Categories and Concepts - Spring 2019

Conceptual combination and creation

Brenden Lake

PSYCH-GA 2207

Conceptual combination

"pet fish"

"apartment dog"

"ocean house"

"engine repair"

Conceptual combination

- *Conceptual combination* concerns understanding "complex concepts" that require more than one word
  - Related but distinct from conceptual creation (which may not use words), which we talk about later in the lecture.

- Murphy: Many theories of concepts make similar predictions when classifying simple stimuli, but may be easier to tell apart through conceptual combination

- Candidates for a theory of conceptual combination..
  - naive prototype combination
  - extensional theory
  - feature reweighting
  - concept specialization

Naive prototype averaging X

complex concept XY is an average of prototypes for X and Y

But this can't explain how a typical "pet fish" is a very atypical pet and an atypical fish

"pet"

"fish"

"apartment"

"dog"

? ?

"pet fish"

? ?

"apartment dog"

Can we just find common exemplars? (Extensional theory)

complex concept XY is the intersection of all things X and all things Y

“pet”

“fish”

“pet fish”

Can we just find common exemplars? (Extensional theory) X

complex concept XY is the intersection of all things X and all things Y

"apartment"

"dog"

"apartment dog"

X

Feature reweighting

complex concept X (adjective) Y (noun) reweights the features of Y, without the need to access world knowledge (Smith & Osherson, 1984). *Only applies to adjective-noun compounds.*

"apple"

- "round" (weight 1.0)
- "edible" (weight 0.8)
- "sweet" (weight 0.7)
- ….
- "green" (weight 0.5)
- "red" (weight 0.5)

+ "red"

"red apple"

- "round" (weight 1.0)
- "edible" (weight 0.8)
- "sweet" (weight 0.7)
- ….
- "green" (weight 0)
- "red" (weight 1.0)

Challenges to feature reweighting

The same adjective doesn't always highlight the same feature

- corporate stationery
- corporate account
- corporate car
- corporate donor
- corporate lawyer

corporate stationery

vs

corporate lawyer

Concept specialization using world knowledge

Complex concept XY is a modification of the schema representation of Y (slots and fillers), where the modification may rely on world knowledge and intuitive theories (Cohen & Murphy, 1984). *Unfortunately, this theory is vaguer than the others.*

"dog"

Simplified Schema for the Concept *dog*.

NAME:	"dog"	
BODY PARTS:	LEGS:	4, 3
	HEAD:	1
	HAIR	
	EYES:	2
	COLOR:	brown, white, black . . .
	(etc.)	
HABITAT:	home, streets	
FUNCTIONS:	best friend, guard home	
BEHAVIORS:	bark, bite, eat, sleep, chase cats, . . .	

"apartment dog"

Simplified Schema for the Concept *dog*.

NAME:	"dog"	
BODY PARTS:	LEGS:	4, 3
	HEAD:	1
	HAIR	
	EYES:	2
	COLOR:	brown, white, black . . .
	(etc.)	
HABITAT:	~~home, streets~~ apartment	
FUNCTIONS:	best friend, guard home	
BEHAVIORS:	bark, bite, eat, sleep, chase cats, . . .	
+ SIZE:	small	

Challenges to any theory of conceptual combination

The different uses of “ocean” here don’t fill the same slots, or highlight the same feature. Also, salient properties of oceans (blue, wet, large etc.) are not carried onto the noun it is attached to:

- ocean road
- ocean cruise
- ocean view
- ocean bird
- ocean creature
- ocean house

“ocean creature”

(lives under water)

“ocean house”

(does not live underwater)

COGNITIVE SCIENCE 12, 529-562 (1988)

Comprehending Complex Concepts

GREGORY L. MURPHY
Brown University

Recent theories of concepts have raised the issue of how people combine simple concepts (like *engine* and *repair*) to form complex concepts (like *engine repair*). This article approaches this issue by asking how people comprehend modified noun phrases of this sort. One explanation of how complex concepts are understood (the feature weighting model) provides a simple mechanism in which the primary feature of the modifying concept is made more salient in the modified concept. Another explanation focuses on how world knowledge directs the combination process. The two explanations are compared in their ability to account for the interpretation of various kinds of noun phrases. Two experiments are reported which evaluate the feature weighting model's predictions for adjective-noun phrases. These contrasts suggest that the combination process does require reference to world knowledge. The consequences of accepting such an account are discussed.

The issue of complex concepts has become an important one in the psychology of concepts. Although theories of concepts may make similar predictions about the structure of simple concepts (e.g., Hintzman & Ludlam, 1980), they may not be equally facile at explaining how concepts are combined to form new, more complex concepts. Thus, an account of complex concepts may be crucial for evaluating the many theories of concepts now extant (see Smith & Medin, 1981).

The creation of complex concepts is a fascinating example of a high level cognitive process that people can perform very quickly. People are likely to create novel noun-noun phrases in their conversations, and listeners are

Murphy : Ex 1 - features of adjective-noun compounds

- Is "feature weighting" sufficient for adjective-noun compounds? Could it be a closed operation does not require world knowledge?
- Key assumption of experiments: if a feature is atypical of both components of the compound, it should also be atypical of the compound
- Task: participants rated each feature's typicality of either the noun, the adjective, as well as the adjective-noun compound compound
- features were chosen to be particularly typical of compound

"red apple"

- "round" (weight 1.0)
- "edible" (weight 0.8)
- "sweet" (weight 0.7)
- ….
- "green" (weight 0)
- "red" (weight 1.0)

Murphy : Ex 1 - features of adjective-noun compounds - results

Complex Concepts and Properties Used in Experiment 1

Concept	Property	Most Typical Part[a]
Smelly trucks	emits lots of black smoke	Noun
Sliced apples	is cooked in a pie	Complex concept
Casual shirts	is pulled over your head	Complex concept
Small couches	seats only 2 people	Complex concept
Split canteloupes	on a plate	Noun
Uncaged canaries	lives in South America	Complex concept
Round tables	used at a conference	Complex concept
Standing ostriches	calm	Complex concept
Unshelled peas	long	Complex concept
Yellow jackets	worn by fishermen	Complex concept
Green bicycles	painted green	Complex concept
Convertible drills	can be used as a screwdriver	Noun
Overturned chairs	is on a table	Complex concept
Short pants	exposes knees	Complex concept
Ancient saws	rusty	Complex concept
Russian novels	(originally) written in Russian	Complex concept
Empty stores	lose money	Complex concept
Dirty bowls	sticky	Complex concept

[a] This entry tells whether subjects judged the indicated property to be most typical of the adjective, noun, or complex concept.

- **For 15 of 18 compounds, the property was more typical of the compound than it was of either component**
- Thus, conceptual combination is clearly not closed, and seems to rely on world knowledge

Examples

empty stores - likely to lose money, more than its components
green bicycles - likely to be painted green, more than its components
unshelled peas - likely to be long, more than its components

Murphy : Ex 2 - defining adjective-noun combos

- More unbiased pairings: Used all pairings of 10 common adjectives and 10 common nouns
  - ✳ adjectives: new, good, long, social, important, human, political, etc.
  - ✳ nouns: year, people, world, life, hand, house, etc.

- task: participants were asked to define each compound

- hypothesis: adjectives have different meanings when combined with different nouns (and thus inconsistent with closed feature weighting)

“red apple”

- “round” (weight 1.0)
- “edible” (weight 0.8)
- “sweet” (weight 0.7)
- ….
- “green” (weight 0)
- “red” (weight 1.0)

Murphy : Ex 2 - defining adjective-noun combos - results

- The same adjective (e.g., “long”) has different meanings when combined with different nouns

Examples of the "Meanings" of Two Adjectives from Experiment 2

One subject's meanings for the adjective *long*.
(year) seeming to pass slowly
(people) *tall
(world) *covering a great distance
(life) lasting for years
(hand) expressed in complete sentences and without abbreviations
(house) with large dimensions, especially in one direction
(problem) whose solution takes a long time
(word) with many syllables
(eye) *towards the future
(city) *with large dimensions in one direction

- Each adjective had 6.95 meanings on average, thus about 70% of meanings were unique
- Again, conceptual combination is not closed and seems to rely on world knowledge

Implications for theories of concepts

- "concept specialization" theory admits many more kinds of conceptual combinations than other approaches

"plastic truck"

- but an even broader and more powerful theory is needed -- one that can radically change the concepts involved (a plastic truck is not a real truck, but a toy)

- All theories require some summary representation like a prototype (or schema), and exemplar models are essentially ruled out due to failure of extensional theory

- Murphy: Conceptual combination is a solid win for prototype theory and the knowledge view

REVIEW

The neural basis of combinatory syntax and semantics

Liina Pylkkänen[1,2]

Human language allows us to create an infinitude of ideas from a finite set of basic building blocks. What is the neurobiology of this combinatory system? Research has begun to dissect the neural basis of natural language syntax and semantics by analyzing the basics of meaning composition, such as two-word phrases. This work has revealed a system of composition that involves rapidly peaking activity in the left anterior temporal lobe and later engagement of the medial prefrontal cortex. Both brain regions show evidence of shared processing between comprehension and production, as well as between spoken and signed language. Both appear to compute meaning, not syntactic structure. This Review discusses how language builds meaning and lays out directions for future neurobiological research on the combinatory system.

When exposed to a familiar language, our brains automatically compose the individual words together into larger meanings. Even without language input, our brains do something similar: We create new meanings in our thoughts and even comprehend our own creations. This is the internal "chatter" that, for most humans, is hard to shut down. Although combining meanings is instinctive and automatic, our minds actually perform some rather complex mental gymnastics while doing so. Consider these seemingly simple sentences:

Sally baked the black beans.
Sally baked the beans black.

In the first sentence, black describes a property of the beans prior to the baking, a so-called modifier reading. In the second, the blackness of the beans is caused by the baking, a resultative reading. We can even make the meaning ambiguous by adding a modifier to the adjective itself:

Sally baked some beans black enough
to look like licorice.

In this sentence, Sally could either be baking beans

term memory traces corresponding to these structures and is therefore able to evaluate incoming language against this knowledge. This much is uncontroversial. But what are the online computations that serve to merge or combine words during language processing, such that the relationship between the composed representations and our knowledge of syntax can be evaluated? This Review focuses on our current understanding of this question.

Comprehension: Rapid concept composition in the left anterior temporal cortex

Just as biologists prefer to study small animals with fewer cells when trying to understand living organisms mechanistically, it has proven productive to start simple in the neurobiology of meaning composition as well. When it comes to brain mechanisms of language, a full sentence is like an elephant. A short, two-word phrase is a more tractable representational unit, and thus research has begun to characterize the composition of these minimal phrases (*1–10*) (Fig. 1). The goal is to functionally decompose a perisylvian brain network implicated for the processing of full sentences (*11*).

When we understand language, dozens of

many different types of combinatory routines (Fig. 2). For example, to account for the syntactic and semantic behavior of a simple phrase such as "black cat," linguistics and cognitive psychology hypothesize at least three types of structures: (i) the syntax, in which the categories "noun" and "adjective" join to form a noun phrase (*12*); (ii) the logical semantics, in which the properties of blackness and catness intersect to yield a representation of entities that possess both properties (*13*); and (iii) the conceptual structure, in which the features of the two concepts combine (*14*). Because these representations may be built simultaneously in parallel, a problem for a mechanistic understanding of composition is to determine whether these possibly distinct representations dissociate in neural activity.

First, we need a characterization of what happens in the brain when a person hears or reads a word in a minimal combinatory context. Most studies on minimal phrases have used adjective–noun combinations. Results show that in English, this type of basic composition increases activity in the left anterior temporal lobe (LATL) at 200 to 250 ms after noun onset, as compared with the processing of the same noun in a context in which it cannot combine with the preceding word (*1*, *2*). If the language has the reverse word order, noun before adjective, the same result is seen with respect to the processing of the adjective (*6*). About 200 ms later, another activity increase often occurs in the ventromedial prefrontal cortex (vmPFC) (Fig. 1). What aspects of composition do these neural traces reflect?

Syntactic effects are difficult to distinguish from semantic effects, because in natural language, syntactic changes usually alter the meaning of the expression. Consequently, many results on the neural basis of basic syntactic composition are actually results on the processing of artificial stimuli that are intended to lack meaning. Most commonly, such expressions are made up of nonsense words strung

Conceptual creation and exemplar generation

(unlike previous section, this isn't necessarily language driven)

How do people create whole new concepts?

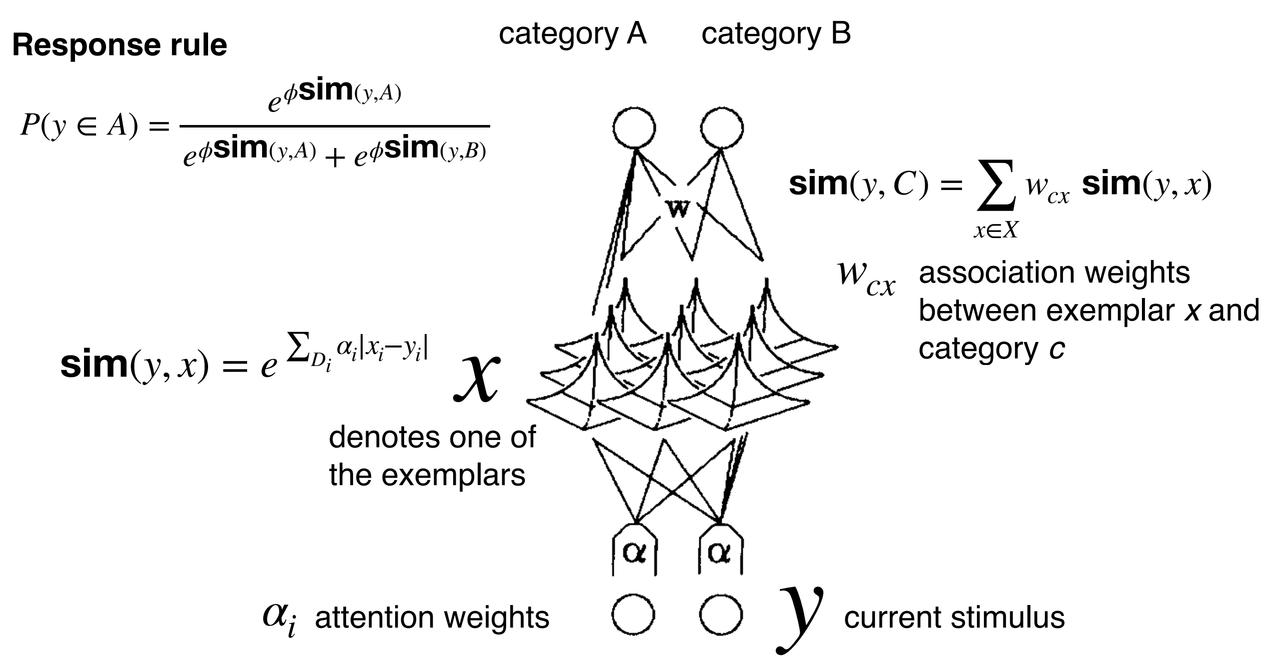
How do people create new exemplars of existing concepts?

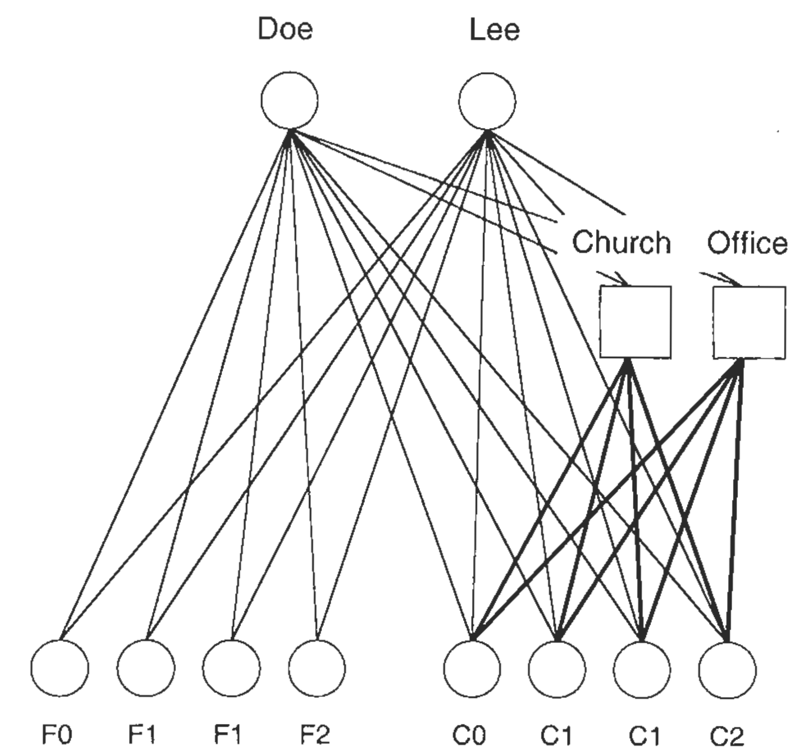
Architects design new houses

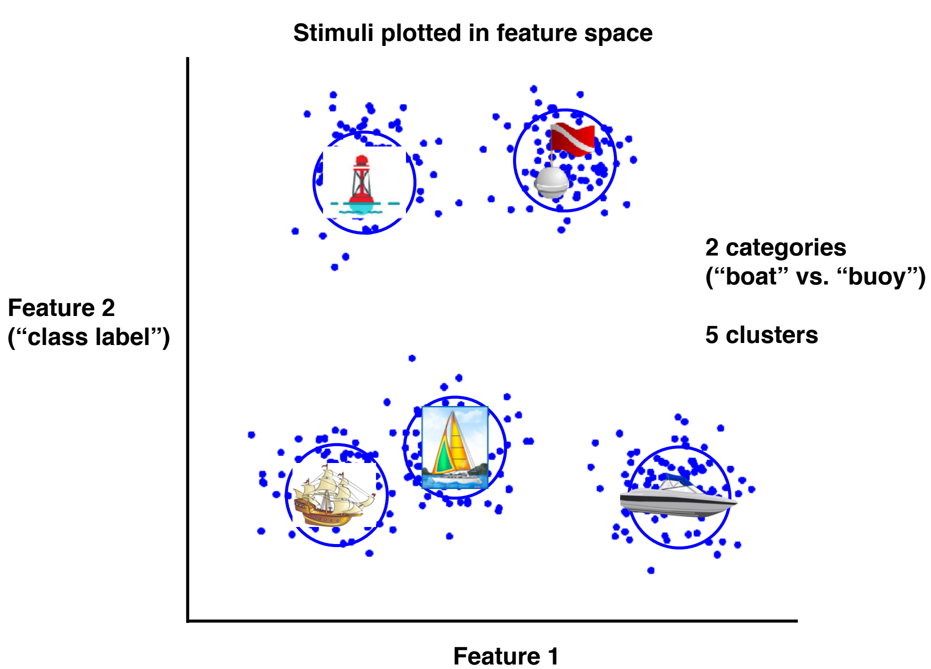
Chefs create new recipes

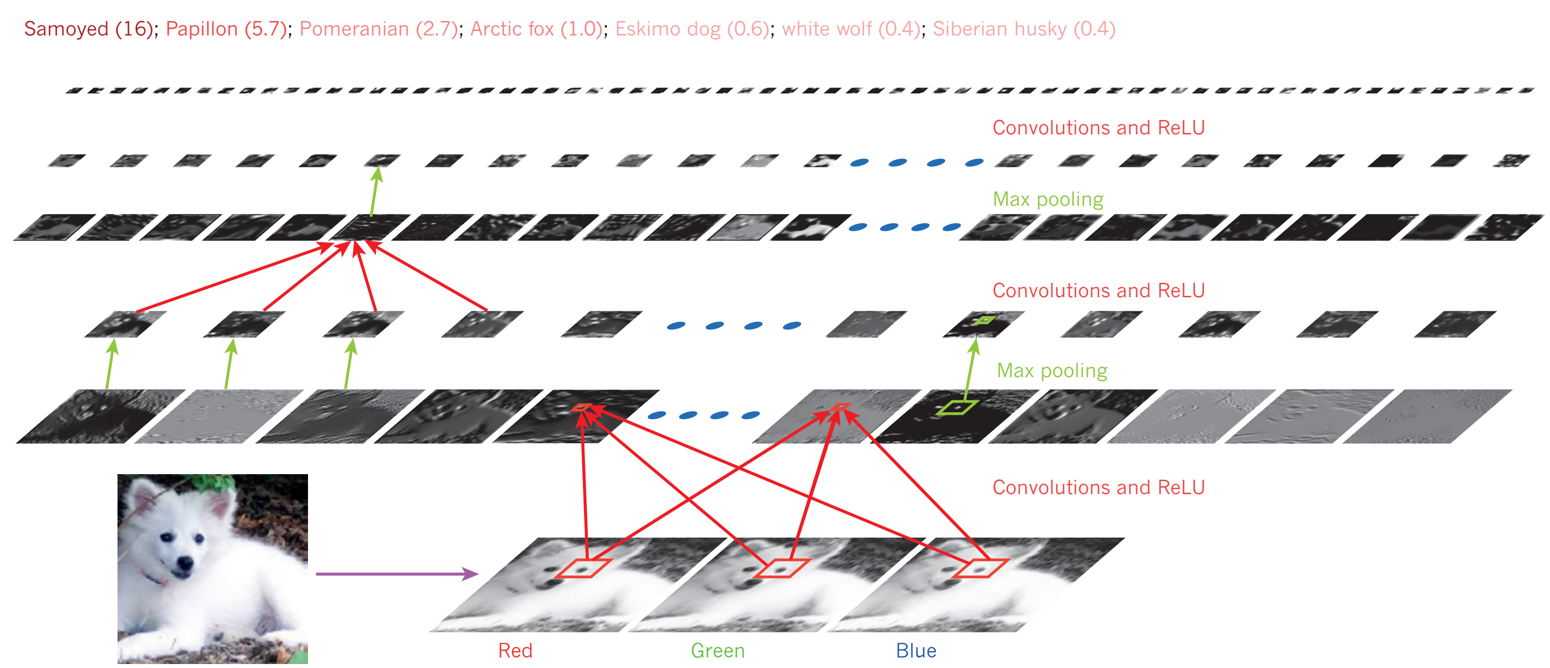
Entrepreneurs create new ideas

Computers / computational models are not known for their creativity









Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY

Vol. 41, No. 2, June 1997
Printed in Belgium

The Structure of Perceptual Categories

Jacob Feldman

Department of Psychology, Center for Cognitive Science, Rutgers University

When presented with a small set of sample objects, human observers have the striking capacity to induce a more general class. Generalization can even proceed from a single object ("one-shot categorization"). The inference is apparently guided by the principle that a good categorical hypothesis is one in which the observed object would be a typcal, "non-accidental," or *generic* example; this idea is formalized here as the Genericity Constraint. In the theory proposed here, each categorical hypothesis is a "generative model," a sequence of transformations by which the object is interpreted as having been created; objects are considered to be in the same category if they were created by the same set of operations. The set of all available category models can be explicitly enumerated in a lattice, an explicit structure that partially orders the models by their degree of regularity or genericity—more abstract models are higher in the lattice, and more regular or constrained models are lower. The Genericity Constraint dictates that among all the models on the lattice that apply, the observer should choose the one in which the observed object is generic, which is simply the lowest in the partial order. A series of experiments are reported in which subjects are asked to generalize from simple figures. The results corroborate the role of the lattice and the Genericity Constraint in subjects' interpretations.

1. THE STRUCTURE OF A GENERALIZATION

This paper investigates the way human observers, presented with a small set of objects, can generalize to a larger class, which constitutes the intuitively "natural" category from which the examples appear to have been drawn.

Address correspondence and reprint requests to: Jacob Feldman, Department of Psychology, Center for Cognitive Science, Rutgers University, Busch Campus, New Brunswick, New Jersey, 08903. E-mail: jacob@ruccs.rutgers.edu.

This paper is derived from a doctoral dissertation in the M.I.T. Department of Brain and Cognitive Sciences. (The original dissertation is available as Technical Report 6 of the Rutgers Center for Cognitive Science.) I am especially grateful to Whitman Richards, who supervised the original dis-

Human observers exhibit a remarkable degree of unanimity in this sort of perceptual abstraction, even when using only a single object as an example (*one-shot categorization*). Since such an inductive problem is inherently ambiguous to an extreme degree, this competence suggests the presence of enormous constraint on the generalization faculty. Yet even a few simple examples suffice to show that observers seem to pick and choose quite selectively about which properties of an observed object are fit to be abstracted over, and which retained, in the inferred general class. The constraints on generalization are heavy, but they seem to be deft.

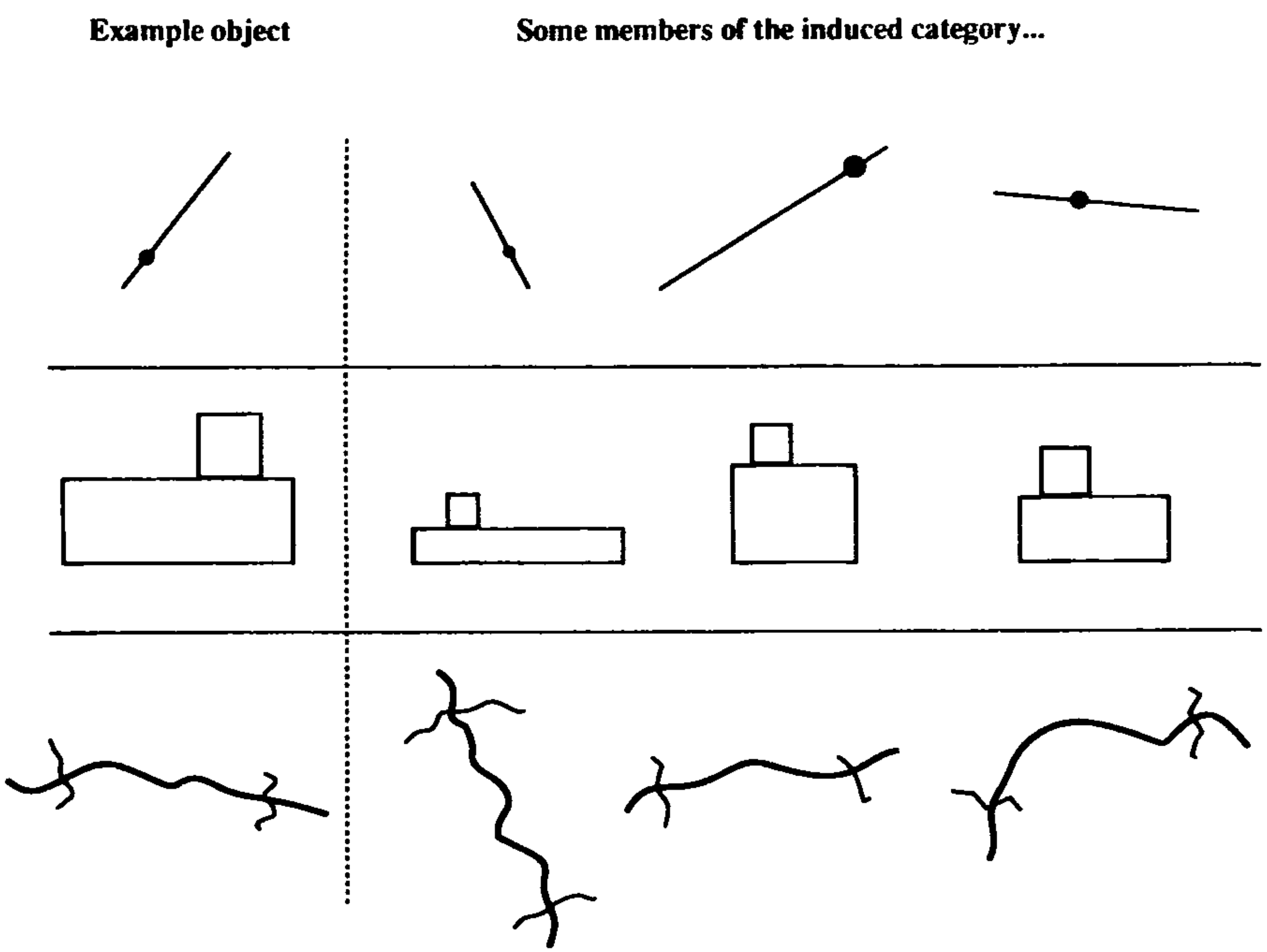
The induction problem investigated in this paper is a narrow one. The objects are all simple, readily parameterizable geometric figures, and there is usually only one sample object (three in some experiments). Yet this problem captures many of the essential difficulties common to all induction problems, deriving from the complete absence of a priori constraint on the solution. The goal of the paper is to construct an explicit inference theory for the problem, specifying the logic by which inductive hypotheses are constructed and selected. The logic here takes the tangible form of a lattice, an explicit algebraic enumeration and ordering of categorical hypotheses. Selection of a hypothesis from the lattice is dictated by the Genericity Constraint, a rule based on the principle that sound generalizations are those that minimize unexplained coincidences in the observations.

A few examples illustrate the approach (Fig. 1). In each row, consider the example object given on the left. What is the category? Some possible "intuitive" answers are given toward the right.

