Categories and Concepts - Spring 2019

Conceptual development

Brenden Lake

PSYCH-GA 2207

Last week: Categorization in pre-verbal infants (mostly 3-9 months)

<u>Habituation or Familiarization Trials</u>

Trial 1 bunny1 

Trial 2 bunny2 

Trial 3 bunny3 

Trial 4 bunny4 

Trial 5 bunny5 

Trial 6 bunny6 

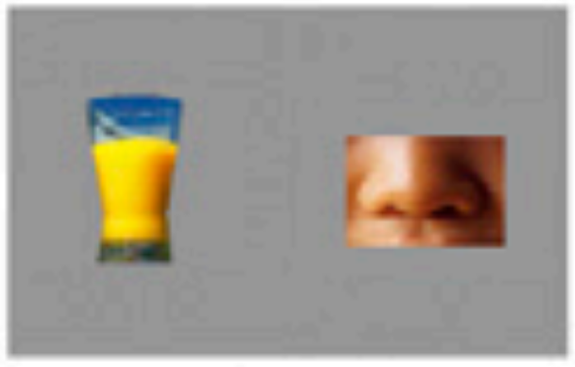
<u>Test Trial</u>

Trial 7 bunny 7 (control group)
or rat1 (experimental group)

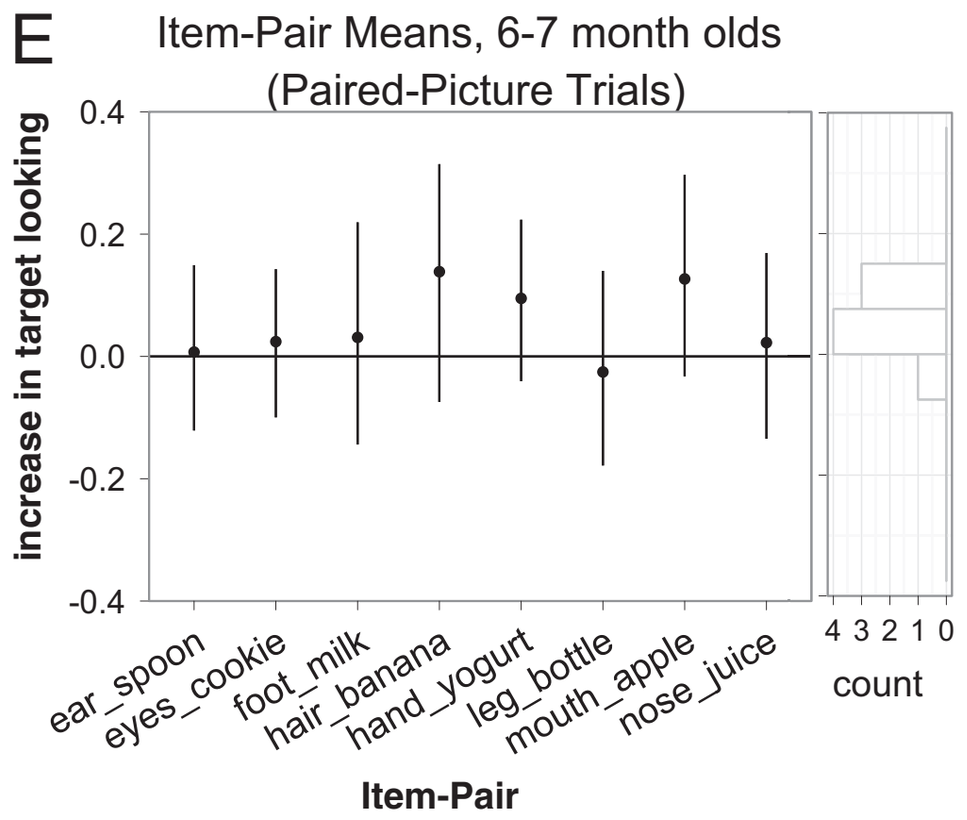
Last week: Evidence that pre-verbal 6-9 mo infants know meanings of many common nouns

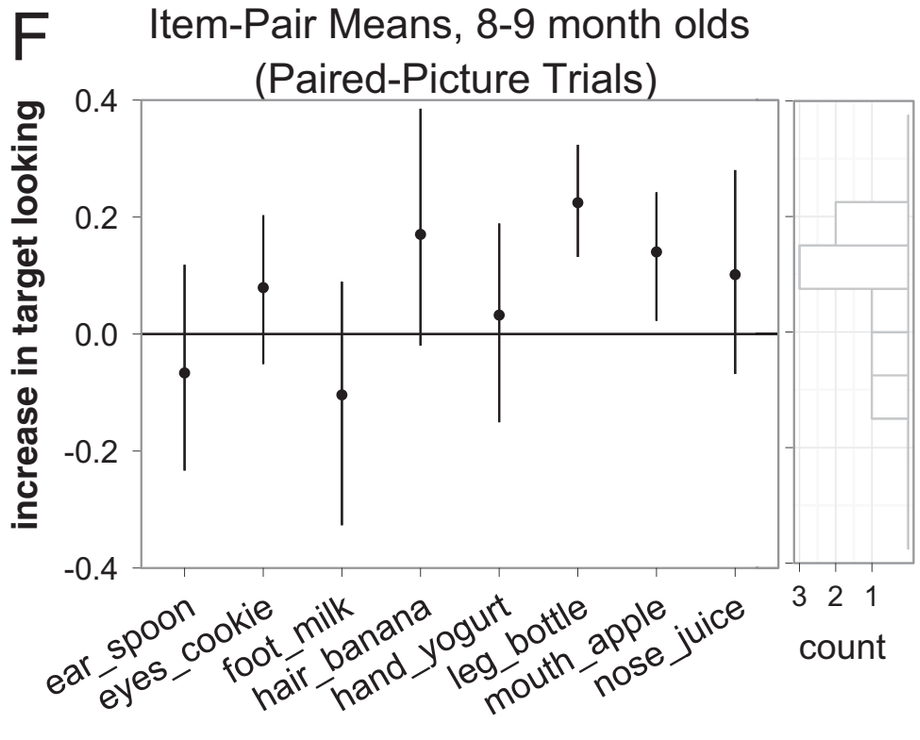
"Look at the mouth"

"Look at the juice"



(Bergelson & Swingley, 2012)





This week: older children (verbal; 1-4 years)

Word learning as a window into conceptual development

- Word learning is one of the most heavily researched and controversial topics in cognitive development
  - ✳ literature often does not make distinction between word learning and concept learning

- Strong argument that the representation of word meanings is based on concepts (see Big Book Chapter 11, which isn't covered in class but worth reading)

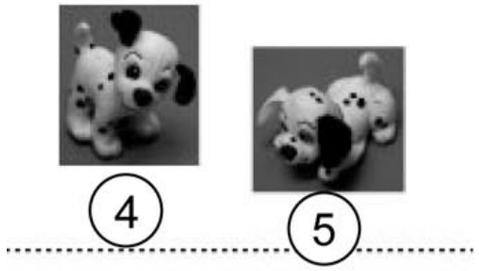
- In cog. dev., there are few concept learning experiments of the sort done with adults, since children are unwilling to do long category learning experiments with artificial categories
  - ✳ e.g., exemplar vs. prototype debate isn't active in this literature

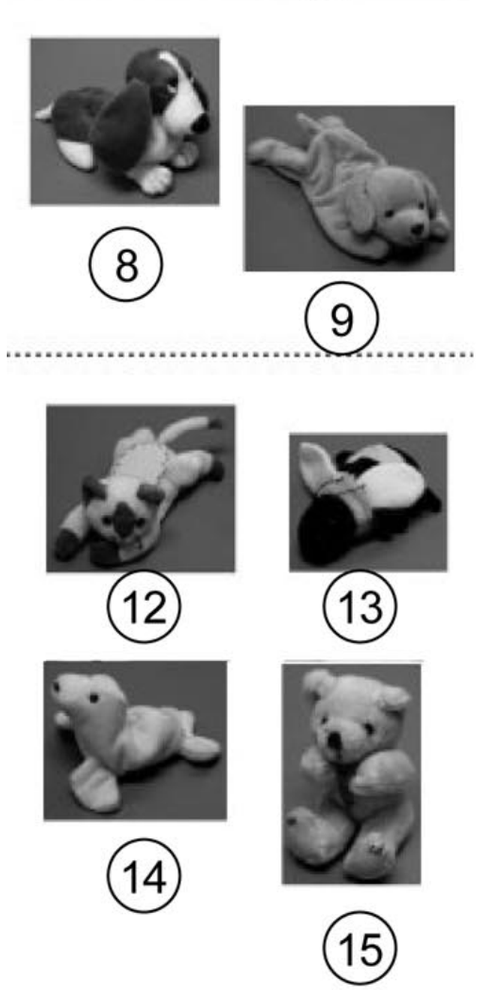
Review: Xu and Tenenbaum (2007): Word learning experiment with 3-4 year olds

Help Mr. Frog who speaks a different language pick out the objects he wants.

