0 dotから120 dotまで8 dot間隔で16個作成した。

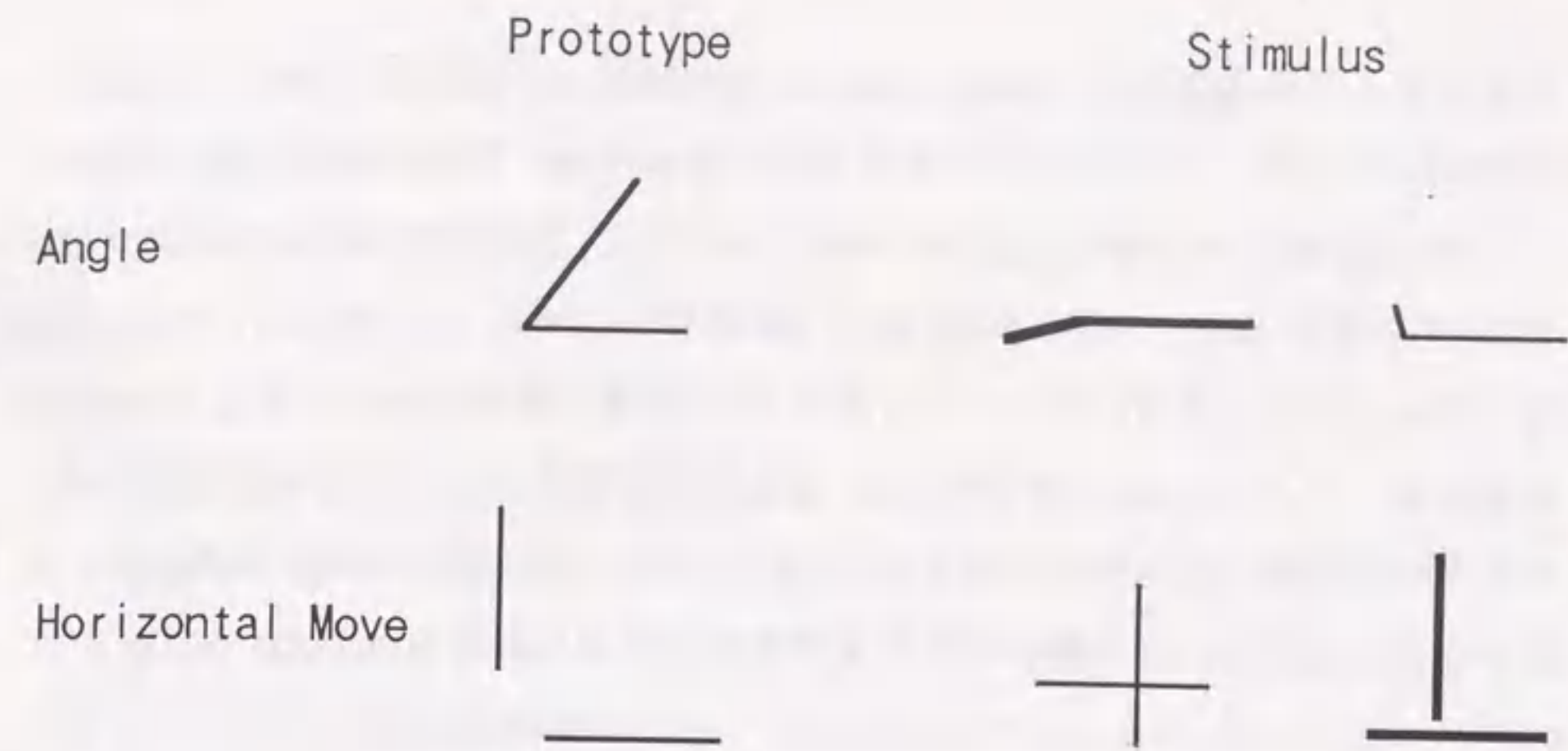


Fig.1.4. Materials in Experiment 1.

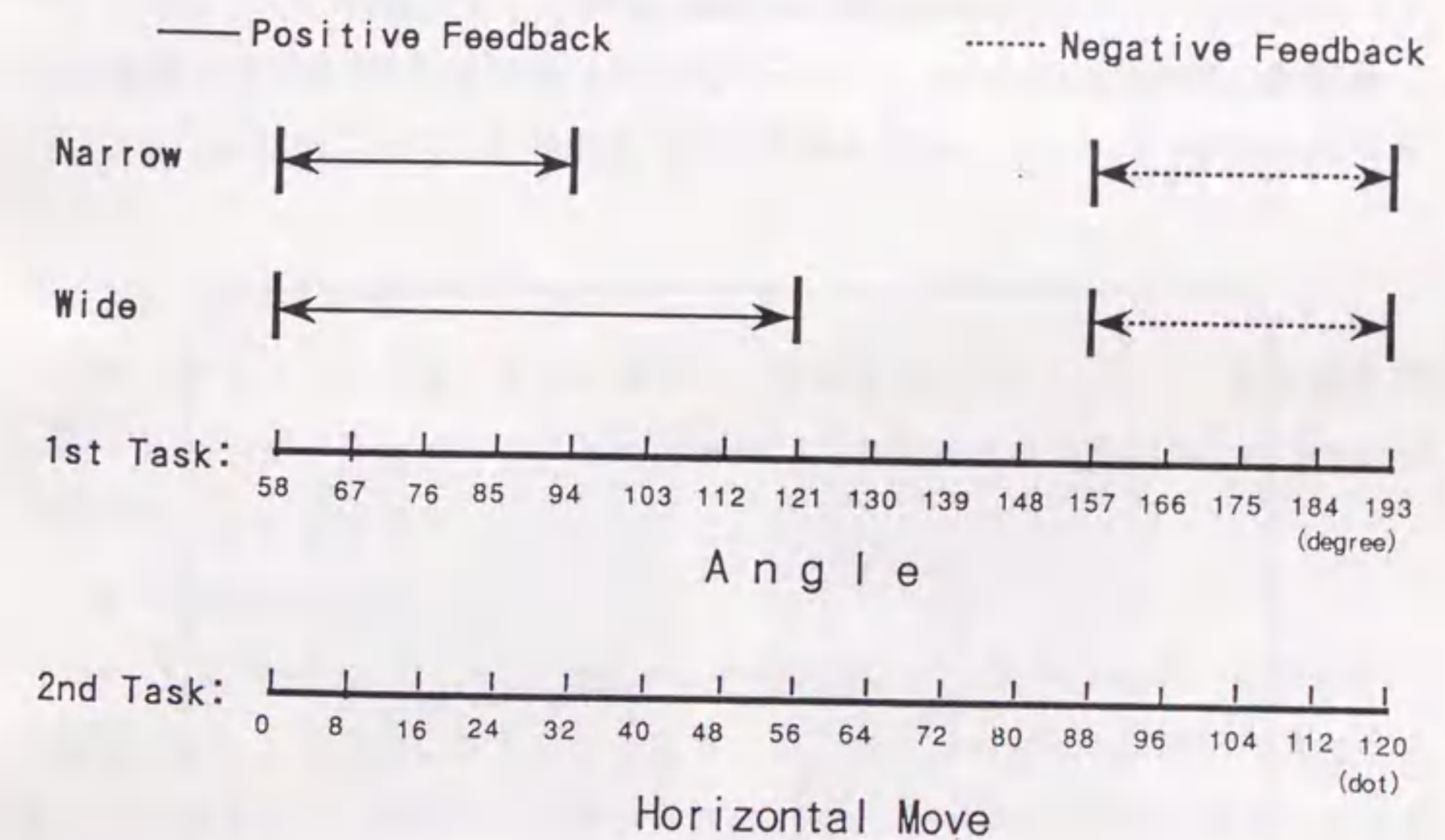


Fig.1.5. Positive and negative feedback ranges of both conditions in the two tasks. Narrow: narrow condition, Wide: wide condition.



そして、学習段階での文脈操作として、典型例との主属性値の差異が、小さいものだけに正のフィードバック(以後、PFB: positive feedback)を与える条件(以後、狭条件, Narrow: narrow condition)と、ある程度大きなものまで含めてPFBを与える条件(以後、広条件, Wide: wide condition)の、2つの条件を設けた(Fig. 1.5)。なお、負のフィードバック(以後、NFB: negative feedback)の範囲は統一されている。したがって、各条件の学習段階での学習図形の種類は、狭条件が10種類、広条件が13種類となる。ただし、この後のテスト段階では、両条件とも16種類すべての属性値を持つ刺激図形が用いられる。したがって、フィードバック(以後、FB: feedback)を受けなかった属性値を持つ図形へのテスト段階での分類反応が、文脈操作の有効性の指標となるとともに、認識の仕方への文脈の影響の指標として最も注目すべき部分となる。

主属性以外のダミー属性については、比率の5種類(1, 0.8, 0.6, 0.4, 0.2)と縦の直線の垂直移動の5種類(0, 18, 36, 48, 72 dot), 線の太さの2種類をランダムに変化させた。これは課題を難しくすると同時に、課題数を増やすためである。

典型例と刺激図形は、Fig. 1.6に示すように、対にしてCRTの左右に提示された。典型例の左右、及び、対にされる刺激図形の順序はランダムである。

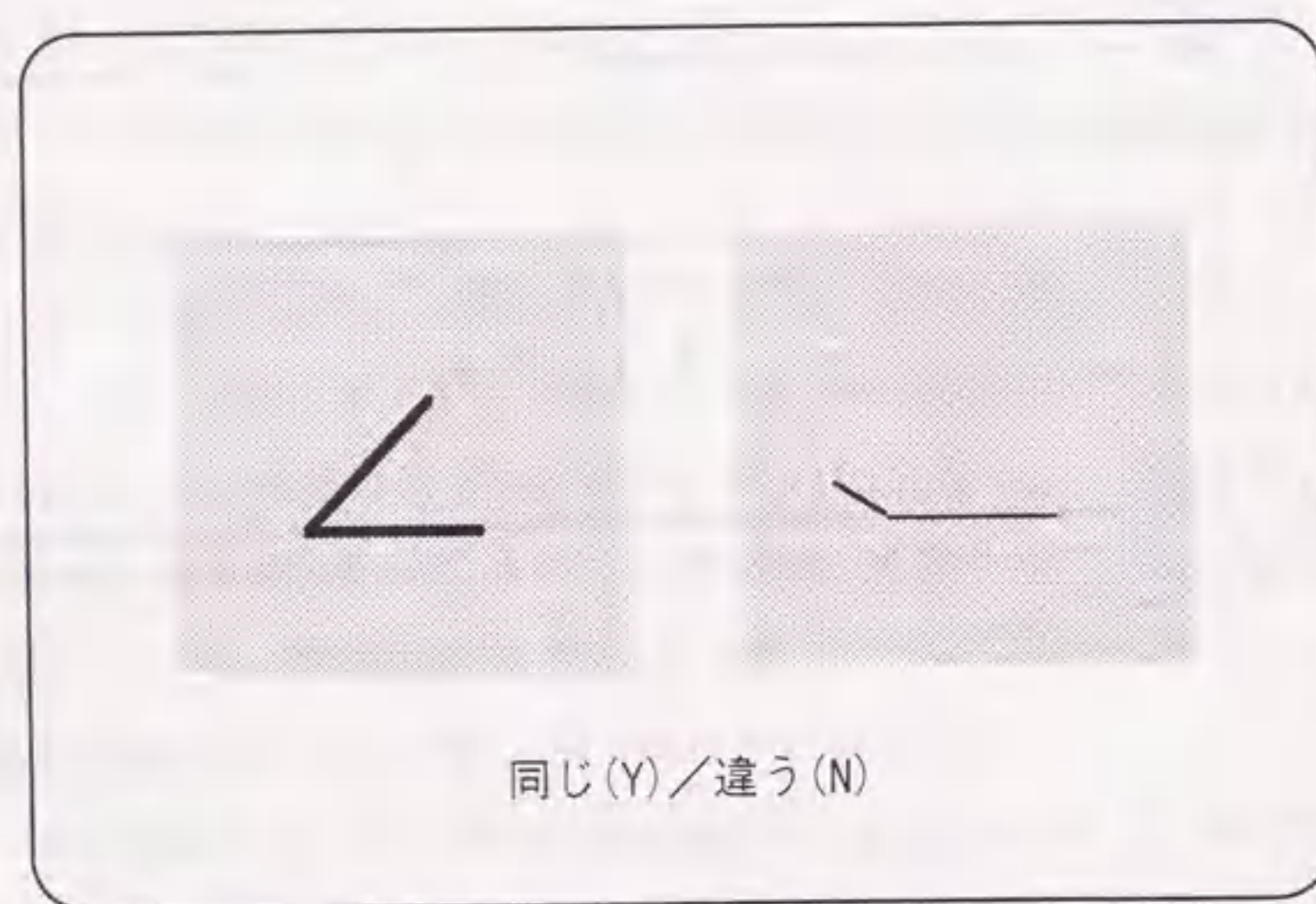


Fig.1.6. An example of paired presentation of a prototype and a stimulus figure.