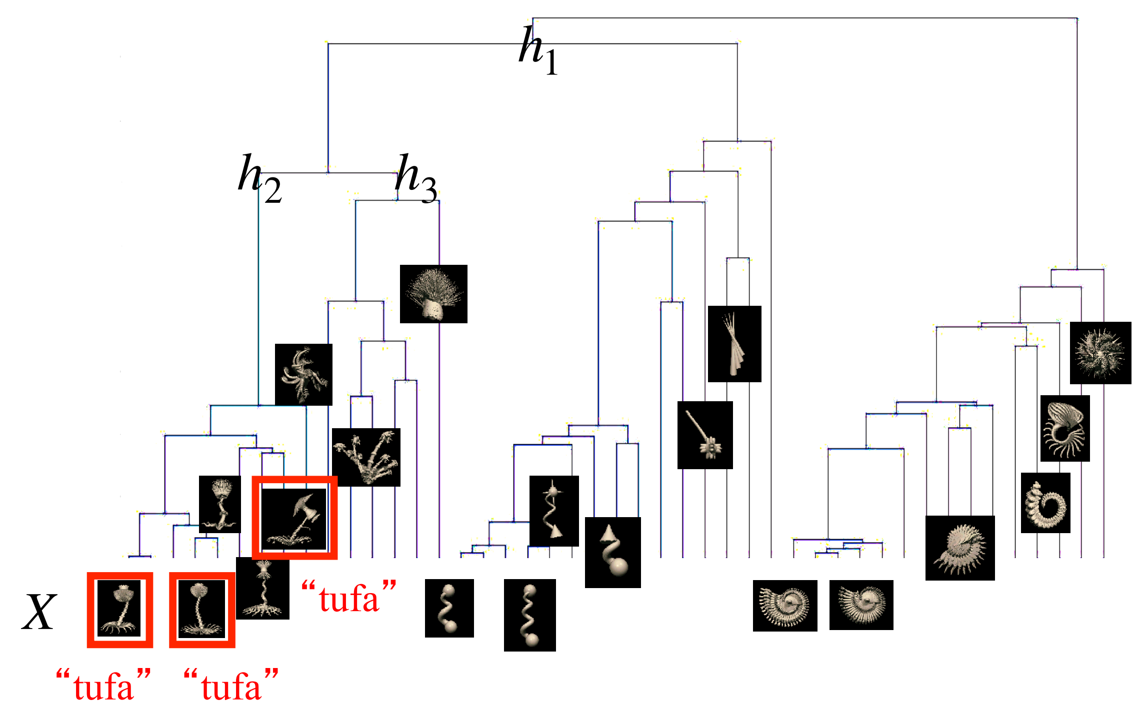
In the top row, it seems natural to infer the category "dot on line." That is, we infer that in other examples of the same category, the dot will always appear somewhere *on* the segment, rather than at some arbitrary point in nearby space, although its position *along* the line might vary (as might the

One-shot exemplar generation experiments

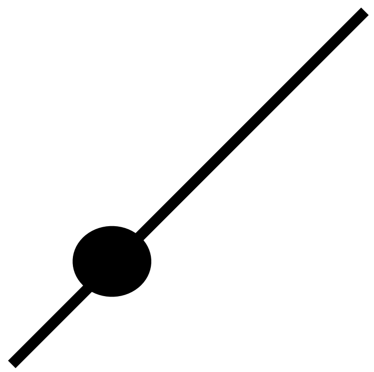
Feldman's one-shot exemplar generation



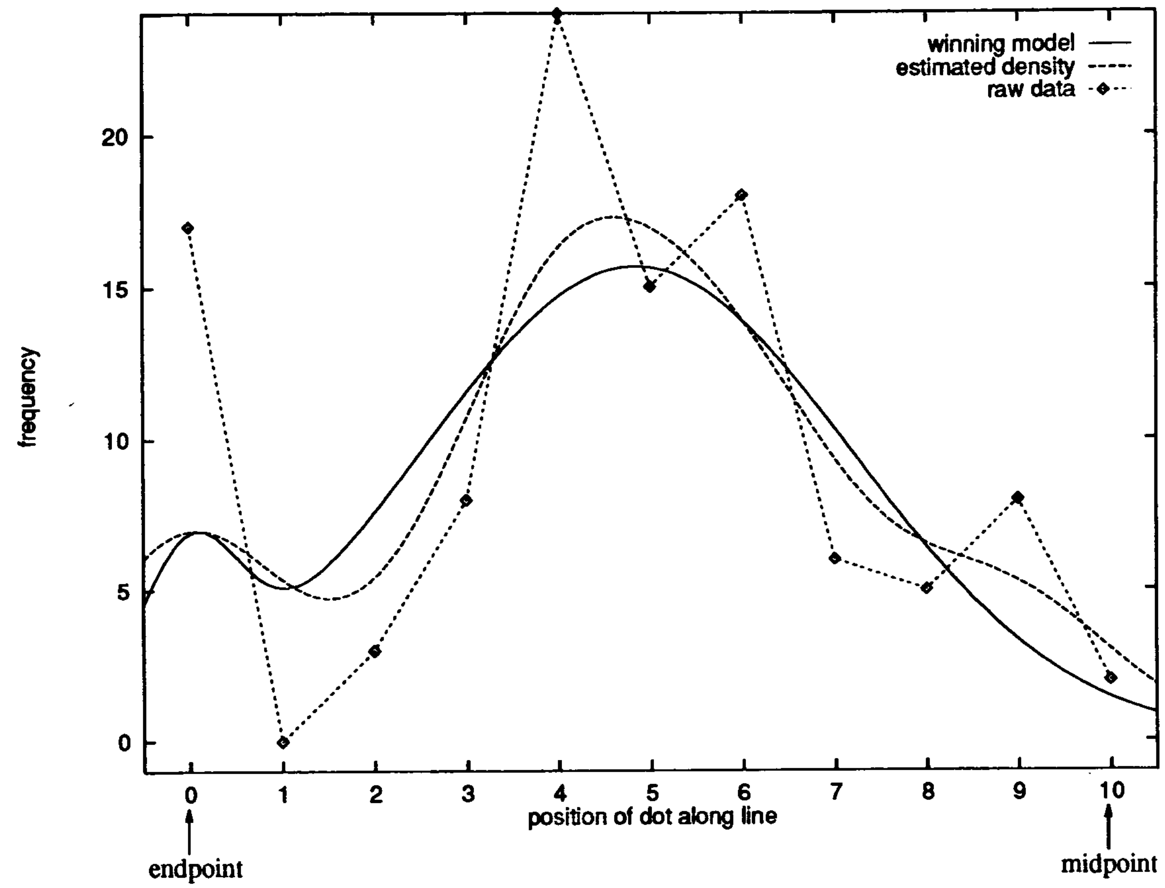
Review: one-shot classification



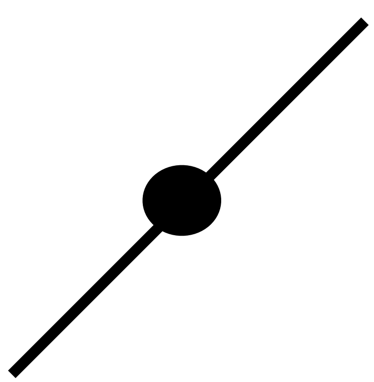
“Here is a blicket”



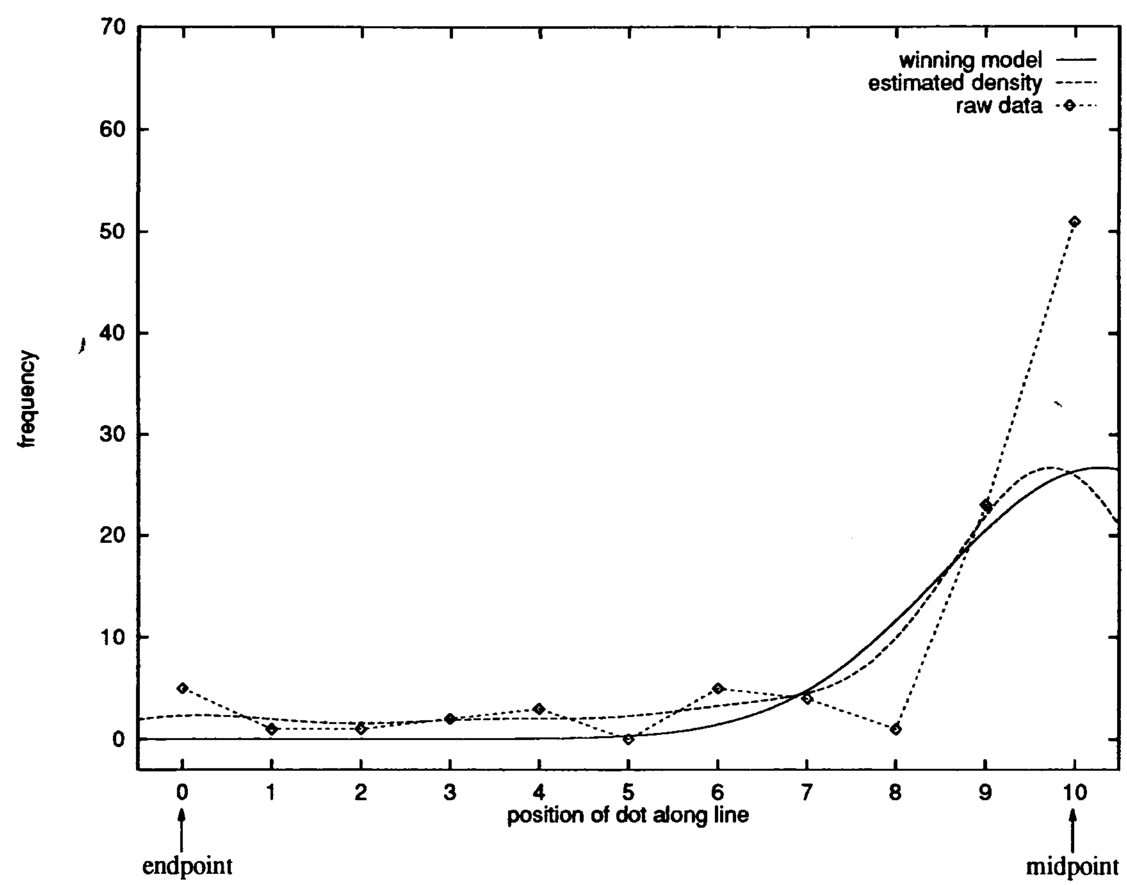
Draw some more blickets



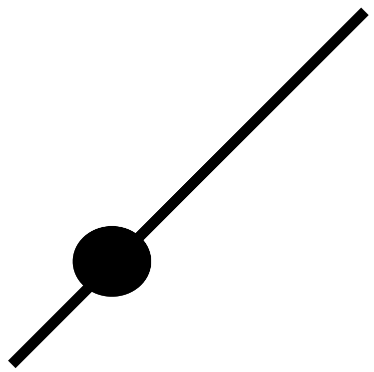
“Here is a blicket”



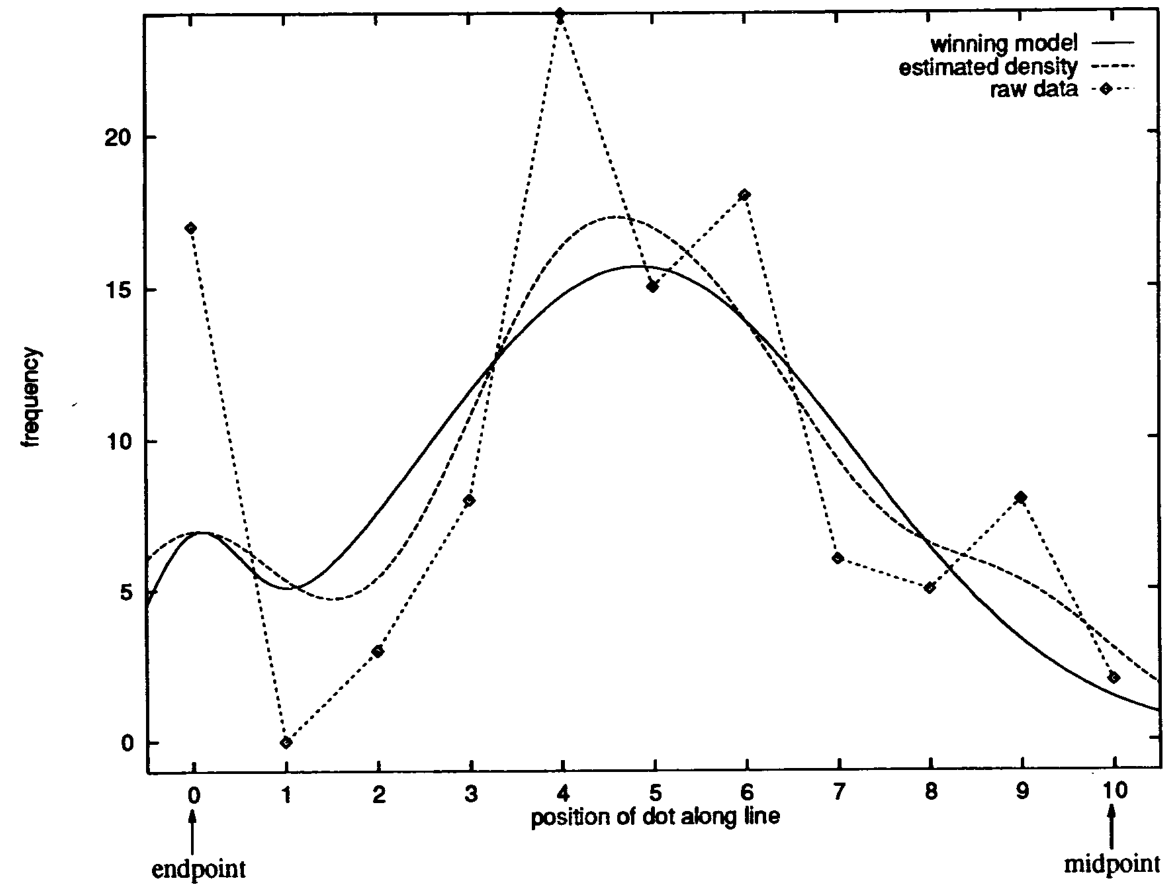
Draw some more blickets



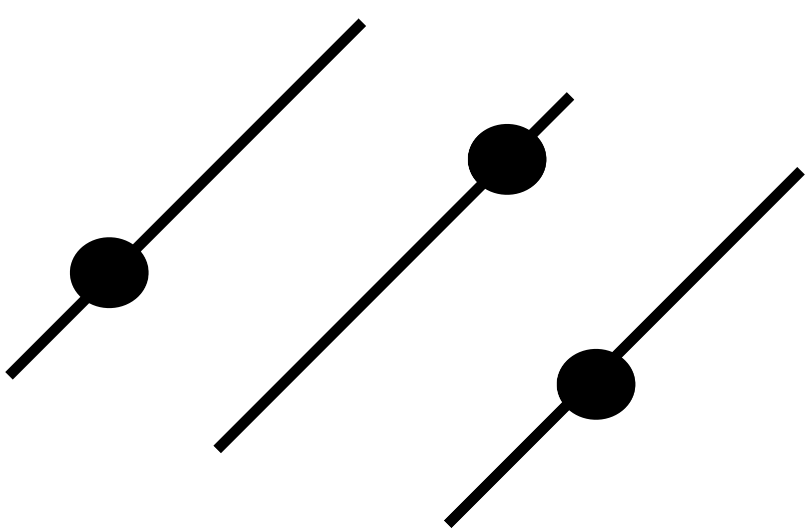
“Here is a blicket”



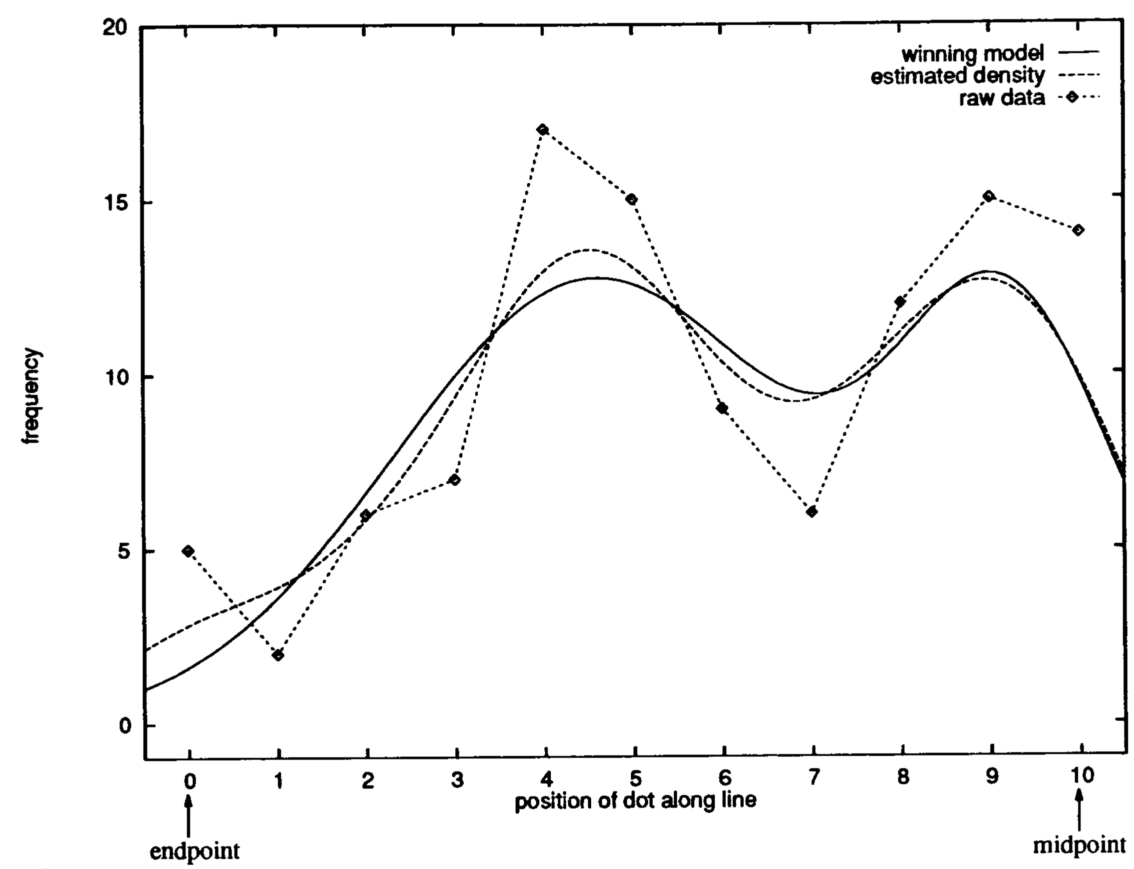
Draw some more blickets



“Here are some blickets”



Draw some more blickets



Cognitive Psychology 66 (2013) 85–125

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

A probabilistic account of exemplar and category generation

Alan Jern *, Charles Kemp

Carnegie Mellon University, Department of Psychology, 5000 Forbes Ave. Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Accepted 14 September 2012

Keywords:
Categorization
Generation
Sampling
Bayesian modeling

ABSTRACT

People are capable of imagining and generating new category exemplars and categories. This ability has not been addressed by previous models of categorization, most of which focus on classifying category exemplars rather than generating them. We develop a formal account of exemplar and category generation which proposes that category knowledge is represented by probability distributions over exemplars and categories, and that new exemplars and categories are generated by sampling from these distributions. This sampling account of generation is evaluated in two pairs of behavioral experiments. In the first pair of experiments, participants were asked to generate novel exemplars of a category. In the second pair of experiments, participants were asked to generate a novel category after observing exemplars from several related categories. The results suggest that generation is influenced by both structural and distributional properties of the observed categories, and we argue that our data are better explained by the sampling account than by several alternative approaches.

1. Introduction

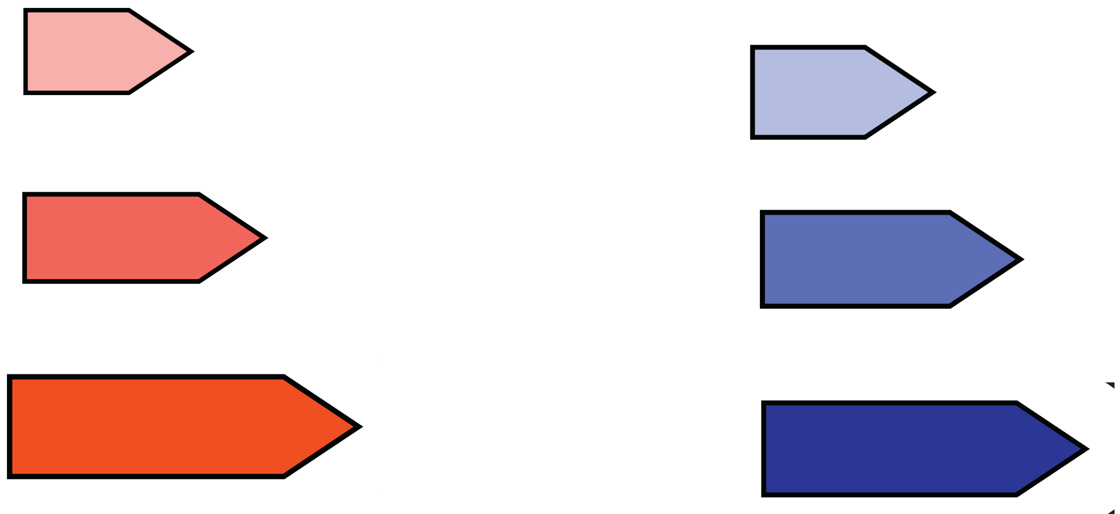
Late one night, Robert Cobb, the owner of the Brown Derby restaurant in Hollywood, went rummaging through his restaurant's kitchen to make himself something to eat. By mixing together a handful of available ingredients including an avocado, chopped lettuce, bacon, and a hard-boiled egg, he created the Cobb salad, which can now be found in restaurants worldwide (Cobb & Willems, 1996). This event is just one demonstration of people's ability to conceive of objects and concepts that they have never encountered. Other demonstrations are found in the history of science. For example, evolutionary biologists predicted the existence of the species later named *Tiktaalik* by conceiving of

Jern & Kemp : Generating new categories— Ex 3

Training Category 1

Training Category 2

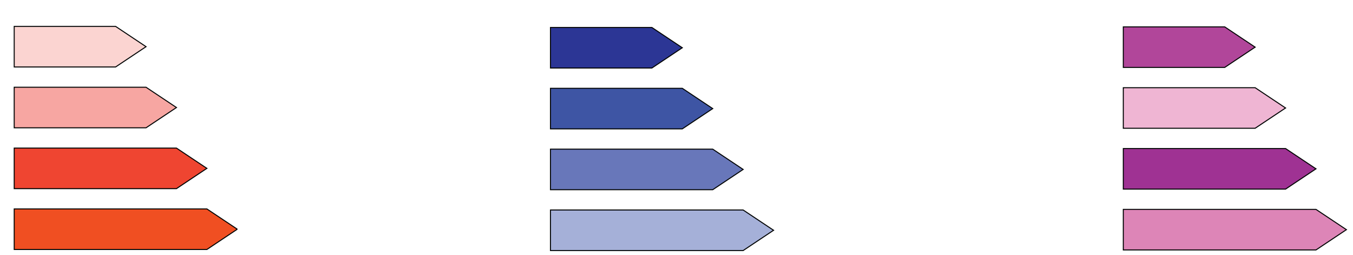
What does Category 3 look like?



- People created "types of crystals" from another planet, which varied in hue(color), length, and saturation
- Task was to generate 6 crystals of a new type

3 conditions (different training examples):

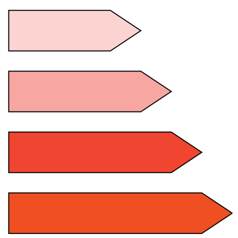
pos. correlation | neg. correlation | uncorrelated



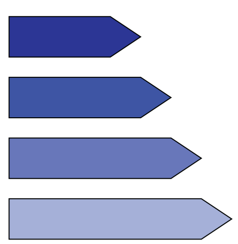
Jern & Kemp : Ex 3 results

correlational structure of novel generated category

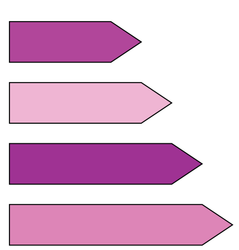
pos. correlation

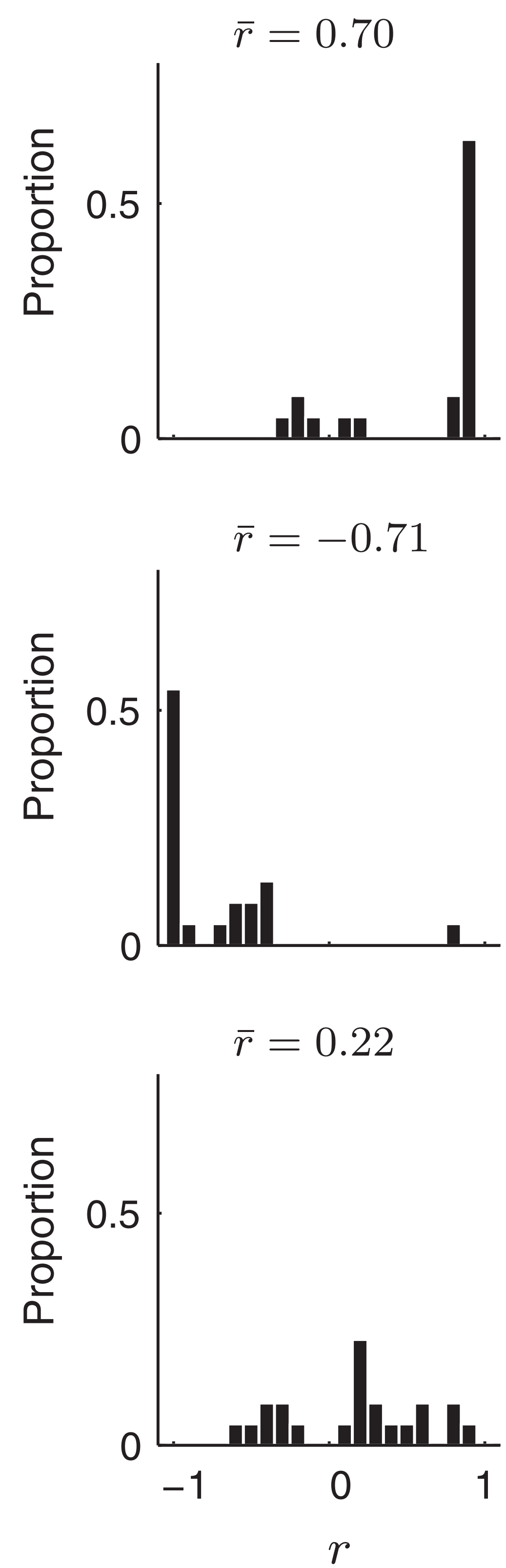


neg. correlation



uncorrelated





People can generate coherent categories that preserve feature relationships in previous categories

COGNITIVE PSYCHOLOGY 27, 1–40 (1994)

Structured Imagination: The Role of Category Structure in Exemplar Generation

THOMAS B. WARD

Texas A&M University

College students imagined animals that might live on a planet somewhere else in the galaxy. In the first experiment, they provided drawings and descriptions of their initial imagined animal, another member of the same species, and a member of a different species. The majority of imagined creatures were structured by properties that are typical of animals on earth: bilateral symmetry, sensory receptors, and appendages. Subjects also allowed shape, appendages and sense receptors to vary often across species but rarely within species. In Experiment 2, subjects' creations were influenced by correlated attributes; those told that the animal was feathered were more likely to produce creatures with wings and beaks, and those told it lived in water and had scales were more likely to produce creatures with fins and gills relative to subjects who were told the animal was furry or who were given no specific features. Experiments 3 and 4 revealed that many subjects approach the task by retrieving exemplars of known earth animals, but that instructions and task constraints can lead to greater use of broader knowledge frameworks. Experiment 5 revealed that the structuring found in college students' imagined animals also holds for extraterrestrials developed by science fiction writers. The results are consistent with the idea that similar structures and processes underlie creative and noncreative aspects of cognition, and are discussed in terms of the concept of *structured imagination*. That is, when subjects create a new member of a known category for an imaginary setting, their imagination is structured by a particular set of properties that are characteristic of that category.

Category structures and processes play a central role in many human activities. Consequently, they have been studied heavily over the past three decades, with particular attention being given to the topics of category learning, classification, and inductive inference. Although these are important cognitive activities, categories serve many other functions which have yet to be studied as thoroughly. One of these is in the use of

Portions of the data from Experiment 1 were reported at the meeting of the Psychonomic Society, November 1991, and portions of the data from Experiments 3 and 4 were reported at the meeting of the Psychonomic Society, November 1992. I thank Ron Finke, Nancy Rhodes, Steve Smith, and the Creative Cognition Research Group for valuable comments throughout various phases of this research. I also gratefully acknowledge the comments of two anonymous reviewers. Please address correspondence and reprint requests to Thomas B. Ward, Department of Psychology, Texas A&M University, College Station, TX 77843. e-mail: TBW7959@TAMVENUS.TAMU.EDU.

Structured imagination of new categories

- Although people most often study learning, classification, and inference, **generation and imagination also draw on category knowledge**

- How can we best conceptualize the process by which people generate and imagine new entities?
  - ✳ Is generation characterized by the same processes and representations as classification?

- Neither prototype, exemplar, or knowledge views offer much specific guidance about how generation would work

- IMHO, generative tasks are a more direct look into representation
  - ✳ Why? Classification is easier (and many models make same predictions), but not for generation

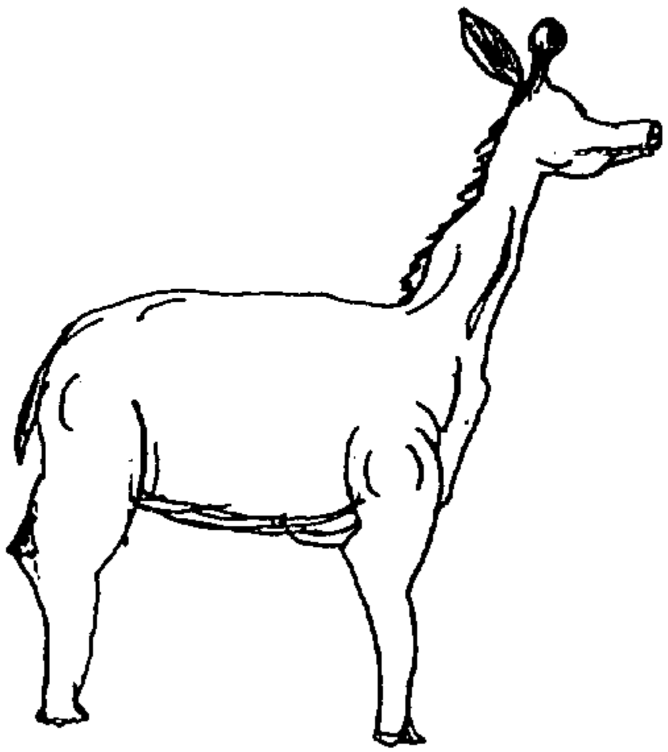
Ward Ex 1 : Imagining new animals

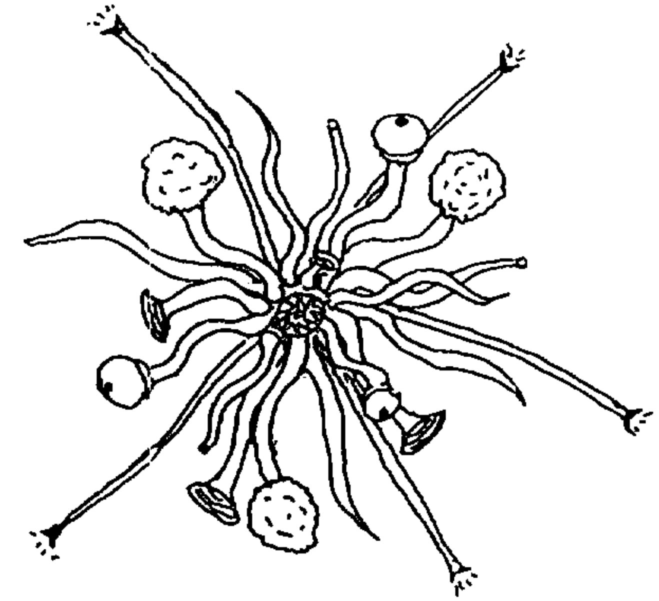
- People imagined, drew, and described (non-obvious properties about) an animal living on a planet very different from earth

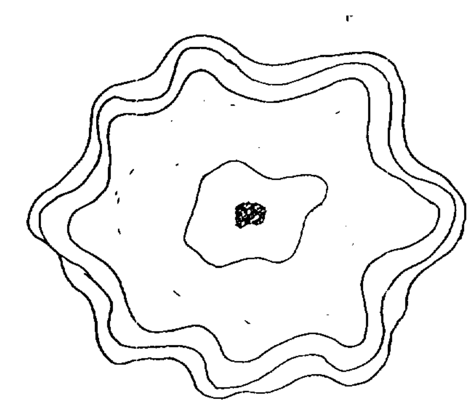
  Steps
  1) Draw imaginary animal from another planet
  2) Draw another example of the same species
  3) Draw a different specifies of animal

- Coded for typical properties for earth creatures (sensory organs, appendages), as well as atypical properties like unusual sense organs, unusual appendages, and other differences with earth animals

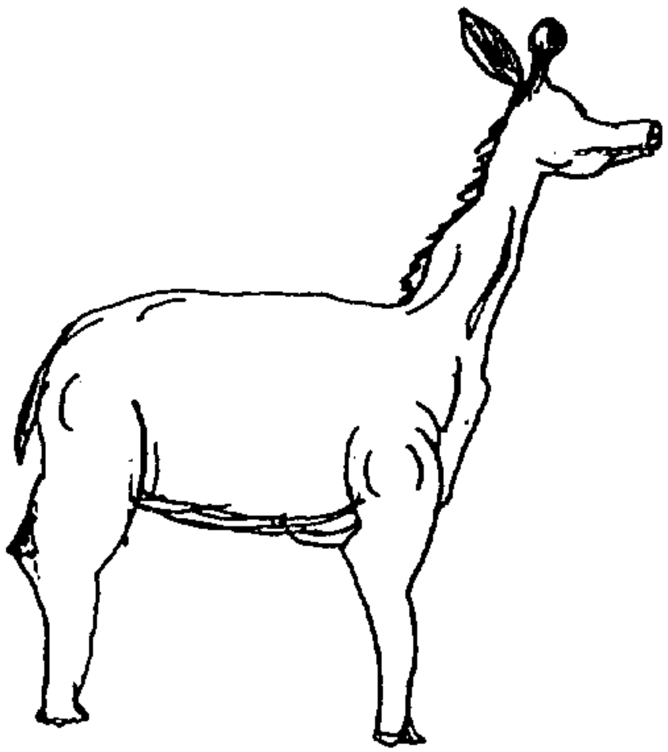
Ward Ex 1 : Results of imagining new animals

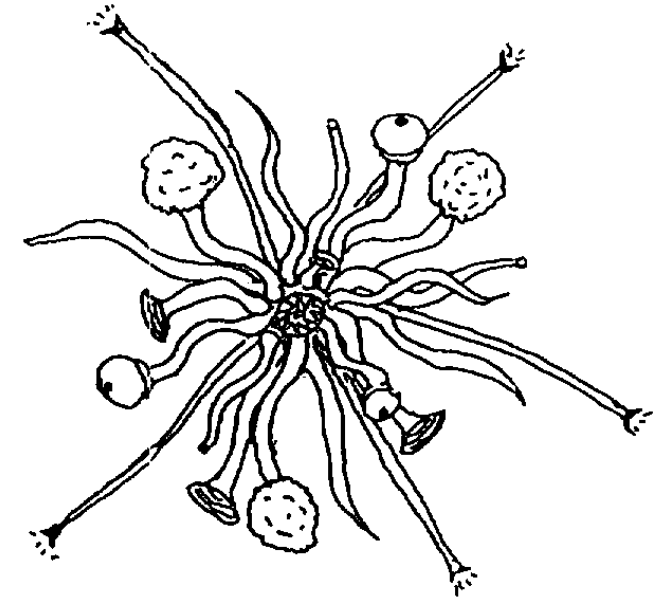


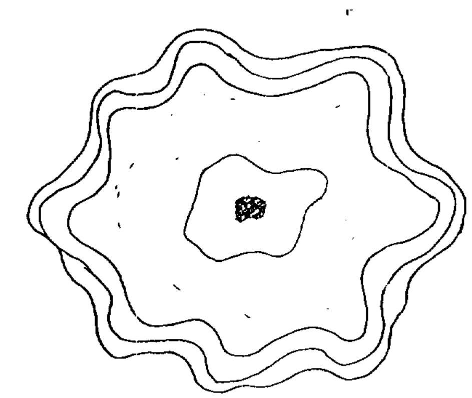




Ward Ex 1 : Results of imagining new animals



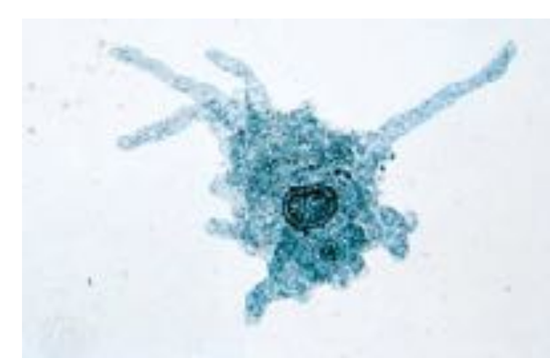




Is this an Okapi?