"Here is a fep"

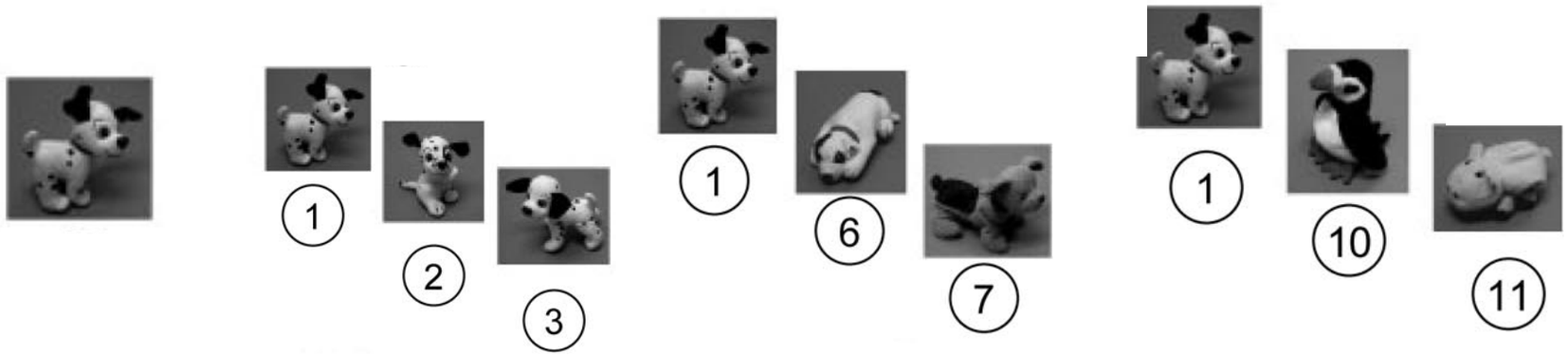
Which others are feps?
(YES/NO for each)



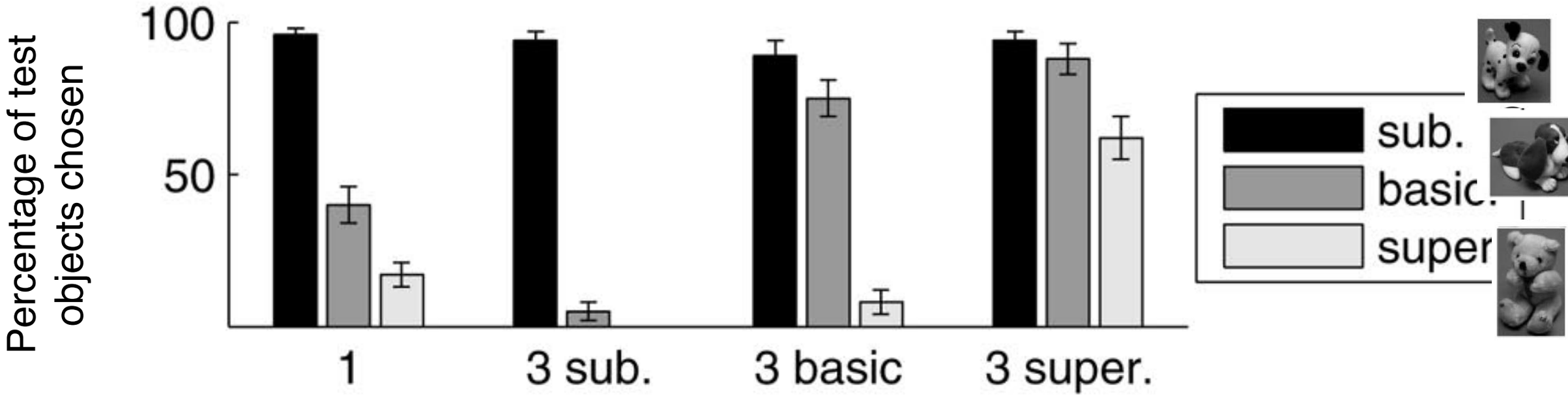


- children learn from sparse, positive examples of a new word
- "one shot learning" or "few shot learning"
- References: Carey & Bartlett, 1978; Markman, 1989; Xu & Tenenbaum, 1999; Bloom, 2000; Smith et al., 2002

Review: Xu and Tenenbaum Exp 3



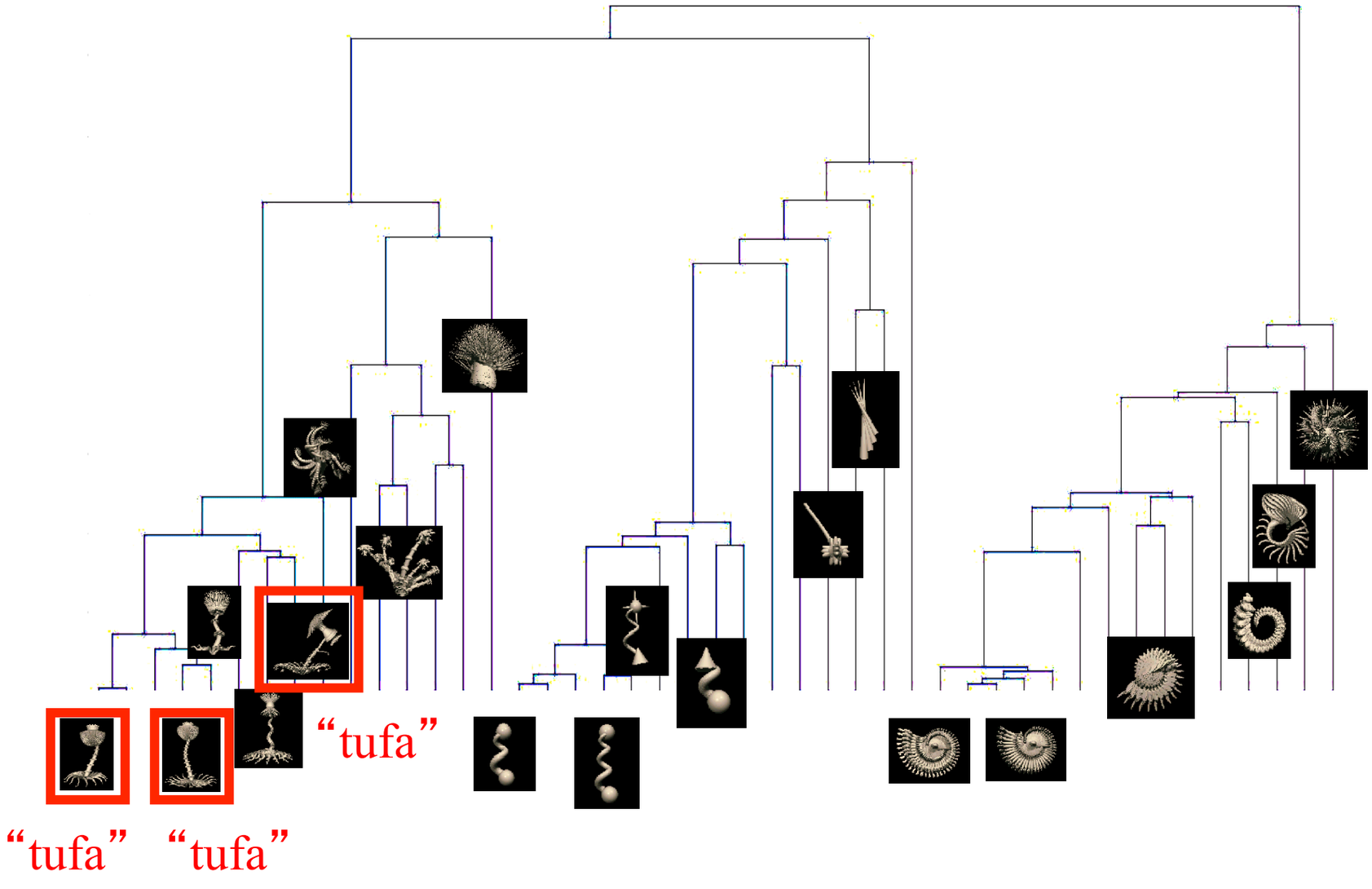
Children



This is **inductive reasoning,** which involves *probabilistic reasoning* from premises that supply some evidence for the truth of the conclusion.

As opposed to **deductive reasoning,** which involves *logical reasoning* from one or more statements (premises) to reach a certain conclusion

Word learning is inductive: "Here are some ‘tufas', where are the others?”



The problem of induction

“gavagai”

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

The problem of induction

A bunny?

An animal?

A bunny *in* the forest?

An object?

A white bunny in the forest?

3 pm?

Ears?

“Wet forest smell”?

Food?

Cute?

“gavagai”

Detached bunny parts?

Location?

Those huckleberries are ripe!

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

The problem of induction

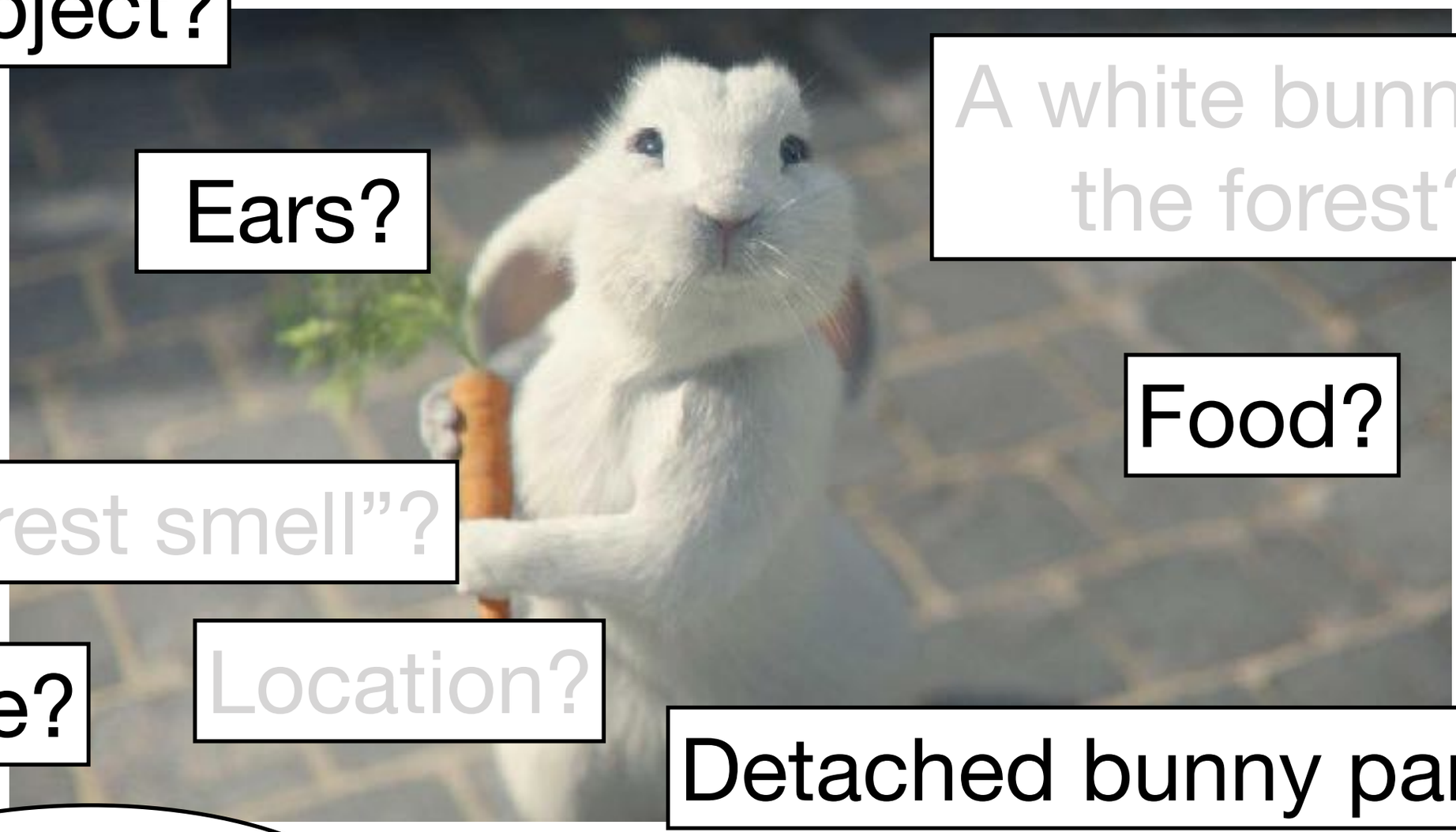
now you get more data…

A bunny?