手続き マイコンによる約40分の個別実験で、教示及び刺激はCRTに提示された。課題は、練習の後、まず、角度課題の学習段階とテスト段階、次に、水平移動課題の学習段階とテスト段階、と続く。いずれも刺激図形が正事例だと判断すればYes, 異なればNoというスイッチを押す。反応はI/Oボード(I-O DATA PIO-9020)を介し記録され、反応時間はカウンター(IWATSU USC-7201)及び40 MHzオシロスコープ(KENWOOD CS-1045)を用いて誤差0.01 ms以下で測定された。

文脈操作のための学習段階では、正解なら1210 Hzと970 Hzの高音が250 msずつ、誤りなら50 Hzの低音が500 msの間提示されると同時に、900 msの視覚FBが与えられて、すぐに次の図形が提示される。学習範囲内の刺激図形のすべてがランダム順で提示されるごとに1セッションとし、正答率が90%以上に達するまで、10セッションを上限として繰り返された。この学習段階は、文脈操作に用いられるだけで、分析の対象とはならない。

分析の対象となるのは、学習後のテスト段階である。この段階ではFBは行われず、両条件とも、16種すべての主属性値を持つ刺激図形がランダム順に提示されるごとに1セッションとし、これを3セッション、計48回の異同判断を行った。反応とともに図形は消され、次の図形提示までの間隔は450 msであった。

#### 4.1.2 結果と考察

分析の対象となるのは、学習段階での分類正答率が90%以上となった被験者だけの、テスト段階の反応である。その人数は、角度課題が狭条件14名、広条件16名、移動課題がそれぞれ11名と13名であった。

##### [1] 分類反応確率

Fig. 1.7, Fig. 1.8 に角度課題、水平移動課題における狭条件、広条件のテスト段階での正の分類反応確率、すなわち、刺激図形が典型例の正事例であると判断した確率を示す。横軸は、典型例と刺激図形の主属性値の差異である。アスタリスクは、狭条件と広条件の分類反応確率に差があるかどうかを、主属性値の差異ごとに $\chi^2$ 検定を行った結果を示す。アスタリスクが2つ(\*\*)ついている部分では、 $\chi^2$ 検定の結果1%水準で有意差が見られ、アスタリスク1つ(\*)の部分では、5%水準で有意差が見られ、そして、ダッガー(+ )のついた部分では、10%水準で傾向があった。これらを見るとわかるように、学習段階でFBを受けなかった部分



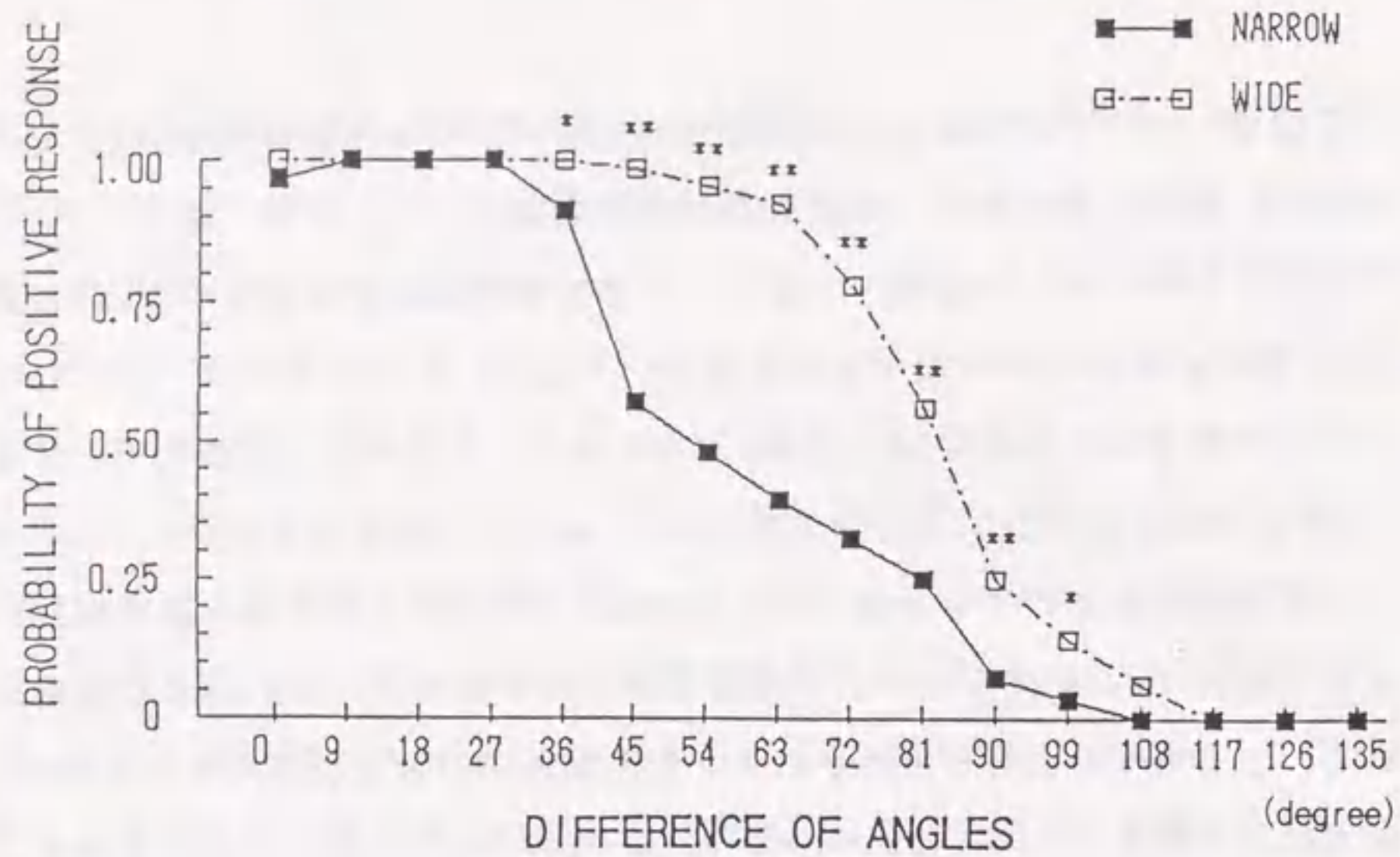


Fig.1.7. Probability of positive response at each difference of angles on each condition and the result of  $\chi^2(1)$  tests.

\*\* $p < .01$ , \* $p < .05$ , + $p < .1$ .

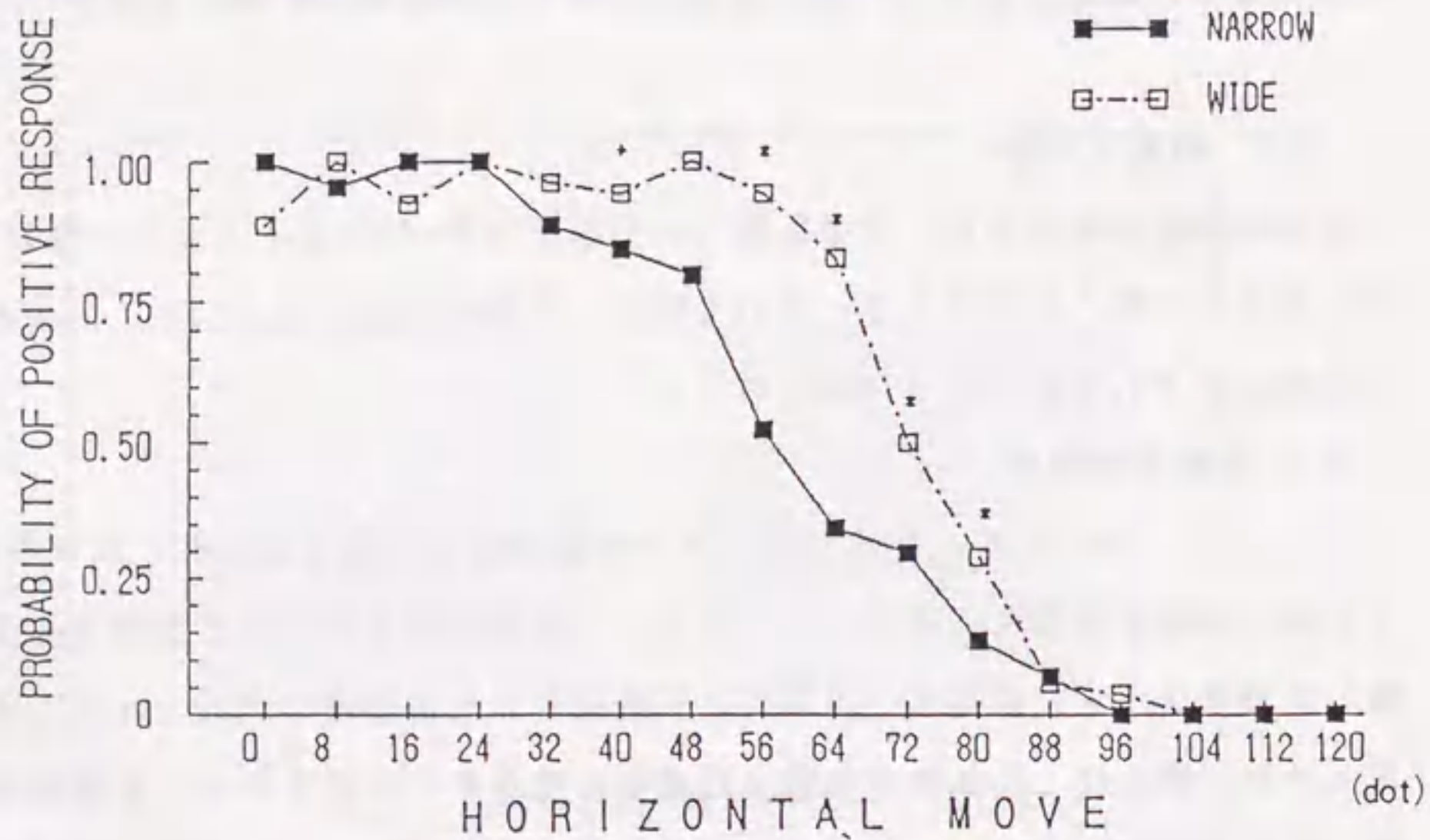


Fig.1.8. Probability of positive response at each horizontal move on each condition and the result of  $\chi^2(1)$  tests.

の $\chi^2$ 検定の結果は、1つの例外を除いてすべて広条件で有意に高く、両条件の分布の形状が異なることを示しており、学習段階での文脈操作が有効だったことを示す。また、FB区間の境界周辺で分類確率が1.0あるいは0になっていないこと、特に角度課題では両条件ともPFBの与えられた差異36'と、NFBの与えられた差異99'の部分でも有意差が見られたことも、確率が1.0と0を結んだ直線的な関係ではなく、曲線的な関係になっていることを示唆している。

このように、両課題に対する分類反応確率の分布が、2種類の非線形関数で表されることは、これらが異なる注目度を持った適合度によって表現される可能性を示している。そしてこの可能性については、次のシミュレーションとの比較によって詳細に検討する。

## [2] 反応時間

Fig. 1.9, Fig. 1.10に両課題、両条件の主属性値の差異ごとの平均反応時間、すなわち、刺激図形を正事例(Yes)、あるいは、負事例(No)だと判断するのに要した時間の平均と、両条件の反応時間が有意に異なるかどうかを見るための、対数変換後の $t$ 検定の結果を示す。どちらの課題でも、分類反応確率が0.5付近での反応時間が最も長く、なおかつ、 $t$ 検定の結果は、いずれかの条件で分類反応確率が急激に低下した部分で有意差があり、分類反応確率と反応時間の密接な関係を示唆していた。そこで、この点を確認するために、分類反応確率の傾きと反応時間との相関係数を求めたところ、両課題とも有意な相関があった(角度： $r = .530$ ,  $t(13) = 2.251$ ,  $p < .05$ ; 水平移動： $r = .939$ ,  $t(13) = 9.828$ ,  $p < .01$ )。

以上の結果から、分類反応確率は、典型例と事例との主属性値の差異の認識の仕方の指標として、十分な信頼性を有することが確認された。

## 4.2 1回目のシミュレーション

### 4.2.1 方法

学習段階では、CORESに、実験の学習段階と同じ2種類の課題文脈条件でFBを与え、かつ、基本的なカテゴリー初頭の典型例だけを用いた学習をさせる。もしこの設定で得られた適合度が、実験で得られた分類反応確率と一致しなければ、実験1で被験者が別の典型例を追加していた可能性を示すことができるからである。



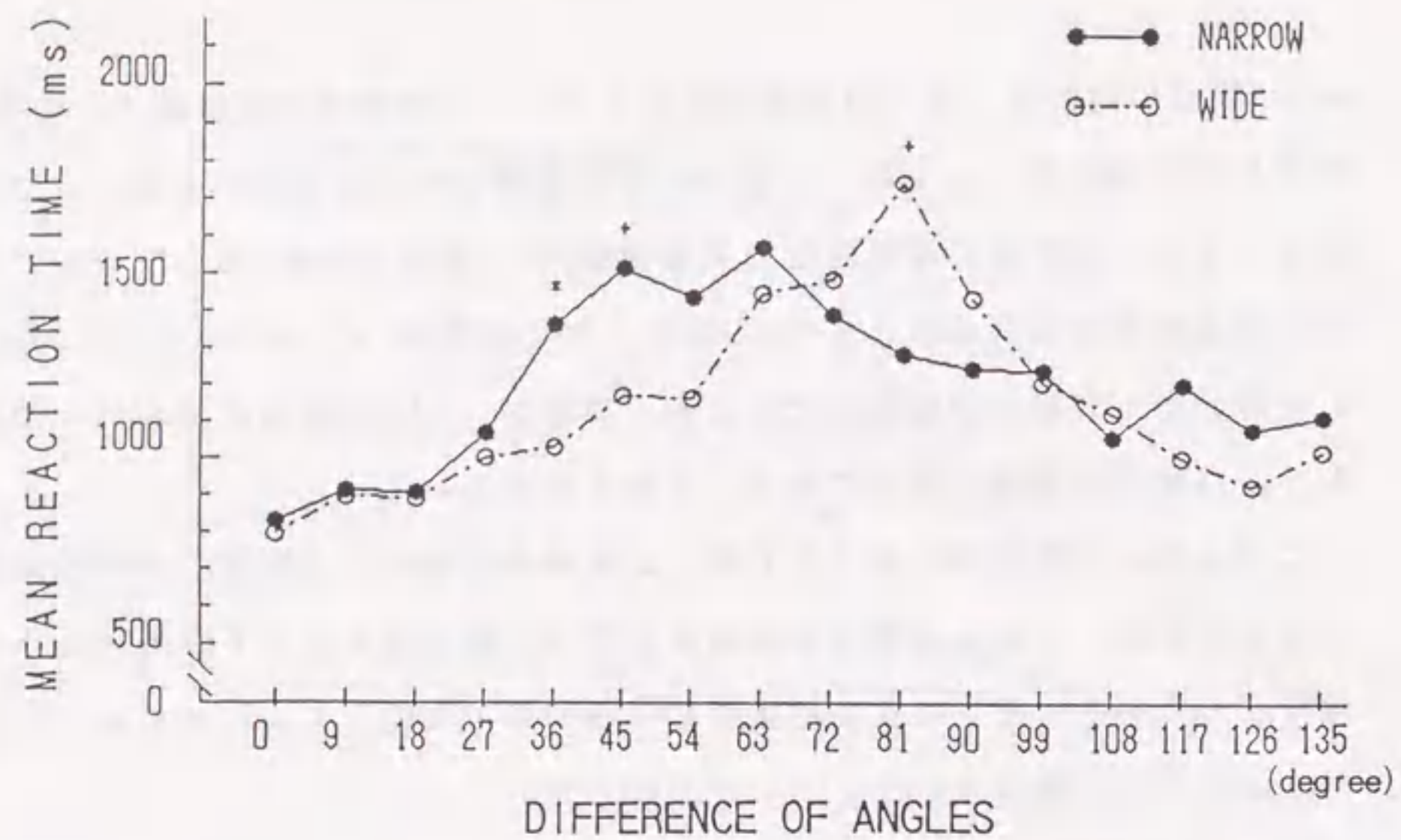


Fig.1.9. Mean reaction time at each difference of angles on each condition and the results of t(28) tests.