ALF?

Amoeba?



Ward Ex 1 : Results of imagining new animals

Strong tendency to produce creatures that differ from earth creatures in only a limited number of ways

- most had bilateral symmetry (89%), one major sense organ (92%; eyes, nose, mouth, etc.), and one major appendage (84%; legs, arms. etc.)
- but not just copies of earth creatures; a majority (65%) had at least one unusual sensor organ or appendage

TABLE 2
Number of Imagined Animals That Had Arms, Legs, Wings, Eyes, Ears, and Noses, and Typical Values of Each of Those Attributes

Type of attribute	Present in creature	Having typical value
Arms	11	9
Legs	29	24
Wings	4	4
Eyes	33	24
Ears	19	19
Nose	22	22

Note. Maximum possible number is 37. Creatures were counted as having a typical value of an attribute if they had two arms, wings, eyes and ears, one nose, and two or four legs.

- e..g, 89% have eyes, and 65% have two eyes; 65% have two legs

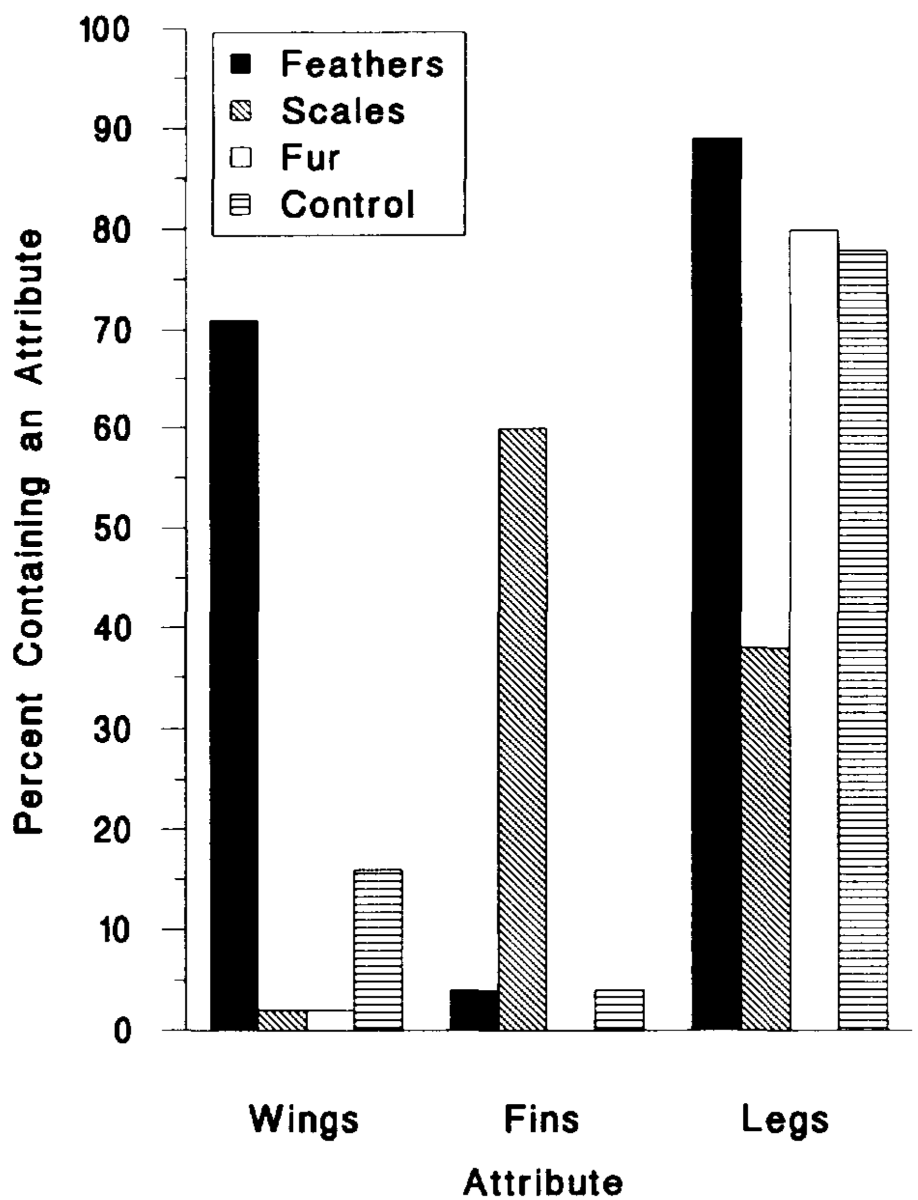
TABLE 3
Percentages of Subjects Who Varied the Shape, Sense Organs, Appendages, or Size of the Creatures within and between Species

Attribute	Varied within species	Varied across species
Shape	11	95
Sense organs	24	59
Appendages	16	76
Size	54	40

- e..g, shape, sense organs, and appendages vary across — but not within — species

Ward Ex 2 : Do features co-occur in usual ways?

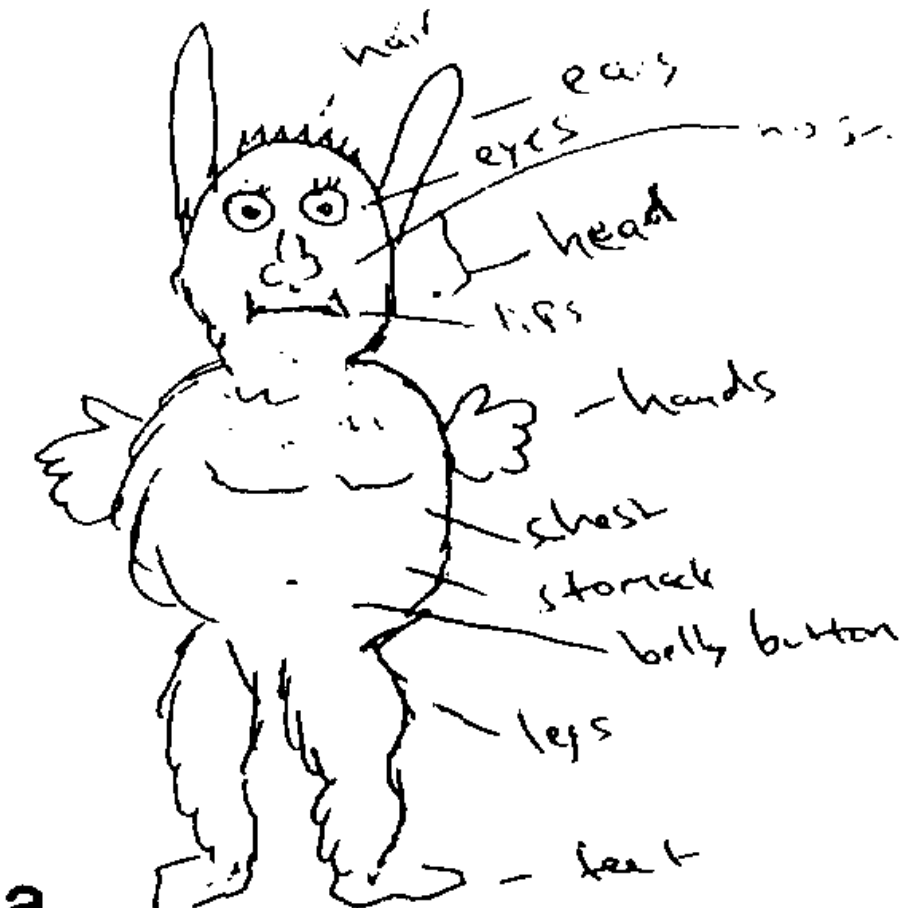
- We know that features cluster together in our concepts of earth animals (e.g., Murphy & Medin)
- People were asked to imagine animals in one of four conditions
  1) “have feathers"
  2) “is furry”
  3) “has scales (and lives in water)”
  4) control / no special information



Results: Features co-occur as they do in earth animals

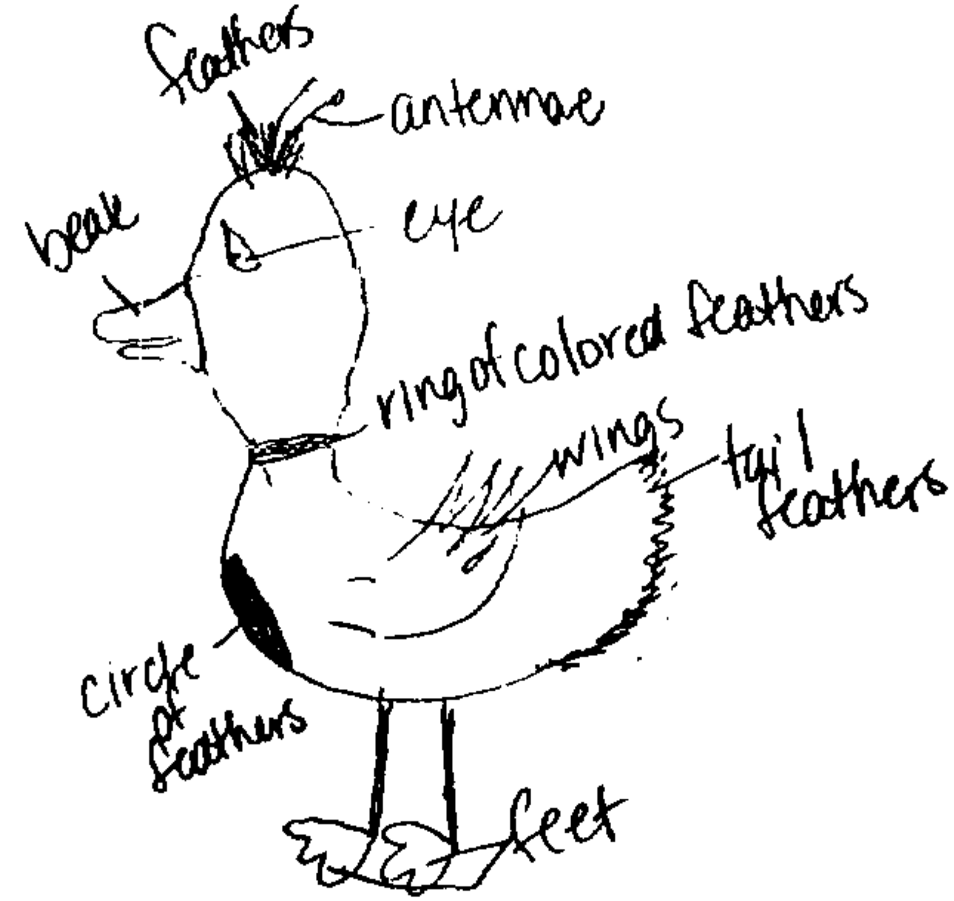
- most with *feathers* have *wings* (70%)
- most with *scales* have *fins* (60%)
- most with *fur* have *legs* (80%), and all other conditions too except scales

Some artistic drawings in Exp 2, but not particularly creative



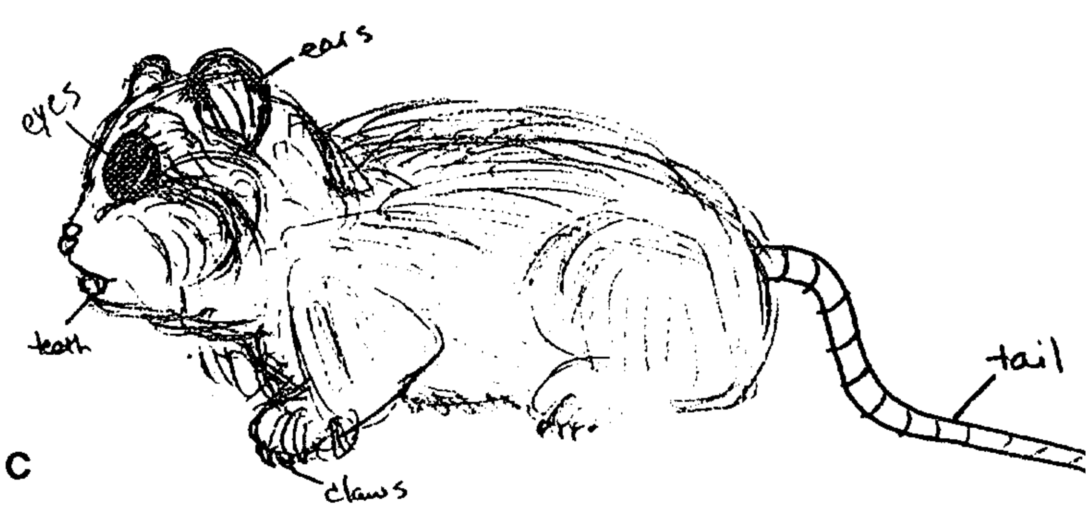
a

control condition



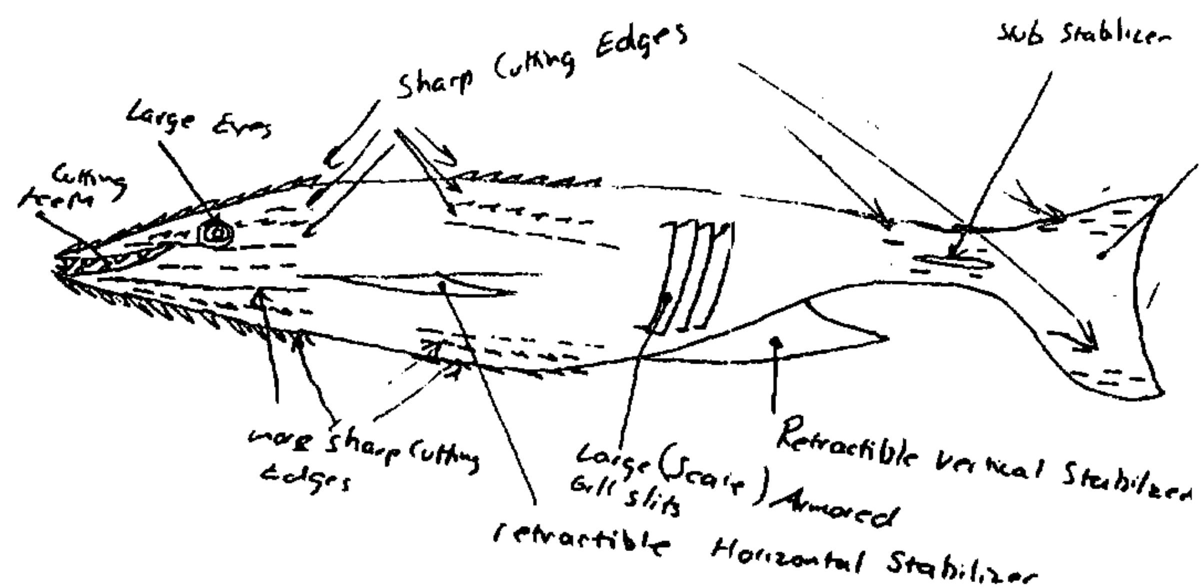
b

feathers condition



c

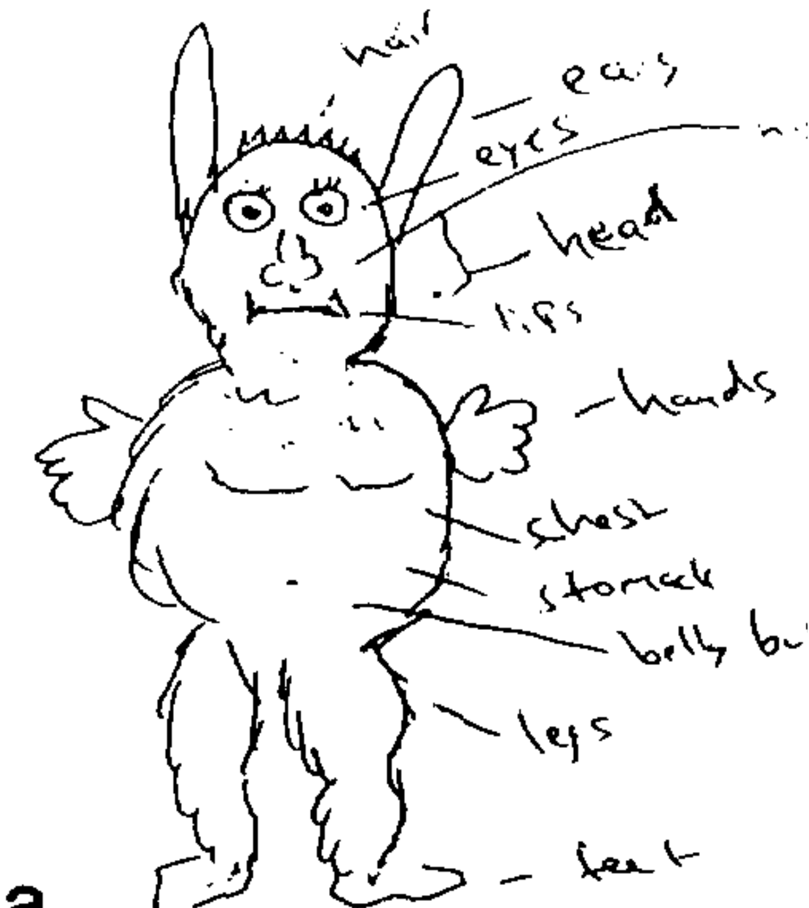
fur condition



d

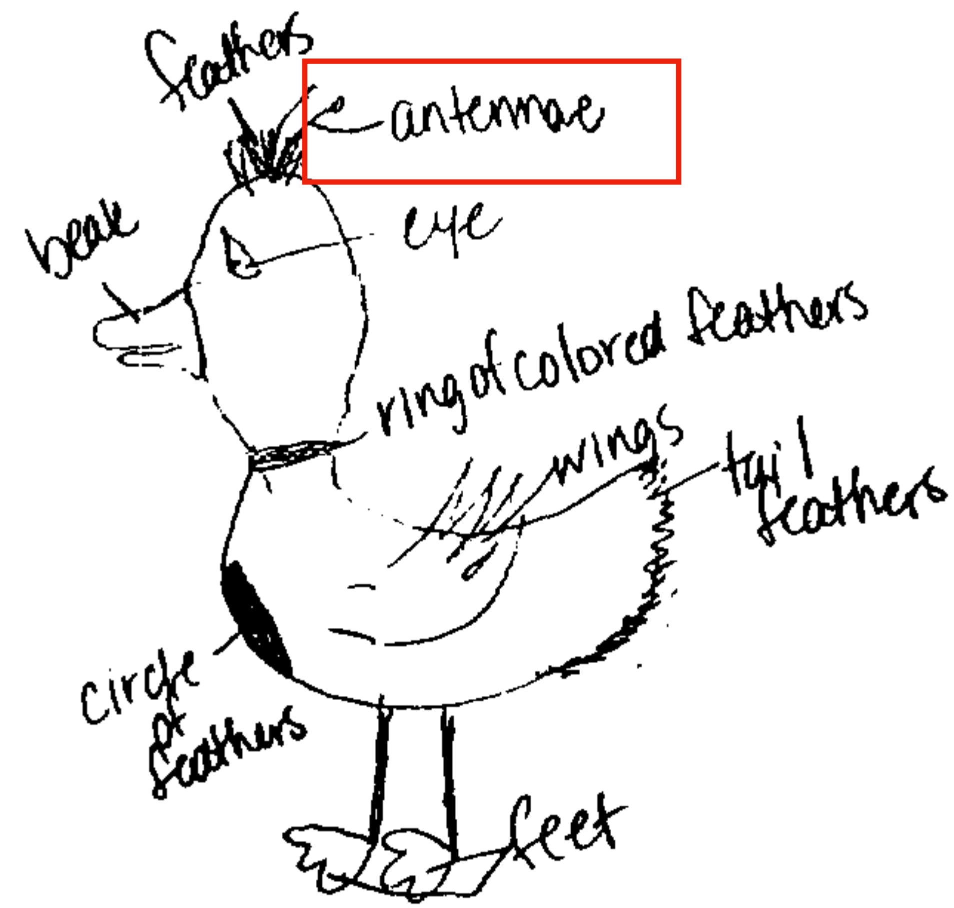
scales condition

Some artistic drawings in Exp 2, but not particularly creative



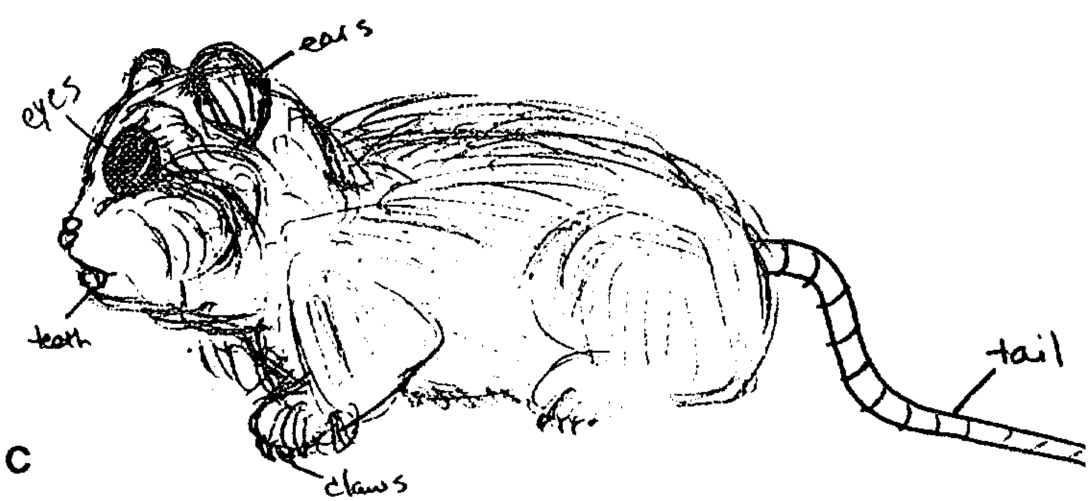
a

control condition



b

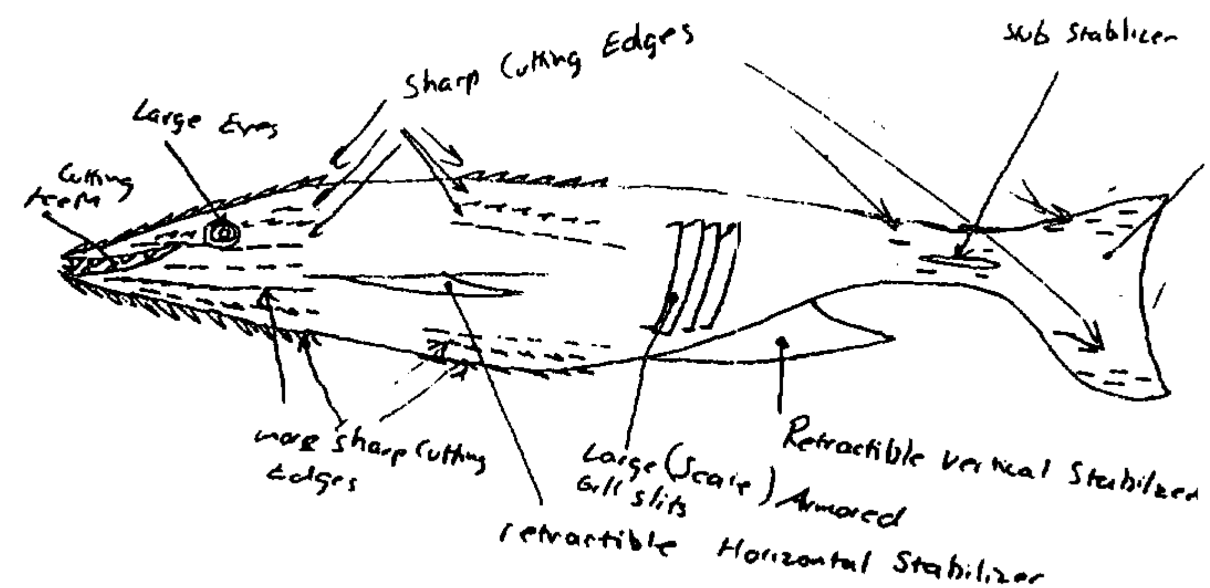
feathers condition



c

fur condition

d



scales condition

Ward Ex 3 : The role of knowledge and naive theories

- Does the creation process activate background knowledge and our naive theories? (which would explain feature correlations)

- Can people do more than activate and modify stored exemplars? (exemplars would also explain feature correlations)

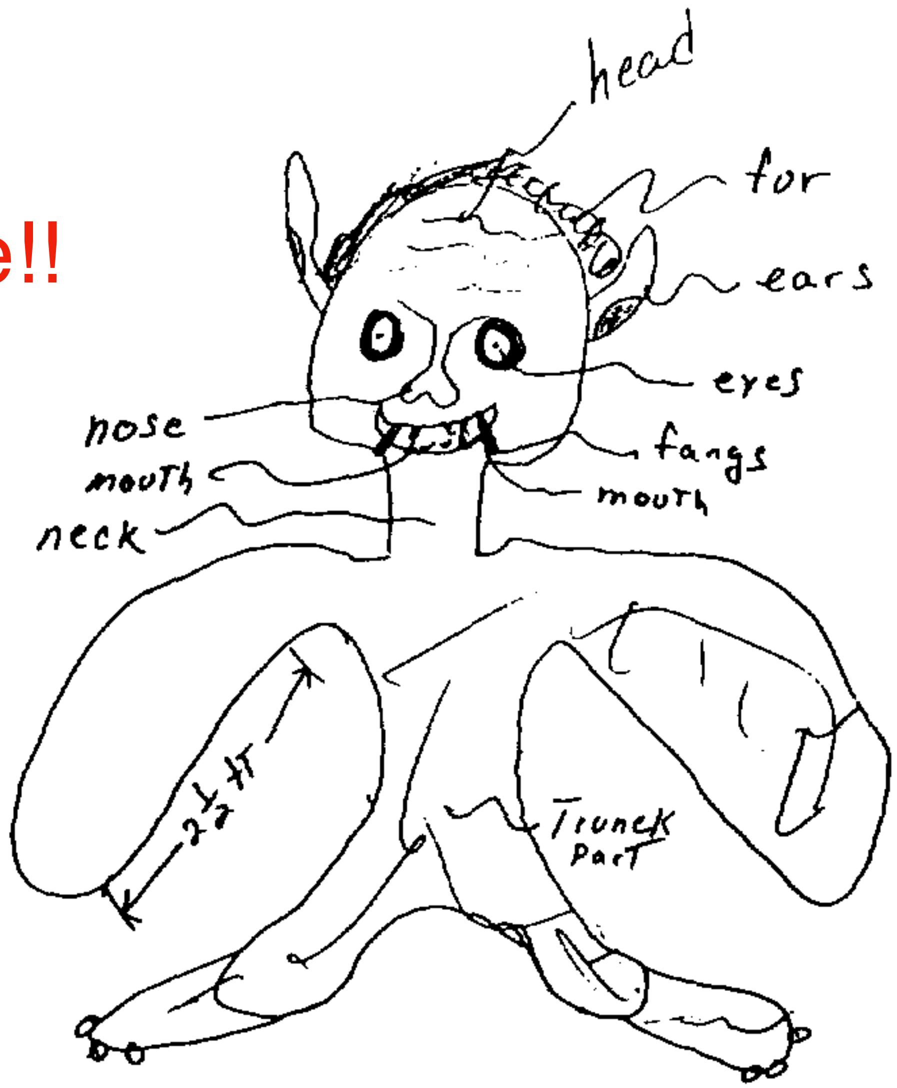
Factor 1 (2 x 2 design)
- lives on planet with “mostly molten rock" with just a few islands, the ability to travel between islands was important
- extremely violent winds above surface, and avoiding winds is important for survival

Factor 2 (same as previous experiment)
- “has fur” vs. “has feathers”

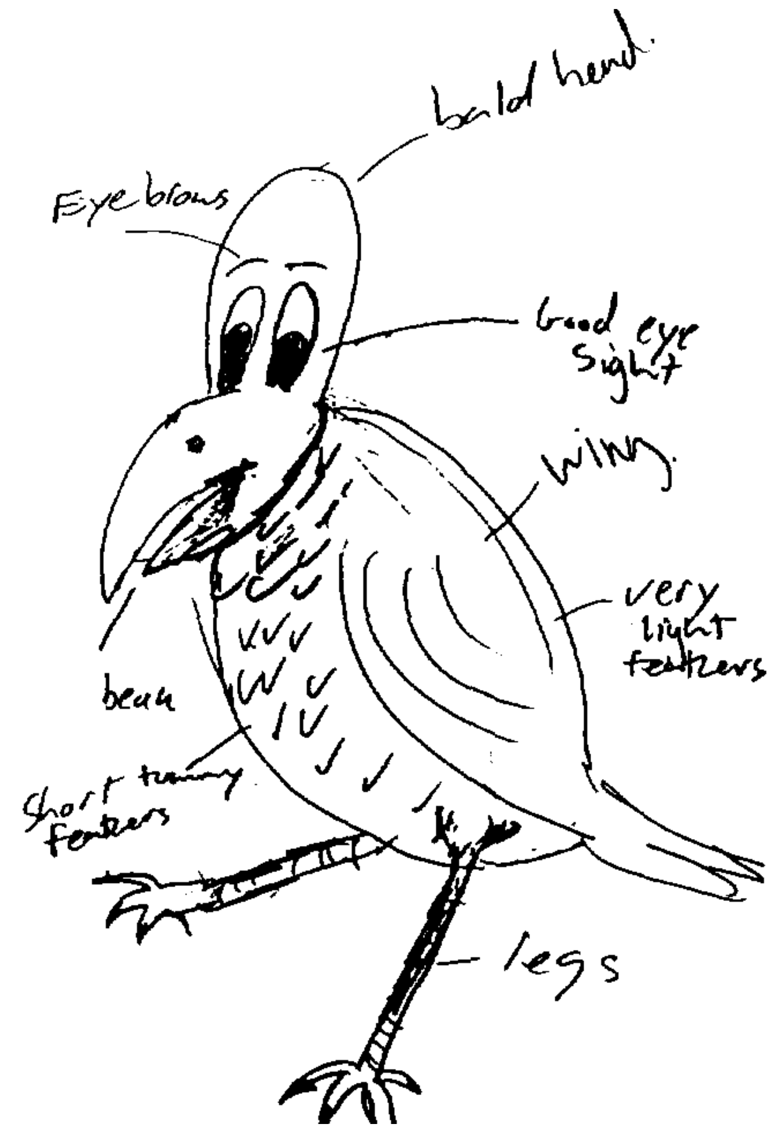
Ward Ex 3 : Results for the role of knowledge and naive theories

Molten-Fur condition

Cute!!