An animal?

A bunny *in* the forest?

An object?



A white bunny in the forest?

Ears?

3 pm?

Food?

“Wet forest smell”?

Location?

Cute?

Detached bunny parts?

“gavagai”

Those huckleberries are ripe!

How can children learn new concepts from just one or a handful of examples?

- To account for the average adult vocabulary, children must learn about 10 words per day from when they start speaking to the end of high school (Bloom, 2000)

- If our inductive inferences go beyond the data given, then something must be making up the difference…

- Developmental psychologists have studied **constraints and biases** that allow children to make inferences that go beyond the data

- (You can also interpret these constraints and biases as **priors** in a Bayesian model of concept learning; Xu & Tenenbaum)

Review: Biases and constraints in Bayesian concept learning

(Xu & Tenenbaum, 2007)

$h \in H$: **hypothesis about meaning of word (e.g., node in tree structure)**

X : **data (often just labels of positive examples)**

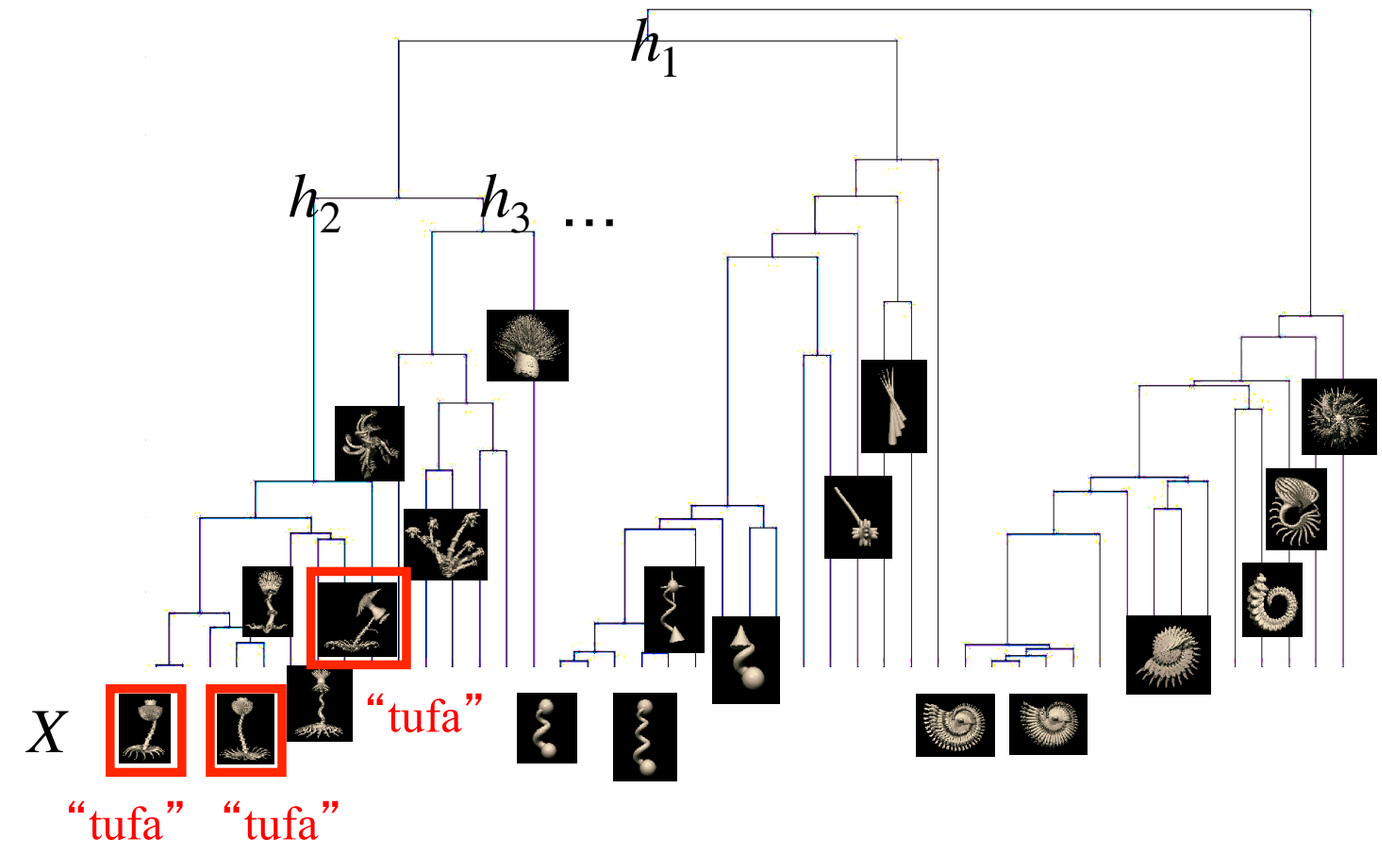
n : **number of examples**

Posterior over word meanings

$$P(h \mid X) = \frac{P(X \mid h)P(h)}{P(X)}$$

Likelihood (e.g., the size principle)

$$P(X \mid h) = [\frac{1}{\textbf{size}(h)}]^n$$



Prior

$P(h)$

The prior determines which hypotheses should be favored, or equivalently which constraints and biases most likely govern generalization

Review: Biases and constraints in Bayesian concept learning

(Xu & Tenenbaum, 2007)

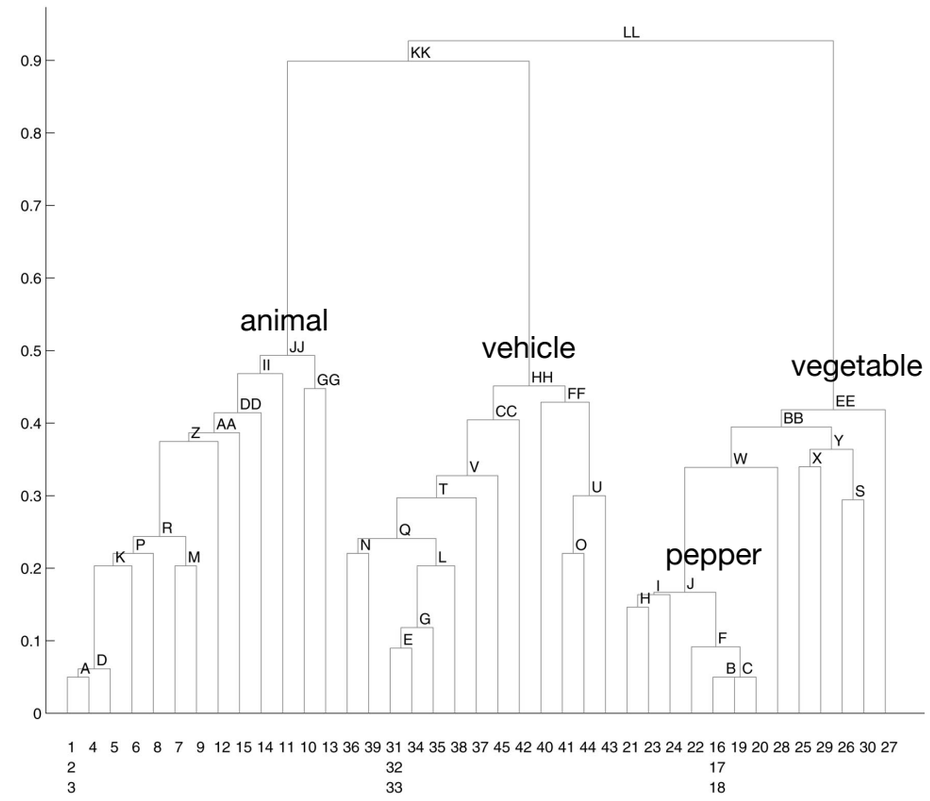
$h \in H$: **hypothesis about the meaning of word (node in tree structure)**

X : **1 or 3 positive examples**

n : **number of examples**

Posterior over word meanings

$$p(h \mid X) = \frac{P(X \mid h)P(h)}{P(X)}$$



Likelihood

$$P(X \mid h) = [\frac{1}{\textbf{size}(h)}]^n \approx [\frac{1}{\textbf{height}(h) + \epsilon}]^n$$

(height is the average within-node distance between examples)

Prior

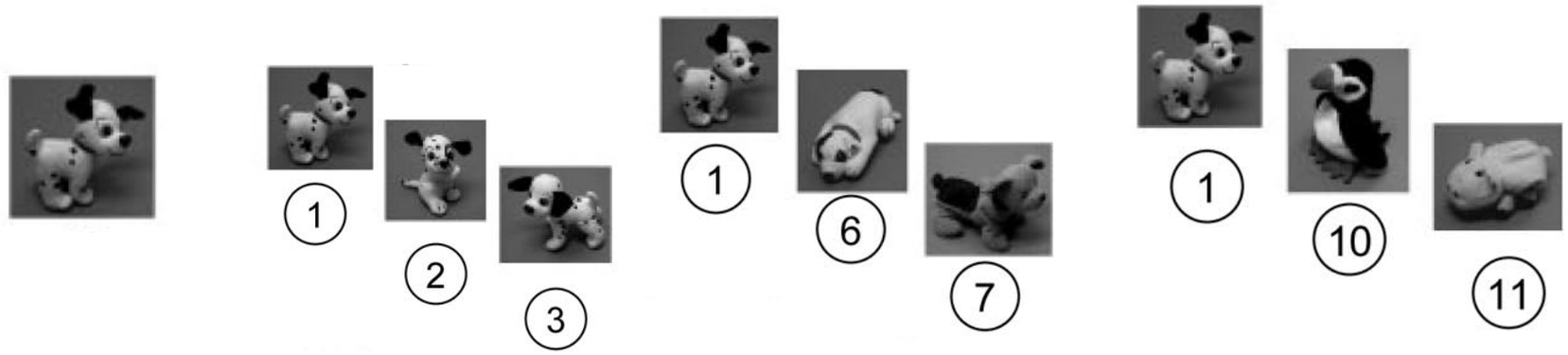
$$P(h) \propto \textbf{height}(\textbf{parent}[h])) - \textbf{height}(h)$$

favors more distinctive nodes,
or favor nodes at a certain level….