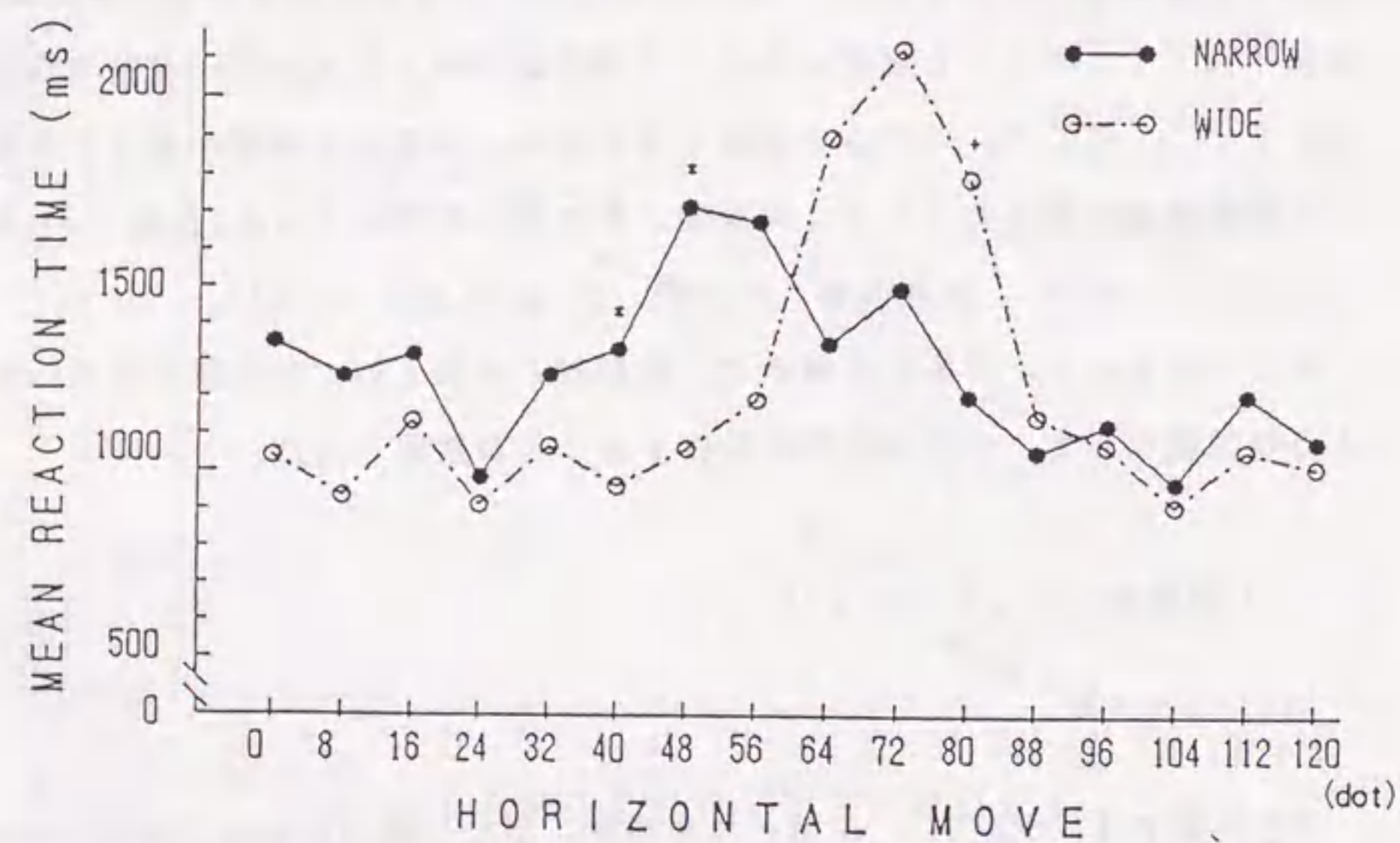


Fig.1.10. Mean reaction time at each horizontal move on each condition and the results of t(22) tests.

基本的な正事例の典型例は、実験と同様最小の主属性値、すなわち、角度課題 58', 水平移動課題 0 dot の事例とし、これを最初に入力する。負事例の典型例は、CORES の仮定通り、実験の学習段階で最初に現れた負事例とし、その主属性値は角度課題 184', 水平移動課題 112 dot である。その上で、実験の学習段階の事例（狭条件 10 個、広条件 13 個）と系列をそのまま用いたシミュレーションを何回か行い（これが学習である）、適合度がほぼ収束した時点を実験の成立とみなす。

テスト段階では、この学習の成立時点での重みと注目度をそのまま使用してシミュレーションを行い、16 事例すべての適合度を算出する。そして、このテスト段階の 2 つの文脈条件での適合度の形が、実験で得られたそれぞれの条件での分類反応確率の分布と一致するかどうか、そして、先に算出された注目度がそれぞれの条件の文脈による違いを表すと言えるのかを調べることで、注目度の妥当性を検証するとともに、典型例を追加する必要があるかどうかを検討する。

なお、共有度基準増加量 ((1.3) 式,  $par_{13}$ )、重み基準増加量 ((1.4) 式,  $par_{10}$ )、修正前の注目度に対する優先度 ((1.6) 式,  $par_{20}$ ) の 3 つのパラメータは、予備的なシミュレーションで学習速度と安定性の観点から吟味され、あらかじめ、0.5, 0.1, 0.2 に設定された。また、主属性と 2 つのダミー属性の重みの初期値は、すべて等しく、かつ、和が 1.0 となるように、そして、注目度の初期値はすべて 1.0 に設定された。

#### 4.2.2 結果と考察

実験で課題間の反応傾向に大きな相違はなく、シミュレーションの基本的過程も同じなので、これ以下は、角度課題のシミュレーションについてのみ報告する。

学習段階のシミュレーションで適合度がほぼ収束したのは狭条件が 7 セッション目、広条件が 6 セッション目であり、それぞれこの時点を実験の終了とみなして次のテスト段階のシミュレーションを行った。

両条件の、学習の終了時点での各属性の注目度 (degree of attention) と重み (weight) を Table 1.2 に、テスト段階のシミュレーションで算出された全事例の適合度 (degree of correspondence) を Fig. 1.11 に示す。これらから明らかなように、適合度の形にも注目度にも、ほとんど条件差が認められなかった。しかも、重みについては、線分の長さの比率はランダム化されているので、線の幅と同様、両



Table 1.2

Degree of Attention and Weight of Each Attribute on Each Condition: First Simulations in Simulation 1

Attribute	Degree of Attention		Weight	
	Narrow	Wide	Narrow	Wide
Width	0.000	0.000	0.012	0.021
Ratio	0.071	0.051	0.191	0.132
Angle	0.060	0.059	0.797	0.847

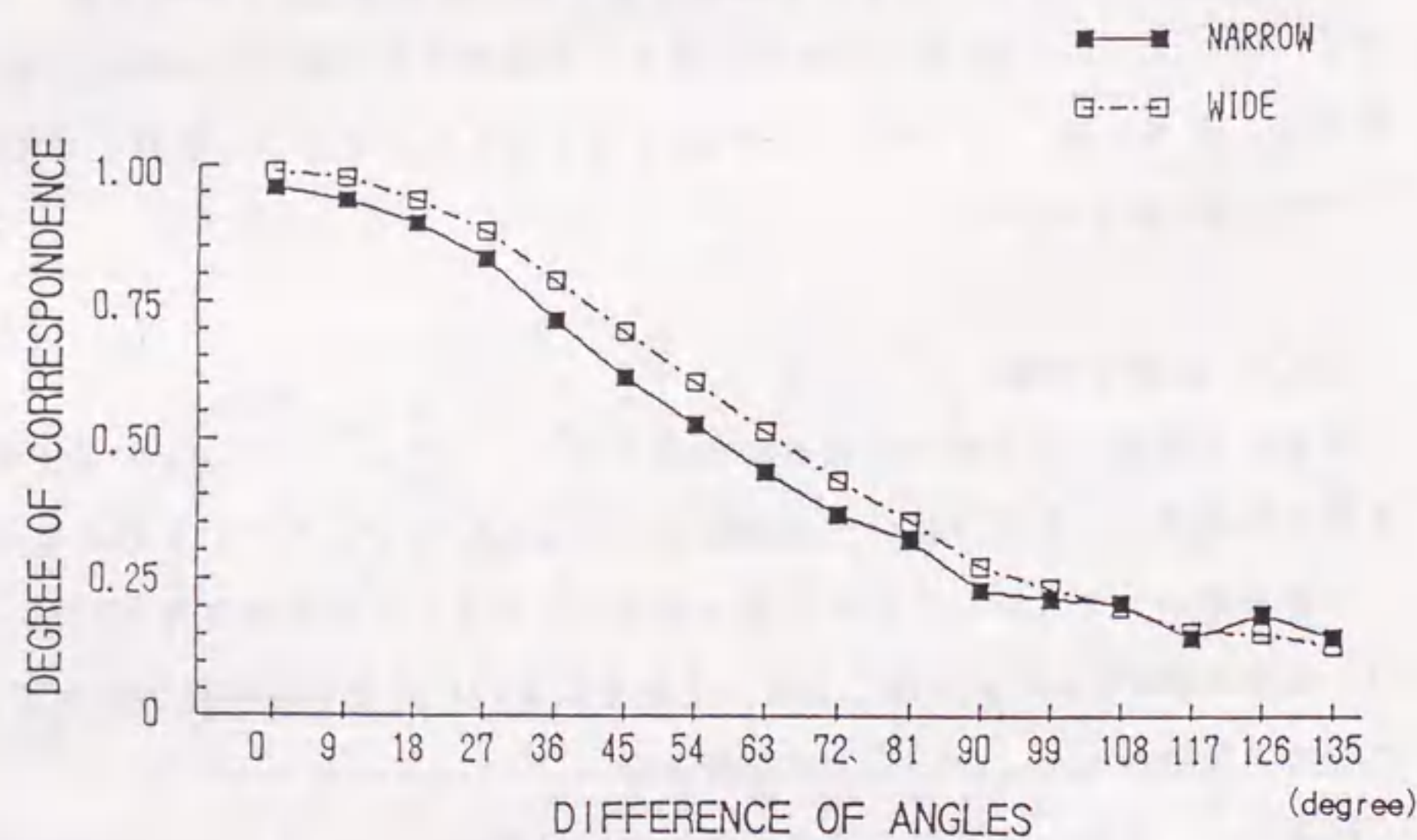


Fig.1.11. Degree of correspondence at each difference of angles on each condition with the original prototype.