Molten-Feather condition

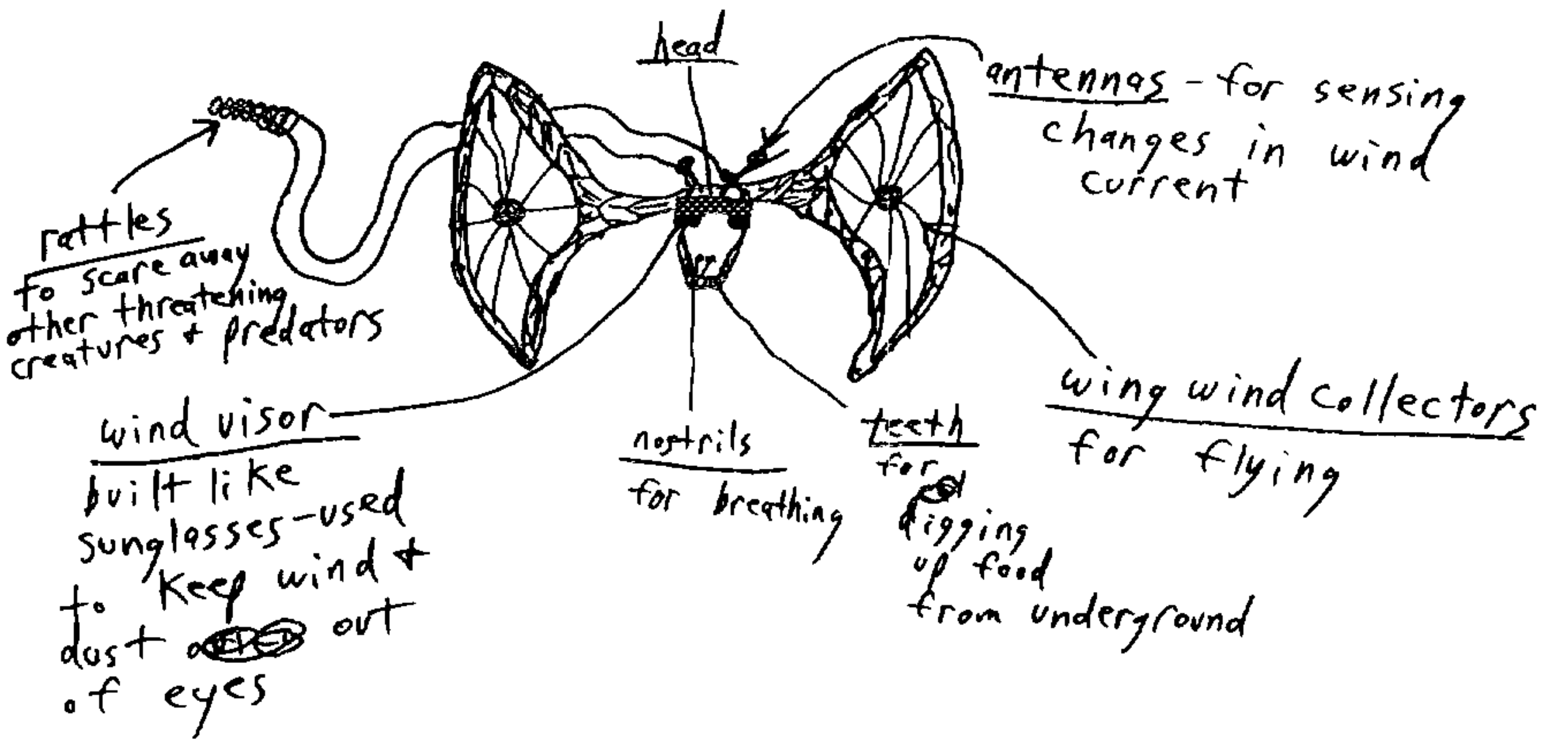


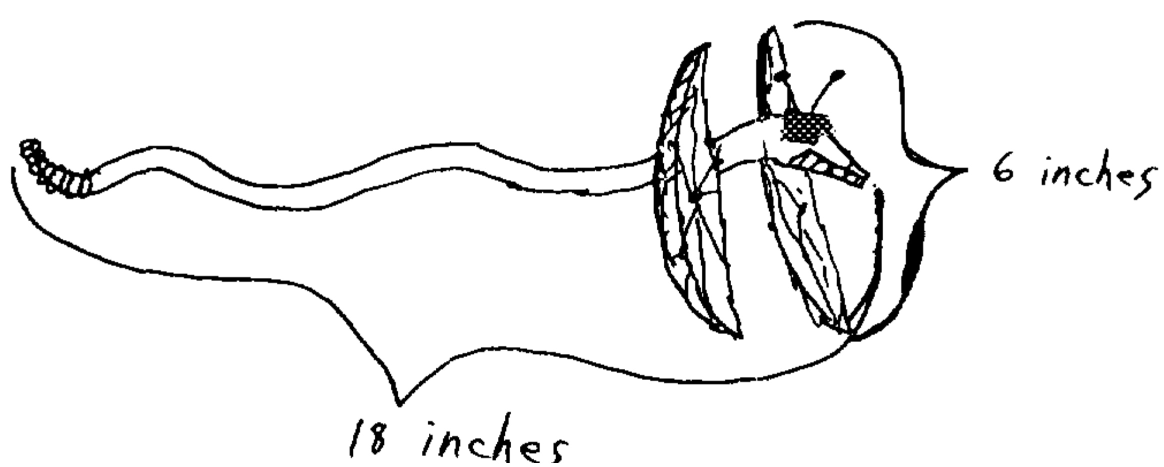
Percent of drawings with these properties

Property	Molten fur	Molten feather
Wings	60	100
Flight	70	90

Vast majority of creatures fly in molten environment — even furry creatures —suggesting environmental constraints outweigh correlational tendencies; less flying in windy environment

Ward Ex 3 : Knowledge + constraints leads to creativity



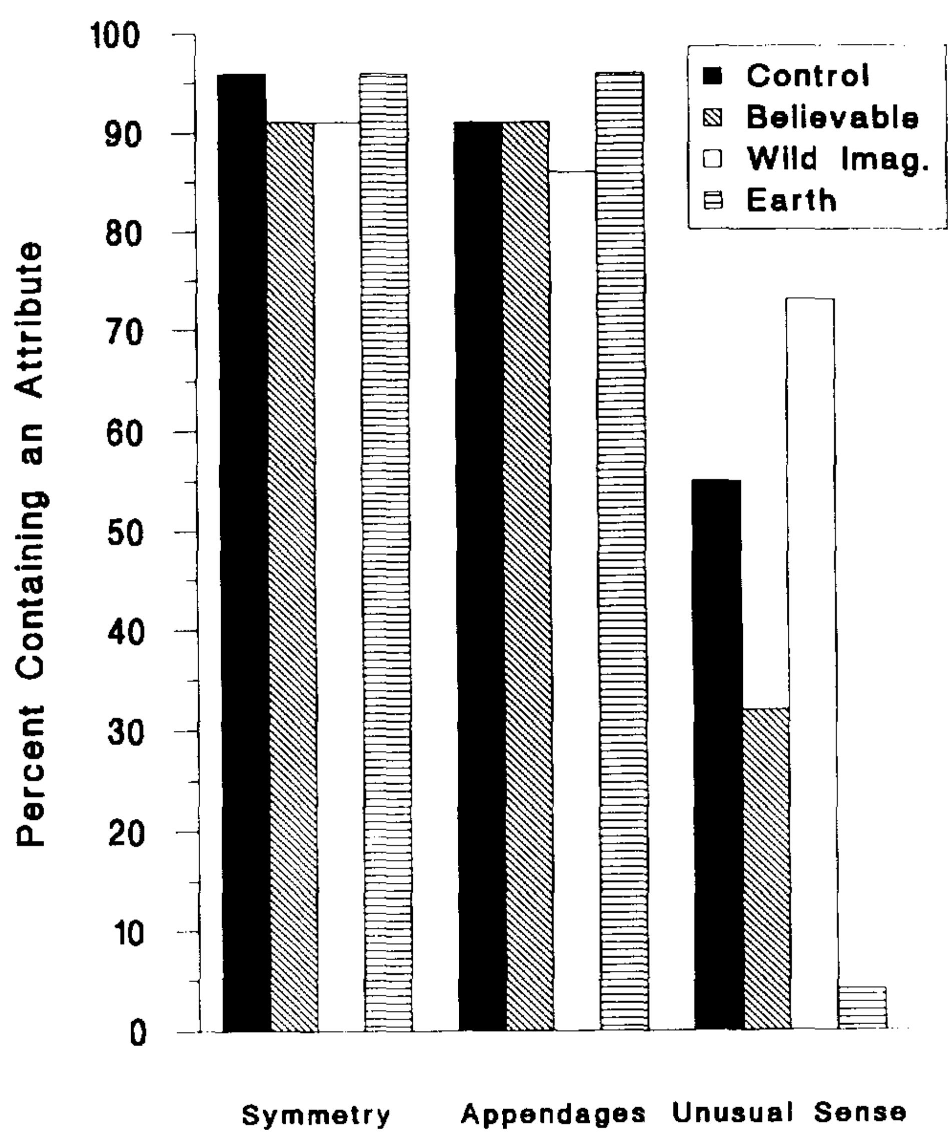


Lots of creative adaptations!

- many people in windy condition developed creatures with suction cups, claws, etc.
- many people in Molten condition developed creatures that had heat-resistant feet (25%)

Ward Ex 4 : What if people used their "wild imaginations?"

- Would people abandon more world knowledge if encouraged to do so?
- People were asked to imagine animals in one of four conditions
  1) use your "wildest imagination", can be unbelievable
  2) draw believable and realistic
  3) draw new species discovered on earth
  4) control

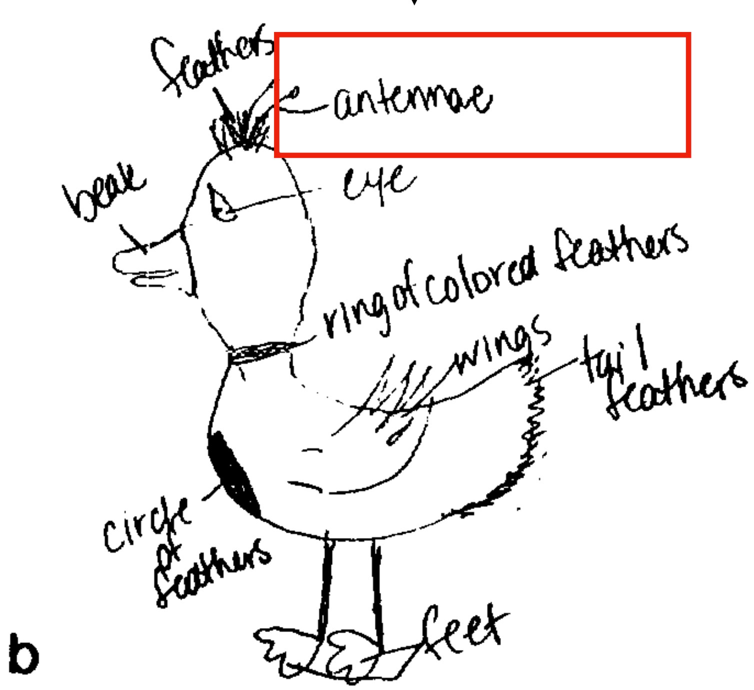


Results:

- People were not that creative, even when using their "wildest imagination" (less creative than Ex 3)
- The largest change was the number of unusual sense organs (75%)

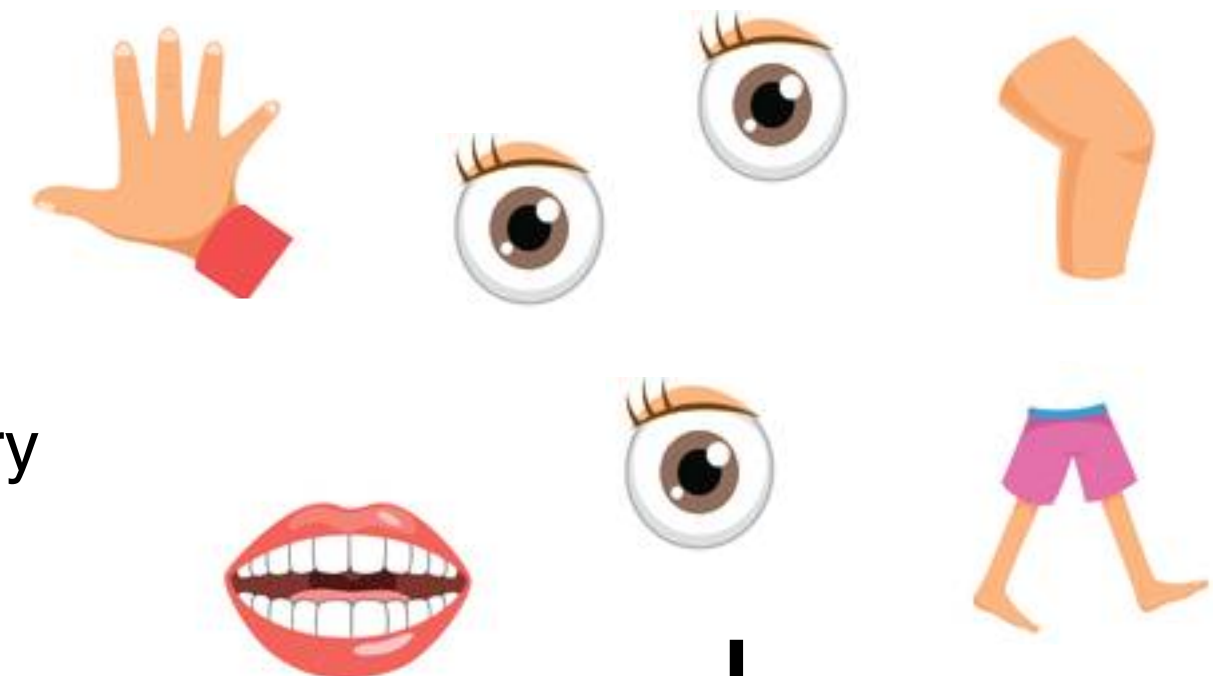
Possible accounts of generation

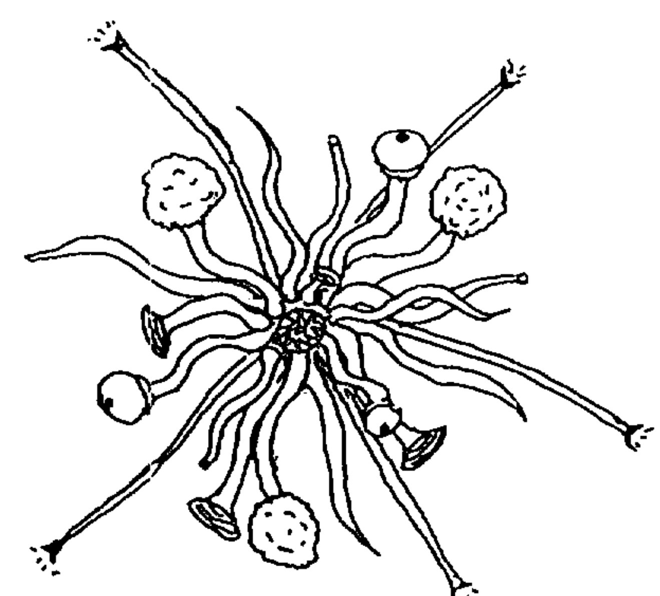
“Sample exemplar and modify”
(Ward’s proposal)



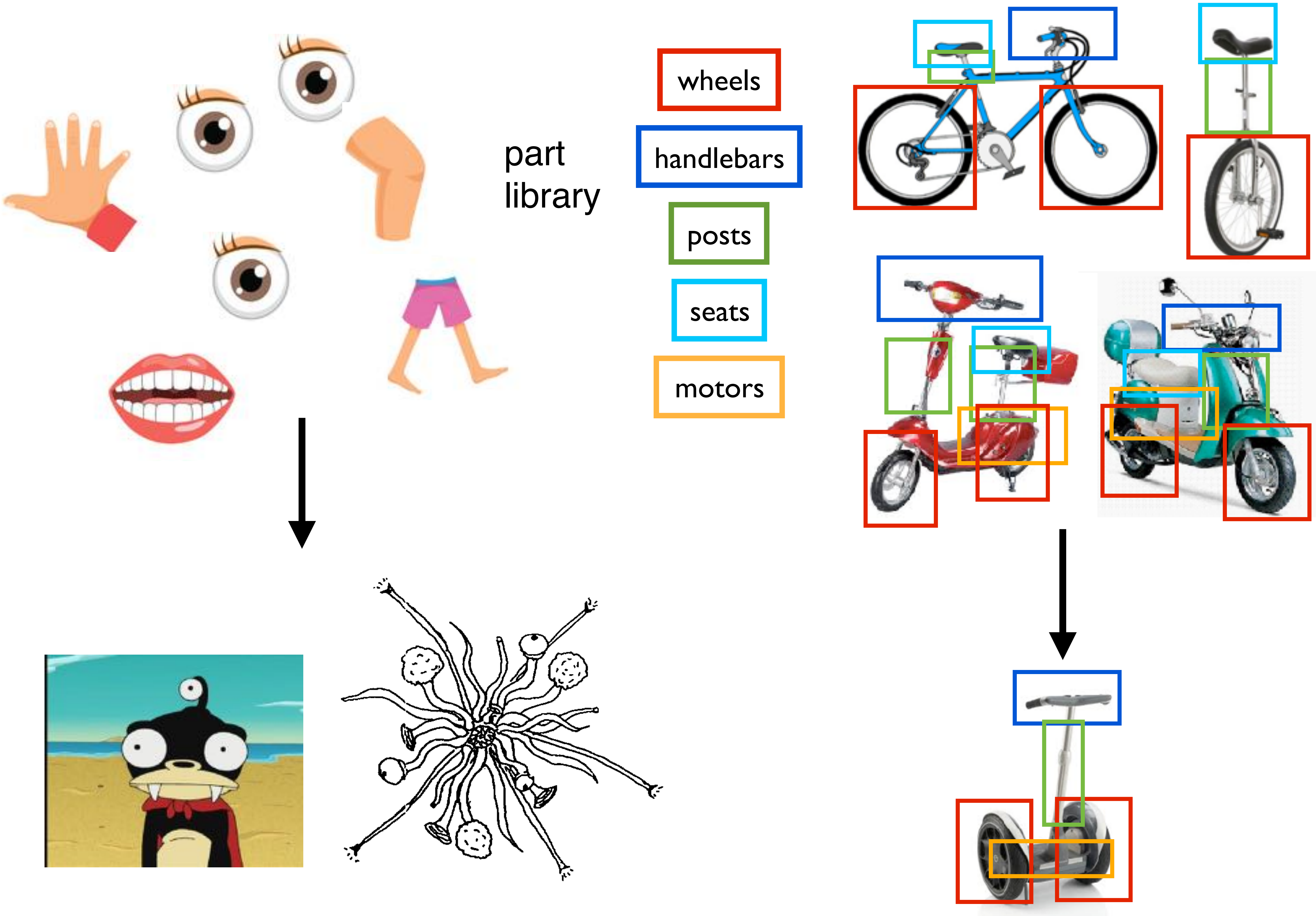
“Compositional generation”

part
library





Compositional generation



Conclusions : Structured imagination of new categories

- When people imagine novel animals, the properties of their creations are predictable from research on non-creative aspects of categorization

- How can we best conceptualize the process by which people generate and imagine new entities?
  - ✳ People often take path of least resistance, even when using “wildest imagination”
  - ✳ “sample exemplar and modify” or “compositional parts” are candidate models for task
  - ✳ Clearly relied more on knowledge and naive theories when necessary, creating the most unusual creatures when given constraints
  - ✳ This ability doesn’t fit perfectly into any current theory of concepts

- Generative tasks provide a rich look into conceptual representation!
  - ✳ Why? Many different representations can lead to the same classification performance

RESEARCH ARTICLES

COGNITIVE SCIENCE

Human-level concept learning through probabilistic program induction

Brenden M. Lake,[1*] Ruslan Salakhutdinov,[2] Joshua B. Tenenbaum[3]

People learning new concepts can often generalize successfully from just a single example, yet machine learning algorithms typically require tens or hundreds of examples to perform with similar accuracy. People can also use learned concepts in richer ways than conventional algorithms—for action, imagination, and explanation. We present a computational model that captures these human learning abilities for a large class of simple visual concepts: handwritten characters from the world's alphabets. The model represents concepts as simple programs that best explain observed examples under a Bayesian criterion. On a challenging one-shot classification task, the model achieves human-level performance while outperforming recent deep learning approaches. We also present several "visual Turing tests" probing the model's creative generalization abilities, which in many cases are indistinguishable from human behavior.

Despite remarkable advances in artificial intelligence and machine learning, two aspects of human conceptual knowledge have eluded machine systems. First, for most interesting kinds of natural and man-made categories, people can learn a new concept from just one or a handful of examples, whereas standard algorithms in machine learning require tens or hundreds of examples to perform similarly. For instance, people may only need to see one example of a novel two-wheeled vehicle (Fig. 1A) in order to grasp the boundaries of the new concept, and even children can make meaningful generalizations via "one-shot learning" (*1–3*). In contrast, many of the leading approaches in machine learning are also the most data-hungry, especially "deep learning" models that have achieved new levels of performance on object and speech recognition benchmarks (*4–9*). Second, people learn richer representations than machines do, even for simple concepts (Fig. 1B), using them for a wider range of functions, including (Fig. 1, ii) creating new exemplars (*10*), (Fig. 1, iii) parsing objects into parts and relations (*11*), and (Fig. 1, iv) creating new abstract categories of objects based on existing categories (*12*, *13*). In contrast, the best machine classifiers do not perform these additional functions, which are rarely studied and usually require specialized algorithms. A central challenge is to explain these two aspects of human-level concept learning: How do people learn new concepts from just one or a few examples? And how do people learn such abstract, rich, and flexible representations? An even greater challenge arises when putting them together: How can learning succeed from such sparse data yet also produce such rich representations? For any theory of

[1]Center for Data Science, New York University, 726 Broadway, New York, NY 10003, USA. [2]Department of Computer Science and Department of Statistics, University of Toronto, 6 King's College Road, Toronto, ON M5S 3G4, Canada. [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
*Corresponding author. E-mail: brenden@nyu.edu

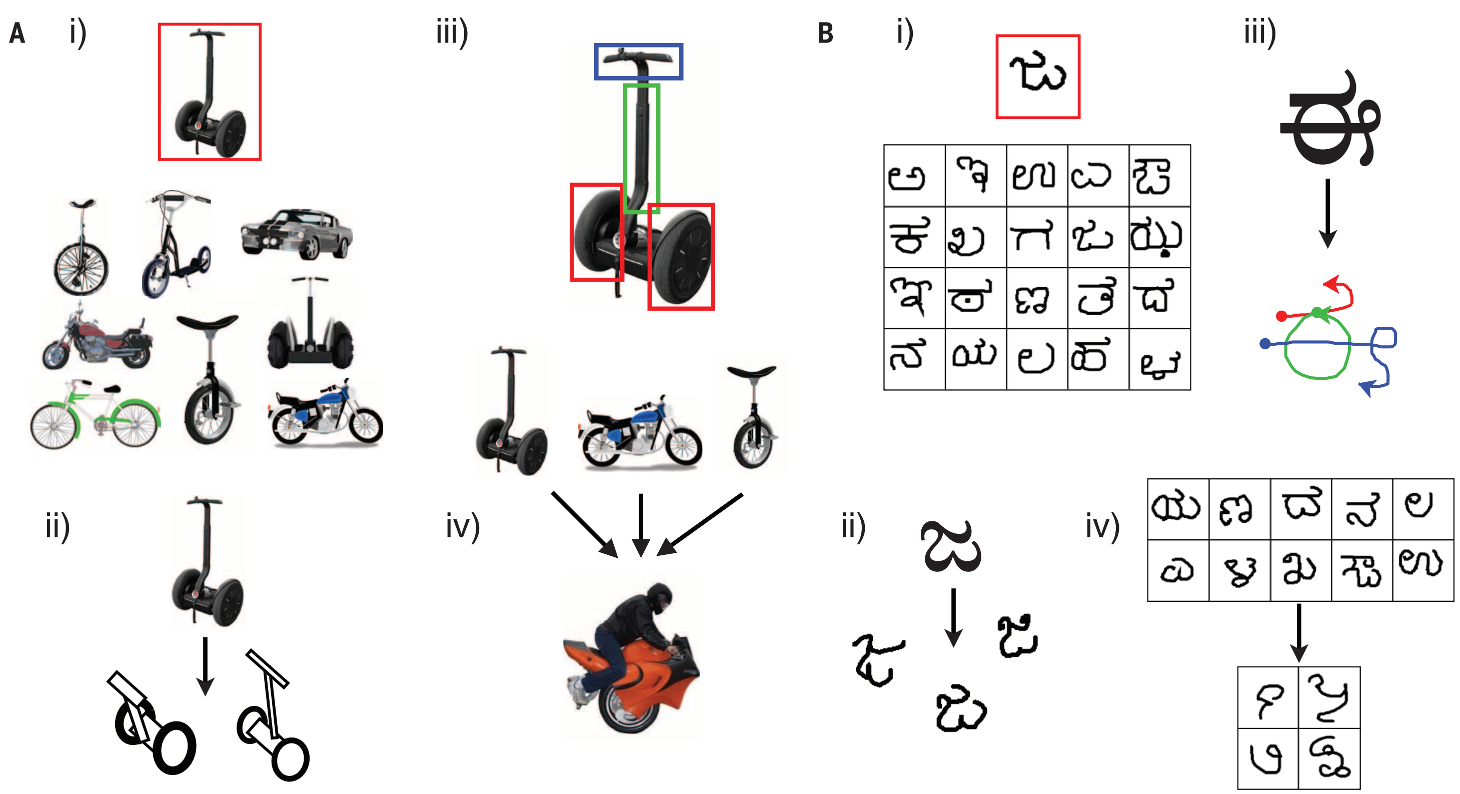
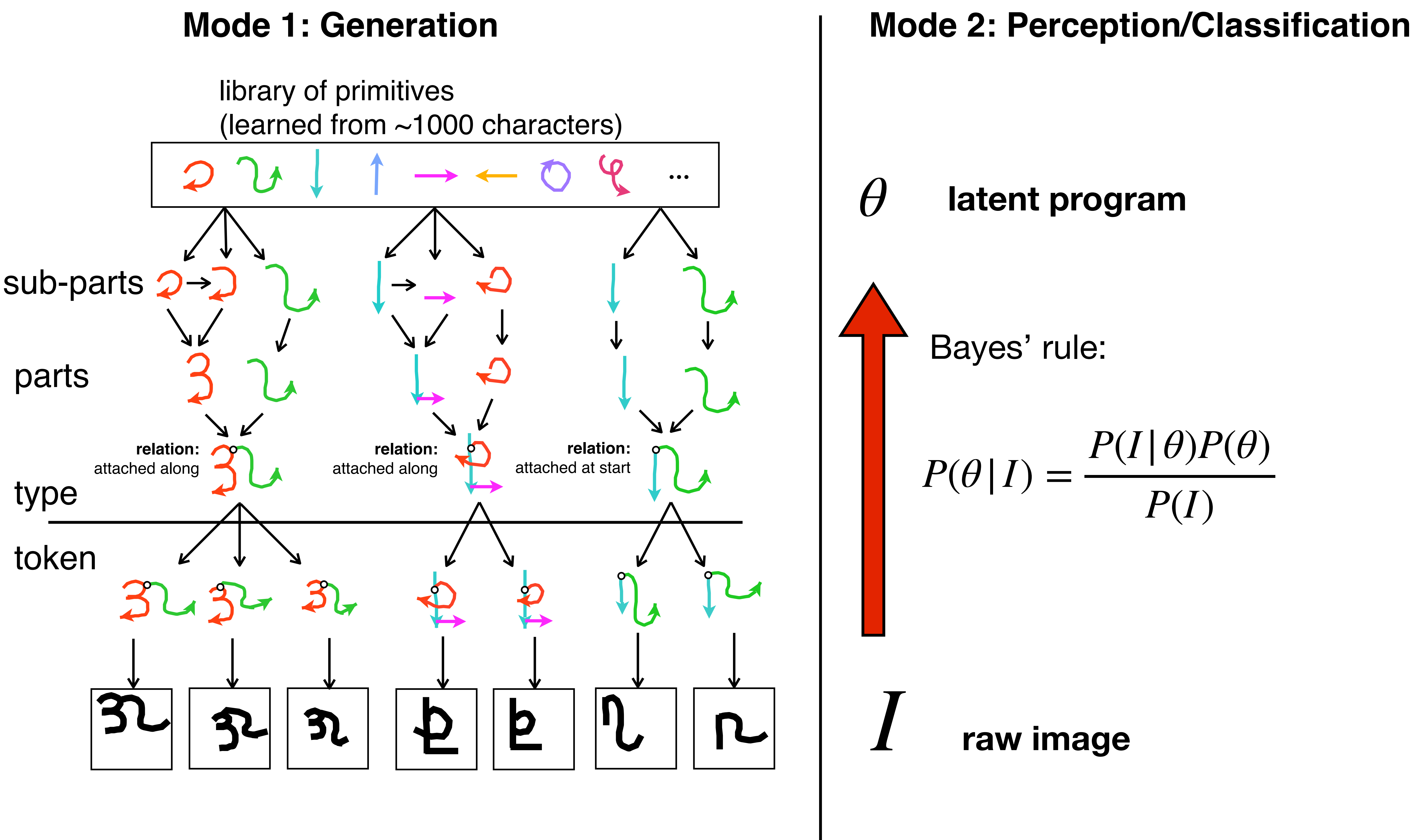


Fig. 1. People can learn rich concepts from limited data. (**A** and **B**) A single example of a new concept (red boxes) can be enough information to support the (i) classification of new examples, (ii) generation of new examples, (iii) parsing an object into parts and relations (parts segmented by color), and (iv) generation of new concepts from related concepts. [Image credit for (A), iv, bottom: With permission from Glenn Roberts and Motorcycle Mojo Magazine]

Compositional generative model of new concepts

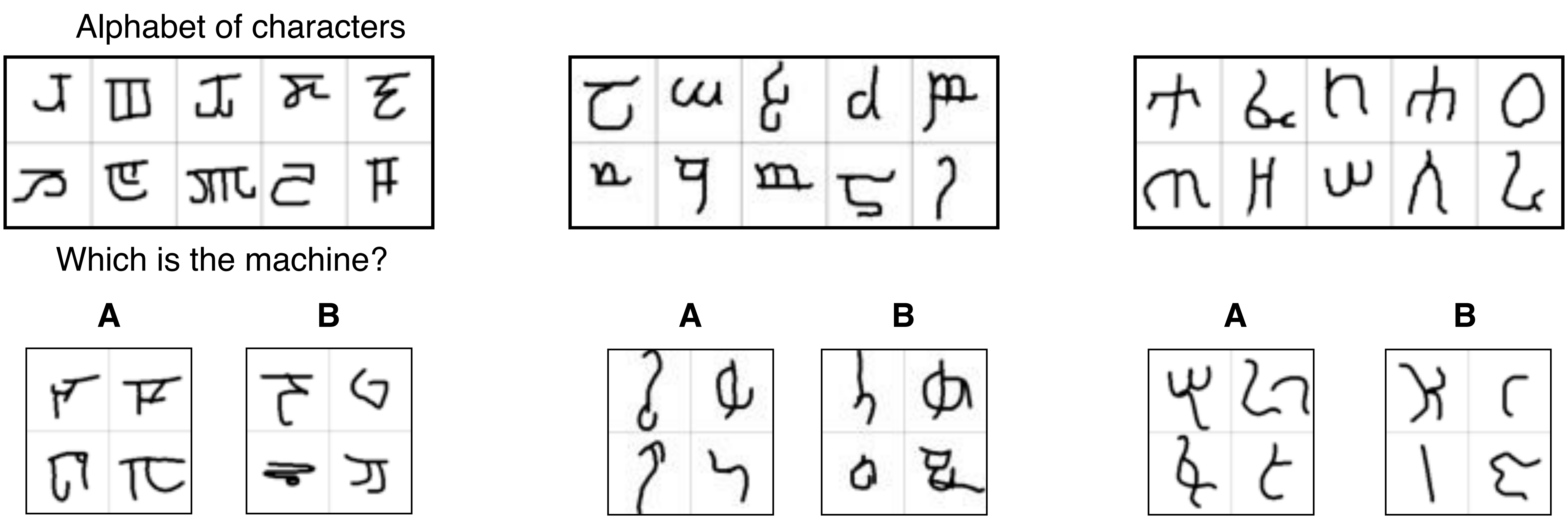
Builds upon Bayesian concept learning (Xu & Tenenbaum), concepts as formal expressions/programs (Rational Rules); concepts as explanations (Murphy & Medin), causal-model theory (Rehder), analysis-by-synthesis models