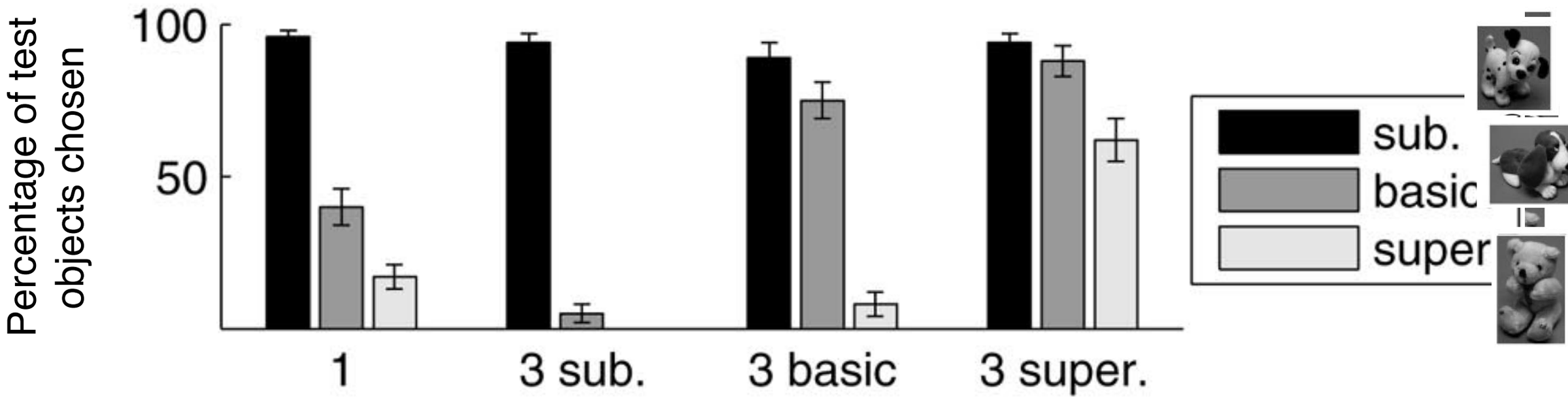
Generalizing to a new example *y*

$$p(y \in C \mid X) = \sum_{h \in H} P(y \in C \mid h) p(h \mid X)$$

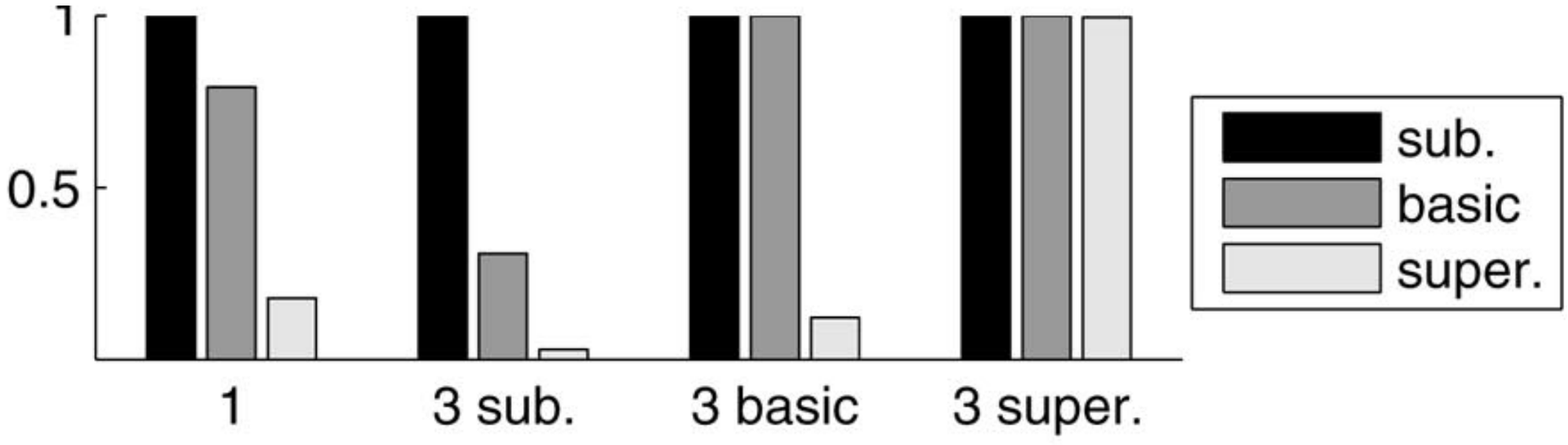
Xu and Tenenbaum : model results



Children



Bayesian model



Key biases and constraints on word learning in the developmental literature

Already covered

- Basic level bias

Today

- Taxonomic bias
- Whole object bias
- Mutual exclusivity bias
- Shape bias

Let’s test biases in word learning!

General experimental strategy: Trials do not have enough information to deduce a “right answer.” Instead, they are inductive and highly unconstrained, probing biases and constraints

In different fields we call this by different names:

- developmental psychology: “constraints and biases”
- machine learning: “inductive biases”
- statistics: “priors”

Here is a “dax”

Which is another “dax”?

Taxonomic bias

Novel words refer to taxonomic rather than thematic categories

Here is a “dax”

Which is another “dax”?

Taxonomic bias

COGNITIVE PSYCHOLOGY 16, 1–27 (1984)

Children's Sensitivity to Constraints on Word Meaning: Taxonomic versus Thematic Relations

ELLEN M. MARKMAN AND JEAN E. HUTCHINSON

Stanford University

A major problem in language learning is to figure out the meaning of a word given the enormous number of possible meanings for any particular word. This problem is exacerbated for children because they often find thematic relations between objects to be more salient than the objects' taxonomic category. Yet most single nouns refer to object categories and not to thematic relations. How do children learn words referring to categories when they find thematic relations so salient? We propose that children limit the possible meanings of nouns to refer mainly to categorical relations. This hypothesis was tested in four studies. In each study, preschool children saw a series of target objects (e.g., dog), each followed by a thematic associate (e.g., bone) and a taxonomic associate (e.g., cat). When children were told to choose another object that was similar to the target ("See this? Find another one."), they as usual often selected the thematic associate. In contrast, when the instructions included an unknown *word* for the target ("See this fep? Find another fep."), children now preferred the taxonomic associate. This finding held up for 2- and 3-year-olds at the basic level of categorization, for 4- and 5-year-olds at the superordinate level of categorization, and 4- and 5-year-olds who were taught new taxonomic and new thematic relations for unfamiliar objects. In each case, children constrained the meaning of new nouns to refer mainly to categorical relations. By limiting the hypotheses that children need to consider, this constraint tremendously simplifies the problem of language learning.

One of the major problems confronting someone learning a language is to figure out the meaning of a word given the enormous number of possible meanings for any particular word. Children commonly learn their

This paper was completed while E. M. Markman was at the Center for Advanced Study in the Behavioral Sciences, which received support from NSF Grant BNS8206304 and the Spencer Foundation. This research was supported in part by a Stanford University Fellowship and NIMH Traineeship to J. E. Hutchinson. Portions of the research were presented at meetings of the Society for Research in Child Development, Detroit, 1983, and the Western Psychological Association, San Francisco, 1983. We thank the directors and staffs

Taxonomic bias : Markman and Hutchinson Ex 1

Novel words refer to taxonomic rather than thematic categories

- 2-3 year olds
- concerned basic level categories like “dog” and chair”
- two conditions: “no word” and “novel word”
- triad task

no word condition

"See this?”

“find another one that is the same as this"

59% (ns)

41%

novel word condition

"See this? It is a dax”

“find another dax that is the same as this dax”

83% (significant)

17%

Taxonomic bias : Markman and Hutchinson Ex 1

Novel words refer to taxonomic rather than thematic categories

TABLE 1
Stimulus Materials for Experiment 1

Standard object	Taxonomic choice	Thematic choice
Police car	Car	Policeman
Tennis shoe	High-heeled shoe	Foot
Dog	Dog	Dog food
Straight backed chair	Easy chair	Man in sitting position
Crib	Crib	Baby
Birthday cake	Chocolate cake	Birthday present
Blue jay	Duck	Nest
Outside door	Swinging door	Key
Male football player	Man	Football
Male child in swimsuit	Female child in overalls	Swimming pool

Taxonomic bias : Markman and Hutchinson Ex 4

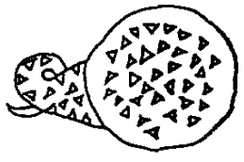
- Do children use abstract knowledge about words rather than just specific known meanings?
- 4-6 year olds; using novel objects; two conditions

Background context

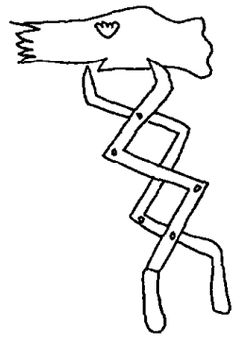
Picture 1

"This swims in water"

"This swims in water"



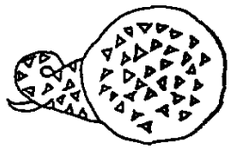
Picture 2: "This catches this"



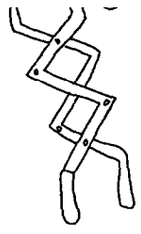
no word test

"See this?"

"can you find another one?"



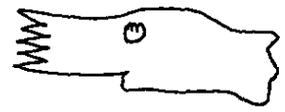
taxonomic choice
37%



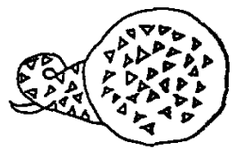
thematic choice
63%

novel word test

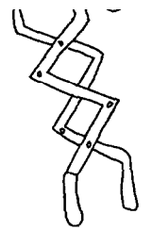
"See this dax?