条件とも本来は0に近くなるべきにも関わらず、比較的大きい値であり、その傾向は注目度にも認められた。そこでそれぞれの原因を以下のように考察した。

まず、適合度と注目度に条件差が認められなかったことについて考察する。実験で得られた分類反応確率のFB区間内の分布は、シミュレーションによる適合度の形とは異なり、FBの境界の近くまで0, 1.0の2値をとっていた。また、分類反応確率の分布の傾斜は、広条件の方が大きく、CORESの仮定に従えば、注目度が高いはずである。以上の不一致から考察して、FBの境界の近くでは、正事例、あるいは、負事例と、実際に用いられたそれぞれの典型例との距離が、基本的な典型例との距離よりも近かった、すなわち、カテゴリー初頭の典型例だけでなく、FBの境界近くの属性値を持つ図形も、典型例として追加されたと考えられる。

ここで、CORESの仮定通り、典型例が変更されたり平均化されたりしたのではなく、追加されたと考える根拠は、以下の通りである。仮に最初の典型例を否定して別の典型例に変更したとすると、正事例で言えば主属性値の小さい方の、負事例で言えば大きい方の図形の分類反応確率は、より小さくなってはいたはずである。しかし、Fig. 1.7を見るとわかるように、実験結果はそうではなかった。よって次のシミュレーションでは、典型例を追加した場合を実験データと比較し、CORESの仮定の妥当性を検証する。

次に、比率の重みが予想に反して大きかったことについて考察する。比率に対してはランダム配置を行ったが、学習回数の問題から、太さ以外はすべての組み合わせを学習対象にできなかった。その結果、主属性の角度と長さの比率の特定の組み合わせが多くなり、これが比率の注目度及び重みを高める結果となったのではないかと考えられる。この問題に関しては、人間の場合の入力情報の一般化という観点から後に考察するが、現時点で簡単に修正できる性質のものではない。よって2回目のシミュレーションではさしあたり、この偏りを実験者が修正するものとする。

#### 4.3 2回目のシミュレーション

##### 4.3.1 方法

実験で示唆された典型例を追加し、正事例が入力されるごとに、2つの典型例



との適合度を算出し、適合度の高いものをその事例の典型例として処理を進める。具体的には、実験結果では正・負事例ともにFB境界から1つだけ内側で両群に有意差があったことから(Fig. 1.7参照)、これを実現するために、正事例、負事例それぞれについて、両条件ともFB境界から2つ内側の主属性値を持つ典型例を追加するものとする。したがって、正事例の典型例は、狭条件が85°、広条件が112°となり、負事例の典型例は、両条件とも166°となる(Fig. 1.5参照)。

入力情報に関しては、角度、比率、及び太さの組み合わせに偏りが無いような理想的入力に変更する。したがって、(1.5)式の注目度の修正式に従えば、正事例から負事例の境界までの区間は広条件の場合の方が狭く、注目度は高くなること予想される。また、特定の組み合わせに偏りが無いので、ダミー属性の比率と幅の重みは0に近くなること予想される。

#### 4.3.2 結果と考察

以上の修正を施した学習段階のシミュレーションでは、両条件とも10セッション目を学習の終了とみなした。この時点での両条件の各属性の重みと注目度は、Table 1.3に示す通りで、主属性の角度の注目度は、広条件の方が高く、重みは、両条件とも高かった。また、ダミー属性についてはいずれもほぼ0となり、予想と一致した。

適合度の分布はFig. 1.12に示す通りで、この両群の違いは上の注目度の違いによるものであり、Fig. 1.11と比べて、Fig. 1.7の実験データともより類似している。そこで、それぞれの条件の適合度と実験での分類反応確率との相関を求めたところ、狭条件が $r=0.991(t(14)=28.81, p<.001)$ 、広条件が $r=.981(t(14)=19.12, p<.001)$ と、ともに有意に高い相関があり、適合度が分類反応確率をよく近似していると言える。

以上から、分類反応確率の文脈による違いは、適合度における注目度の違いで表されることが明らかとなるとともに、カテゴリ初頭の典型例に加えて、境界近くの典型例が追加されていたことが示唆された。

#### 4.4 総合的考察

Table 1.3

Degree of Attention and Weight of Each Attribute on Each Condition: Second Simulations in Simulation 1

Attribute	Degree of Attention		Weight	
	Narrow	Wide	Narrow	Wide
Width	0.000	0.000	0.007	0.006
Ratio	0.000	0.000	0.014	0.011
Angle	0.079	0.105	0.979	0.984

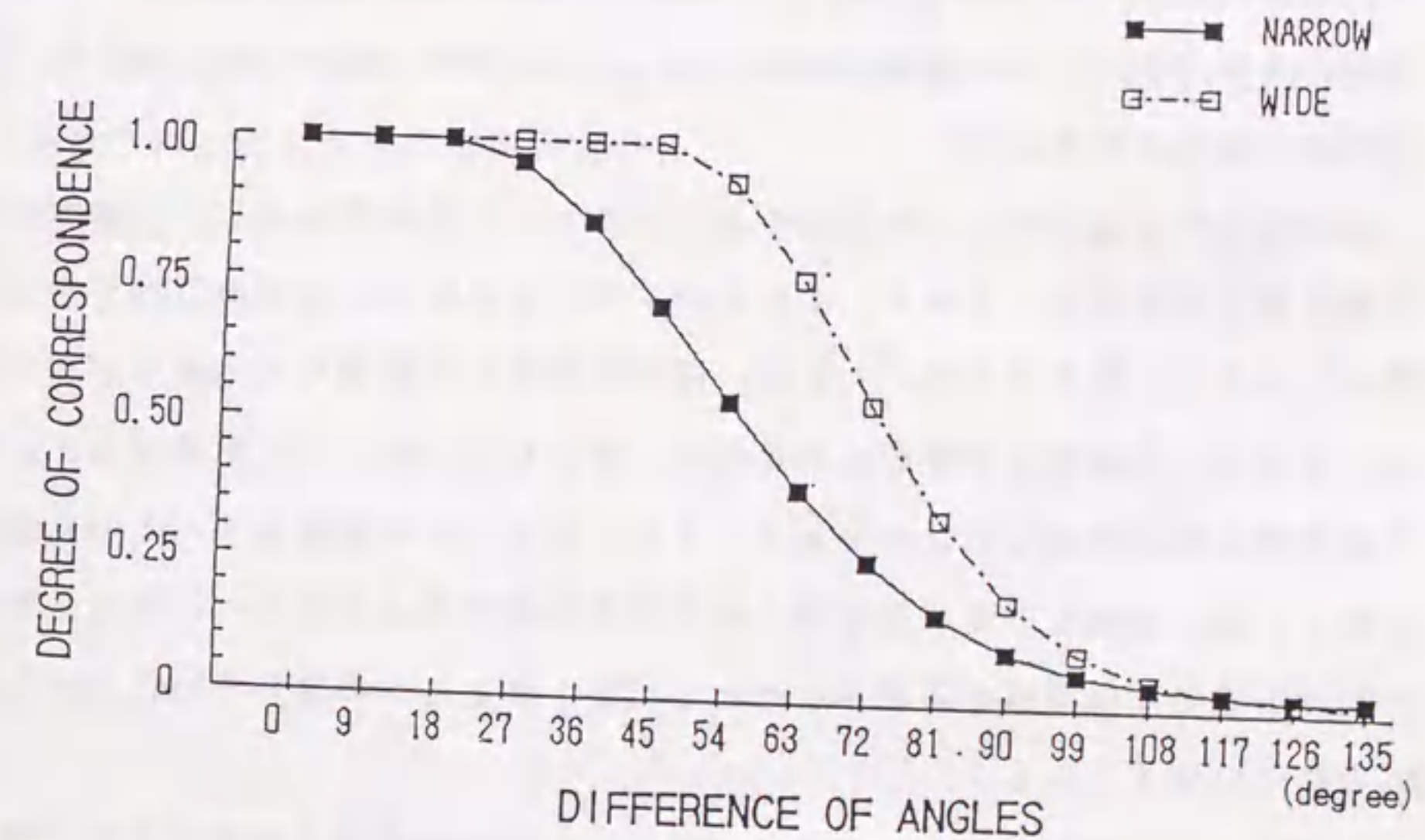


Fig.1.12. Degree of correspondence at each difference of angles on each condition with both the original and the additional prototypes.



文脈の影響と記憶の過負荷の2つの問題は、CORES独自の注目度の変数と典型例の設定によって解決され、その基本的妥当性が示された。以下には、残された2つの問題、すなわち、人間の用いる内的な典型例と、人間独自の入力情報の一般化の問題について述べる。

典型例については、実際に用いられている典型例を直接観察することは不可能でも、本研究で用いたようなパラダイムによって、FBがない場合は、カテゴリー初頭の事例を典型例として選ぶことの妥当性を、そして、実験1のようにFBがあって、しかも、もとの典型例だけではFB通りの正しい分類が不可能な場合には、カテゴリー初頭の典型例に加えて、新たな典型例が追加されることを示唆することができた。特に後者については、何よりも、典型例は1つであるという従来の典型モデルの固定観念をくずす証拠となったという点で意義があると言えよう。そして、現実には、このように単一の典型例では処理しきれない場合は、2.1.1で述べた宇宙人の例のように、比較的多いとも考えられる。

このようなメカニズムを実現するために、CORESでは、入力された正事例とそれまで用いていた典型例の適合度が極めて小さくFBに対応しきれない場合、すなわち、適合度がある閾値よりも低く、なおかつ、重みや注目度を修正してもその適合度を大きくすることが困難な場合には、その正事例を新たな典型例として追加するよう設定されている。

このような方法で新たな典型例を追加すると、入力系列によっては非常に多くの典型例が記憶され、しかも、その中のすべてが必要不可欠な典型例であるとは限らないことが懸念される。しかし、最適な学習には最適な入力順序があるように、適切な入力順序で学習を行わせれば、適切な典型例だけが記憶され、不適切な典型例は形成されないはずである。また、入力された事例をすべての典型例と比較し、最も適合度の高い典型例をその事例の典型例とするモデルを考えれば、一旦は不適切な典型例が形成されたとしても、ほとんどの場合それは2度とは用いられないはずである。

したがって、次に必要となるのは、用いられない典型例を忘却させる方法である。そのためには、各典型例に、参照頻度や時間の経過に基づいた活性度のような指標を持たせることが考えられる。そしてこれについては、第2部の自然言語処理において実現されることとなる。

情報の一般化については、機械は、入力情報のすべてを利用し、かつ、それだけに基づいて処理を進めるので、入力情報がたとえランダムであっても、属性の組み合わせのわずかな偏りが大きく影響するのに対し、人間は、入力に偏りがあっても、一般化、つまり、入力された事例以外にも欠落した事例を補って処理を進めているらしいことが示唆された。

このような入力情報の一般化は、人間が、限られた情報から効率よく合理的に知識を形成するための1つのヒューリスティックスであり、帰納的な思考の根元であるのかもしれない。しかし、これを機械で実現する上では、2つの問題がある。1つは、情報を補うためには、どのような情報を与えておく必要があるのかが明らかでないこと、もう1つは、たとえ情報を与えておくことができたとしても、どのような場合に一般化を行うかというメカニズムについては、ほとんど明らかにされていないこと、である。

しかし、今回の実験とシミュレーション結果から判断しても、このメカニズムが時として非常に重要であることは明らかである。ただし、人間の行う一般化は決して単純なメカニズムによるものではない。なぜなら、情報の偏り自体が、ある意味では、重要な情報だからである。その偏りが重要であるか否かを判断し、重要でない場合には一般化するようなメカニズムを機械において実現するためには、今後さらにこのような心理実験とシミュレーションの比較による知見を蓄積していく必要があるだろう。



第4章では、属性を1つに絞った厳密に統制された課題によって、注目度と典型例を中心とした、CORESの基本的仮定の妥当性を検証した。次に、CORESの包括的な妥当性を検証するためには、これらを含めたCORESのすべての処理が、属性が複数存在する複雑な課題の場合も有効であるかどうか、そして、ここで設定された文脈以外にも適用可能かどうかなど、より幅広い角度から検討する必要がある。

従来のモデルのほとんどは、分類が文脈によって変化することを無視し、一通りの分類しかできなかった。しかし分類が、文脈の1つである課題文脈によって異なることは、すでに実験1でも示された。そこで、CORESがその注目度や典型例を用いて、他の、より多様な文脈のもとでの分類を再現できるかどうかを確認することが、CORESの独自性を立証する上でも、最も重要な方法であり手段であると考えた。

そこで、本章では、より複雑で現実に近い課題を用いて、CORESが、系列差、反復差、個人差、を含めた幅広い文脈への柔軟な適応性を有することを示し、その包括的な妥当性と独自性を検証する。

## 5.1 実験2

本実験は、以下に述べる系列、及び、反復、という文脈の違いによって、実際の人間の分類結果が異なることを検証するとともに、その分類過程をCORESの処理と比較するための詳細なデータを得ることを目的とする。特に、ここで得られる被験者すべての個人データは、続く被験者別のシミュレーションというまったく新しい試みにおいて用いられる。

材料は、Fig. 1.13に示すように、非常に複雑で非組織的な、一般的な図形である。そして、実験1とは異なり、FBは与えられず、分類される事例の数も、カテゴリーの数も制限されない。よって、仮に新たに入力された事例のこれまで用いた典型例に対する適合度が低くても、そのカテゴリーの典型例として追加されるのではなく、最初のCORESの仮定通り、別の、新しいカテゴリーの典型例になると

予想される。

実験における文脈操作と、その影響に関する予想は、以下の通りである。

### [1] 系列効果

2つの異なる入力系列が、分類結果にどのような違いを生むかを明らかにする。ここでCORESから予想されることは、これまでのカテゴリーに属しないと判定されて新たなカテゴリーの初頭となる典型例は、系列順序によって異なってくると考えられるので、分類の結果も異なるであろうということである。

### [2] 反復効果

同じ入力系列による分類を3回反復させることで、分類過程の変化をみる。これは、CORESがあくまでも学習システムであり、反復によって重み、注目度が増加することを想定しているからである。

予想されることは、1回目の試行では、学習が不十分で、各時点での重み、注目度が適切でないため、系列の影響が大きく、分類結果が系列で大きく異なること、一方、3回目の試行になると、学習が進展して重みや注目度がより適切になると考えられるため、分類自体も適切なものになると同時に、系列による影響が少なくなり、系列が異なっても分類結果が類似してくるであろう、ということである。また、続くシミュレーションとの比較においても、十分学習を行った3回目の試行の分類結果の方が、反復の後の適切な重みや注目度の初期値を用いたシミュレーション結果と、より一致したものになるであろうとも予想される。

#### 5.1.1 方法

**被験者** 大学生20名(男子13名, 女子7名)を、ランダムに10名ずつ、A群、B群の2群に配置した。

**材料** Fig. 1.13に示す2次元図形を、各々、150 mm × 105 mmのカードに印刷して使用した。

**手続き** 別の材料で練習試行を行った後、上の材料をA、Bの2群に異なるランダム順で、系列的にカードを提示し分類させた(Table 1.4)。典型例は、カテゴリー初頭に出現する事例だと想定されている。よって、カードをカテゴリーごとに別の山に置かせ、各山の先頭のカードを典型例とし、それ以外は見えないようにするために、逐次その下にカードを置かせた。系列の影響に加えて反復の影響を