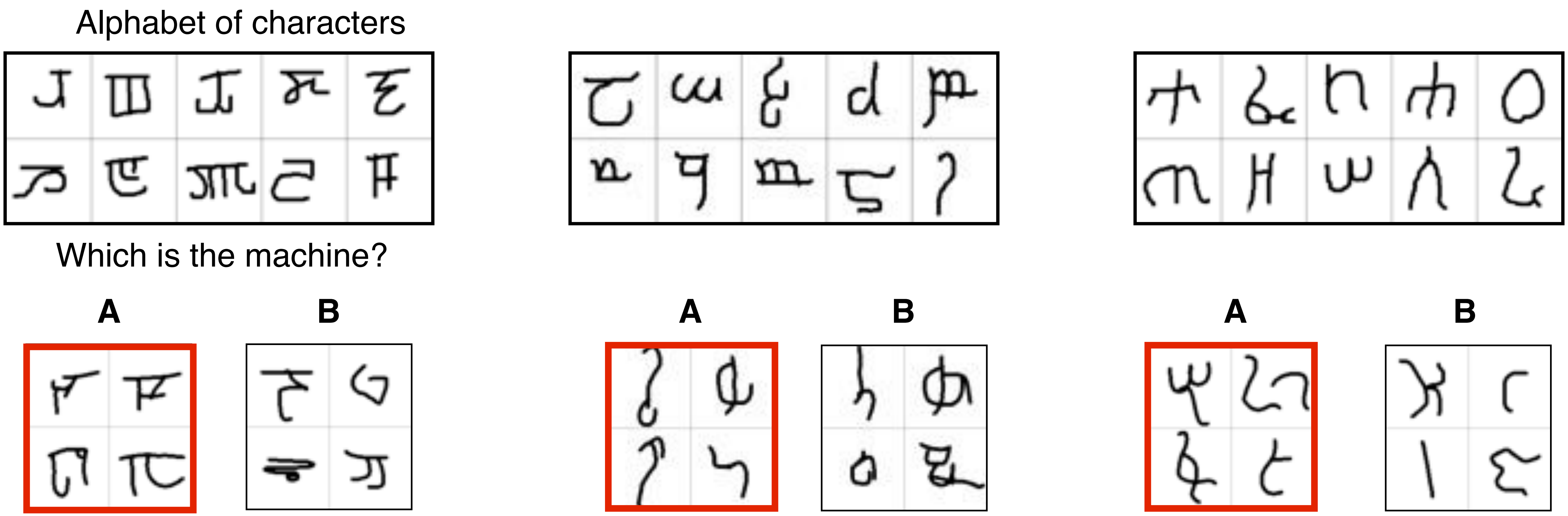
Task: “Design a new character from the same alphabet”

3 seconds remaining

A “visual Turing test” for generating new concepts

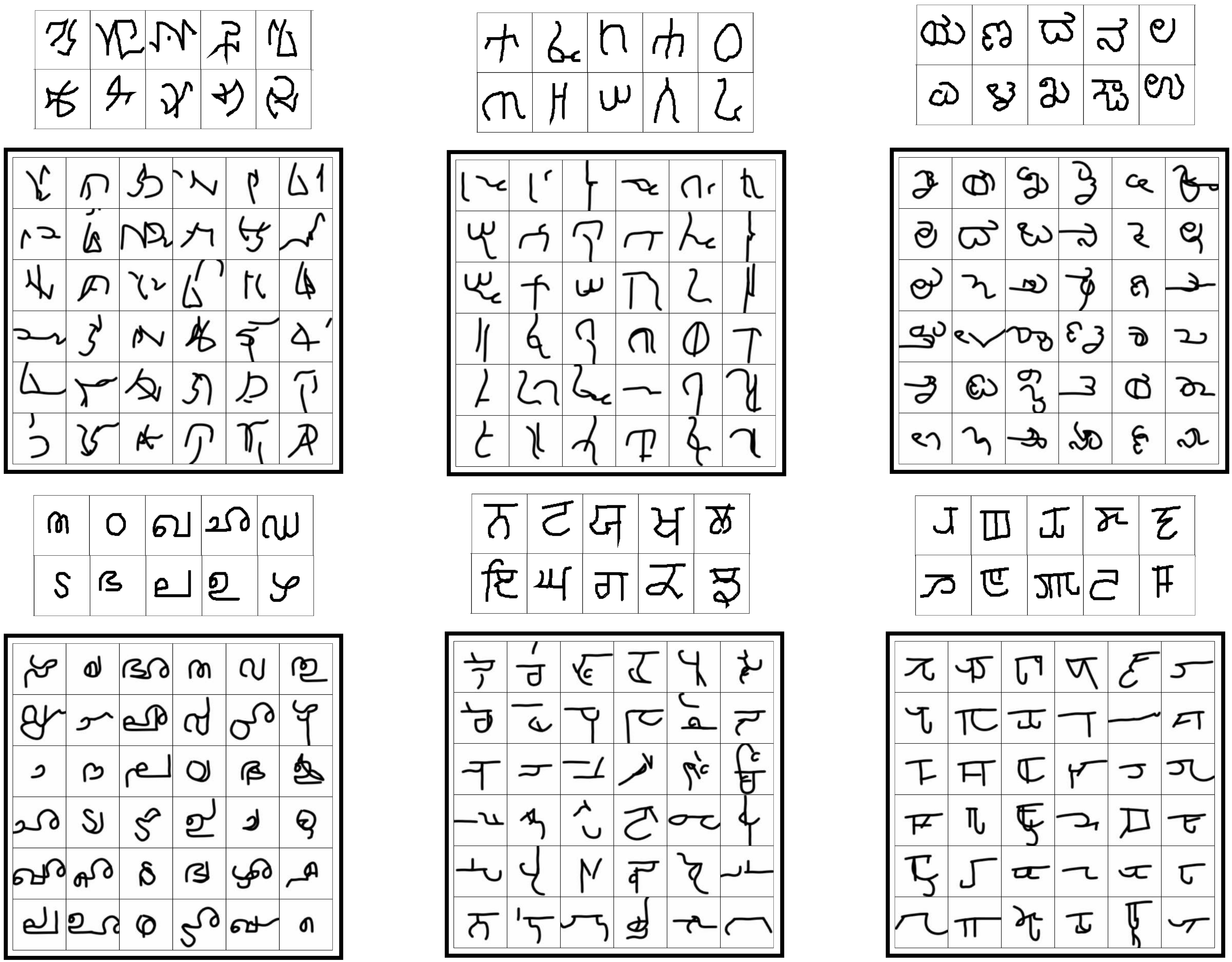


A “visual Turing test” for generating new concepts



machine generated

49% ID level (8/35 judges above chance)



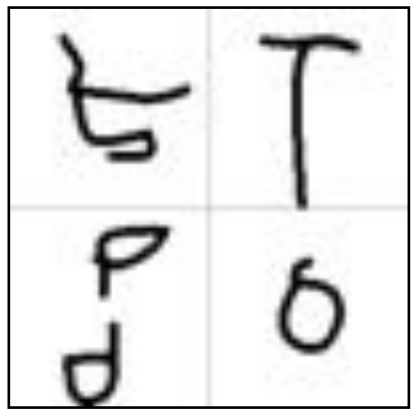
Task: “Design a new handwritten character”

3 seconds
remaining

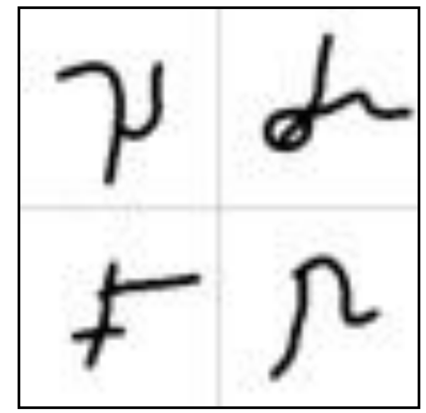
A “visual Turing test” for generating new concepts

Which is the machine?

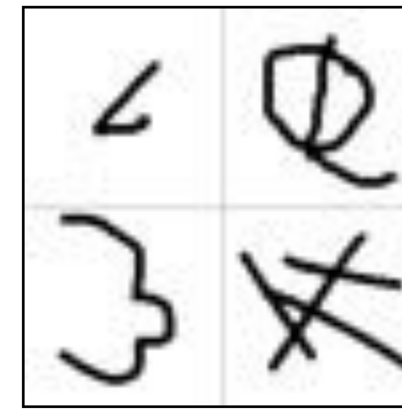
A



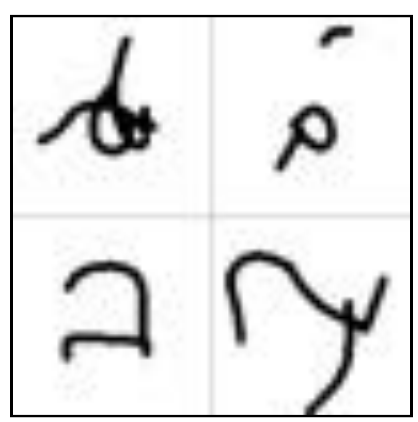
B



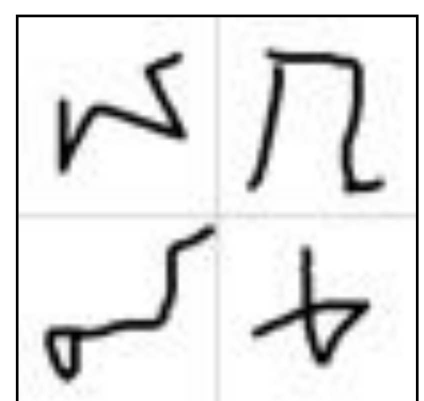
A



B



A

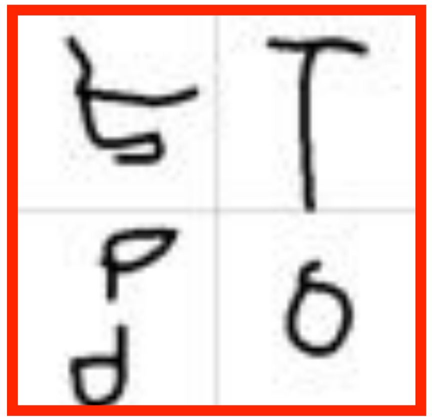


B

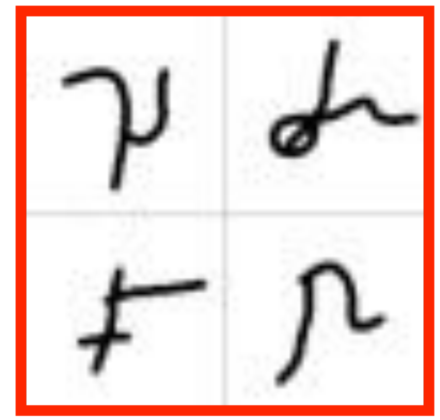
A “visual Turing test” for generating new concepts

Which is the machine?

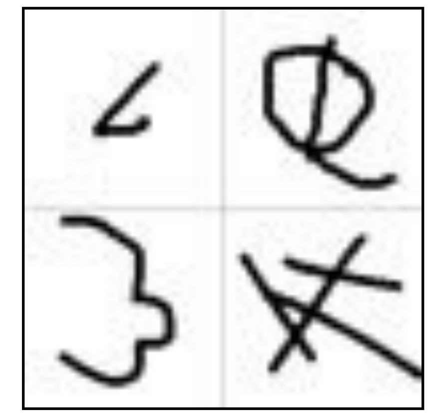
A



B



A

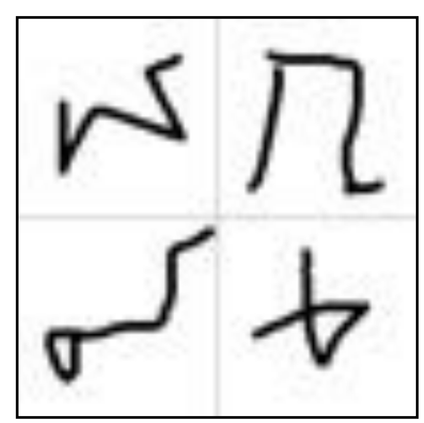


B

machine generated

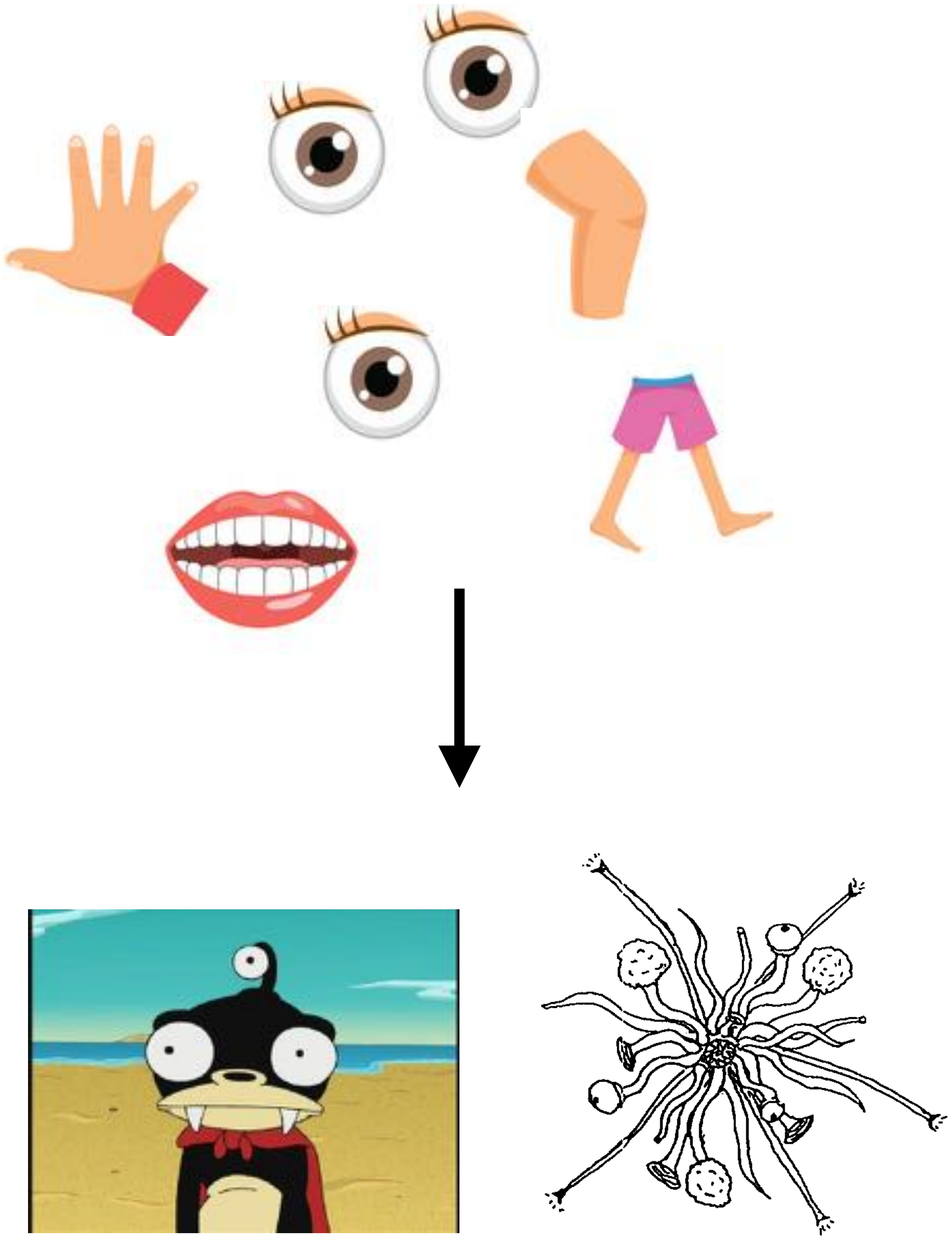
51% ID level (2/25 judges above chance)

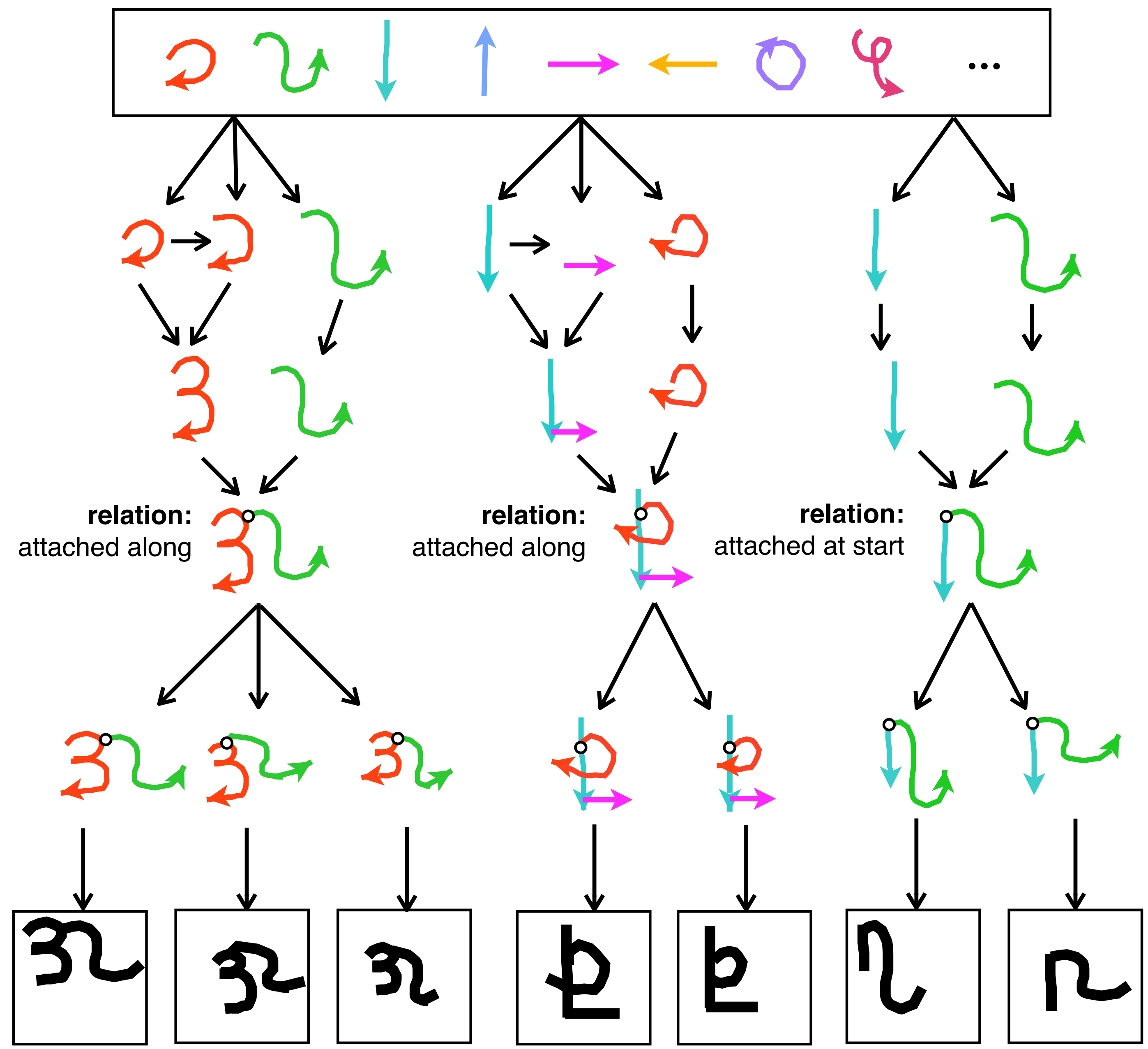
A



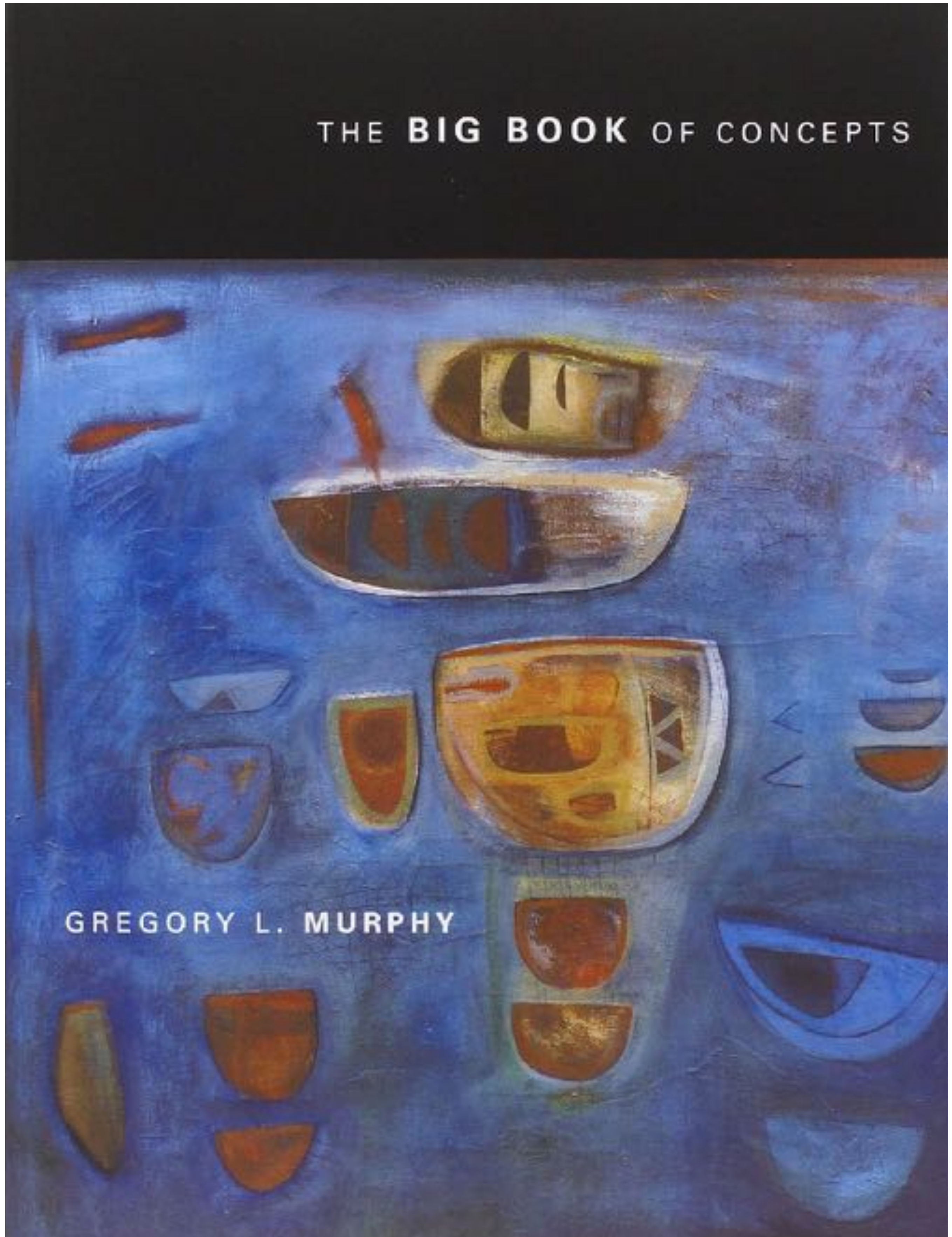
B

Compositional generation





Where have we been, and where to go next?



COGNITIVE PSYCHOLOGY 7, 573–605 (1975)

Family Resemblances: Studies in the Internal Structure of Categories

ELEANOR ROSCH AND CAROLYN B. MERVIS

University of California, Berkeley

Six experiments explored the hypothesis that the members of categories which are considered most prototypical are those with most attributes in common with other members of the category and least attributes in common with other categories. In probabilistic terms, the hypothesis is that prototypicality is a function of the total cue validity of the attributes of items. In Experiments 1 and 3, subjects listed attributes for members of semantic categories which had been previously rated for degree of prototypicality. High positive correlations were obtained between those ratings and the extent of distribution of an item's attributes among the other items of the category. In Experiments 2 and 4, subjects listed superordinates of category members and listed attributes of members of contrasting categories. Negative correlations were obtained between prototypicality and superordinates other than the category in question and between prototypicality and an item's possession of attributes possessed by members of contrasting categories. Experiments 5 and 6 used artificial categories and showed that family resemblance within categories and lack of overlap of elements with contrasting categories were correlated with ease of learning, reaction time in identifying an item after learning, and rating of prototypicality of an item. It is argued that family resemblance offers an alternative to criterial features in defining categories.

As speakers of our language and members of our culture, we know that a chair is a more reasonable exemplar of the category *furniture* than a radio, and that some chairs fit our idea or image of a chair better than others. However, when describing categories analytically, most tradi-

Psychological Review
1978, Vol. 85, No. 3, 207–238

Context Theory of Classification Learning

Douglas L. Medin
Rockefeller University

Marguerite M. Schaffer
Barnard College

Most theories dealing with ill-defined concepts assume that performance is based on category level information or a mixture of category level and specific item information. A context theory of classification is described in which judgments are assumed to derive exclusively from stored exemplar information. The main idea is that a probe item acts as a retrieval cue to access information associated with stimuli similar to the probe. The predictions of the context theory are contrasted with those of a class of theories (including prototype theory) that assume that the information entering into judgments can be derived from an additive combination of information from component cue dimensions. Across four experiments using both geometric forms and schematic faces as stimuli, the context theory consistently gave a better account of the data. The relation of the context theory to other theories and phenomena associated with ill-defined concepts is discussed in detail.

One of the major components of cognitive behavior concerns abstracting rules and forming concepts. Our entire system of naming objects and events, talking about them, and interacting with them presupposes the ability to group experiences into appropriate classes. Young children learn to tell the difference between dogs and cats, between clocks and fans, and between stars and street lights. Since few individual instances or exemplars are related to the superordinate concept. Although there is general agreement that natural categories are structured so that exemplars within a category are more similar to one another than to exemplars from alternative categories, there is disagreement concerning the rigidity of this structure. One extreme view is that all natural concepts are characterized by simple sets of

Psychological Review

VOLUME 92 NUMBER 3 JULY 1985

The Role of Theories in Conceptual Coherence

Gregory L. Murphy
Brown University

Douglas L. Medin
University of Illinois

The question of what makes a concept coherent (what makes its members form a comprehensible class) has received a variety of answers. In this article we review accounts based on similarity, feature correlations, and various theories of categorization. We find that each theory provides an inadequate account of conceptual coherence (or no account at all) because none provides enough constraints on possible concepts. We propose that concepts are coherent to the extent that they fit people's background knowledge or naive theories about the world. These theories help to relate the concepts in a domain and to structure the attributes that are internal to a concept. Evidence of the influence of theories on various conceptual tasks is presented, and the possible importance of theories in cognitive development is discussed.

Why is a given set of objects grouped together to form a category? That is, why is it that some groupings are informative, useful, and efficient, whereas others are vague, absurd, or useless? The current surge of interest in people's concepts has provided much in- ciated with the abominations of Leviticus, which produce the categories *clean animals* and *unclean animals.* Why should camels, ostriches, crocodiles, mice, sharks, and eels be declared unclean, whereas gazelles, frogs, most fish, grasshoppers, and some locusts be

Memory & Cognition
1983, 11 (3), 211-227

Ad hoc categories

LAWRENCE W. BARSALOU
Emory University, Atlanta, Georgia 30322

People construct ad hoc categories to achieve goals. For example, constructing the category of "things to sell at a garage sale" can be instrumental to achieving the goal of selling unwanted possessions. These categories differ from common categories (e.g., "fruit," "furniture") in that ad hoc categories violate the correlational structure of the environment and are not well established in memory. Regarding the latter property, the category concepts, concept-to-instance associations, and instance-to-concept associations structuring ad hoc categories are shown to be much less established in memory than those of common categories. Regardless of these differences, however, ad hoc categories possess graded structures (i.e., typicality gradients) as salient as those structuring common categories. This appears to be the result of a similarity comparison process that imposes graded structure on any category regardless of type.

The study of natural categories has been limited mostly to common categories such as "birds," "furniture," and "fruit." However, the use of highly specialized and unusual sets of items pervades everyday living. Some examples are "things to take on a camping trip," "possible costumes to wear to a Halloween party," and "places to look for antique desks." Since categories like these often appear to be created spontaneously for use in specialized contexts, I refer to them as ad hoc categories. Theories of natural categories primarily reflect (1981) and Smith and Medin (1981), the discovery of graded structure has had a major impact on theories of categorization. Graded structure has three aspects. First, some instances are better examples of a category than are others; "chair" is a more typical example of "furniture" than is "bookcase." This aspect of graded structure has been found in all common categories investigated so far. Rosch (1973, 1975b) found typicality in color categories (e.g., red, green). Rips, Shoben, and Smith (1973) and Rosch (1973, 1975a) found

Psychological Review
1992, Vol. 99, No. 1, 22–44

0033-295X/92/$3.00

ALCOVE: An Exemplar-Based Connectionist Model of Category Learning

John K. Kruschke
Indiana University Bloomington

ALCOVE (attention learning covering map) is a connectionist model of category learning that incorporates an exemplar-based representation (Medin & Schaffer, 1978; Nosofsky, 1986) with error-driven learning (Gluck & Bower, 1988; Rumelhart, Hinton, & Williams, 1986). ALCOVE selectively attends to relevant stimulus dimensions, is sensitive to correlated dimensions, can account for a form of base-rate neglect, does not suffer catastrophic forgetting, and can exhibit 3-stage (U-shaped) learning of high-frequency exceptions to rules, whereas such effects are not easily accounted for by models using other combinations of representation and learning method.

This article describes a connectionist model of category learning called ALCOVE (attention learning covering map). Any model of category learning must address the two issues of what representation underlies category knowledge and how that representation is used in learning. ALCOVE combines the exemplar-based representational assumptions of Nosofsky's (1986) generalized context model (GCM) with the error-driven learning assumptions of Gluck and Bower's (1988a, 1988b) network models. ALCOVE extends the GCM by adding a learning mechanism and extends the network models of Gluck and Bower by allowing continuous dimensions and including explicit dimensional attention learning. ALCOVE can be con- dient descent on error, it is unlike standard back propagation in its architecture, its behavior, and its goals. Unlike the standard back-propagation network, which was motivated by generalizing neuronlike perceptrons, the architecture of ALCOVE was motivated by a molar-level psychological theory, Nosofsky's (1986) GCM. The psychologically constrained architecture results in behavior that captures the detailed course of human category learning in many situations where standard back propagation fares less well. Unlike many applications of standard back propagation, the goal of ALCOVE is not to discover new (hidden-layer) representations after lengthy training but rather to model the course of learning itself by determining which

Psychological Review
1991, Vol. 98, No. 3, 409–429

0033-295X/91/$3.00

The Adaptive Nature of Human Categorization

John R. Anderson
Carnegie Mellon University

A rational model of human categorization behavior is presented that assumes that categorization reflects the derivation of optimal estimates of the probability of unseen features of objects. A Bayesian analysis is performed of what optimal estimations would be if categories formed a disjoint partitioning of the object space and if features were independently displayed within a category. This Bayesian analysis is placed within an incremental categorization algorithm. The resulting rational model accounts for effects of central tendency of categories, effects of specific instances, learning of linearly nonseparable categories, effects of category labels, extraction of basic level categories, base-rate effects, probability matching in categorization, and trial-by-trial learning functions. Although the rational model considers just 1 level of categorization, it is shown how predictions can be enhanced by considering higher and lower levels. Considering prediction at the lower, individual level allows integration of this rational analysis of categorization with the earlier rational analysis of memory (Anderson & Milson, 1989).

Anderson (1990) presented a rational analysis of human cognition. The term *rational* derives from similar "rational-man" analyses in economics. Rational analyses in other fields are sometimes called *adaptationist analyses.* Basically, they are efforts to explain the behavior in some domain on the assumption that the behavior is optimized with respect to some criteria of adaptive importance. This article begins with a general characterization of how one develops a rational theory of a particu- steps involved in a research program that attempts to understand cognition in terms of its adaptation to the environment:

1. The first task is to specify what the system is trying to optimize. Perhaps such models are ultimately to be justified in terms of maximizing some evolutionary criterion like number of surviving offspring. However, this is not a very workable criterion in most applications. Thus, economics uses wealth as the variable to be optimized; optimal foraging theory (Stephens

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

1 Introduction

Current approaches to object recognition make essential use of machine learning methods. To improve their performance, we can collect larger datasets, learn more powerful models, and use better techniques for preventing overfitting. Until recently, datasets of labeled images were relatively

Cognitive Science 32 (2008) 108–154

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210701802071

A Rational Analysis of Rule-Based Concept Learning

Noah D. Goodman[a], Joshua B. Tenenbaum[a], Jacob Feldman[b], Thomas L. Griffiths[c]

[a]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*
[b]*Department of Psychology, Rutgers University*
[c]*Department of Psychology, University of California, Berkeley*

Abstract

This article proposes a new model of human concept learning that provides a rational analysis of learning feature-based concepts. This model is built upon Bayesian inference for a grammatically structured hypothesis space—a concept language of logical rules. This article compares the model predictions to human generalization judgments in several well-known category learning experiments, and finds good agreement for both average and individual participant generalizations. This article further investigates judgments for a broad set of 7-feature concepts—a more natural setting in several ways—and again finds that the model explains human performance.