"can you find another dax?"



taxonomic choice
63%



thematic choice
37%

Historical note: Children's responses in sorting tasks (Piaget & Inhelder; Vygotsky)

Sorting task : "Put together the objects that are alike or go together"

Results (preschoolers)

- Children often make spatial arrays or scene constructions
- Often use thematic groupings, e.g., dog and bowl
- Researchers' (incorrect) conclusion: Children's concepts are all messed up (Piaget et al.). Probably they're based on thematic and associative relations
- Jerry Fodor's response: If this were true, we wouldn't be able to talk with children at all!

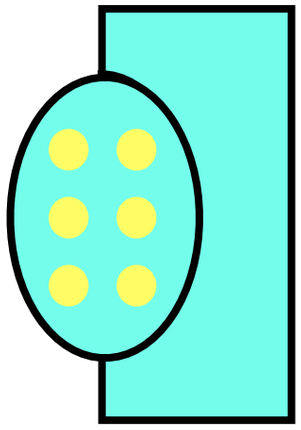
Let’s test our next bias!

General experimental strategy: Trials do not have enough information to deduce a “right answer.” Instead, they are inductive and highly unconstrained, probing biases and constraints

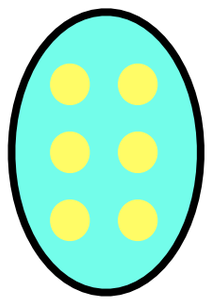
In different fields we call this by different names:

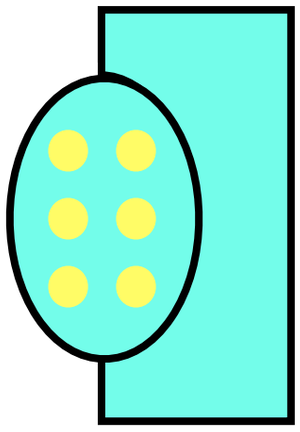
- developmental psychology: “constraints and biases”
- machine learning: “inductive biases”
- statistics: “priors”

Here is a “dax”



Which is the “dax"?

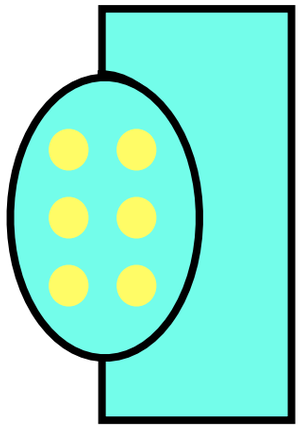




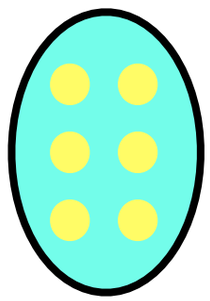
Whole object bias

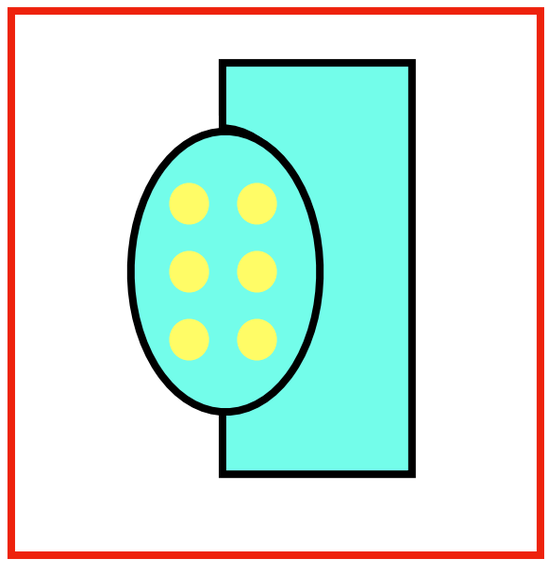
Novel words refer to whole objects, rather than properties, actions, events, etc.

Here is a “dax”



Which is the “dax"?





Developmental Psychology
2007, Vol. 43, No. 5, 1051–1061

0012-1649/07/$12.00 DOI: 10.1037/0012-1649.43.5.1051

Young Children Associate Novel Words With Complex Objects Rather Than Salient Parts

George Hollich
Purdue University

Roberta M. Golinkoff
University of Delaware

Kathy Hirsh-Pasek
Temple University

How do children learn associations between novel words and complex perceptual displays? Using a visual preference procedure, the authors tested 12- and 19-month-olds to see whether the infants would associate a novel word with a complex 2-part object or with either of that object's parts, both of which were potentially objects in their own right and 1 of which was highly salient to infants. At both ages, children's visual fixation times during test were greater to the entire complex object than to the salient part (Experiment 1) or to the less salient part (Experiment 2)—when the original label was requested. Looking times to the objects were equal if a new label was requested or if neutral audio was used during training (Experiment 3). Thus, from 12 months of age, infants associate words with whole objects, even those that could potentially be construed as 2 separate objects and even if 1 of the parts is salient.

Keywords: word learning, constraints, whole object bias

As Quine (1960) observed, the task of learning a word presents an infinite array of possible word–referent links. A word such as *bottle* could refer to the nipple, to the plastic base, or to the whole bottle including both of these parts. It also could refer to sucking or even the process of feeding (see Bloom, 2000). Nonetheless, despite many possible misinterpretations, children generally sort out the correct meanings. How?

One seemingly obvious solution to this problem is that children will make an educated guess. Indeed, one branch of research in developmental psychology has sought to identify the heuristics that children use to limit their hypotheses about the meaning of a

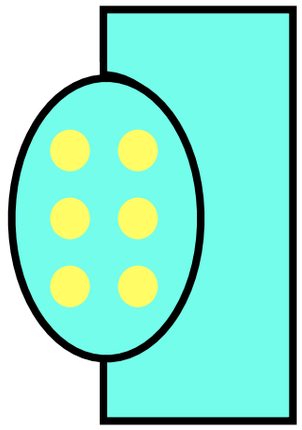
objects rather than actions, attributes, or parts of objects. Markman (1989) called this heuristic the *whole object bias*. For example, in a seminal study, Woodward (1993) presented 18-month-old children with a novel word and two possible referents. One referent was a visually attractive display representing an event (e.g., brightly colored dye diffusing through water); the other was a novel object in a static display. Despite a salience preference for the event, the children looked at the object more when they heard a novel noun.

There are a few problems with this whole object bias as a solution to the word-learning dilemma. First, most of the evidence

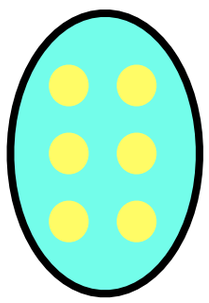
Whole object bias: Results

- First experiments by Markman and Wachtel (1988)
- Hollich et al. (2007) used preferential looking time with 12 mo

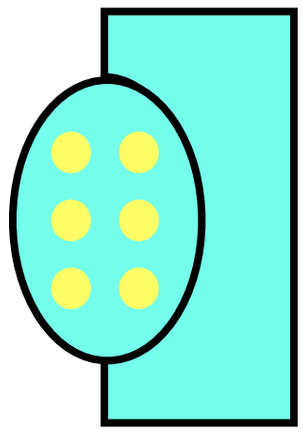
Here is a “dax”



Look at the “dax”!



1.79 seconds

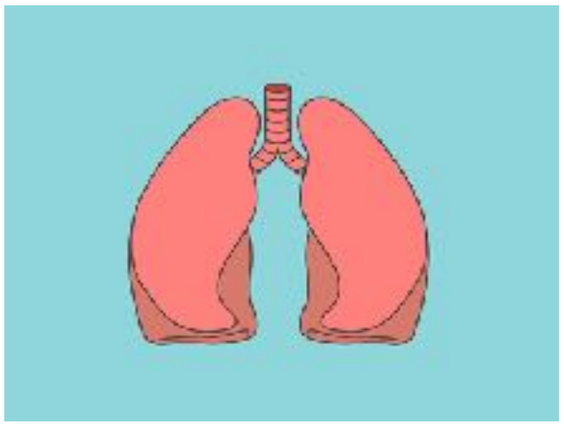


2.73 seconds
(significant)

Markman and Wachtel (1988) Ex 2

- 3-4 year olds, using unfamiliar words (for 3-4 year olds)

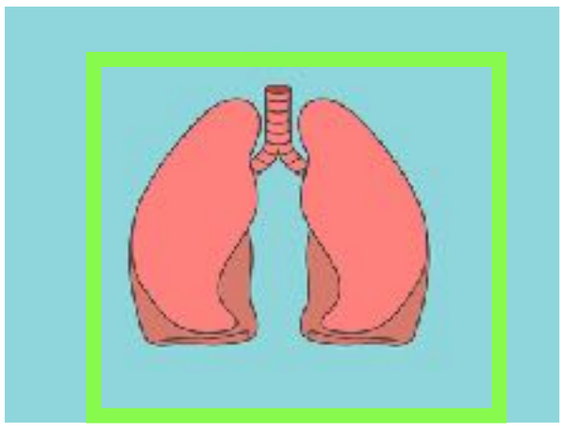
"See this? It is a lung”



Object	Novel label for part
*current detector	detector
pipe tool	damper
*ritual implement	crescent
*pagoda	finial
microscope	platform
*lung	trachea

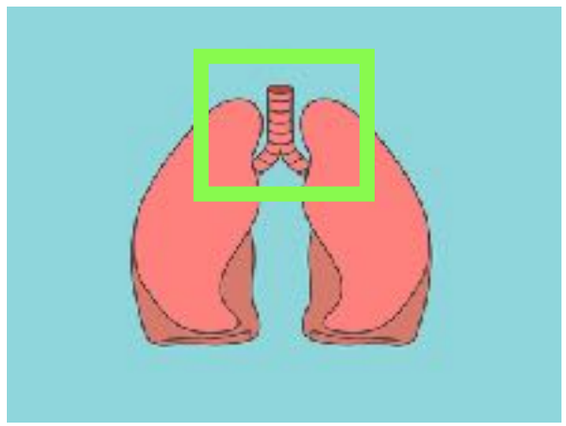
“now, which is the lung?"

“This whole thing”



80%

“or just this part”



20%

Let’s test our next bias!