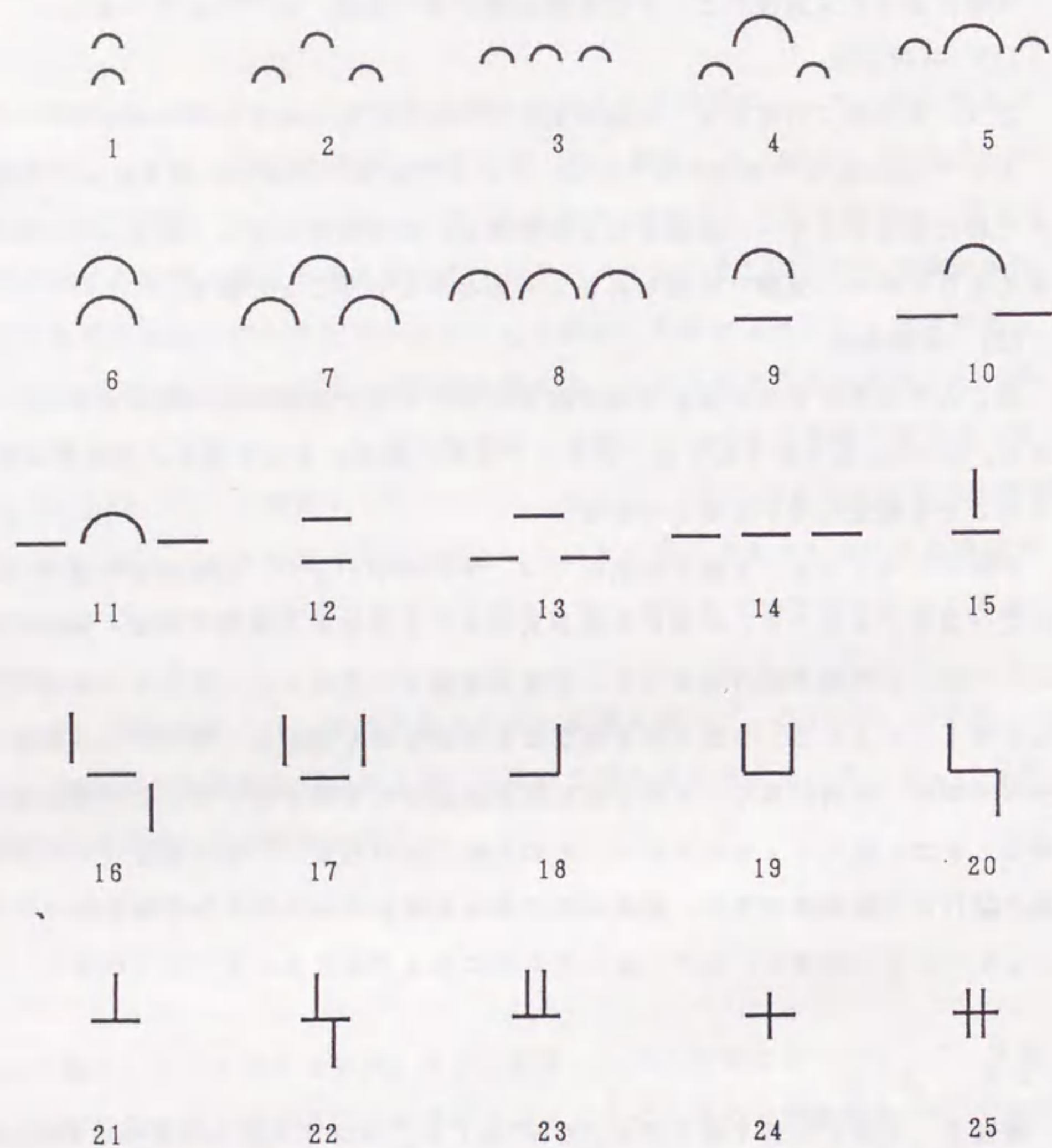


Fig.1.13. Materials for Experiment 2.

Table 1.4

Input Order of Group A and Group B

Figure Numbers	
Group A;	16, 21, 8, 5, 15, 4, 23, 2, 1, 20, 17, 6, 13, 12, 9, 7, 10, 18, 22, 14, 3, 11, 25, 19, 24
Group B;	14, 13, 19, 18, 12, 4, 3, 2, 1, 23, 22, 21, 11, 25, 10, 9, 8, 7, 6, 5, 16, 20, 17, 24, 15

見るために、試行は連続して3回行わせた。そして、最後に各カテゴリーの分類基準のプロトコルをとるとともに、実験全般についての内省報告をさせた。

#### 5.1.2 結果と考察

##### [1] 系列効果

ここでは、情報や知識がなく、系列の効果が最も正確にとらえられる1回目の試行の分類結果を分析した。

両系列群のカテゴリー数は、A系列11.3(1.08:以下、括弧内はSD)、B系列9.2(1.70)で、その差をt検定で比較したところ、有意な差があり( $t(18)=7.10$ ,  $p<.001$ )、A群のカテゴリー数がB群のものより多く、系列によって異なることが確認された。

次にこの系列による違いが、単純な系列位置だけに影響された結果であるかどうかを見るために、系列上の距離と同一カテゴリーに分類される頻度との相関係数を求めたところ、両群とも有意ではなく( $r=.16$ ,  $ns$ : non-significance;  $r=.09$ ,  $ns$ )、系列上の距離が近いために同じカテゴリーに分類されたわけではないことが確認された。

そこで、これが典型例の違いによるものかどうかを検討するために、まず、A・B両群で典型例となる頻度の高かった図形を比較した。典型例となる確率が80%



以上だった図形を系列の前から見ていくと(Table 1.4参照), A群では, 図形16, 図形21, 図形17, そして図形13, B群では, 図形14, 図形19, そして図形4とすべて異なり, 重複は1つもなかった. さらに, これらの典型例(Fig. 1.13参照)とTable 1.4とを照らし合わせると, 典型例と典型例以前に分類された図形とは, 例えば, A系列の典型例の図形21とその前に提示された図形16, 典型例の図形13と図形6, といったように, どれも非常に異なっており, 与えられた対象が, それまでのカテゴリーのいずれにも分類されない場合は, それが新たな典型例として記憶されるというCORESの仮定の妥当性を示唆していた.

次に, この点をより客観的に, 詳しく検討するために, 両群の分類に対してクラスター解析を行った. Fig. 1.14にはその結果が示してあり, 縦が図形番号, 横軸が類似性で, 図形番号の近くでクラスタリングされているほど, その図形同士が同じカテゴリーに分類される傾向が高いことを示す. その結果, A, B両群のクラスターの違いは, 特に図形15から図形25において顕著であり, 以下のような典型例の影響が認められた. 例えば, 図形23は, A群では図形15や図形21と同じクラスターにあり, これは図形21が高い頻度で典型例に選ばれ, これらの間に, 1つの構成要素の中央で接近し, 2つの構成要素の成す角度が90°である, といった特徴が共通の属性となったためだと考えられる. 一方B群では, その図形23は図形17や図形19と同じクラスターにあり, 上と同様, これは図形19が典型例となる頻度が高く, 3つの構成要素から成り, そのうち2つの構成要素の成す角度が90°である, といった特徴が共通属性となったためだと考えられる.

また, 次のようなケースも認められた. 図形20はほとんどすべての属性において, 図形16と極めて類似しており, 一見して同じカテゴリーに属する. しかし, B群では, 図形20は図形25と同じクラスターにあった. これは, B群では図形16が典型例として選ばれる頻度が低かったためだと考えられ, 適切な典型例が選択されないと, 非常によく似た図形さえも同じカテゴリーに分類されないという, 典型例の決定的な影響を示すものだと考えられる.

しかしながら, 系列や典型例の違いによって影響されなかったクラスターも存在した. それらは, 図形1と6, 図形2, 4と7, 及び図形3, 5と8から成る3つのクラスターで, いずれの図形も構成要素に曲線を含んでいた. これから示唆されることは, 曲線は, 他の特徴に比べて識別上極めて重要で, しかも, その重要性は

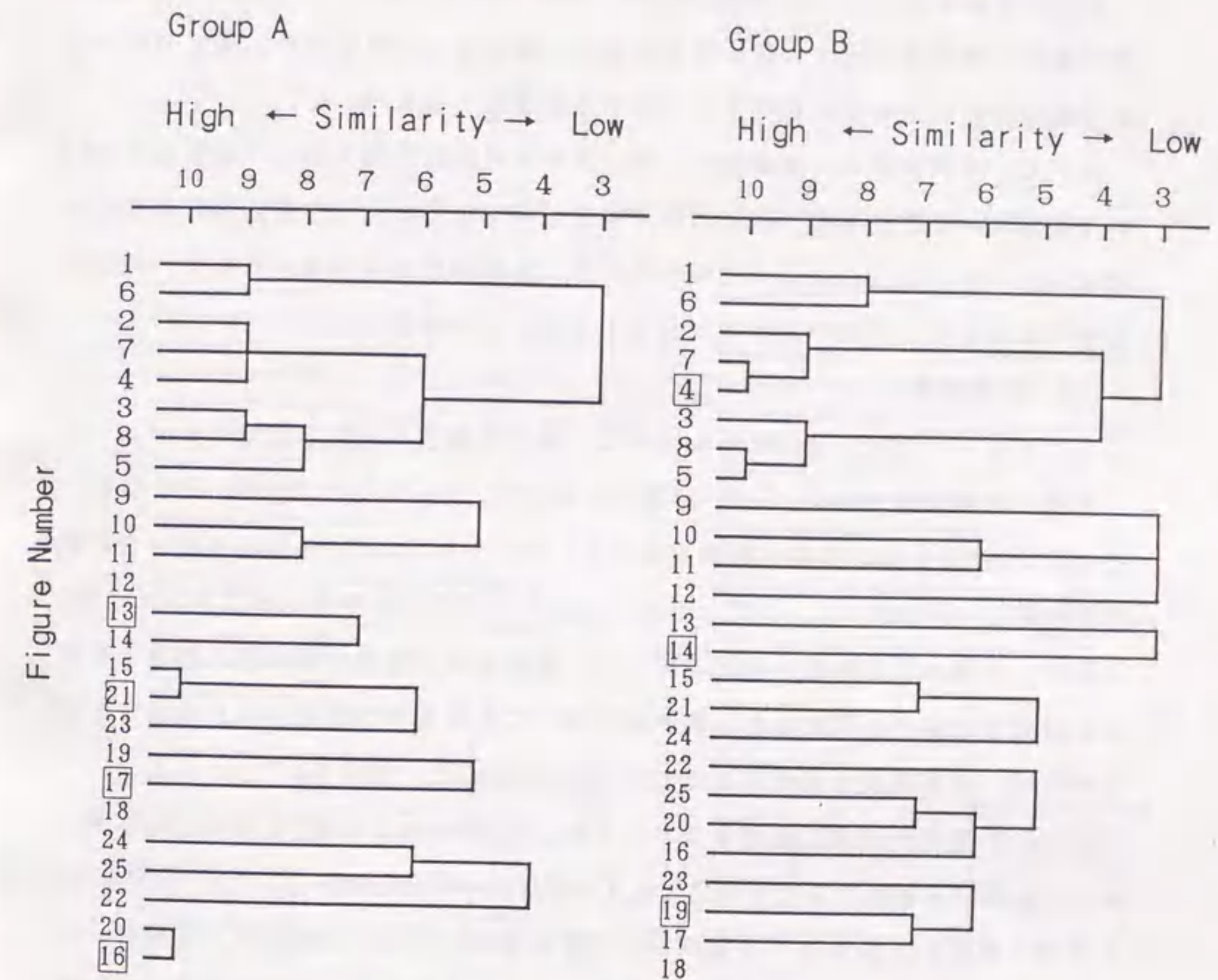


Fig.1.14. Clusters of the first incremental categorizations. Framed figure numbers indicate that the probabilities of their being selected as prototypes were more than 0.8.



人間の認識において本質的なものだということである。そしてこれを、CORESに照らして考えれば、曲率という属性の重みや直線か曲線かという曲率の属性値の差異への注目度の初期値が、他に比べて高い、という説明が可能になる。

よってこれらはすべて、CORESの仮定、特に、各カテゴリーの初頭の事例が典型例となり、その典型例との類似性が分類を決定するという仮定や、属性の重みや注目度の存在についての仮定を、支持する結果だと言える。

ただし、各群の20名の被験者のうち、まったく同じ分類を行った被験者は1名もいなかった。したがって、統計的検定によって上記のような一般的傾向の存在は認められても、個人によって分類が異なることは明らかであり、これを一通りの分類しかできない従来のモデルで説明することは不可能である。

### [2] 反復効果

ここでは、3回目の分類試行を分析し、先の1回目の分類と比較する。

3回目の両群のカテゴリー数はA系列8.0(.88)、B系列7.9(.75)で、先と同様、その差をt検定したところ、有意差はなく( $t(18)=.26, ns$ )、クラスタリングの結果も収束しており(Fig. 1.15)、1回目に比べて系列による違いが減少していた。これは、反復による学習の進行に伴って、被験者の全属性の評価がより適切なものになったためだと考えられ、典型例が違って重みや注目度がより適切なものになれば、分類結果は類似するというCORESの仮定と一致する。

また、1回目に系列の影響を受けにくかった図形1から図形8までから成るクラスターには、3回目でも大きな変化は見られなかった(Fig. 1.14, Fig. 1.15参照)。これは、曲線か直線かという視点が、学習する以前から、分類に大きく関わっていたことを示唆しており、曲率という属性や曲線と直線の違いは潜在的に高い重みと注目度を有していたのではないかという、1回目の分類試行における考察を裏づける結果だと言える。

ただし、3回目の分類ですら個人間の差異は大きく、まったく同じ分類をした被験者は1名もいなかった。したがって、属性の評価が全般的に適切なものに変化しても、依然として個人差は存在することが明らかとなった。