Psychological Review
2007, Vol. 114, No. 2, 245–272

0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.2.245

Word Learning as Bayesian Inference

Fei Xu
University of British Columbia

Joshua B. Tenenbaum
Massachusetts Institute of Technology

The authors present a Bayesian framework for understanding how adults and children learn the meanings of words. The theory explains how learners can generalize meaningfully from just one or a few positive examples of a novel word's referents, by making rational inductive inferences that integrate prior knowledge about plausible word meanings with the statistical structure of the observed examples. The theory addresses shortcomings of the two best known approaches to modeling word learning, based on deductive hypothesis elimination and associative learning. Three experiments with adults and children test the Bayesian account's predictions in the context of learning words for object categories at multiple levels of a taxonomic hierarchy. Results provide strong support for the Bayesian account over competing accounts, in terms of both quantitative model fits and the ability to explain important qualitative phenomena. Several extensions of the basic theory are discussed, illustrating the broader potential for Bayesian models of word learning.

Keywords: word learning, Bayesian inference, concepts, computational modeling

Learning even the simplest names for object categories presents a difficult induction problem (Quine, 1960). Consider a typical dilemma faced by a child learning English. Upon observing a competent adult speaker use the word *dog* in reference to Max, a particular Dalmatian running by, what can the child infer about the meaning of the word *dog*? The potential hypotheses appear endless. The word could refer to all (and only) dogs, all mammals, all animals, all Dalmatians, this individual Max, all dogs plus the

underdetermination, even 2- or 3-year-olds seem to be remarkably successful at learning the meanings of words from examples. In particular, children or adults can often infer the approximate extensions of words such as *dog* given only a few relevant examples of how the word can be used and no systematic evidence of how words are not to be used (Bloom, 2000; Carey, 1978; Markman, 1989; Regier, 1996). How do they do it?

Two broad classes of proposals for how word learning works

KNOWLEDGE SELECTION IN CATEGORY LEARNING

Evan Heit
Lewis Bott

I. Introduction

In our ordinary experience, we make countless observations every hour, with no observation perfectly resembling a previous case. We face a daily parade of unique events. Every time we walk into a building, for example, the building is unlike any other building in many ways. Even one particular building itself would be constantly undergoing various small changes. It has been suggested that to make better use of past experiences and simplify the processing that would be required for so many unique events, we learn about equivalence classes or categories of observations (e.g., Markman, 1989). For example, rather than treating every built architectural structure as being a unique construction, we form equivalence classes such as houses,

12

Essentialism as a Generative Theory of Classification

Bob Rehder

It is obvious that we classify the objects we encounter by their appearance, that is, by the particular features, aspects, or characteristics that they display. But, after a moment's reflection, it becomes clear that appearance is sometimes not all there is to it, that there are other factors not available to immediate inspection that might contribute to an object's identity. A study

apparently, there is something more to category membership than just how an object appears. In fact, there is evidence that even children as young as 3 years old believe that the "insides" of objects are relevant in determining its class membership (Gelman & Wellman, 1991; also see Gelman, 2003; Keil, 1989).

The idea that different aspects or characteristics of

COGNITIVE PSYCHOLOGY 8, 382–439 (1976)

Basic Objects in Natural Categories

ELEANOR ROSCH, CAROLYN B. MERVIS, WAYNE D. GRAY, DAVID M. JOHNSON, AND PENNY BOYES-BRAEM

University of California, Berkeley

Categorizations which humans make of the concrete world are not arbitrary but highly determined. In taxonomies of concrete objects, there is one level of abstraction at which the most basic category cuts are made. Basic categories are those which carry the most information, possess the highest category cue validity, and are, thus, the most differentiated from one another. The four experiments of Part I define basic objects by demonstrating that in taxonomies of common concrete nouns in English based on class inclusion, basic objects are the most inclusive categories whose members: (a) possess significant numbers of attributes in common, (b) have motor programs which are similar to one another, (c) have similar shapes, and (d) can be identified from averaged shapes of members of the class. The eight experiments of Part II explore implications of the structure of categories. Basic objects are shown to be the most inclusive categories for which a concrete image of the category as a whole can be formed, to be the first categorizations made during perception of the environment, to be the earliest categories sorted and earliest named by children, and to be the categories most codable, most coded, and most necessary in language.

The world consists of a virtually infinite number of discriminably different stimuli. One of the most basic functions of all organisms is the cutting up of the environment into classifications by which nonidentical

COGNITIVE PSYCHOLOGY 23, 457–482 (1991)

Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?

JAMES W. TANAKA AND MARJORIE TAYLOR

University of Oregon

Classic research on conceptual hierarchies has shown that the interaction between the human perceiver and objects in the environment specifies one level of abstraction for categorizing objects, called the basic level, which plays a primary role in cognition. The question of whether the special psychological status of the basic level can be modified by experience was addressed in three experiments comparing the performance of subjects in expert and novice domains. The main findings were that in the domain of expertise (a) subordinate-level categories were as differentiated as the basic-level categories, (b) subordinate-level names were used as frequently as basic-level names for identifying objects, and (c) subordinate-level categorizations were as fast as basic-level categorizations. Taken together, these results demonstrate that individual differences in domain-specific knowledge affect the extent that the basic level is central to categorization.

In a series of important experiments, Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) established that a basic level of abstraction has special significance in human categorization (also see Brown, 1958). The basic level was shown to be the most inclusive level at which a generalized shape of category exemplars is identifiable and imaginable. In addition, basic categories elicit similar motor programs and basic-level cate-

Psychological Review
1990, Vol. 97, No. 2, 185–200

0033-295X/90/$00.75

Category-Based Induction

Daniel N. Osherson
Massachusetts Institute of Technology

Ormond Wilkie
Massachusetts Institute of Technology

Edward E. Smith
University of Michigan

Alejandro López
University of Michigan

Eldar Shafir
Princeton University

An argument is categorical if its premises and conclusion are of the form *All members of C have property P,* where *C* is a natural category like FALCON or BIRD, and *P* remains the same across premises and conclusion. An example is *Grizzly bears love onions. Therefore, all bears love onions.* Such an argument is psychologically strong to the extent that belief in its premises engenders belief in its conclusion. A subclass of categorical arguments is examined, and the following hypothesis is advanced: The strength of a categorical argument increases with (a) the degree to which the premise categories are similar to the conclusion category and (b) the degree to which the premise categories are similar to members of the lowest level category that includes both the premise and the conclusion categories. A model based on this hypothesis accounts for 13 qualitative phenomena and the quantitative results of several experiments.

The Problem of Argument Strength

Fundamental to human thought is the confirmation relation, joining sentences $P_1 \ldots P_n$ to another sentence *C* just in case belief in the former leads to belief in the latter. Theories of confirmation may be cast in the terminology of argument strength, because $P_1 \ldots P_n$ confirm *C* only to the extent that $P_1 \ldots P_n/C$ is a strong argument. We here advance a partial theory of argument strength, hence of confirmation.

To begin, it will be useful to review the terminology of argument strength. By an *argument* is meant a finite list of sentences, the last of which is called the *conclusion* and the others its *premises.* Schematic arguments are written in the form $P_1 \ldots P_n/C$, whereas real arguments are written vertically, as in

belief in the conclusion of an argument (independently of its premises) is not sufficient for argument strength. For this reason, Argument 1 is stronger than Argument 2 for most people, even though the conclusion of Argument 2 is usually considered more probable than that of Argument 1. An extended discussion of the concept of argument strength is provided in Osherson, Smith, and Shafir (1986). It will be convenient to qualify an argument as strong, without reference to a particular person *S,* whenever the argument is strong for most people in a target population (e.g., American college students). We also say that $P_1 \ldots P_n$ confirm *C* if $P_1 \ldots P_n/C$ is strong.

An illuminating characterization of argument strength would represent a long step toward a theory of belief fixation

Psychological Review
2009, Vol. 116, No. 1, 20–58

0033-295X/09/$12.00 DOI: 10.1037/a0014282

Structured Statistical Models of Inductive Reasoning

Charles Kemp
Carnegie Mellon University

Joshua B. Tenenbaum
Massachusetts Institute of Technology

Everyday inductive inferences are often guided by rich background knowledge. Formal models of induction should aim to incorporate this knowledge and should explain how different kinds of knowledge lead to the distinctive patterns of reasoning found in different inductive contexts. This article presents a Bayesian framework that attempts to meet both goals and describe 4 applications of the framework: a taxonomic model, a spatial model, a threshold model, and a causal model. Each model makes probabilistic inferences about the extensions of novel properties, but the priors for the 4 models are defined over different kinds of structures that capture different relationships between the categories in a domain. The framework therefore shows how statistical inference can operate over structured background knowledge, and the authors argue that this interaction between structure and statistics is critical for explaining the power and flexibility of human reasoning.

Keywords: inductive reasoning, property induction, knowledge representation, Bayesian inference

Humans are adept at making inferences that take them beyond the limits of their direct experience. Even young children can learn the meaning of a novel word from a single labeled example (Heibeck & Markman, 1987), predict the trajectory of a moving object when it passes behind an occluder (Spelke, 1990), and choose a gait that allows them to walk over terrain they have never before encountered. Inferences like these may differ in many respects, but common to them all is the need to go beyond the information given (Bruner, 1973).

Two different ways of going beyond the available information can be distinguished. Deductive inferences draw out conclusions

This article describes a formal approach to inductive inference that should apply to many different problems, but we focus on the problem of property induction (Sloman & Lagnado, 2005). In particular, we consider cases where one or more categories in a domain are observed to have a novel property and the inductive task is to predict how the property is distributed over the remaining categories in the domain. For instance, given that bears have sesamoid bones, which species is more likely to share this property: moose or salmon (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975)? Moose may seem like the better choice because they are more similar biologically to bears, but different properties

Development of Subordinate-Level Categorization in 3- to 7-Month-Old Infants

Paul C. Quinn

Visual preference procedures were used to investigate development of perceptually based subordinate-level categorization in 3- to 7-month-old infants. Experiments 1 and 2 demonstrated that 3- to 4-month-olds did not form category representations for photographic exemplars of subordinate-level classes of cats and dogs (i.e., Siamese vs. Tabby, Beagle vs. Saint Bernard). Experiments 3 though 5 showed that 6- and 7-month-olds formed a category representation for Tabby that excluded Siamese and a category representation for Saint Bernard that excluded Beagle, but they did not form a category representation for Siamese that excluded Tabby or a category representation for Beagle that excluded Saint Bernard. The findings are consistent with a differentiation-driven view of early perceptual category development from global to basic to subordinate levels.

Categorization refers to equivalent responding to discernibly different instances from a common class (Bruner, Goodnow, & Austin, 1956). It is considered to be an adaptive mental process that allows for organized storage of information in memory, efficient retrieval of that information, and the capability of responding with familiarity to an indefinitely large number of instances from a variety of classes, most of which have not been previously encountered (Murphy, 2002). Without categorization, each experienced entity would be unrelated to all represented entities, and no represented entity would be related to any other (Smith & Medin, 1981).

Categorization must begin at some point during development, and recent evidence indicates that

Empirical studies examining the development of category representations during the first year of life have investigated the age and means by which individuated representations can be formed for narrowly tuned basic-level and more broadly inclusive global-level classes (e.g., cat vs. dog, mammal vs. furniture; Mandler & McDonough, 1993; Quinn, Eimas, & Rosenkrantz, 1993). Of particular concern has been whether basic-level representations cohere to form global (superordinate-level) representations in accord with a constructionist perspective or whether basic-level representations evolve from original global representations in accord with a differentiation perspective. Much of this work has been in response to the theory of Rosch and Mervis, which

Concept Formation in Infancy

Jean M. Mandler
University of California, San Diego
MRC Cognitive Development Unit, London

Laraine McDonough
University of California, San Diego

Four experiments investigated conceptual categorization in 7- to 11-month-old infants. Experiments 1 and 2 showed that 9- and 11-month-olds differentiated the global domains of animals and vehicles. Within the animal domain no subcategorization was found: the infants did not differentiate dogs from fish or from rabbits. Within the vehicle domain infants differentiated cars from both airplanes and motorcycles. Experiment 3 showed similar, although weaker, categorization for 7-month-olds. Experiment 4 showed that categorization of animals and vehicles was unaffected by degree of between-category similarity. Birds and airplanes were treated as different even though the exemplars from both categories had similar shapes, including outstretched wings, and were of the same texture. These data, showing global differentiation of animals and vehicles, with lack of differentiation of "basic-level" categories within the animal domain, contrast with data from studies designed to assess perceptual categorization. Even younger infants differentiate various animal subcategories perceptually. However, the results presented here suggest that infants may not respond to such perceptual differences as being conceptually relevant.

We are concerned in this article with the process of forming concepts, or conceptual categories, in infancy. Little is known about this topic, because most of the

Chapter 2
Acquisition of Category Terms

To begin thinking about what children must accomplish in learning how to categorize objects in their world let us consider what is involved in learning labels for object categories.

Quine's Problem of Induction

A simple, straightforward way in which young children learn new category labels is through ostensive definition. That is, an adult or other teacher points to an object and labels it. It is important to consider how much can be learned by way of ostensive definition, because very young children who are learning their first language may have little else to rely on. We cannot explain or describe to the young child what "rabbit" means, because the child does not yet understand the language needed for the explanation. At first, then, adults are limited to pointing to an object (such as a rabbit) and labeling it ("rabbit"). What is required for the child to conclude, in this situation, that the word "rabbit" is the label for a rabbit? First, the child must be able to interpret the pointing gesture correctly. The child's gaze must follow the direction of the pointing finger (see Churcher and Scaife 1982, Murphy and Messer 1977, Scaife and Bruner 1975 for discussion of how children make use of eye gaze and pointing). Then the child must assume that the word is somehow related to what he or she views. How is it that the child settles on an interpretation? At first glance this would seem to be quite a simple problem, and in fact children correctly make hundreds of such inferences when acquiring new vocabulary. But this apparent simplicity belies an incredibly difficult inferential problem. To illustrate the problem, I will summarize the example used by Quine (1960) to make his well-known argument about translation.

Quine asks us to imagine that a linguist visits an unknown country and attempts to learn the native language. A rabbit passes by and a native of the country says, "Gavagai." How is the linguist to figure out what "Gavagai" means? Like us, the linguist hypothesizes that

Advanced Review

Categorical perception

Robert L. Goldstone* and Andrew T. Hendrickson

Categorical perception (CP) is the phenomenon by which the categories possessed by an observer influences the observers' perception. Experimentally, CP is revealed when an observer's ability to make perceptual discriminations between things is better when those things belong to different categories rather than the same category, controlling for the physical difference between the things. We consider several core questions related to CP: Is it caused by innate and/or learned categories, how early in the information processing stream do categories influence perception, and what is the relation between ongoing linguistic processing and CP? CP for both speech and visual entities are surveyed, as are computational and mathematical models of CP. CP is an important phenomenon in cognitive science because it represents an essential adaptation of perception to support categorizations that an organism needs to make. Sensory signals that could be linearly related to physical qualities are warped in a nonlinear manner, transforming analog inputs into quasi-digital, quasi-symbolic encodings. *WIREs Cogn Sci*

When we look at a rainbow, we tend to see about seven distinct bands of color, even though we know from physics that the dominant wavelength of light that meets one's eye changes smoothly from the top to bottom of the rainbow. Although the rainbow presents itself to us with a continuous and full range of visible wavelengths of light, we tend to see it in terms of distinct colors such as red, yellow, blue, and violet. This effect is a striking example of categorical perception (CP). According to this phenomenon, we tend to perceive our world in terms of the categories that we have formed. Our perceptions are warped such that differences between objects that belong in different categories are accentuated, and differences between objects that fall into the same category

these systems. We humans do not simply base our categories on the outputs of perceptual systems independent of feedback. Instead, our perceptual systems become customized to the task-useful categories that we acquire, slowly at the evolutionary timescale or quickly at the timescale of individual learning.

Another reason why CP is theoretically important is that offers a potential account for how the apparently symbolic activity of high-level cognition can be grounded in perception and action.[1] A basic feature of human symbolic thought is that people form equivalence classes. In the classical notion of an equivalence class, distinguishable stimuli come to be treated as the same thing once they have been placed

Journal of Experimental Psychology: General
1994, Vol. 123, No. 2, 178–200

Influences of Categorization on Perceptual Discrimination

Robert Goldstone

Four experiments investigated the influence of categorization training on perceptual discrimination. Ss were trained according to 1 of 4 different categorization regimes. Subsequent to category learning, Ss performed a Same–Different judgment task. Ss' sensitivities (d's) for discriminating between items that varied on category-(ir)relevant dimensions were measured. Evidence for acquired distinctiveness (increased perceptual sensitivity for items that are categorized differently) was obtained. One case of acquired equivalence (decreased perceptual sensitivity for items that are categorized together) was found for separable, but not integral, dimensions. Acquired equivalence within a categorization-relevant dimension was never found for either integral or separable dimensions. The relevance of the results for theories of perceptual learning, dimensional attention, categorical perception, and categorization are discussed.

Psychologists have long been intrigued by the possibility that the concepts that people learn influence their perceptual abilities. It may be that the way people organize their world into categories alters the actual appearance of their world. The purpose of the present research is to investigate influences of concept learning on perception.

The notion that experience and expectations can influence perception can be traced back to the "New Look" movement of the 1940s and 50s (J. A. Bruner & Postman, 1949). Evidence suggests that experts perceive structures in X rays (Norman, Brooks, Coblentz, & Babcook, 1992), beers (Peron & Allen, 1988), and infant chickens (Biederman & Shiffrar, 1987) that are missed by novices. As the experts in

learning has come a long way since J. S. Bruner et al.'s study (Estes, 1986; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Reed, 1972), vestiges of this earlier work are apparent in current research. Specifically, many researchers have investigated concept learning using stimuli that have clear-cut dimensions with clearly different values on these dimensions. Although such stimuli are mandatory in many cases for experimental control and precision, they do not require subjects to perceptually learn new dimensions or finer discriminations. In the present described concept learning tasks, subjects had to make fine discriminations along dimensions or isolate dimensions that normally are fused together. In both cases,

Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 3, 681–696

Categorization Creates Functional Features

Philippe G. Schyns
University of Glasgow

Luc Rodet
University of Grenoble

Many theories of object recognition and categorization claim that complex objects are represented in terms of characteristic features. The origin of these features has been neglected in theories of object categorization. Do they form a fixed and independent set that exists before experience with objects, or are features progressively extracted and developed as an organism categorizes its world? This article maintains that features can be learned flexibly as a consequence of categorizing and representing objects. All 3 experiments reported in this article used categories of unfamiliar, computer-synthesized 2-dimensional objects ("Martian cells"). The results showed that varying the order of category learning induced the creation of different features that changed the perceptual appearance and the featural representation of identical category exemplars. Network simulations supported a flexible rather than a fixed-feature interpretation of the data.

Many theories of object recognition and categorization assume that objects are represented in memory as groups of components. To classify an object, one must first identify its components and then compare them to memory representations. For example, when a person sees a cup, he or she might first identify a container or a handle before categorizing the object properly. Of course, not all components of an object are necessary for its categorization, but many of them are probably identified during the recognition process. Componential accounts that embody this general approach include theories of object recognition and categorization (see, among others, Biederman, 1987; Marr & Nishihara, 1978; Rosch & Mervis, 1975; E. E. Smith & Medin, 1981; Treisman & Gelade, 1980).

the ever-changing retinal input. Even though our sophisticated visual apparatus probably comes equipped with a priori ways of analyzing and organizing retinal images, there are occasions when a relevant perceptual analysis is not readily available. For example, complete novices reading chest X-rays (e.g., Christensen et al., 1981), sexing chickens (Biederman & Shiffrar, 1987), and categorizing dermatoses (Norman, Brooks, Coblentz, & Babcock, 1992) have little understanding of the relevant dimensional structure of these categories. Even when told what the signs of different diagnosis are, novices are not always able to see the features experts use to organize the input. If one takes a developmental perspective, it seems clear that infants and young children are not always able to analyze objects by using all the

COGNITIVE SCIENCE 12, 529–562 (1988)

Comprehending Complex Concepts

GREGORY L. MURPHY
Brown University

Recent theories of concepts have raised the issue of how people combine simple concepts (like *engine* and *repair*) to form complex concepts (like *engine repair*). This article approaches this issue by asking how people comprehend modified noun phrases of this sort. One explanation of how complex concepts are understood (the feature weighting model) provides a simple mechanism in which the primary feature of the modifying concept is made more salient in the modified concept. Another explanation focuses on how world knowledge directs the combination process. The two explanations are compared in their ability to account for the interpretation of various kinds of noun phrases. Two experiments are reported which evaluate the feature weighting model's predictions for adjective-noun phrases. These contrasts suggest that the combination process does require reference to world knowledge. The consequences of accepting such an account are discussed.

The issue of complex concepts has become an important one in the psychology of concepts. Although theories of concepts may make similar predictions about the structure of simple concepts (e.g., Hintzman & Ludlam, 1980), they may not be equally facile at explaining how concepts are combined to form new, more complex concepts. Thus, an account of complex concepts may be crucial for evaluating the many theories of concepts now extant (see Smith & Medin, 1981).

The creation of complex concepts is a fascinating example of a high level cognitive process that people can perform very quickly. People are likely to create novel noun-noun phrases in their conversations, and listeners are adept at understanding them (E. Clark, Gelman, & Lane, 1985; H. Clark, 1983; Downing, 1977). In fact, Clark, et al. (1985) found that children as young as 3 years old could understand noun compounds as well as create novel compounds like *lion door* and *apple car* to describe objects that they had never seen before (see also Nelson, 1976, p. 23). The ability to understand a novel phrase like *United States Senate Michigan bean soup* (an example by

This research was supported by NSF grant BBS 83-15145. The author gratefully acknowledges the comments of Hiram Brownell, Ed Smith, and anonymous reviewers on earlier versions of the article.

Correspondence should be sent to Gregory L. Murphy, Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912.

COGNITIVE PSYCHOLOGY 27, 1–40 (1994)

Structured Imagination: The Role of Category Structure in Exemplar Generation

THOMAS B. WARD
Texas A&M University

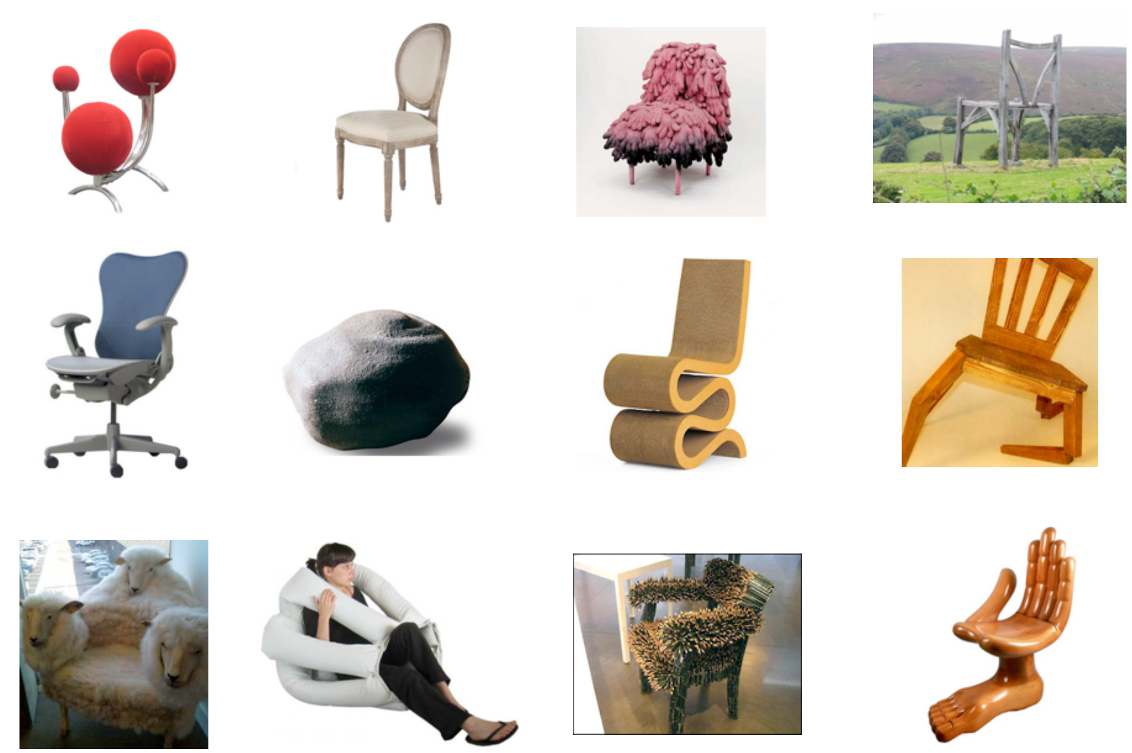
College students imagined animals that might live on a planet somewhere else in the galaxy. In the first experiment, they provided drawings and descriptions of their initial imagined animal, another member of the same species, and a member of a different species. The majority of imagined creatures were structured by properties that are typical of animals on earth: bilateral symmetry, sensory receptors, and appendages. Subjects also allowed shape, appendages and sense receptors to vary often across species but rarely within species. In Experiment 2, subjects' creations were influenced by correlated attributes; those told that the animal was feathered were more likely to produce creatures with wings and beaks, and those told it lived in water and had scales were more likely to produce creatures with fins and gills relative to subjects who were told the animal was furry or who were given no specific features. Experiments 3 and 4 revealed that many subjects approach the task by retrieving exemplars of known earth animals, but that instructions and task constraints can lead to greater use of broader knowledge frameworks. Experiment 5 revealed that the structuring found in college students' imagined animals also holds for extraterrestrials developed by science fiction writers. The results are consistent with the idea that similar structures and processes underlie creative and noncreative aspects of cognition, and are discussed in terms of the concept of *structured imagination*. That is, when subjects create a new member of a known category for an imaginary setting, their imagination is structured by a particular set of properties that are characteristic of that category.

Category structures and processes play a central role in many human activities. Consequently, they have been studied heavily over the past three decades, with particular attention being given to the topics of category learning, classification, and inductive inference. Although these are

Classical view

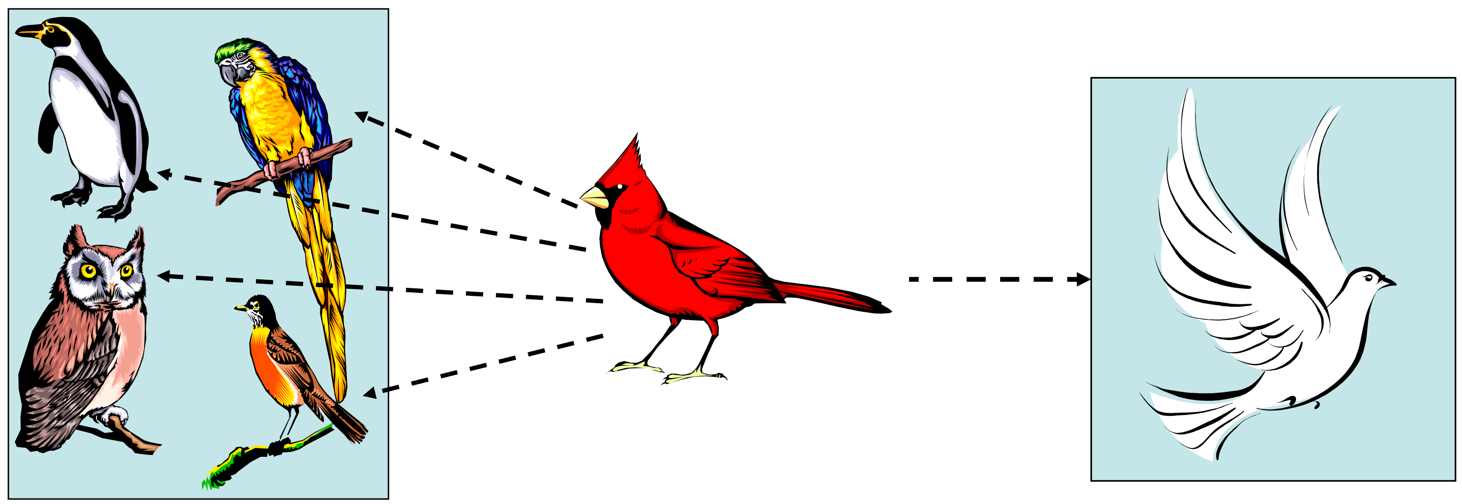
- *bachelor*: an unmarried man
- *prime number*: a number not divisible by any number besides “1” and itself
- *triangle*: a polygon with three sides

What is a chair?



Exemplar vs. prototype views

exemplar theory | **prototype theory**



Birds
You've Seen

Bird?

Prototypical
Bird

family resemblance view

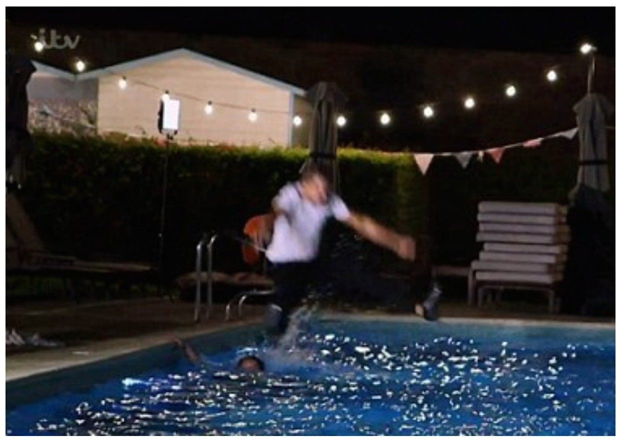
famous "5-4" category structure

TRAINING STIMULI

"A" STIMULI							"B" STIMULI						
STIMULUS NUMBER	DIMENSION VALUES C	F	S	N	FE	RAT-ING	STIMULUS NUMBER	DIMENSION VALUES C	F	S	N	FE	RAT-ING
4	1	1	1	0	4.9	4.8	12	1	1	0	0	5.5	5.0
7	1	0	1	0	3.3	5.4	2	0	1	1	0	5.2	5.1
15	1	0	1	1	3.2	5.1	14	0	0	0	1	3.9	5.2
13	1	1	0	1	4.8	5.2	10	0	0	0	0	3.1	5.5
5	0	1	1	1	4.5	5.2							
Prototype:	**1**	**1**	**1**	**1**			**Prototype:**	**0**	**0**	**0**	**0**		

Concepts as theories and the knowledge view

Would you classify this man as “drunk”? (Murphy & Medin)



Flammable?

Flammable?

Ad hoc categories (Barsalou, 1983)

- e.g., “Things to carry out of a burning house”
  - [children, dog, photo albums, computer, etc.]