Show me the “dax”

Mutual exclusivity (ME) bias

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

Show me the “dax”

Children's Use of Mutual Exclusivity to Constrain the Meanings of Words

ELLEN M. MARKMAN

AND

GWYN F. WACHTEL

Stanford University

For children to acquire vocabulary as rapidly as they do, they must be able to eliminate many potential meanings of words. One way children may do this is to assume category terms are mutually exclusive. Thus, if a child already knows a label for an object, a new label for that object should be rejected. Six studies with 3-year-olds tested this hypothesis. Study 1 demonstrated that children reject a second label for an object, treating it, instead, as a label for a novel object. In the remaining studies, this simple novel label-novel object strategy was precluded. If the only object present is familiar, children cannot map a novel term to a novel object. Instead they must analyze the object for some other attribute to label. In Studies 2–6, children were taught either a new part term, e.g., *trachea,* or a new substance term, e.g., *pewter,* by showing them an object and saying, "This is a trachea" or ("It is pewter"). For unfamiliar objects, children tended to interpret the term as a label for the object itself. For familiar objects, they tended instead to interpret it as a part or substance term. Thus, mutual exclusivity motivates children to learn terms for attributes, substances, and parts as well as for objects themselves.

In the first few years of life, children learn new vocabulary at a staggering rate (Carey, 1978). The learning of words can be viewed as an inductive process, where from a limited amount of information, children must figure out the meaning of a novel term. One fundamental problem with induction is that the evidence underdetermines the hypotheses (Peirce, 1957; Quine, 1960). For any set of data there will be an indefinite number of logically possible hypotheses that are consistent with it. How is it, then, that humans so frequently converge on the same hypotheses?

Mutual exclusivity (ME) bias : Markman and Wachtel Ex 1

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

- 3-4 year olds
- used objects familiar to child (cup) and unfamiliar objects (cherry pitter)
- two conditions: “no word” and “novel word”

no word condition

“Show me one”

45%

55% (ns)

novel word condition

“Show me the dax”

18%

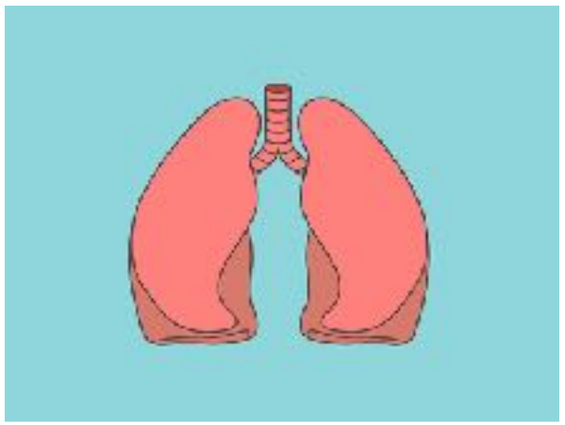
82%

Markman and Wachtel (1988) Ex 2

- 3-4 year olds, using novel objects and familiar objects, in two conditions
- Result: children will use ME even when there is no other object as a possible referent, instead finding another property to attach the label to

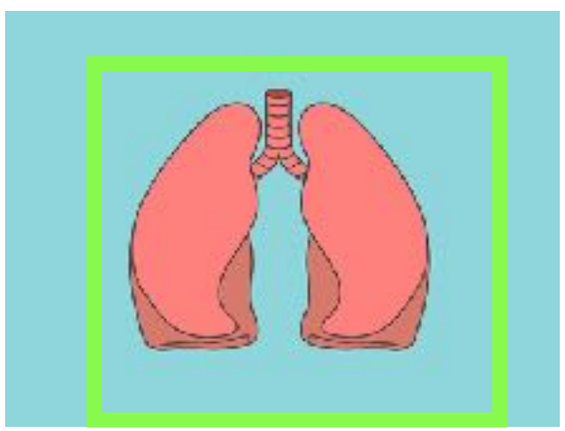
Novel object / novel label condition

"See this? It is a lung”



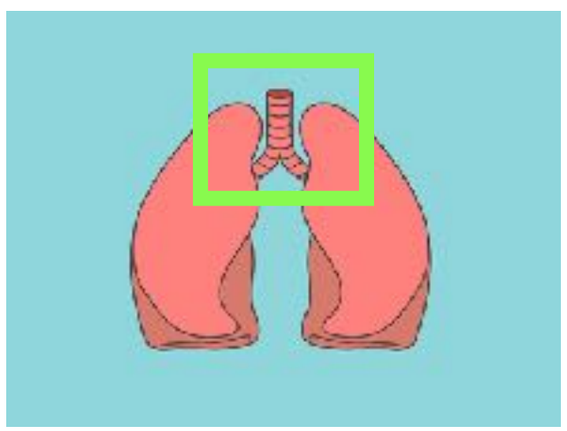
“now, which is the lung?"

“This whole thing”



80%

“or just this part”



20%

Familiar object / novel label condition (ME)

See this? It is a "boom”

Which is the “boom”?

This whole thing

43%

or just this part

57%

Fast mapping

- "Fast mapping” refers to learning a new word (perhaps via ME) and retaining it over an extended period
- Study with 3-4 year olds below (Carey and Bartlett; 1978) …

Exposure (in classroom setting)

You see those two trays over there. Bring me the “chromium" one. Not the red one, the “chromium" one.

Comprehension test 7-10 days later

Choose the chromium one (9 options)…

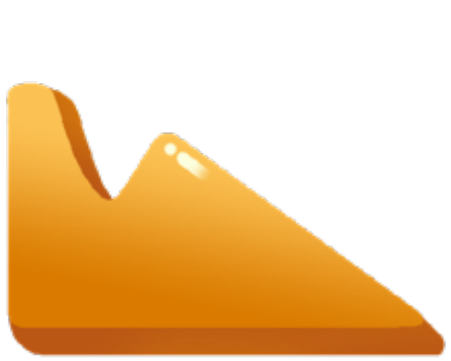
…

- Children were 47% correct after one exposure, and 63% correct after two exposures (10 weeks layer); some even spontaneously said “chromium” during naming test
- Demonstrates some ability for rapid learning and retention

Let’s test our next bias!

Here is a “dax”

Which is the other “dax”?



A

B

C

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”?

A

B

C

Cognitive Development, 3, 299–321 (1988)

The Importance of Shape in Early Lexical Learning

Barbara Landau
Columbia University

Linda B. Smith
Susan S. Jones
Indiana University

We ask if certain dimensions of perceptual similarity are weighted more heavily than others in determining word extension. The specific dimensions examined were shape, size, and texture. In four experiments, subjects were asked either to extend a novel count noun to new instances or, in a nonword classification task, to put together objects that go together. The subjects were 2-year-olds, 3-year-olds, and adults. The results of all four experiments indicate that 2- and 3-year-olds and adults all weight shape more heavily than they do size or texture. This observed emphasis on shape, however, depends on the age of the subject and the task. First, there is a developmental trend. The shape bias increases in strength and generality from 2 to 3 years of age and more markedly from early childhood to adulthood. Second, in young children, the shape bias is much stronger in word extension than in nonword classification tasks. These results suggest that the development of the shape bias originates in language learning—it reflects a fact about language—and does not stem from general perceptual processes.

Within the first few years of life, children learn many hundreds of words for different kinds of natural objects and artifacts. As many have noted, the rapidity and accuracy of this learning present a puzzle: The information objectively

Shape bias : Landau, Smith, & Jones (1998; Ex 3)

Objects with the same name tend to have the same shape

- 2 year olds and 3 year olds
- two conditions: “no word” and “novel word”; triad task
- Result: naming task directs children’s attention to shape

no word condition

“Here is one”

2”; wooden

“which one belongs with it”

same texture vs. same shape

wood   wire

64% (2 yr); 67% (3 yr)

same size vs. same shape

2”   24”

40% (2 yr); 48% (3 yr)

novel word condition

Here is a “dax”

2”; wooden

“which of these is a dax?”

same texture vs. same shape

wood   wire

71% (2 yr); 79% (3 yr)

same size vs. same shape

2”   24”

60% (2 yr); 75% (3 yr)

Research Article

OBJECT NAME LEARNING PROVIDES ON-THE-JOB TRAINING FOR ATTENTION

Linda B. Smith,[1] Susan S. Jones,[1] Barbara Landau,[2] Lisa Gershkoff-Stowe,[3] and Larissa Samuelson[4]

[1]Indiana University, [2]University of Delaware, [3]Carnegie Mellon University, and [4]University of Iowa

Abstract—*By the age of 3, children easily learn to name new objects, extending new names for unfamiliar objects by similarity in shape. Two experiments tested the proposal that experience in learning object names tunes children's attention to the properties relevant for naming—in the present case, to the property of shape—and thus facilitates the learning of more object names. In Experiment 1, a 9-week longitudinal study, 17-month-old children who repeatedly played with and heard names for members of unfamiliar object categories well organized by shape formed the generalization that only objects with similar shapes have the same name. Trained children also showed a dramatic increase in acquisition of new object names outside of the laboratory during the course of the study. Experiment 2 replicated these findings and showed that they depended on children's learning both a coherent category structure and object names. Thus, children who learn specific names for specific things in categories with a common organizing property—in this case, shape—also learn to attend to just the right property—in this case, shape—for learning more object names.*

Learning names for things requires attention to the right object properties. For example, learning which things are called "cup" in English may require that a child attend especially to object shape, because in English, shape is the perceptual property that matters most for determining which objects are included in the category "cup" (Biederman, 1987; Rosch, 1973; Samuelson & Smith, 1999). Young children are remarkably successful at forming object categories organized around the same properties as the categories of the adults in their language communities. But how do children know which properties to attend to? Which properties are the right ones for learning object names?

We have previously suggested that attention gets on-the-job training (Landau, Smith, & Jones, 1988; Smith, 1995). The idea is that learning object names contextually tunes attention, making it skilled in the task of learning object names. Smart attention leads to the more rapid formation of individual categories, and to an accelerated rate of object names asystematically, then generalizing the names for artifacts systematically by shape (e.g., Samuelson & Smith, 1999).

The first 300 nouns that young children learn tend to be names for concrete-artifact categories that adults judge to be well organized by shape (Samuelson & Smith, 1999). Individual exceptions among early learned categories show that shape is not uniformly privileged in defining object categories. Nonetheless, we have shown that shape is a good cue for determining membership in an overwhelming majority of common-object categories (Samuelson & Smith, 1999; see also Biederman, 1987; Rosch, 1973). And there is evidence that young children may learn to use that cue to good effect. Previous research indicates that children's attention to shape co-develops with acceleration in the rate of learning object names.