### [3] 同時入力による分類結果との対比

同時入力は、冒頭に述べたように、多くのモデルで採用されている。しかし、現実には、人間がすべての属性や属性値を分類まで保存しておいて、最後に一度

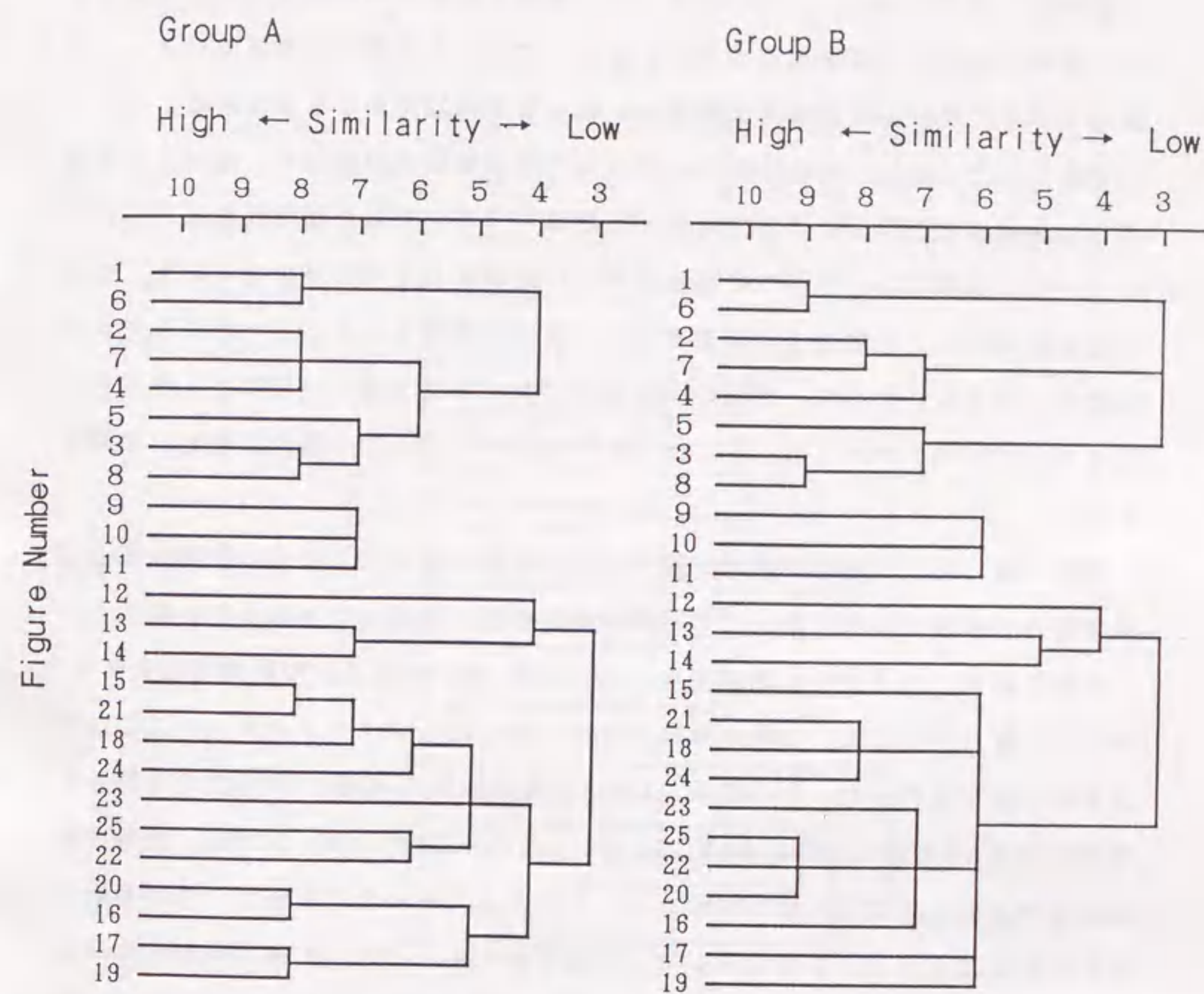


Fig.1.15. Clusters of the third incremental categorizations.



に分類するとは考えられない。よって、この点を確認するために、全材料を一度に提示し、すべてを見せた上で分類させる補足実験を行い、これまでの系列入力による実験結果と比較した。

ただし、人間の場合、上記の理由から、厳密な意味での同時入力は不可能であり、様々な系列入力が非規則的に繰り返された結果の分類だと考えている。したがって、ここで行った同時入力実験は、あくまでも補足的なものである。

予想されたことは、同時入力と系列入力では分類結果が異なり、同時入力の場合の方が個人差が少ないことである。なぜなら、この実験では材料数は限られているので、被験者は、すべての属性を同時に評価するまでは至らなくとも、少なくとも系列的入力の場合よりは公平かつ客観的に評価することはできると考えられたからである。そして、同様の考えから、この予備実験の最後にとられる各被験者の分類基準のプロトコルを、後のシミュレーションでの属性の設定の参考とした。

被験者は、別の大学生36名で、以下の分析では、同時入力による分類結果と、本実験の2種類の異なる系列での系列分類の結果を合わせたものとを比較した。

分析の結果、カテゴリー数は平均 7.8(2.62) で、系列入力の両群を込みにした場合と比較したところ、分散が等質でなかったため、コ克蘭・コックスの法による差の検定を行った。その結果、1回目の両系列群の平均とも3回目のそれとも有意差があり( $t'(19, 35) = 4.85, p < .01$ ;  $t'(19, 35) = 1.76, p < .05$ )、同時分類の結果の方がカテゴリー数は少なく、一応予想と一致した。しかし、その分散からも明らかなように、予想に反して、同時分類だからといって個人差が少ないとは言えず、全く同じ分類をしたのは40名中2組4名だけであった。そしてクラスター解析の結果は(Fig. 1.16)、系列入力のクラスター解析の結果(Fig. 1.14, Fig. 1.15)と比べて、類似性が均等化していた。つまり、極めて類似しているクラスターもあまり類似していないクラスターもなく、どのクラスターも平均化していた。

これらの結果は、同時分類では、平均値としての類似性の評価は均等化した。これは個人内の属性の評価が幅広く分散したためであることを示唆する。

人間の場合の同時入力は、先に述べたように、見方を変えれば、様々な、そして時にはランダムな、系列入力の繰り返しだと考えられる。だとすれば、個人差が少なくなるのも当然なのかもしれない。また、このことからしても、同時

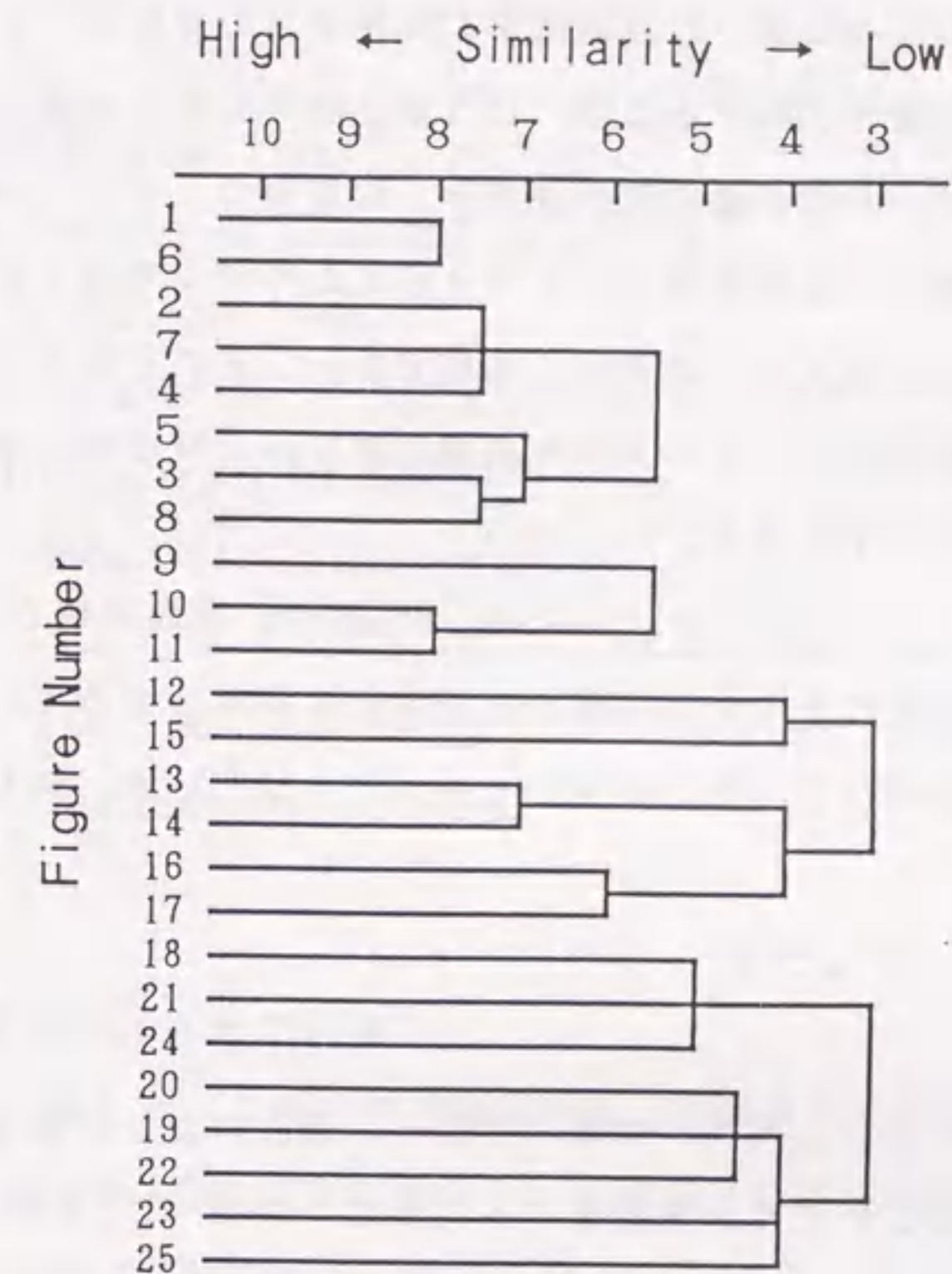


Fig.1.16. Clusters of the simultaneous categorizations.



入力による分類をモデル化することは意味がなく、また不可能でもあると考えられる。

以上、分類の系列による違い、反復による違いだけでなく、入力方式による違いも認められ、文脈による分類の多様性が確認された。しかも分類には個人差が大きいことが改めて確認され、人間の認識がいかに多様であるか、そしてこれを従来のモデルのように単一の評価基準で実現するのが難しいかが端的に示された。

個人個人の分類は多様であるが、どれが正解でもどれが誤りでもないはずである。だとすれば、一般的傾向を反映でき、なおかつ、その各々の分類を再現できるモデルが、本当に分類過程の本質をとらえたモデルではなかろうか。そしてそうするためには、個人がどの重みを重視するか、どの差異に注目するか、そして、どの事例を典型例としてカテゴリーを構成するか、などが反映できるモデルでなければならないはずである。

そこで、次のシミュレーションでは、CORESが、実験で得られた入力系列による分類の違い、反復による分類の違いを再現できるかに加えて、被験者個々の分類をもとに、被験者の個人差を実現できるかどうかという、全く新しい試みを行う。

## 5.2 シミュレーション2

実験2で得られた、系列、反復、そして、個人によって異なる分類が、CORESによって再現できるかどうかを確認することで、CORESの文脈に即した多様な分類への適応力と妥当性を検証する。ここで行うのは、本研究独自のまったく新しい試みとしての個人のシミュレーションである。したがって、試行の反復による分類の違いも個人内比較となる。

CORESによる処理は、A・B系列とも基本的過程にはほとんど違いがない。よって、ここでは、A系列に焦点を絞り、その中から任意の被験者のデータの、1) 教師ありのシミュレーション、2) 教師なしのシミュレーション、の結果を報告し、最後に、すべての個別シミュレーションの集計と個人差について詳述する。

なお、シミュレーションで使用される属性はTable 1.5に示す通りである。これらは先述したように、実験2の同時入力の実験における被験者のプロトコルを参考に選定されたものである。

Table 1.5  
*Supposed Attributes*

No. Attributes
1: Radius
2: Direction
3: Distance from <sup>a)</sup>
4: Crossed or Closest Point to <sup>a)</sup>
5: Direction toward <sup>a)</sup>
6: Difference of Directions <sup>a)</sup>
7: Difference of Radii <sup>a)</sup>
8: Number of Components <sup>b)</sup>

Note: <sup>a)</sup> Decided relative to a certain other component.

<sup>b)</sup> Decided par a figure.