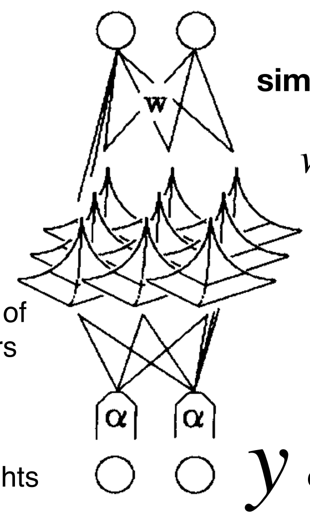
Computational models of category learning

ALCOVE

Response rule

$$P(y \in A) = \frac{e^{\phi \textbf{sim}(y,A)}}{e^{\phi \textbf{sim}(y,A)} + e^{\phi \textbf{sim}(y,B)}}$$

category A   category B



$$\textbf{sim}(y, C) = \sum_{x \in X} w_{cx} \, \textbf{sim}(y, x)$$

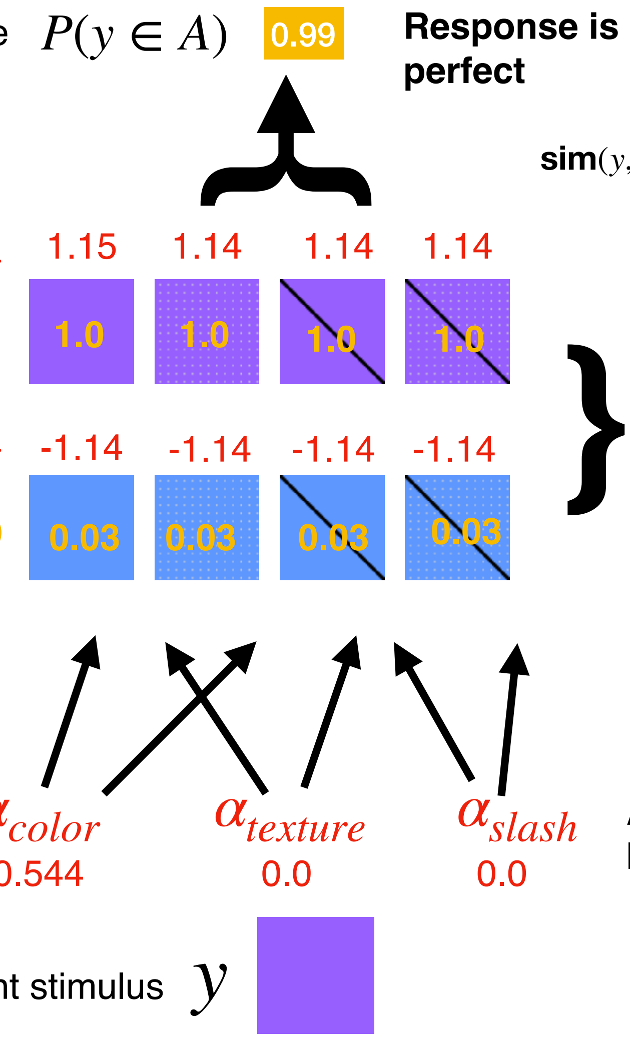
w_{cx} association weights between exemplar *x* and category *c*

$$\textbf{sim}(y, x) = e^{\sum_{D_i} \alpha_i |x_i - y_i|}$$

x denotes one of the exemplars

α_i attention weights y current stimulus

response $P(y \in A)$



Response is now nearly perfect

$$\textbf{sim}(y, C) = \sum_{x \in X} w_{cx} \, \textbf{sim}(y, x)$$

association weights w_{Ax}

$\textbf{sim}(y, x)$

$$\textbf{sim}(y, x) = e^{\sum_{D_i} \alpha_i |x_i - y_i|}$$

w_{Ax}

$\textbf{sim}(y, x)$

Exemplars x

attention weights α_{color} $\alpha_{texture}$ α_{slash}

Attention only looks at color

current stimulus y

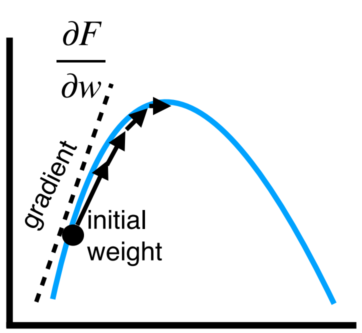
Maximizing likelihood via gradient descent

(the workhorse of modern machine learning, and often cognitive modeling too)

objective function to maximize

$$F(w, \alpha) = \sum_{x \in A} \log P_{w,\alpha}(x \in A) + \sum_{x \in B} \log(P_{w,\alpha}(x \in B))$$

likelihood F

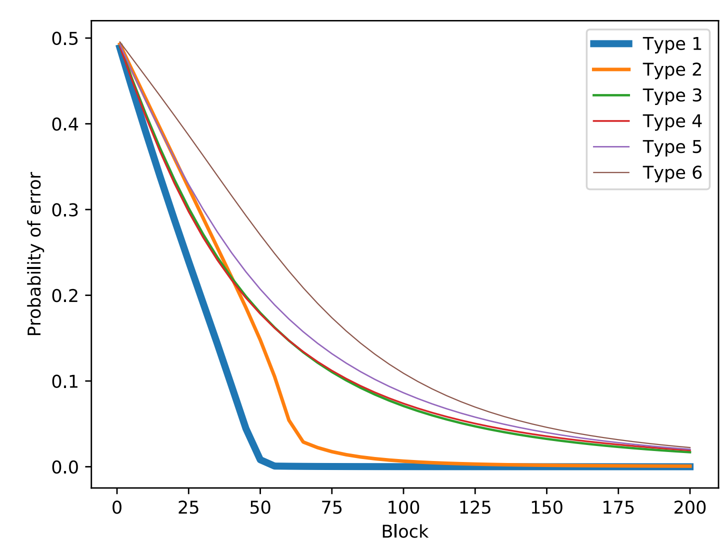


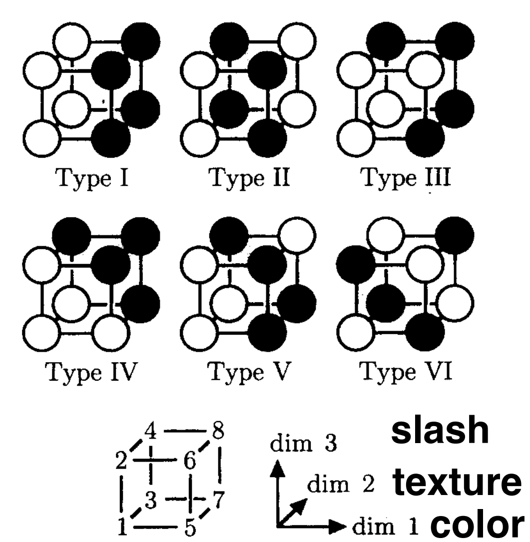
Computing the gradient tells us which direction to go for steepest ascent:

$$w \leftarrow w + \gamma \frac{\partial F}{\partial w}$$

γ : learning rate (small constant)

ALCOVE with attention weights

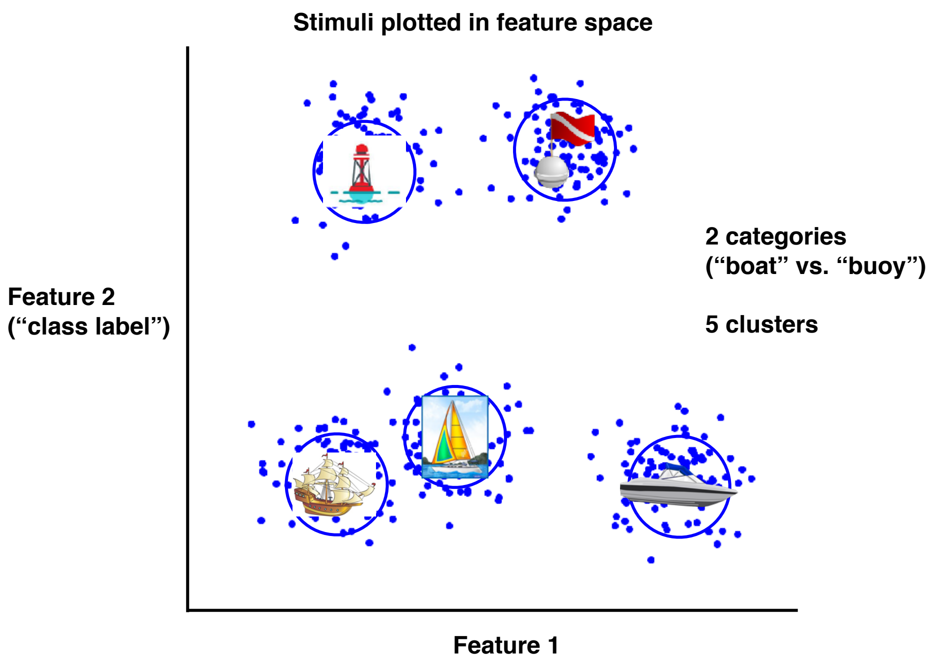




slash
texture
color

Computational models of category learning

Rational Model



Base rate assignment to existing cluster k

$$P(k) = \frac{n_k}{\alpha + n}$$

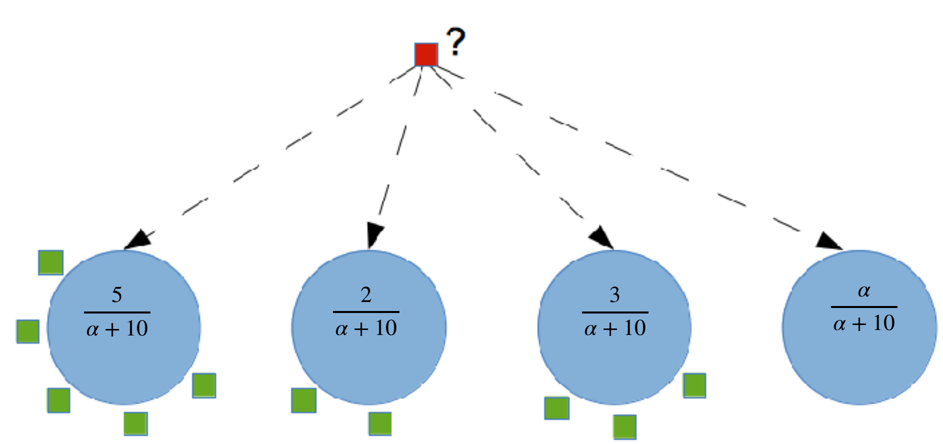
Base rate of forming a new cluster

$$P(\textbf{new}) = \frac{\alpha}{\alpha + n}$$

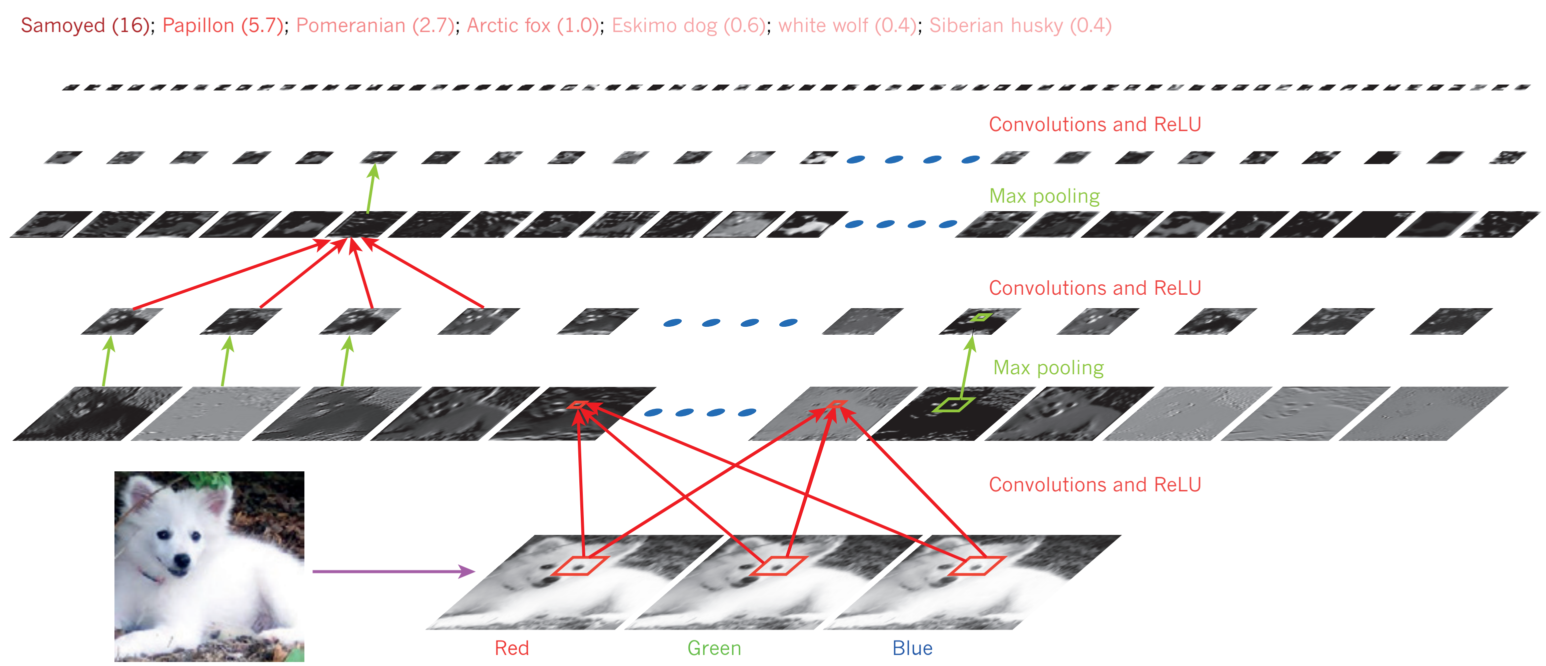
n_k number of stimuli in k

n number of stimuli total

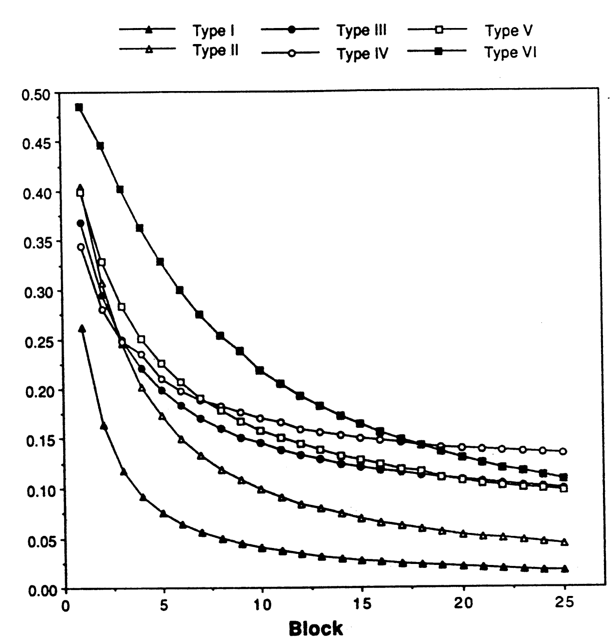
α parameter that controls likelihood of new cluster



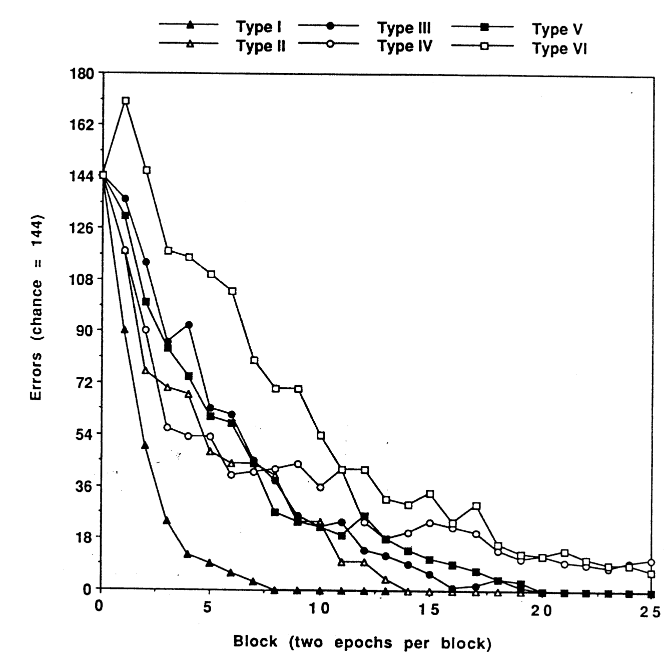
Convolutional neural networks



People

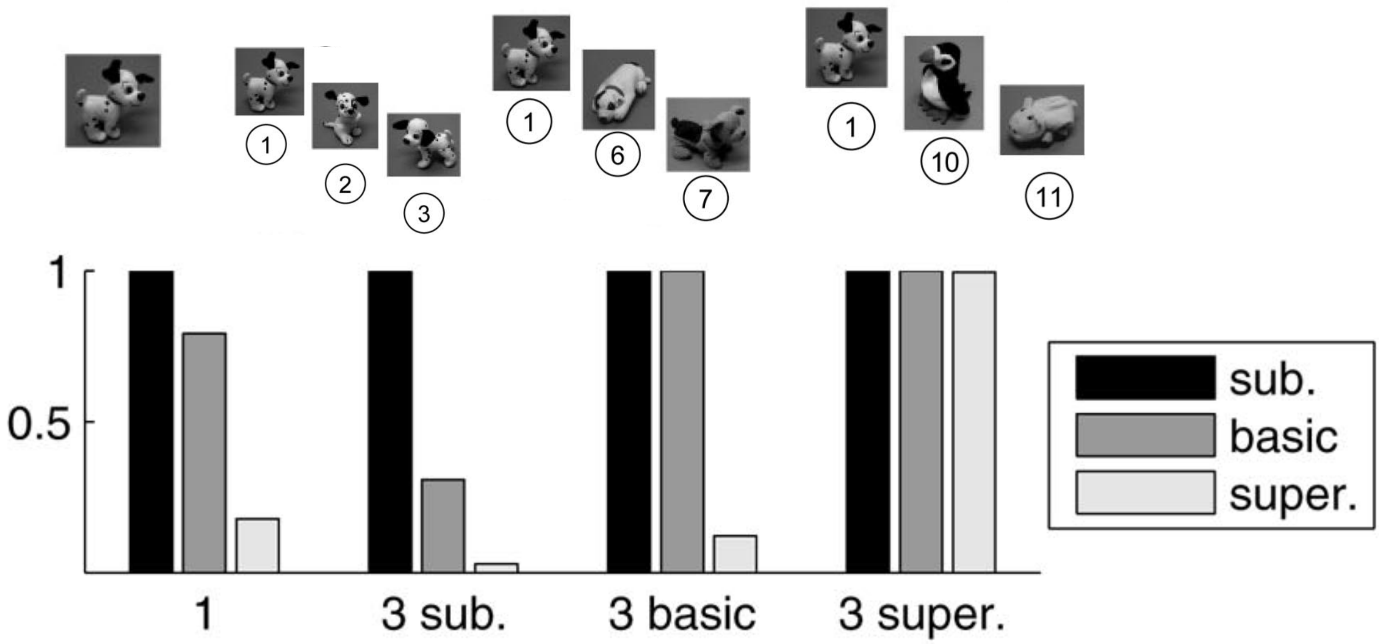


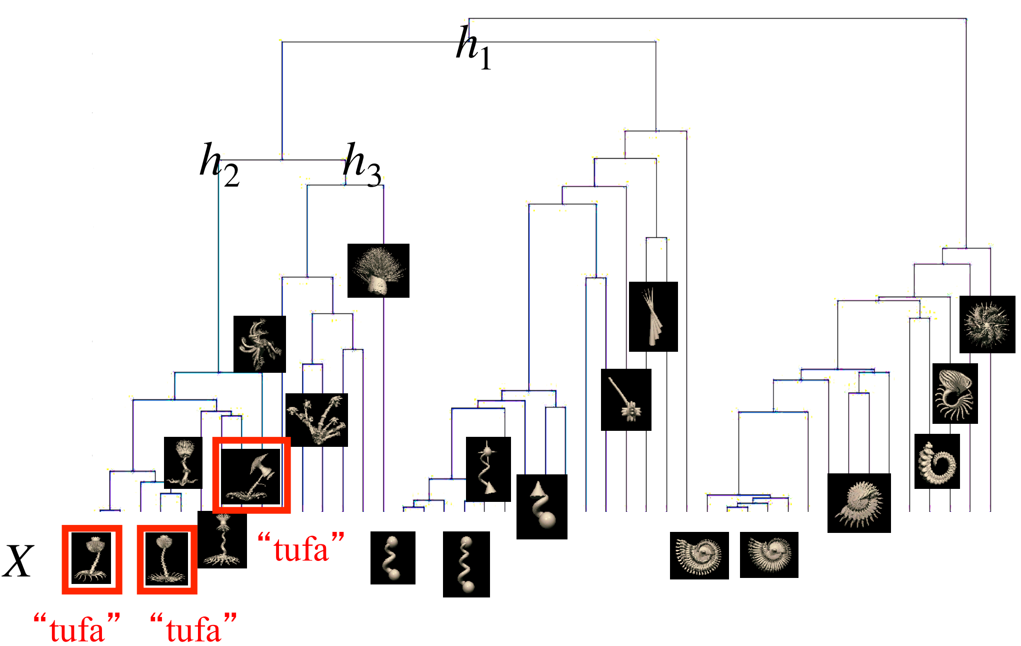
Model



Computational models of category learning

Bayesian word learning





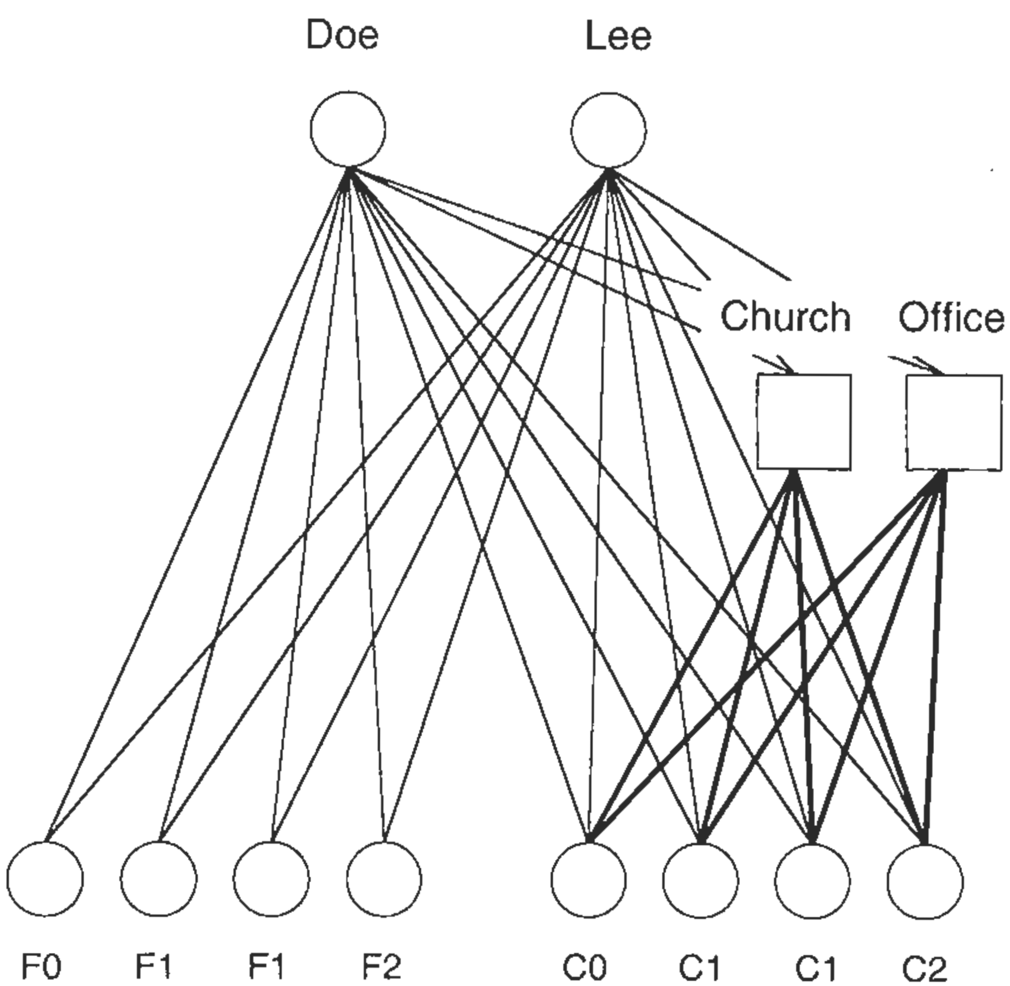
$$
\begin{aligned}
& S \\
S1: \quad & \forall x\ \ell(x) \Leftrightarrow (D) \\
D1: \quad & \forall x\ \ell(x) \Leftrightarrow ((C) \vee D) \\
D2: \quad & \forall x\ \ell(x) \Leftrightarrow ((C) \vee \text{False}) \\
C1: \quad & \forall x\ \ell(x) \Leftrightarrow ((P \wedge C) \vee \text{False}) \\
C2: \quad & \forall x\ \ell(x) \Leftrightarrow ((P \wedge \text{True}) \vee \text{False}) \\
P1: \quad & \forall x\ \ell(x) \Leftrightarrow ((F_1 \wedge \text{True}) \vee \text{False}) \\
F_1 2: \quad & \forall x\ \ell(x) \Leftrightarrow (((f_1(x)=0) \wedge \text{True}) \vee \text{False})
\end{aligned}
$$

Rational rules

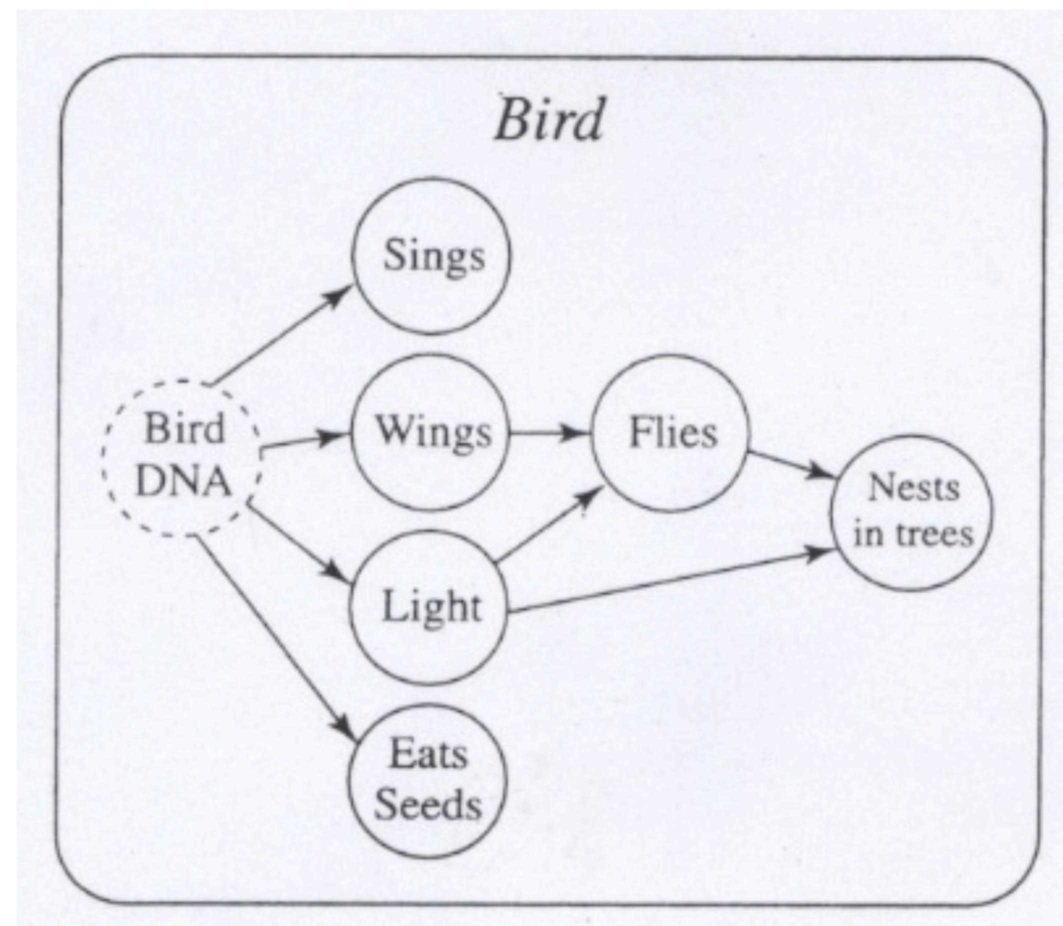
$$
\begin{array}{lrcl}
(\mathrm{S1}) & S & \rightarrow & \forall x\ \ell(x) \Leftrightarrow (D) \\
(\mathrm{D1}) & D & \rightarrow & (C) \vee D \\
(\mathrm{D2}) & D & \rightarrow & False \\
(\mathrm{C1}) & C & \rightarrow & P \wedge C \\
(\mathrm{C2}) & C & \rightarrow & True \\
(\mathrm{P1}) & P & \rightarrow & F_1 \\
 & & \vdots & \\
(\mathrm{PN}) & P & \rightarrow & F_N \\
(\mathrm{F_1 1}) & F_1 & \rightarrow & f_1(x) = 1 \\
(\mathrm{F_1 2}) & F_1 & \rightarrow & f_1(x) = 0 \\
 & & \vdots & \\
(\mathrm{F_N 1}) & F_N & \rightarrow & f_N(x) = 1 \\
(\mathrm{F_N 2}) & F_N & \rightarrow & f_N(x) = 0
\end{array}
$$

Computational models of category learning

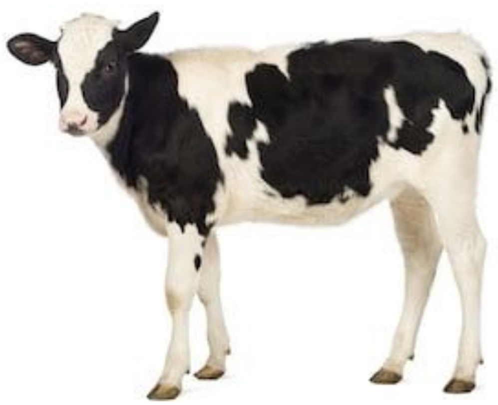
Background knowledge in category learning



Causal-model theory

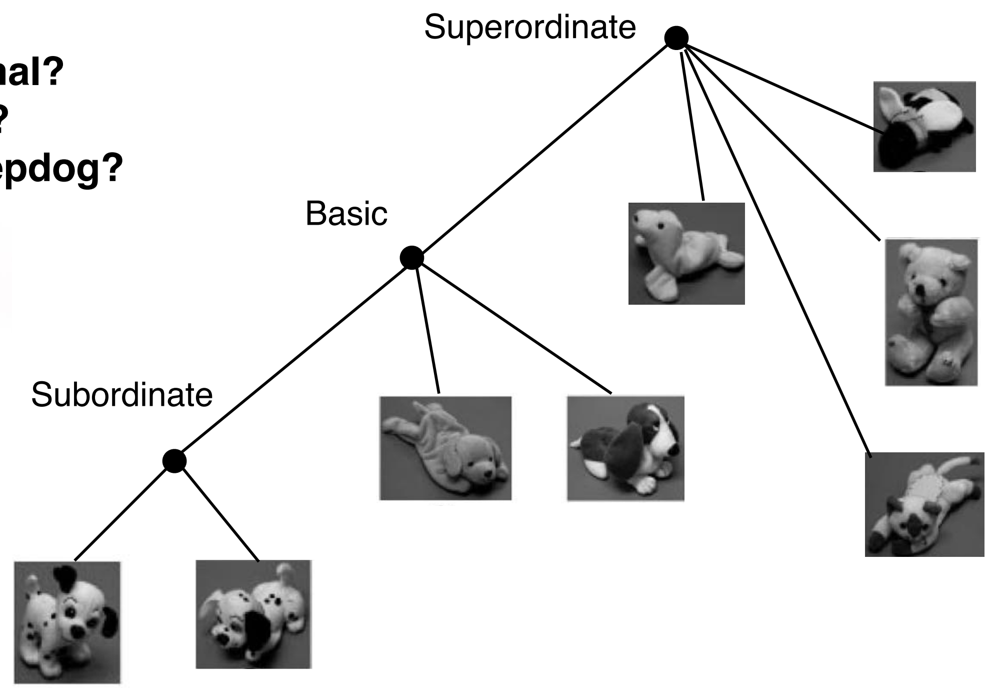


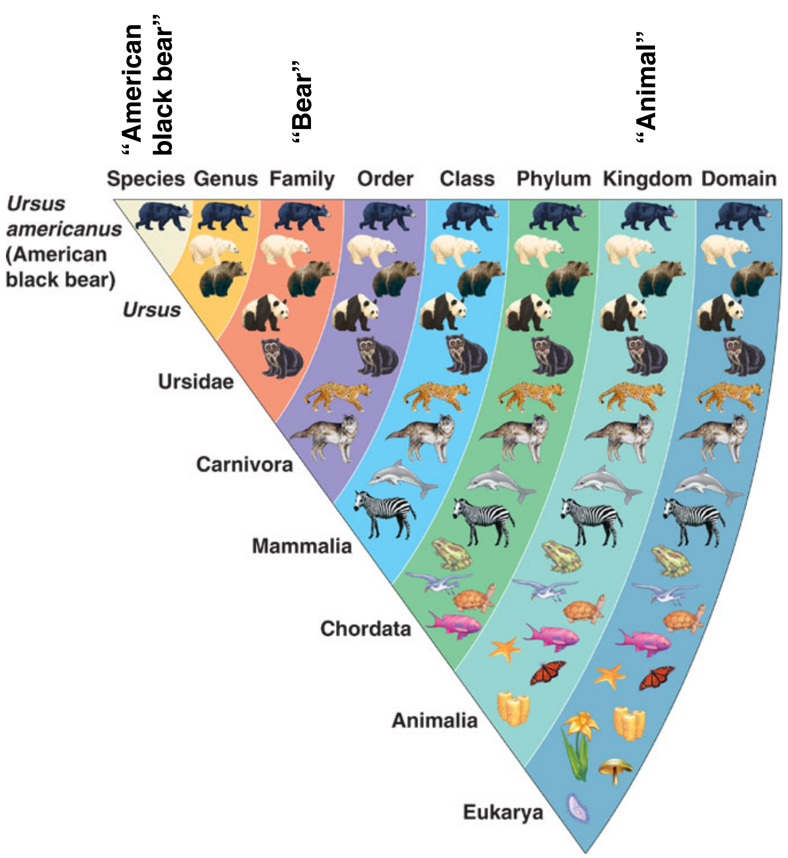
Taxonomic organization and the basic level

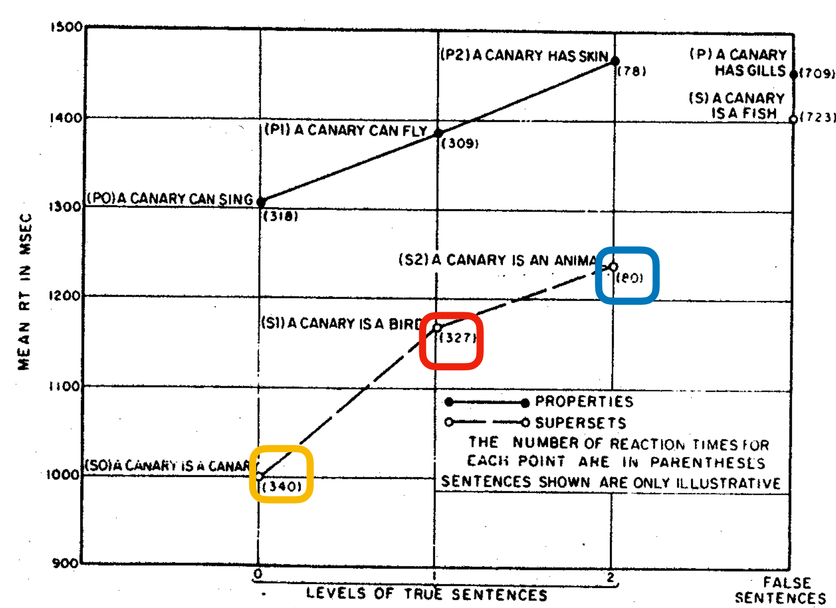


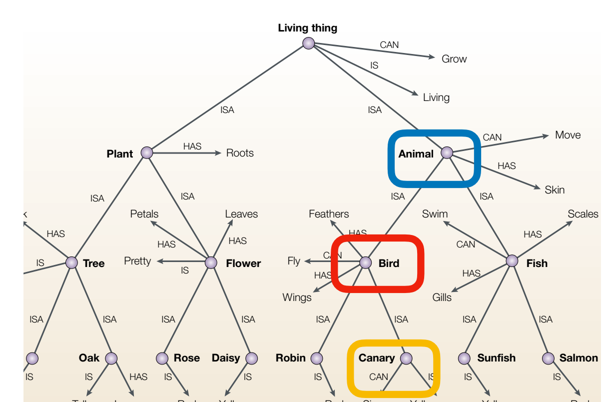
Object?
Living thing?
Animal?
Mammal?
Ungulate?
Bovine?
Cow?
Holstein Friesian cow?
My cow, “Betsy"?