Figure 1 illustrates four proposed steps through which learning object names and attention to shape may be bidirectionally and causally related.[1] Step 1 is mapping names to objects—the name "ball" to a particular ball and the name "cup" to a particular cup, for example. This is done multiple times for each name as a child encounters new instances. The objects that get the same name are likely to be similar in shape (Samuelson & Smith, 1999). This learning of individual names for things thus sets up Step 2—first-order generalizations about the structure of individual categories, that is, the knowledge that balls are round and cups are cup shaped. This first-order generalization should enable the learner to recognize novel balls and cups.

Another higher-order generalization is also possible. Because many of the object categories that children learn are shape based, children could also learn the second-order generalization that object names in general span categories of similarly shaped things. As illustrated in Step 3, this second-order correlation requires generalizations over specific names and specific category structures. But making this higher-order generalization should enable the child to extend any object name, even one encountered for the first time, to new instances by shape. Step 4 illustrates the potential developmental consequence of this higher-

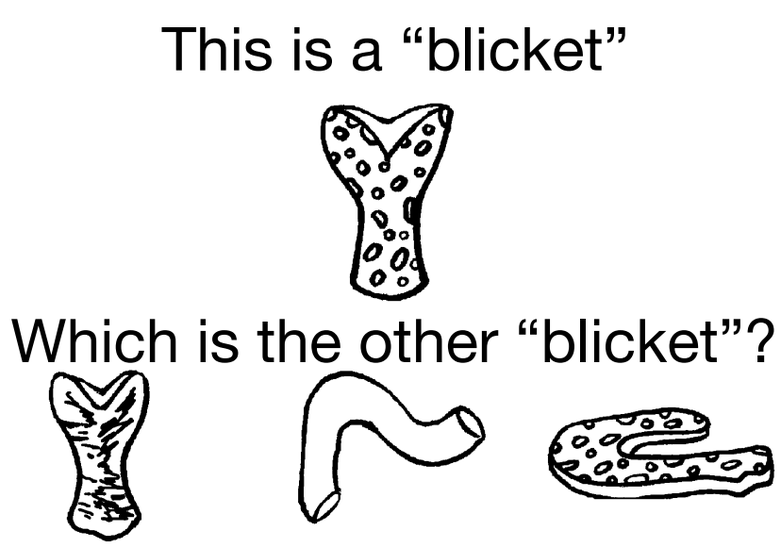
Evidence that shape bias is acquired

- 17 mo children at start of study (who have no shape bias)
- 7 weeks of once-a-week play sessions; children in training group taught four novel names "wif", "zup" "dax" and "lug"

Shape bias training (7 weeks)

"This is a dax. Here is another one. Let’s put the daxes in the wagon"

Shape bias test (using new category)

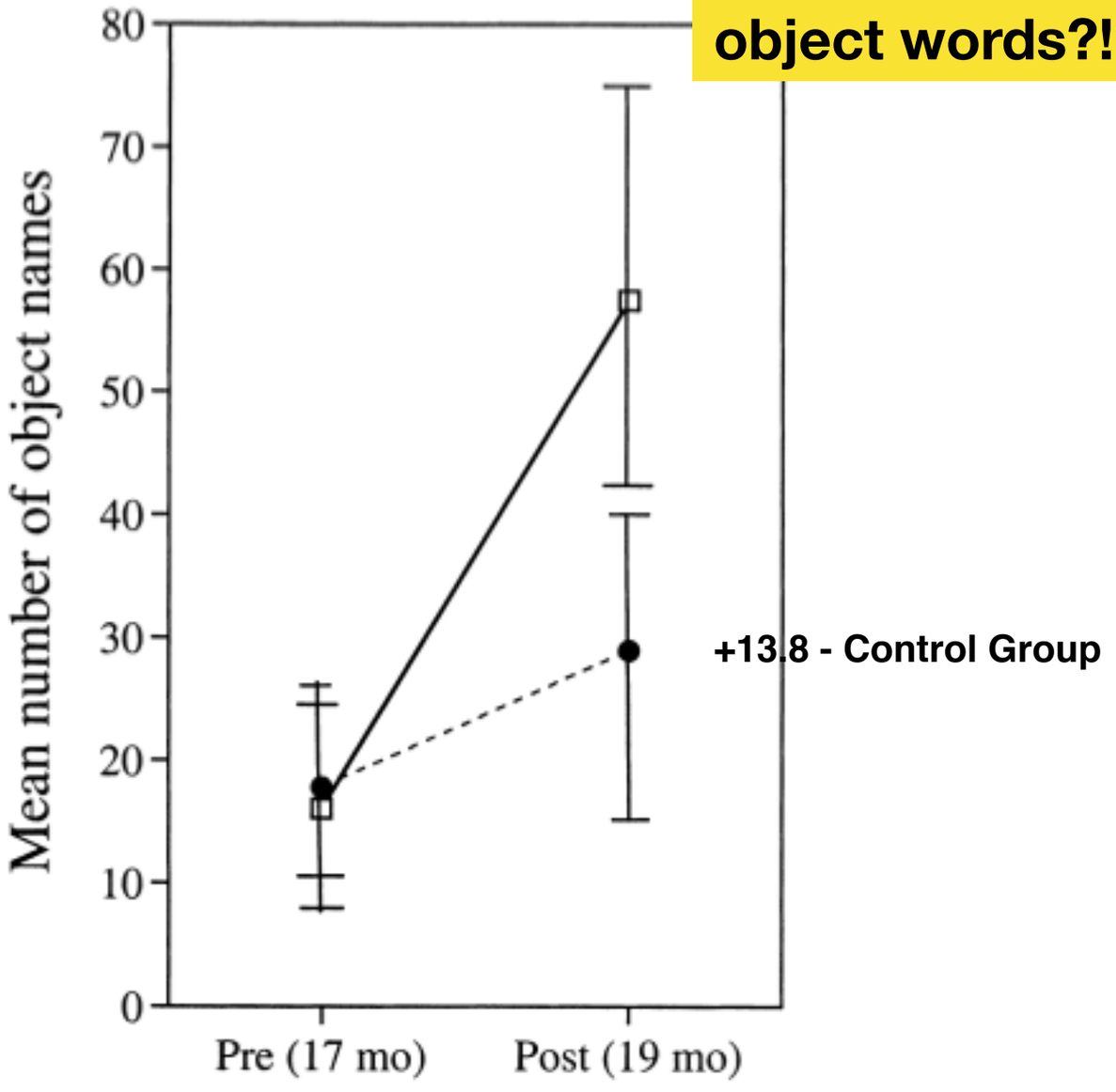


At week 9, the shape bias was tested
- training group: 70% correct
- no-contact control: 34% correct (ns)

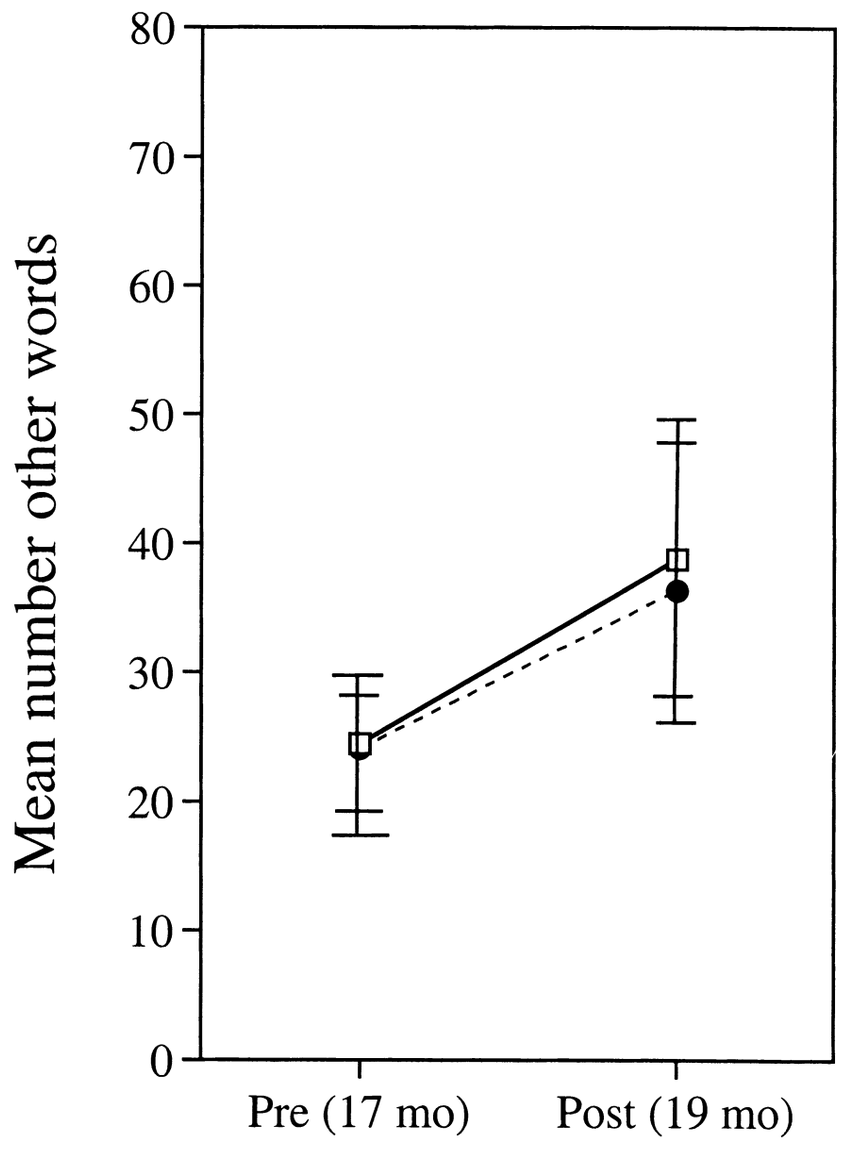
Shape bias training affects real word vocabulary learning

Result : teaching children names for only four artificial categories, each well organized by shape, accelerates object name learning *outside the laboratory*

Learned 3x more object words?!?!