### 5.2.1 教師ありのシミュレーション

まず、A系列で分類を行った任意の被験者Aの、教師ありのシミュレーション (supervised simulation)の結果を報告する。教師ありの場合はCORESに被験者の分類の仕方がフィードバックされる。属性の重みの初期値はすべて等しく、かつ、和が1.0となるように設定され、注目度の初期値はすべて1.0に設定されている。

Table 1.6からTable 1.12は、以下に詳述する様々な設定でのシミュレーション結果で、左のコラムには典型例の図形番号が、中央のコラムにはカテゴリーの成員の図形番号と左の典型例との適合度が書かれている。右のコラムには、中央の適合度よりも他のカテゴリーの典型例との適合度の方が高くなってしまった場合に限って、中央のコラムから右のコラムにかけて、その典型例の図形番号と適合度に下線がついている。したがってこれ以下のすべての表の下線のある部分は、実験データから得られたカテゴリーとシミュレーションとの不一致を示す。最下欄には算出された各属性に対しての注目度が書かれている。



[1] 1回目の試行のシミュレーション

Table 1.6に、重みと注目度を固定した場合のシミュレーション結果を示す。したがってこの場合のパラメータは、共有度基準増加量((1.3)式,  $par_{1g}$ )と重み基準増加量((1.4)式,  $par_{1o}$ )が0, 修正前の注目度に対する優先度((1.6)式,  $par_{2o}$ )が1.0に設定されている。このように、重みと注目度を固定してしまうと、不一致が多く、これらを修正する必要があることがわかる。

そこで、シミュレーション1と同様、これまでのシミュレーション上の学習速度と安定性の観点から、共有度と重みの基準増加量のパラメータを0.1に( $par_{1g}$ ,  $par_{1o}$ ), 修正前の注目度の優先度パラメータを0.2に設定し( $par_{2o}$ ), 重みと注目度を修正しつつ分類した場合の結果がTable 1.7である。Table 1.6に比べて明らかに不一致が少なくなっている。

次に、Table 1.8は、20回学習後の重みと注目度の初期値を導入した場合で、ほとんどの不一致が解消された。このように20回学習させた後の値を使用したのは、1つには、実験2で曲率の重みや注目度が初めから高かったことから示唆されたように、人間の場合は、分類を開始する初期状態においても重みや注目度に対して潜在的バイアスを有しているため、そのような初期状態を学習によって成立させようと考えたからである。また、もう1つは、学習速度の問題である。シミュレーションモデルでは、例えば、重みについても、(1.4)式のように一定量ずつしか変化させることができないために、新しく入力された事例によって大幅な重みの変更は不可能である。しかし、人間の場合は、自分自身の内部に設定された仮説に従い、よりヒューリスティックで大胆な修正を行っている可能性があり、現在のCORESではその修正過程自身をシミュレートすることは困難である。そこで、初期値を、被験者が所有している最終的な値により近いであろう値に設定することによって、この学習速度の差を多少でも減少させようとしたからである(このようなヒューリスティクスに関しては後に詳しく考察する)。

さらに、Table 1.8でも不一致の残ったカテゴリー1とカテゴリー2に関して詳しく検討した結果、この部分に被験者側の分類上の矛盾が見いだされた。それは、1) 図形16と20を同じカテゴリーに分類するのであれば、図形19は図形18ではなく、図形17のカテゴリーに分類する方が妥当であり、2) 構成要素同士が離れているか、接しているかで図形を区別するなら、図形15をカテゴリー2に分類する基準が明

Table 1.6  
Supervised Simulation of First  
Categorization: Fixed Weights  
and Degrees of Attention

No.	Prototype	Fig.	Doc. <sup>a)</sup>	Other Prototype	Doc.
1	16	20	(0.8572)		
2	21	15	(0.8889)		
		23	(0.2460)	16	(0.5768)
		22	(0.2460)	16	(0.6349)
3	8	5	(0.8574)		
		4	(0.7147)		
		2	(0.7147)		
		1	(0.2383)	21	(0.3338)
		6	(0.3333)	21	(0.3337)
		7	(0.8573)		
		3	(0.8574)		
4	17				
5	13	12	(0.3361)	21	(0.5560)
		14	(0.7681)	8	(0.8571)
6	9	10	(0.2070)	13	(0.8560)
		11	(0.2068)	8	(0.8096)
7	18	19	(0.1918)	17	(0.8571)
8	25	24	(0.3360)	21	(0.8889)

Table 1.7  
Supervised Simulation of First  
Categorization: Modification of  
Weights and Degrees of Attention

No.	Prototype	Fig.	Doc.	Other Prototype	Doc.
1	16	20	(0.8704)		
2	21	15	(0.8889)		
		23	(0.2788)	16	(0.5635)
		22	(0.9191)		
3	8	5	(0.8574)		
		4	(0.7395)		
		2	(0.7974)		
		1	(0.3169)	21	(0.3339)
		6	(0.6320)		
		7	(0.9345)		
		3	(0.9420)		
4	17				
5	13	12	(0.7732)		
		14	(0.9135)		
6	9	10	(0.7991)		
		11	(0.7673)		
7	18	19	(0.7579)	21	(0.8574)
8	25	24	(0.9657)		

Note. <sup>a)</sup> Doc.: Degree of correspondence.

Attribute : Degree of Attention  
 1 : 0.13112 2 : 1.00000  
 3 : 0.83293 4 : 0.11046  
 5 : 0.04135 6 : 1.00000  
 7 : 0.65030 8 : 0.01563



確でなく、その場合はむしろ図形15, 16, 17の категорияができるはずである、という2点である(Fig. 1.14参照)。そこで、構成要素が離れているか接しているか(属性3)という分類基準を厳密にして被験者の矛盾を解消し、図形16に図形15と17が分類されるよう分類の仕方を修正した。その修正された分類のシミュレーション結果がTable 1.9である。ここでは、すべての不一致は解消されており、CORESの妥当性が検証されたと言えよう。

また、算出された注目度は、属性値の違いへの注目の度合いを的確に示していた。例えば、Table 1.8において、属性2と6の注目度は1.0で、被験者が図形の構成要素の方向を完全に考慮に入れたことを、属性8の構成要素の数のそれは0で、数に関しては無視されたことを、そして属性3はその中間であり、構成要素が離れているか接しているかは部分的に考慮されたことを示している。そして、この例に限らず、以下でも一貫して、注目度が文脈に応じて微妙な値をとることを考えれば、特定の属性の注目度をあらかじめ設定して固定的な値として入力すべきではないこと、そして、その方法もないことは明らかであり、注目度を内部修正することの必要性和妥当性が示されたと言える。

#### [2] 3回目の試行のシミュレーション

実験2における3回目の分類では、被験者の全属性の評価がより適切なものに収束した。したがって、1回目の分類で見られたような分類上の矛盾は少なくなっていると考えられる。そこで、そのような矛盾がない場合には、実験者による修正がなくてもCORESが正確に分類を再現できることを確認するために、1回目と同じシミュレーションを行った。その結果、1回目のシミュレーションと同様、重みと注目度を固定した場合は不一致が見られ(Table 1.10)、修正の必要性が再確認された。そして、重みを修正したTable 1.11では不一致は少なくなり、学習回数を増すごとに不一致は徐々に解消された。そして最後の、20回学習後の重みと注目度の初期値を使用したTable 1.12ではまったく不一致が認められず、しかも、適合度は一様に0.9を上回り、特に、1回目に矛盾の見られた、離れているか接しているかの属性3の注目度は増し、より明確に基準化されていることを示していた。

以上、3回目のシミュレーションでは実験結果とシミュレーション結果は完全に一致し、CORESの妥当性が再確認された。また、この結果から、1回目の分類デ

Table 1.8  
Supervised Simulation of First Categorization: Modification of Weights and Degrees of Attention with Appropriate Initial Values

No.	Prototype	Fig.	Doc.	Other	Doc.
		Prototype			
1	16	20	(0.9046)	18	(0.9378)
2	21	15	(0.8792)	17	(0.9235)
		23	(0.9706)		
		22	(0.9706)		
3	8	5	(0.9823)		
		4	(0.9551)		
		2	(0.9462)		
		1	(0.9251)		
		6	(0.9609)		
		7	(0.9728)		
		3	(0.9734)		
4	17				
5	13	12	(0.9586)		
		14	(0.9641)		
6	9	10	(0.9556)		
		11	(0.9539)		
7	18	19	(0.9378)		
8	25	24	(0.9756)		

Attribute : Degree of Attention  
 1 : 0.10573 2 : 1.00000  
 3 : 0.47248 4 : 0.11127  
 5 : 0.04518 6 : 1.00000  
 7 : 0.13647 8 : 0.00000

Table 1.9  
Supervised Simulation of Corrected First Categorization: Modification of Weights and Degrees of Attention with Appropriate Initial Values

No.	Prototype	Fig.	Doc.
1	16	15	(0.8705)
		17	(0.9200)
2	21	23	(0.9495)
		22	(0.9495)
3	8	5	(0.9823)
		4	(0.9496)
		2	(0.9395)
		1	(0.9117)
		6	(0.9454)
		7	(0.9673)
		3	(0.9722)
4	20	18	(0.8625)
		19	(0.9121)
5	13	12	(0.9042)
		14	(0.9704)
6	9	10	(0.9486)
		11	(0.9379)
7	25	24	(0.9524)

Attribute : Degree of Attention  
 1 : 0.10770 2 : 1.00000  
 3 : 0.59404 4 : 0.14009  
 5 : 0.11968 6 : 1.00000  
 7 : 0.12946 8 : 0.00000



Table 1.10  
Supervised Simulation of Third  
Categorization: Fixed Weights and  
Degrees of Attention

No.	Prototype	Fig.	Doc.	Other Prototype	Doc.
1	16	15	(0.3334)	21	(0.8888)
		17	(0.8096)		
2	21	23	(0.2460)	16	(0.5768)
		22	(0.2460)	20	(0.7302)
3	8	6	(0.3333)	21	(0.3337)
		7	(0.8573)	2	(0.8574)
4	5	4	(0.8573)		
5	2	1	(0.3334)	21	(0.3337)
		3	(0.7620)	8	(0.8574)
6	20	18	(0.3333)	21	(0.7779)
		19	(0.8096)		
7	13	12	(0.3361)	21	(0.5560)
		14	(0.7681)	8	(0.8571)
8	9	10	(0.2070)	13	(0.8560)
		11	(0.2068)	8	(0.8096)
9	25	24	(0.3360)	21	(0.8889)

Table 1.11  
Supervised Simulation of Third  
Categorization: Modification of  
Weights and Degrees of Attention

No.	Prototype	Fig.	Doc.	Other Prototype	Doc.
1	16	15	(0.3234)	21	(0.8889)
		17	(0.8137)		
2	21	23	(0.3186)	16	(0.6215)
		22	(0.9238)	20	(0.9362)
3	8	6	(0.3818)		
		7	(0.8401)		
4	5	4	(0.8661)		
5	2	1	(0.3494)		
		3	(0.8636)		
6	20	18	(0.7690)	21	(0.9568)
		19	(0.8820)	21	(0.8988)
7	13	12	(0.6899)		
		14	(0.9132)		
8	9	10	(0.6740)	13	(0.7711)
		11	(0.8508)		
9	25	24	(0.9135)	21	(0.9667)

Attribute : Degree of Attention  
 1 : 1.00000 2 : 1.00000  
 3 : 0.70057 4 : 0.11023  
 5 : 0.09559 6 : 1.00000  
 7 : 1.00000 8 : 0.03125

Table 1.12  
Supervised Simulation of Third Categorization:  
Modification of Weights and Degrees of  
Attention with Appropriate Initial Values

No.	Prototype	Fig.	Doc.
1	16	15	(0.9156)
		17	(0.9321)
2	21	23	(0.9608)
		22	(0.9608)
3	8	6	(0.9643)
		7	(0.9875)
4	5	4	(0.9624)
5	2	1	(0.9816)
		3	(0.9814)
6	20	18	(0.9240)
		19	(0.9497)
7	13	12	(0.9478)
		14	(0.9728)
8	9	10	(0.9576)
		11	(0.9486)
9	25	24	(0.9624)

Attribute : Degree of Attention  
 1 : 1.00000 2 : 1.00000  
 3 : 0.74112 4 : 0.13291  
 5 : 0.09362 6 : 1.00000  
 7 : 1.00000 8 : 0.00000



ータでは被験者の分類上の矛盾のためにシミュレーションとの不一致が起こりう  
ると考えられるため、これ以降のシミュレーションでは、3回目のデータを用い  
て検証を進めるものとする。

### 5.2.2 教師なしのシミュレーション

教師なしのシミュレーション(unsupervised simulation)では、被験者の分類情  
報は与えない代わりに、重みと注目度の初期値をあらかじめ入力してやる必要が  
ある。本研究ではそれらを推定する際に、実際の被験者の分類基準についてのプ  
ロトコルを用いた。したがって、プロトコルで示される属性がモデルで設定され  
た属性とほぼ対応し、推定可能なものを選択する必要があった。必然的に対象と  
なる被験者はきわめて限られてしまうが、従来多く行われてきた恣意的なパラメ  
ータ設定を避けたために生じた制約と言え、実験結果に即したシミュレーション  
としての意義は高いと考えた。

よってそれらを推定しやすい被験者Bについて、事前にこれらの初期値を入力し、  
分類情報を与えない教師なしのシミュレーションを行った。教師なしの場合は、  
設定された分類閾値を越える適合度を有する図形が同一カテゴリーに分類される。

実験2で得られた被験者Bの分類結果と分類基準プロトコルをTable 1.13に示す。  
これから、カテゴリー2の分類には属性3の要素間距離、つまり、接しているか  
否かが大きく関与していることがわかる。さらに、カテゴリー1・2の分類には、  
属性1の半径、つまり、曲線か直線かが関与し、カテゴリー4の分類には属性7  
の半径の差、つまり、直線だけからなるか直線と曲線の両方を含むかが関与して  
いる。

そこで、以上を参考に、注目度の初期値を、属性1については曲線と直線を区  
分するだけであるとして0.2に、属性3については接しているか否かの区分ができ  
るよう1.0に、そして属性7は半径の差が小さいものと非常に大きいものを区別す  
るよう0.2に設定した。重みの初期値に関しては、利用されていた属性1, 3, 7  
に対して1.0を均等配分した。また、パラメータについては、それまでのシミュレ  
ーション上の学習速度と安定性から、共有度の基準増加量( $par_{10}$ )は0.3に、重み  
の基準増加量( $par_{10}$ )は0.5に、そして、修正前の注目度の優先度( $par_{20}$ )は0.2に  
設定した。分類の基準とされる適合度の閾値は、同じく前のシミュレーションか

Table 1.13  
Protocols of Categorization Criteria

No. Prototypes	Figures	Protocols of Categorization Criteria
1	16 15 17 13 12 14	Uncrossed lines.
2	21 23 20 18 22 25 19 24	Crossed lines.
3	8 5 4 2 1 6 7 3	Only curves.
4	9 10 11	Both lines and curves.