Animal?
Dog?
Sheepdog?





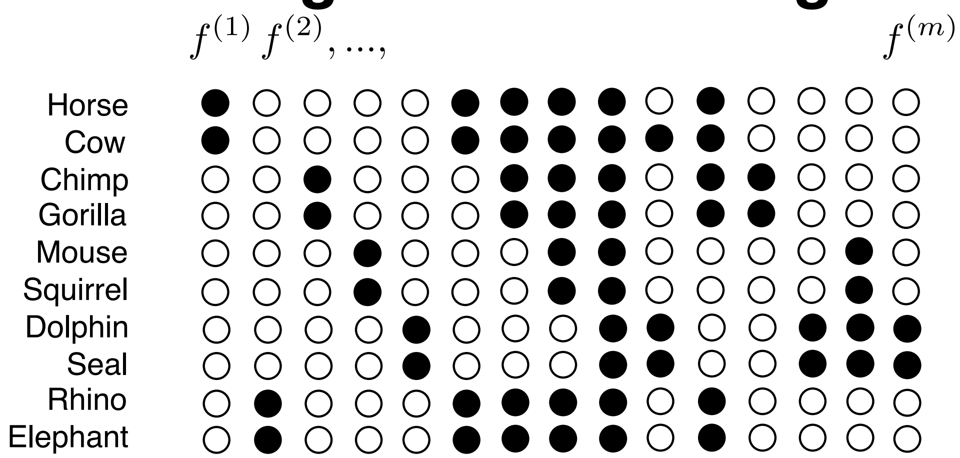




Category-based induction

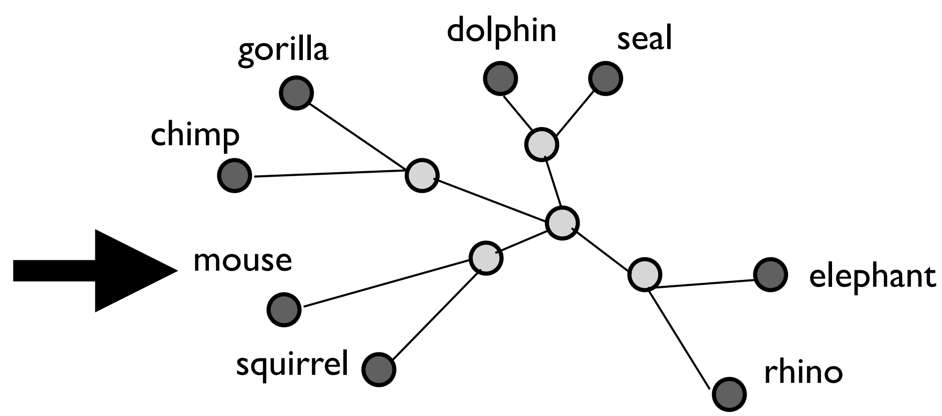
Question: “Given that cows and seals have T9 hormones, how likely is it that horses do?”

background knowledge

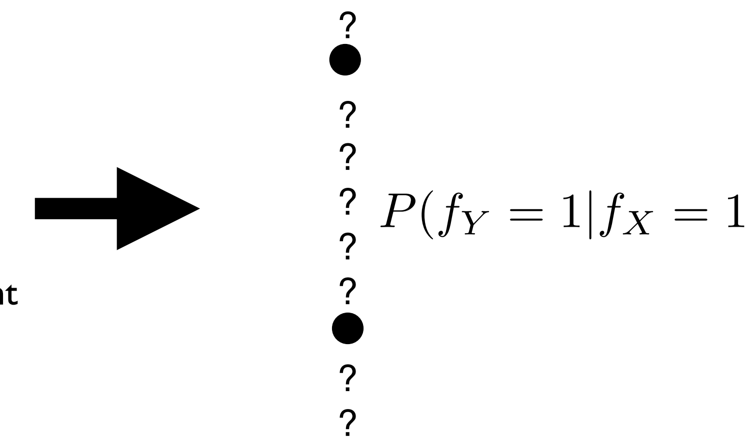


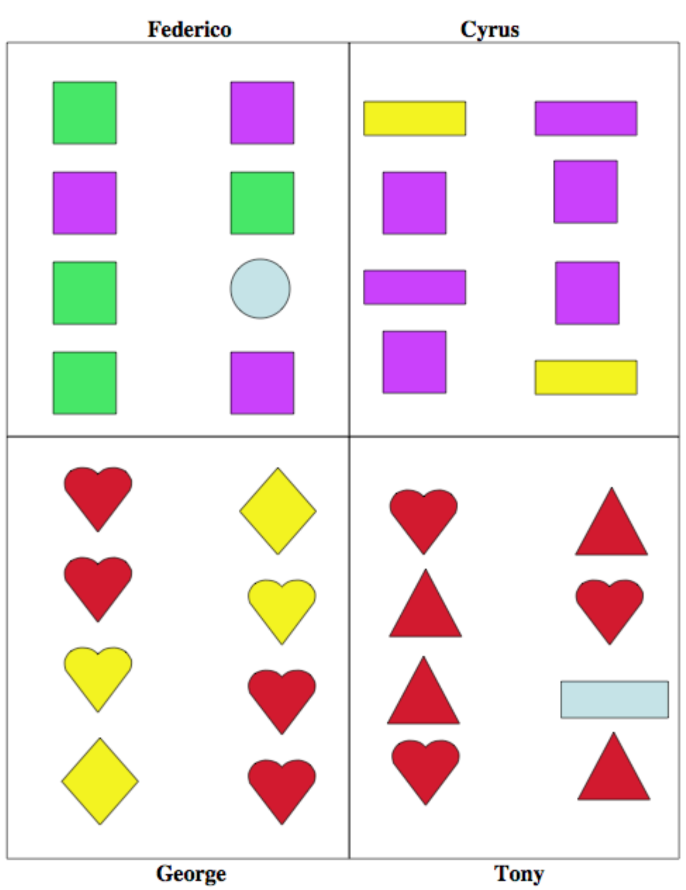
Features

structure

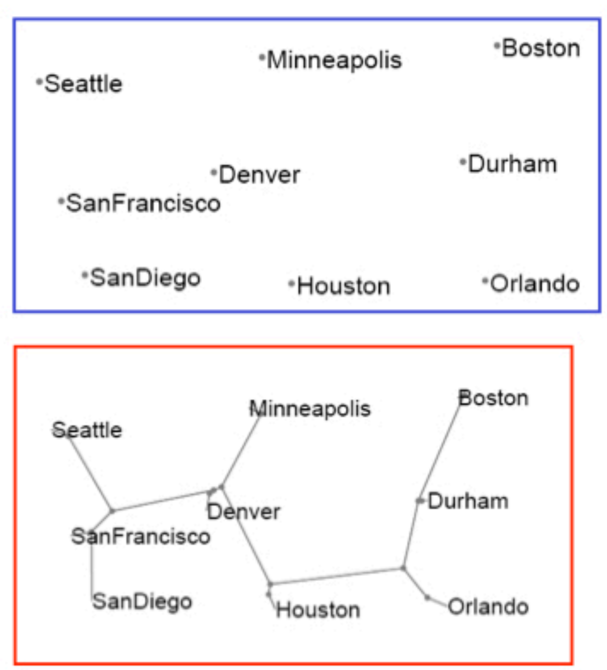


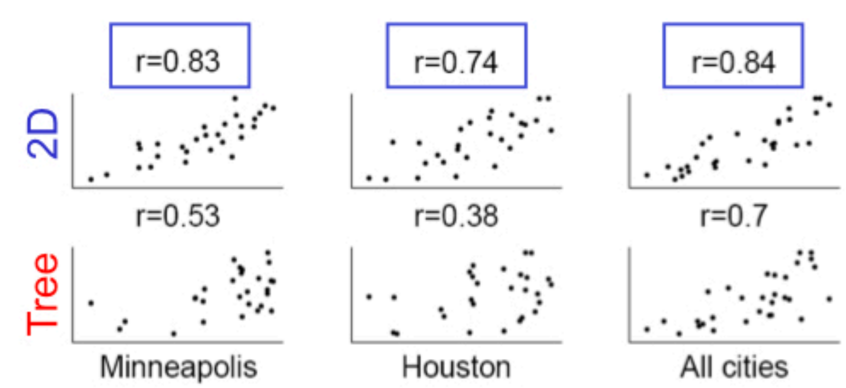
inference



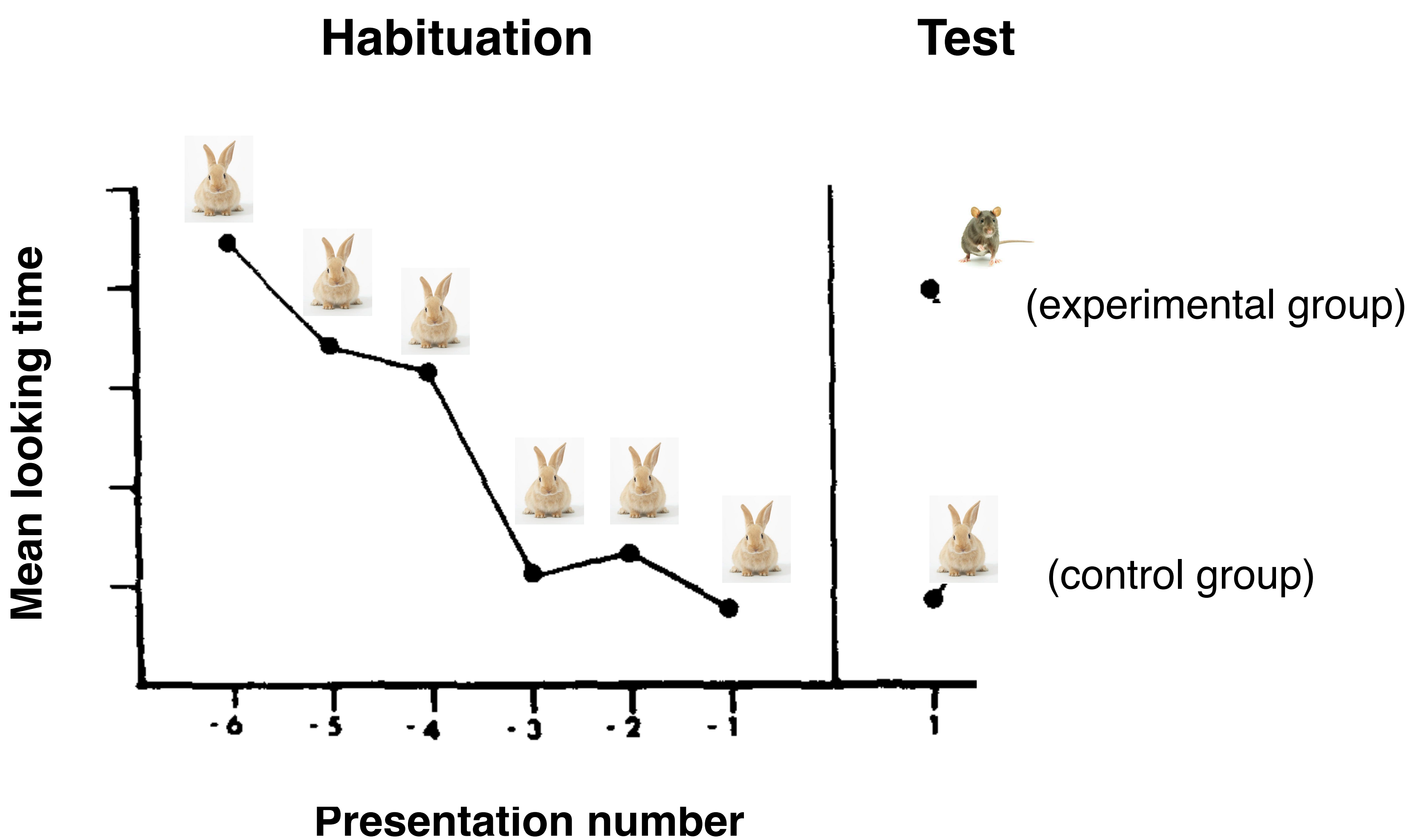


“Given that a certain kind of native American artifact has been found in sites near city X, how likely is the same artifact to be found near city Y?"





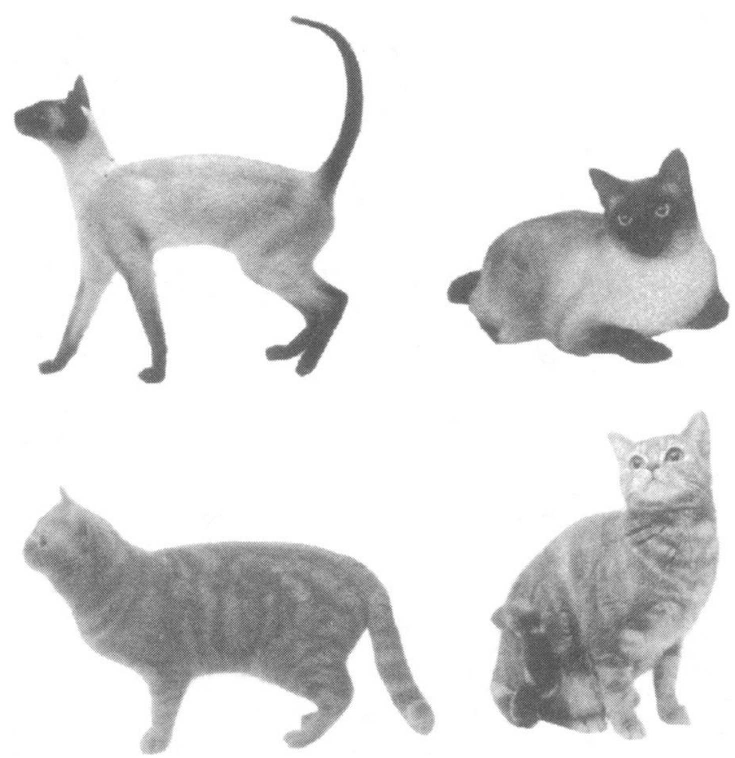
Concepts in infancy



siamese

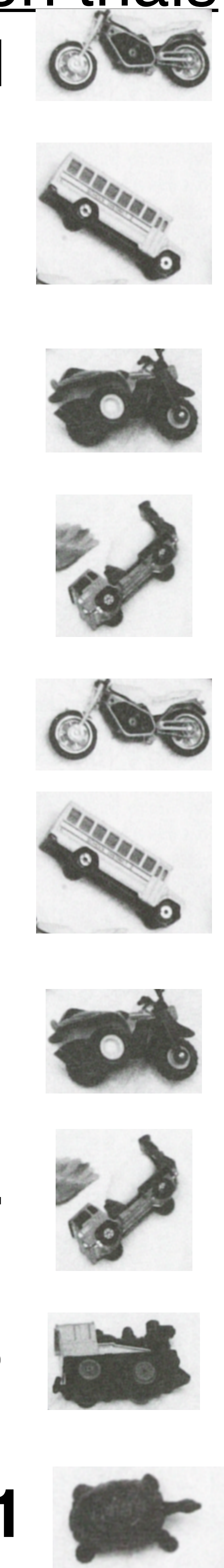
vs.

tabby



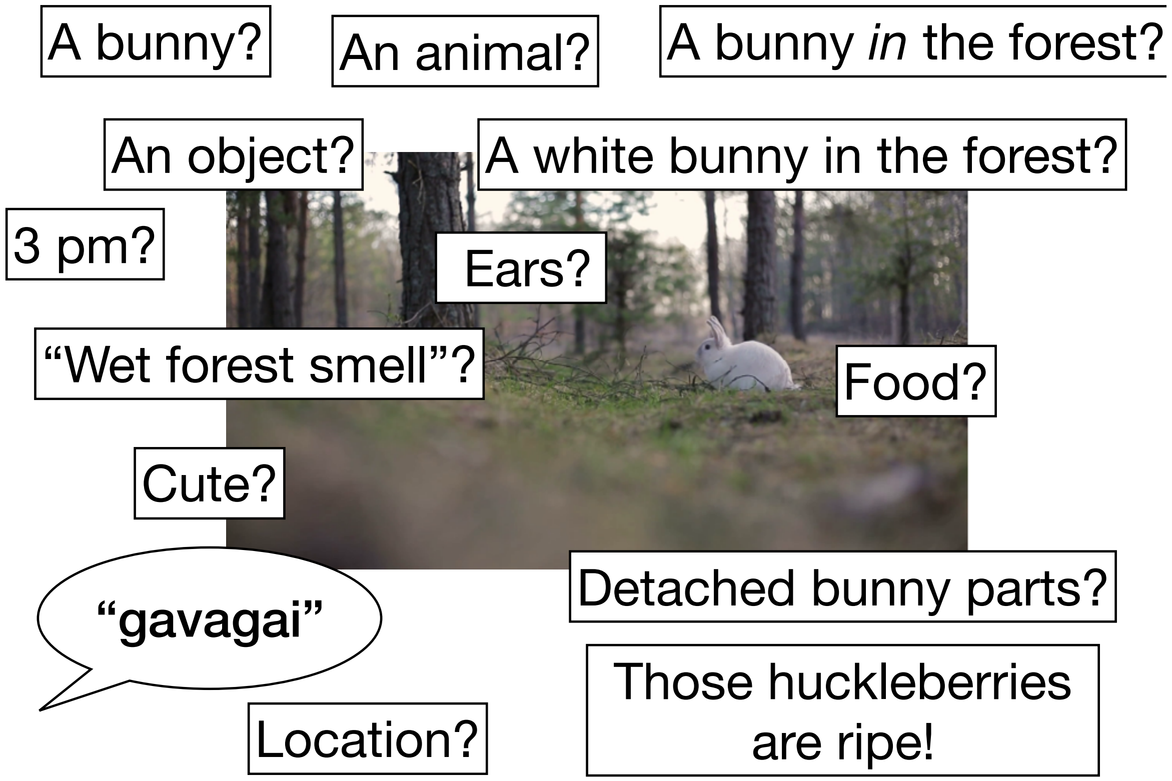
Examination trials

1. vehicle1
2. vehicle2
3. vehicle3
4. vehicle4
5. vehicle1
6. vehicle2
7. vehicle3
8. **vehicle4**
9. **vehicle5**
10. **animal1**



Conceptual development

The problem of induction



Original thought experiment due to W. V. Quine (1960).

Taxonomic bias

Novel words refer to taxonomic rather than thematic categories

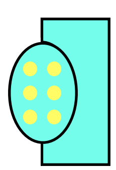
Here is a "dax"

Which is another "dax"?

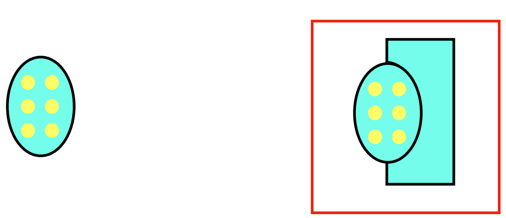
Whole object bias

Novel words refer to whole objects, rather than properties, actions, events, etc.

Here is a "dax"



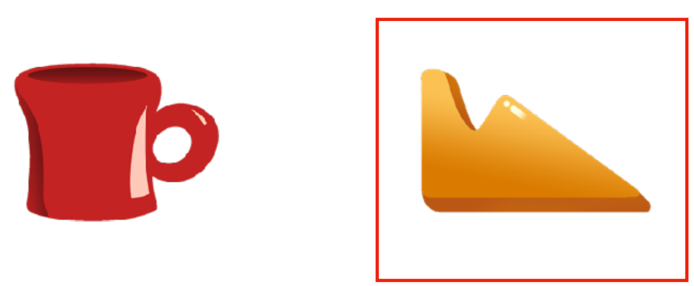
Which is the "dax"?



Mutual exclusivity (ME) bias

Once an object has one label, then it does not need another

Show me the "dax"

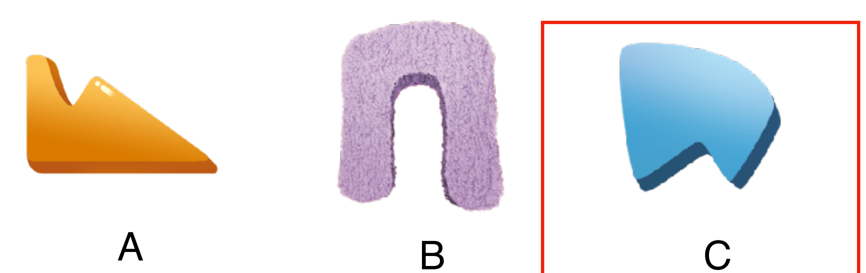


Shape bias

Objects with the same name tend to have the same shape
(as opposed to texture, color, size, etc.)

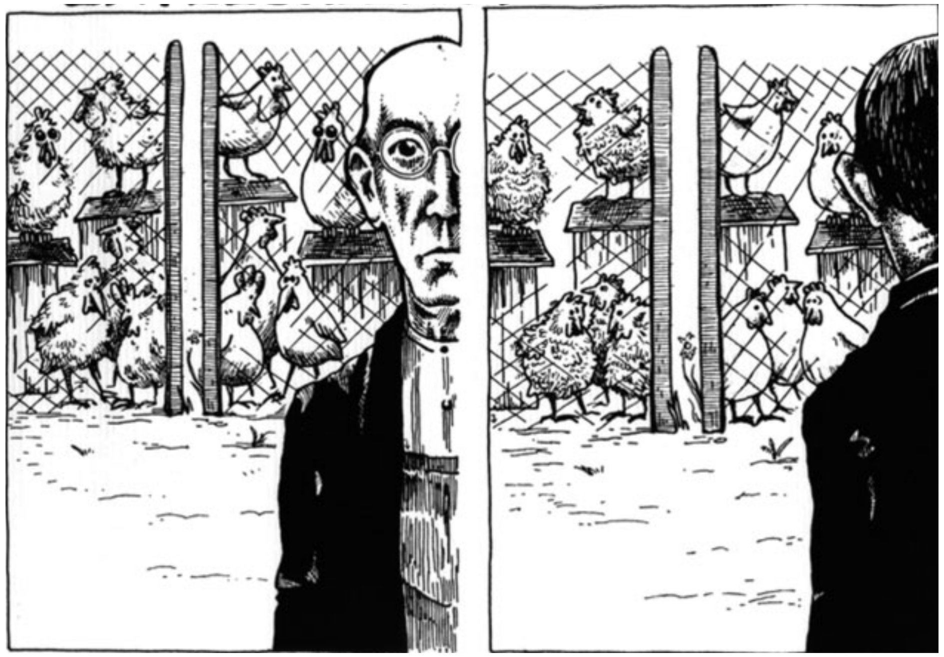
Here is a "dax"

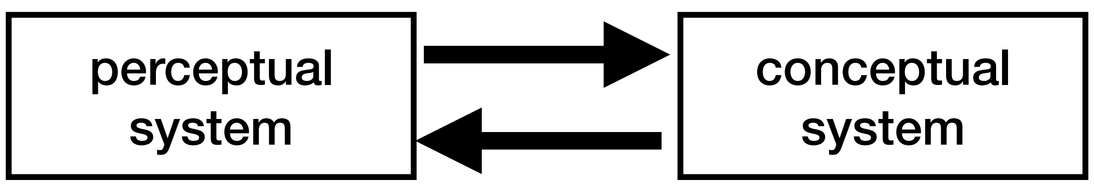
Which is the other "dax"?



How categories influence perception

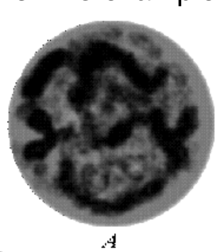
Categorical perception



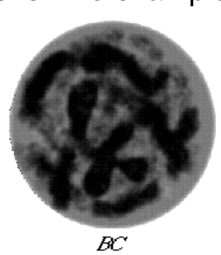


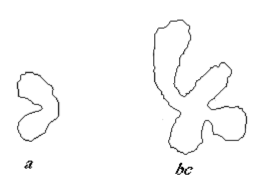
Group A-BC

shown 5 examples



shown 5 examples

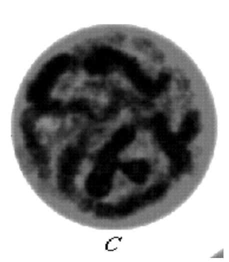


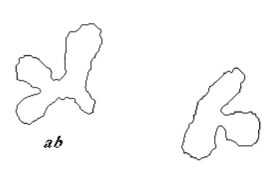


Group AB-C

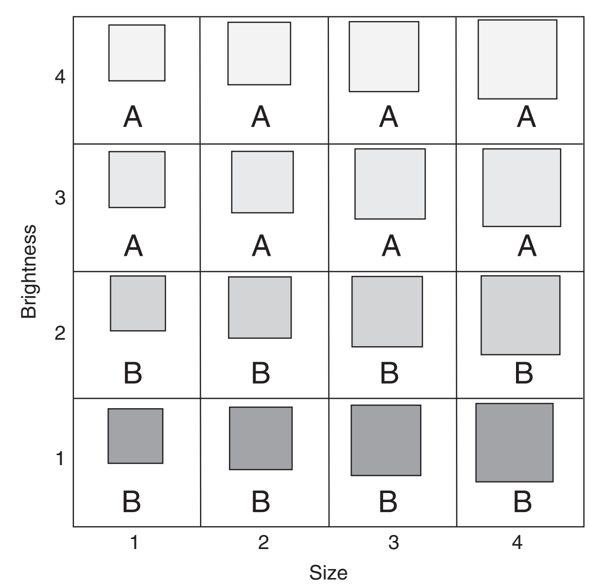
shown 5 examples

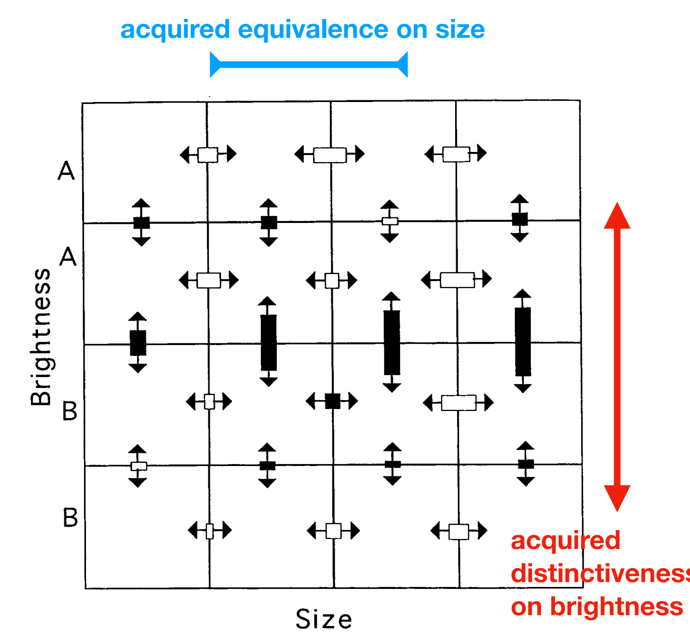
shown 5 examples





Brightness categorizers





Conceptual combination and creation

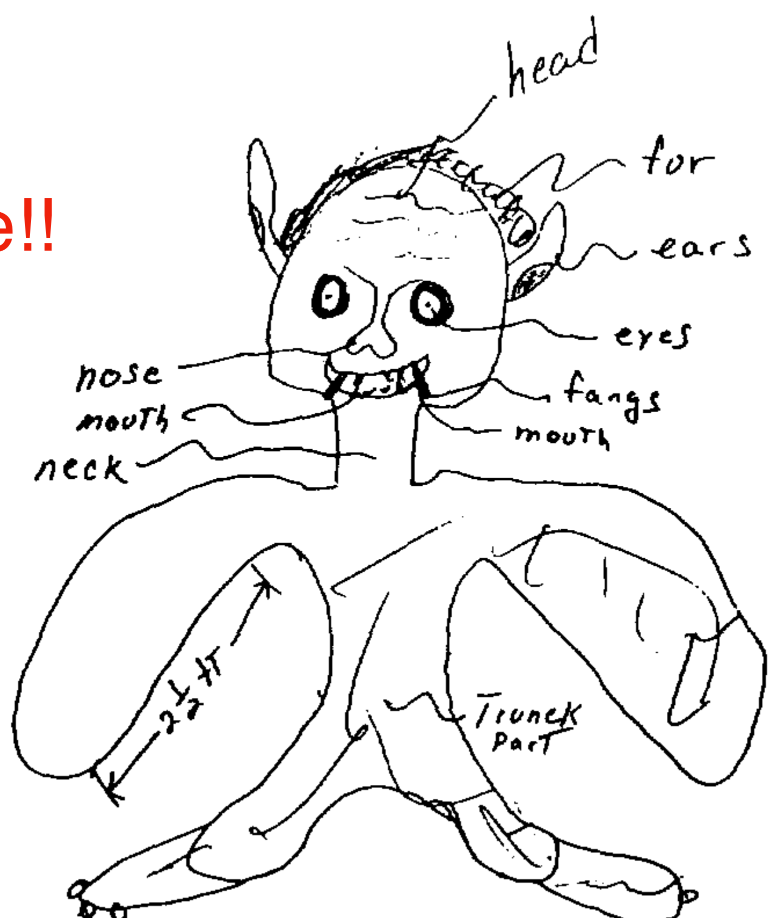
"pet"

"fish"

? ?

"pet fish"

Cute!!



"plastic truck"

A selection of open questions

- There are many successful models for different aspects of concepts, and we need a larger effort to develop models with broad coverage. Models should tackle multiple conceptual abilities.

- How can children learn a new word from just a handful of examples, and how do we replicate this ability in machines? What are their conceptual primitives?

- How are concepts represented in the brain, and what does this tell us about conceptual representation?

- How can we understand individual differences and cultural differences in categories and concepts?

- More work to develop formal models of feature creation, social concepts, and abstract concepts.

- Generation and combination are a clearer window into representation, and we should use these tasks to advance our understanding.

Thank you

A selection of open questions

- There are many successful models for different aspects of concepts, and we need a larger effort to develop models with broad coverage. Models should tackle multiple conceptual abilities.

- How can children learn a new word from just a handful of examples, and how do we replicate this ability in machines? What are their conceptual primitives?

- How are concepts represented in the brain, and what does this tell us about conceptual representation?

- Why are people very good at managing uncertainty in some conceptual tasks (e.g., word learning), but not others (e.g., Murphy and Ross's category-based induction tasks)?

- More work on formal models of feature creation, and models that interface more richly with background knowledge

- Generation and combination are a clearer window into representation, and more work is needed to understand them