ら、典型例同士のどの適合度よりも大きく、かつ、どのカテゴリー成員のその典  
型例に対する適合度よりも小さくなる値である、0.76に設定した。

そのシミュレーション結果を、Table 1.14に示す。他のTableと異なり、中央の  
コラムに典型例と、その典型例と最大の適合度を持つ他の典型例とその適合度の  
値が示してある。これによって分類閾値が先の基準を満たして設定されているこ  
とが確認できる。右のコラムはこれまでと同様、カテゴリー成員の図形番号と、  
典型例との適合度である。このように、Table 1.14にはまったく不一致がなく、  
推測が可能な場合は適切な注目度や重みの初期値、分類閾値等をあらかじめ入力  
することで、教師なしのシミュレーションも可能であることが示され、CORESの本  
来の基本的仮定の妥当性は立証され、その適応力の高さが確認された。

このような実際のプロトコルに基づいたパラメータの入力は、これまでは行わ  
れていない。しかし、今回このような方法で分類過程を再現できたことは、この  
方法が、実験者の恣意的な設定を避け客観的な検証をする上で、有効であること  
を示した。今回の方法とは逆に、教師ありのシミュレーションから得られた重み  
と注目度の初期値をあらかじめ入力するようなことをすれば、教師なしのシミュ  
レーションでも一致した結果が得られるはずで、そのような恣意的な設定の仕方  
と実験者の主観的仮定に基づく設定の仕方の区別は、実質的に不可能になってし  
まう。よってこのような客観的プロトコルを得るためにも、厳密な実験によって



Table 1.14  
Unsupervised Simulation of Third Categorization

No.	Prototype	Other	Doc.	Fig.	Doc.
Prototype					
1	16			15	(1.0000)
				17	(1.0000)
				13	(0.8310)
				12	(1.0000)
				14	(1.0000)
2	21	16	(0.7500)	23	(1.0000)
				20	(1.0000)
				18	(1.0000)
				22	(1.0000)
				25	(1.0000)
				19	(1.0000)
				24	(1.0000)
3	8	16	(0.7279)	5	(0.9263)
				4	(0.7884)
				2	(0.8398)
				1	(0.9484)
				6	(1.0000)
				7	(0.9340)
				3	(0.9547)
				10	(1.0000)
4	9	16	(0.7536)	11	(1.0000)

Attribute : Degree of Attention  
 1 : 0.11207    2 : 0.00000  
 3 : 0.55695    4 : 0.00000  
 5 : 0.00000    6 : 0.00000  
 7 : 0.44237    8 : 0.00000

正確な人間のデータを収集することは今後も一層重要になると考えられる。

### 5.3 総合的考察

最後に、20名の被験者の3回目の分類データすべてに対して個別にシミュレーションを行った集計結果を記す。ただし、先に述べた理由から、教師なしのシミュレーションのための初期値推定を、全員に対して行うことは不可能であり、また、処理過程自体の妥当性は既に教師なしの場合でも確認されているので、全員に対しては、教師ありのシミュレーションが行われた。

実験データとシミュレーション結果が完全に一致したのは、6名分であった。ただし、一致しない原因はすべてモデル上の致命的な欠陥から生じたものではなく、主として次に述べるような被験者側の矛盾等の問題によるものであった。よって、その原因について、さらに以下のような詳細な分析と集計を行った(Table 1.15)。

#### 5.3.1 被験者自身の分類上の矛盾のための不一致

シミュレーションの不一致が、被験者自身の分類上の矛盾のために生じた場合であり、CORESの仮定上の問題を示すものではない。特に、分類基準の見落とし(Oversee of Categorization Criteria)によって矛盾が生じた5名分については、被験者Aの1回目の分類のように、実験者が部分的に修正することで一致させることができた。

#### 5.3.2 CORESの仮定との矛盾のための不一致

以下の2種類の原因間では被験者が一部重複しており、Table 1.15のこれ以下の人数には重複集計を記した。この部分に該当する実質被験者数は6名である。

##### [1] 用いられた属性の違い(属性選択)

これは被験者の分類基準に、仮定した属性では表現が困難な属性が関与した場合である。1) Analogical Figures or Lettersは、例えば、図形24をプラス、図形12をマイナスと呼んで同じカテゴリーに分類するといったように、既存の図形や数字との類似性を利用して分類した場合である。2) Unusual Attributesには、  
 a. 水平な直線の上だけに直線がある場合と下にもある場合、  
 b. 垂直な直線2本



Table 1.15  
*Causes and Numbers of Inconsistencies between Experiments and Simulations*

Results	Causes	Details	Numbers
Consistent			6
Inconsistent			
	<b>5.3.1 Subjects' Contradictions</b>		
		Oversee of Categorization Criteria	5
		Use of Ambiguous Categorization Criteria	2
		Arbitrary Cognitive Change of Figures	1
	<b>5.3.2 Contradictions with CORES' Suppositions</b>		
	[1] Use of Different Attributes		
		1) Analogical Figures or Letters	4 <sup>a)</sup>
		2) Unusual Attributes	3
		3) Parallel Shift	2
	[2] Use of Hierarchical Structures		
		Partial Hierarchy	4

Note: <sup>a)</sup> The following causes were overlapped, and the numbers were aggregated.

が並行でかつ非常に接近している, c. 図形5を山形とみなす, などが見られた。これらはあくまでも個人の特異な認識であり, ここで想定した属性では表現できなかったに過ぎない。したがって, これらを表しうる属性を想定すれば, シミュレーションは不可能ではない。しかし, この実験で用いた単純な図形ですらその属性は無限に想定でき, そのような属性を増すことがCORESの適応力を増すことにつながるとは限らない。したがって, 今後は, 人間の認識する基本的属性について明らかにしていくことも重要な課題であろう。

ただし, 上記, 2), b.の接近した並行直線の認識や, 3)の並行移動(Parallel Shift)は, ゲシュタルト的な見方, 例えば「接近性の原理」を反映している可能性があり, 図形認識を機械的に扱う上で問題となるヒューリスティックな認識の仕方として, 留意する必要があるかもしれない。

#### [2] 用いられた分類構造の違い (階層構造)

階層的な分類とは言っても, 4名分すべてが従来のモデルが扱ってきたような完全階層構造ではなく, 矛盾を含んだ部分的階層構造(Partial Hierarchy)を成していた。その主なものは, まず, ある属性で分類し, 下位構造を作る。次に, 下位構造ごとに異なる属性を基準にして独立の分類をする, という一貫性のない分類方法によるものである。したがって, この場合は被験者の分類方法の非一貫性も不一致の原因であり, 必ずしもCORESの問題だけを示すものではないが, このような分類傾向が見られた以上, 階層的な分類については, その合理性からしても, 十分考慮する必要があるであろう。

しかし, 階層的な分類を考えた場合には, 何を上位の属性, 下位の属性として選択するかという, 先に述べた属性の選択の問題が関与してくることは避けられない。完全な階層的な分類だけでなく, ここで見られたような, 上位の属性で大まかに分類して, まず分類すべき事例の下位集合を小さくし, 次に, 下位分類ごとにそれに適した方法で分類するといった方法や, 一旦分類はするが, 全属性や全体構造を把握してからより適切な構造に修正する, といった合理的プロセスをモデルに組み込むためには, 属性の上下関係や包含関係などの, 属性の性質に応じた分類の仕方や構造的修正が必要となる。したがって今後は, そのような属性の基本的性質を明らかにすることが先決であろう。



## 第6章 まとめ

以上、一連の心理実験によるモデルの修正と吟味、シミュレーションによる検証という認知科学的アプローチによって、人間に即した知識形成モデル、CORESが、文脈によって異なる様々な分類を再現できることが示され、その妥当性が明らかとなった。特に、ここで行われた個別シミュレーションは、他に類を見ない試みであり、これが実現できたのは、CORESが、人間の処理過程にできるだけ忠実に考えられ、柔軟で適応力に富むよう作られたからに他ならない。

以下には、こうした人間の知と機械の知の対比によって明らかとなった重要な3つの知見について詳述し、第1部のまとめとする。

### 6.1 注目度、文脈、及び、知識構造

文脈に関しては、実験1とシミュレーション1から、文脈による属性値の差異の大きさの認識の仕方の違いが、注目度で実現されうること、そして、実験2とシミュレーション2では、分類過程でダイナミックに決定される注目度によって、CORESが、多様な文脈における様々な分類方法を、個人差のレベルまで再現できることが明らかとなった。

文脈の影響については、これまでも、例えば、Medin[Medin 89]やRoth[Roth 83]が、事例の典型性評定が文脈で変化することを指摘しており、また、Barsalou[Barsalou 82]は、文脈によって特定の概念の副次的な特徴が付加的に活性化することを実験的に証明しているが、これを実際にモデル化しようとする試みは極めて少ない。

そのような数少ない研究の1つは、Anderson[Anderson 83]によるものである。彼はこの文脈の影響を、階層構造ノードの再活性化によって実現しようとした。しかし、彼のモデルは厳密な階層構造を仮定しており、そのために、多様な文脈による違いを説明するには無数の階層構造を相互に関係づける必要が生じるため、その中には必然的に矛盾が出てきてしまう。例えば、ぬいぐるみのくまは動物ではないけれども、ぬいぐるみしかない状況では、子供でも、「動物は？」と聞け

ば「くま」を指す。このような概念間の多様な関係を、同一の階層的意味ネットワークで表現することは、ほとんど不可能である。

また、実験1で測定された心理学的距離も、従来の研究のように限定された階層構造内の離散的な概念間の距離ではなく、連続的次元上の概念間の可変的な距離であった。そして、そのような心理学的距離が文脈によって変化したことも、CORESの仮定するような柔軟な構造を想定する必要性を示唆していると言える。

概念は絶対的なものではなく、文脈によって変化するはずである。なぜなら、概念が絶対的なものだとすれば、文脈によって異なる概念や属性値を定義せねばならないからである。そして、その構造は、本質的に厳密な階層構造を成しているのではなく、任意の属性が注目される場合とされない場合、あるいはその認識のされ方だけが異なる場合など、特定の文脈下でのみ成立するものだと考えられる。

CORESは、このような概念の性質を反映し、概念の置かれた文脈下での相対的關係から、その文脈に応じた構造を形成しうる可能性を持ち、しかも、連続的な属性値を用いているために、多様な文脈の影響に対応できる柔軟性を持つ。そして、少なくとも知識探索の効率化の面からも、これらの性質を反映する非階層的知識構造をCORESに導入することが必要であり、これが、第2部におけるCORES拡張の1つの主題でもある（概念間関係の導入）。

ただし、CORESは完全に階層的構造を否定するものではない。記憶された知識（概念）を効率よく探索するためには、階層的構造は有効であり、また、文脈によっては厳密な階層的構造となる場合もありうるはずである。よって、これらの問題に対応するためには、基本的には非階層的構造を仮定するが、重みや注目度のような上位の変数によって擬似的階層構造を形成できるようにすべきだと考えられる。そしてそのためには、シミュレーション2で示唆されたように、人間の用いる基本的属性やその性質を同定することが不可欠であり、今後の重要な課題となろう。

### 6.2 典型例

カテゴリーが形成された後からカテゴリーの特徴を推論するのは、さほど困難



ではない。しかし、問題となるのはその形成過程そのものであって、これが人間の認知活動に他ならない。そのような過程では、記憶の負荷の観点からも、典型例、すなわち、参照され、新しい事例と比較されるものは何か、ということが非常に重要になってくる。

実験1・シミュレーション1では、FBがあり、そのFB通りに分類するのにそれまで用いた典型例では不十分な場合、新たな典型例が追加されることが示された。そして、典型例を追加し続けるのではなく、不必要な典型例を忘却させるための活性度の指標が必要であることが示唆された。

時間の経過に影響される概念の活性度を考慮する必要性は、他の箇所からも示唆されている。1つは、知識の探索の問題である。膨大な知識が存在する場合、先に述べた知識間の構造を利用すると共に、探索の優先順位が非常に重要になってくる。さらに、既に述べたように Kintsch[Kintsch 78] は、新しい命題が初期の上位の命題、あるいは、最近処理された命題と関係づけられることを明らかにしているが、現在のCORESは、前者だけを考慮しているに過ぎない。しかし、すべての事例やその特徴を記憶できないとすれば、後者のような時間的近接性も新しい概念をどの既存概念と関係づけるかの選択に重要な役割を果たしている可能性が高いと考えられる。よって、これらの時間的要因は、特にそれが重要となる第2部の自然言語処理の段階で取り入れられることになる。

### 6.3 ヒューリスティックスの必要性

実験2から、人間の場合は、体系的に分類するだけでなく、接近した並行直線のゲシュタルト的な認識や、その認識に都合のよい並行移動などを行っていることが明らかとなった。このようなゲシュタルト的認識は、ある意味ではヒューリスティックスな認知的操作で、曖昧な世界を合理的に処理する重要な方略なのかもしれない。

さらに、実験2とシミュレーション2の比較からは、人間の学習速度はシミュレーション上の学習速度に比べてはるかに早いことも明らかとなった。この学習速度の速さは、多くの世界的知識に基づいた大胆な修正方略の存在を示唆する。このような過程をCORESで実現しようとするならば、今回行ったような重みや注目度の

初期値の修正というマイクロレベルでの改善ではなく、人間の判断に対応するようなヒューリスティックスを取り入れる必要がある。

以上のようなヒューリスティックスは、必ずしも合理的とは限らない。しかし、多くの場合様々な経験知に導かれて成功し、失敗したとしてもそれを軌道修正するようなメカニズムが働いていると考えられる。処理を合理化する、比較的非形式的なヒューリスティックスの存在は、人間が形式論理的ルールよりもむしろ、実用的論理スキーマに頼って推論することが多いという実験結果によっても示唆されている[Holland 86]。

実用論的スキーマを人工知能モデルにとって不要とする意見もあるかも知れない。しかしそのようなスキーマは機械翻訳等で最終的には必要とされた知識ではなかったか。よって、様々なヒューリスティックスを解明するとともに、いつ、どのようなヒューリスティックスが用いられ、かつ、どのように表現されているのかを明らかにして、積極的にモデル化することは、認知過程を知る上でも、適用範囲の広い合理的なシステムを構築する上でも今後一層重要な課題となろう。