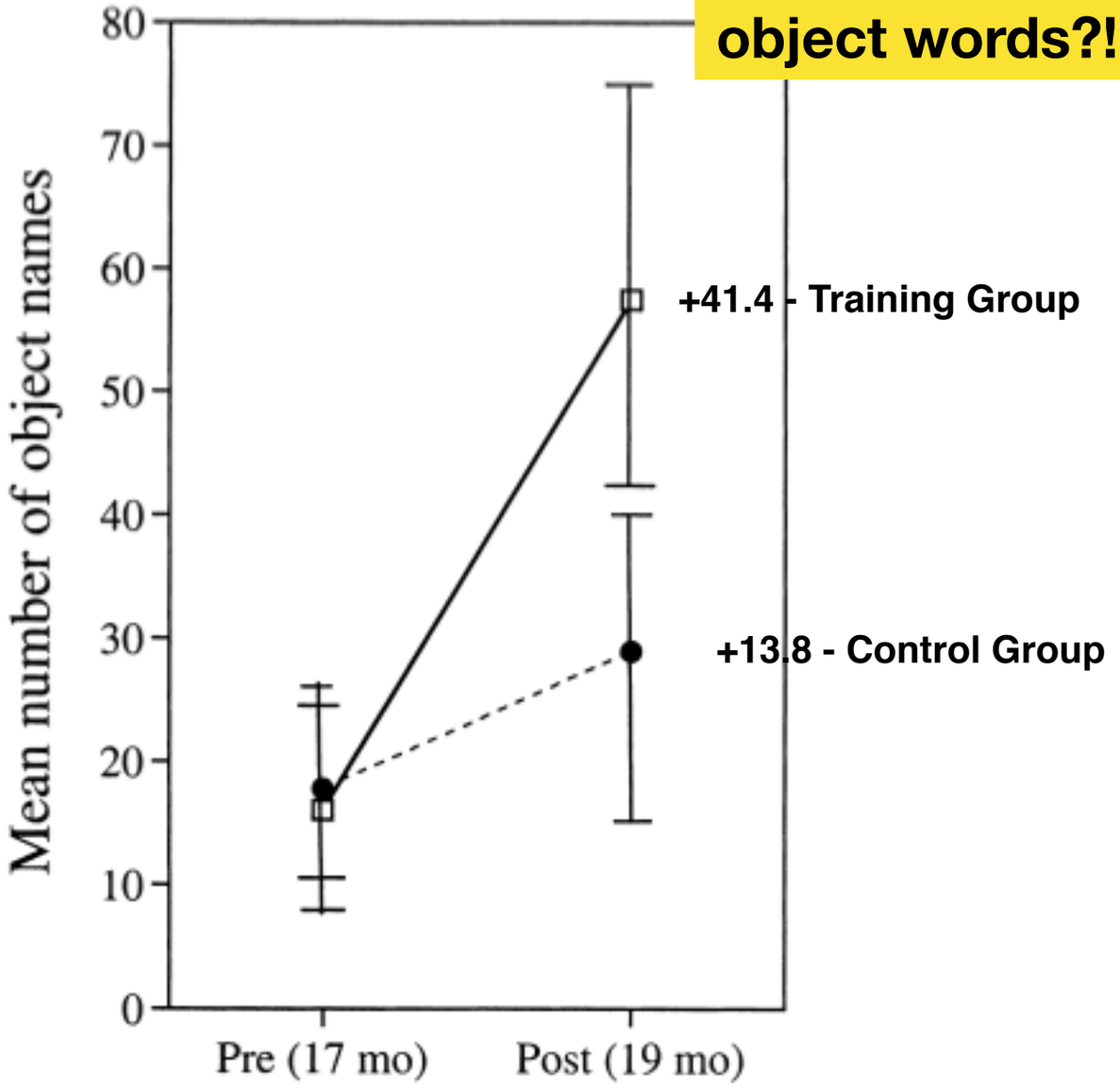
+13.8 - Control Group

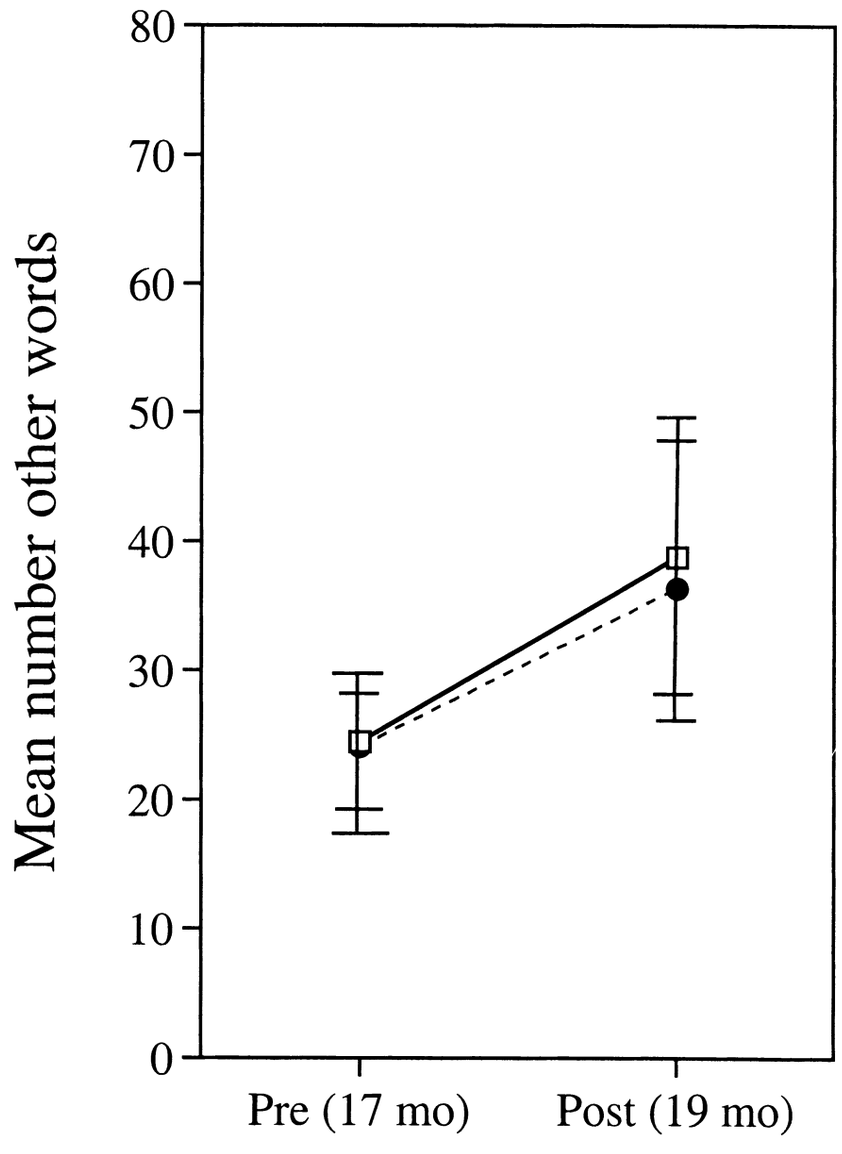


Shape bias training affects real word vocabulary learning

Result : teaching children names for only four artificial categories, each well organized by shape, accelerates object name learning *outside the laboratory*

Learned 3x more object words?!?!



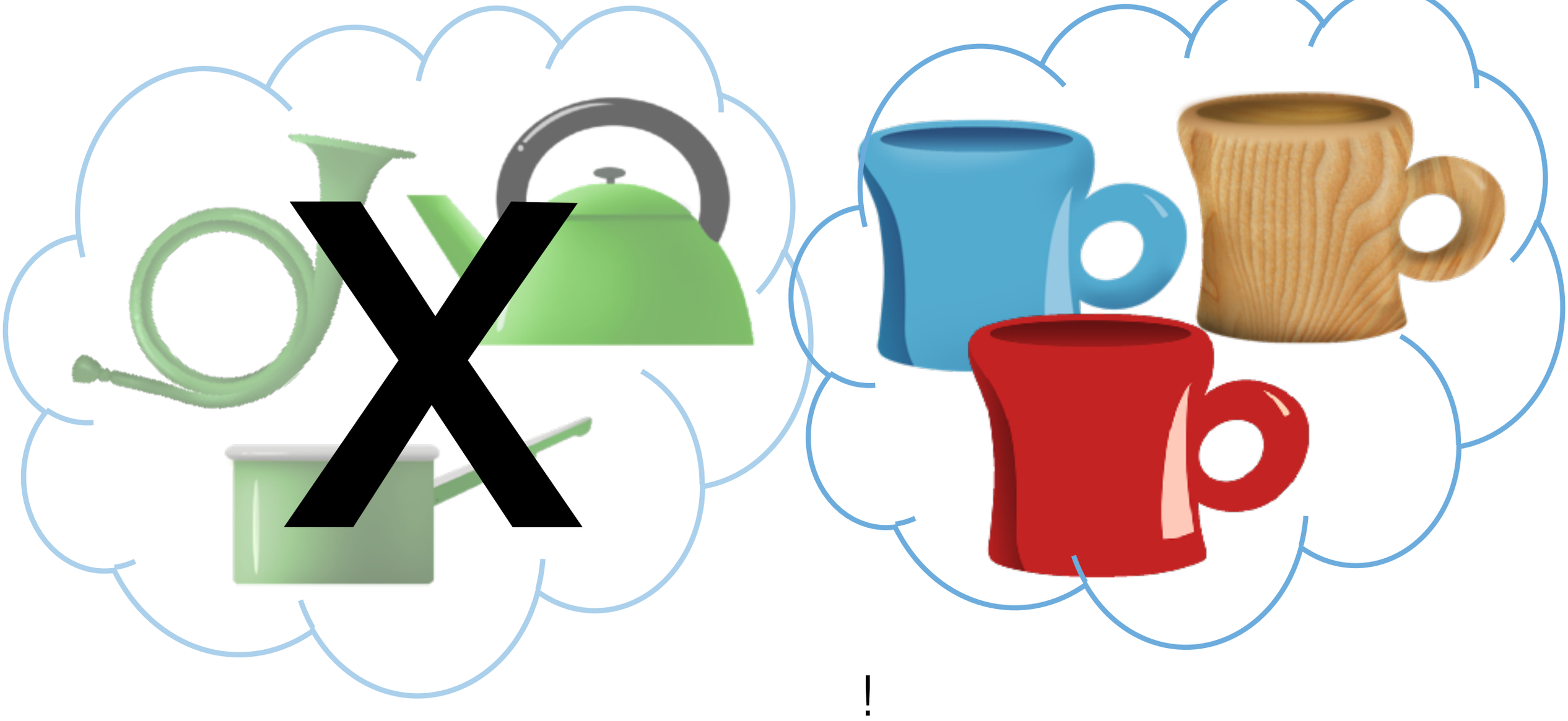



It’s a cup!

It’s a cup!

??

It’s a cup!



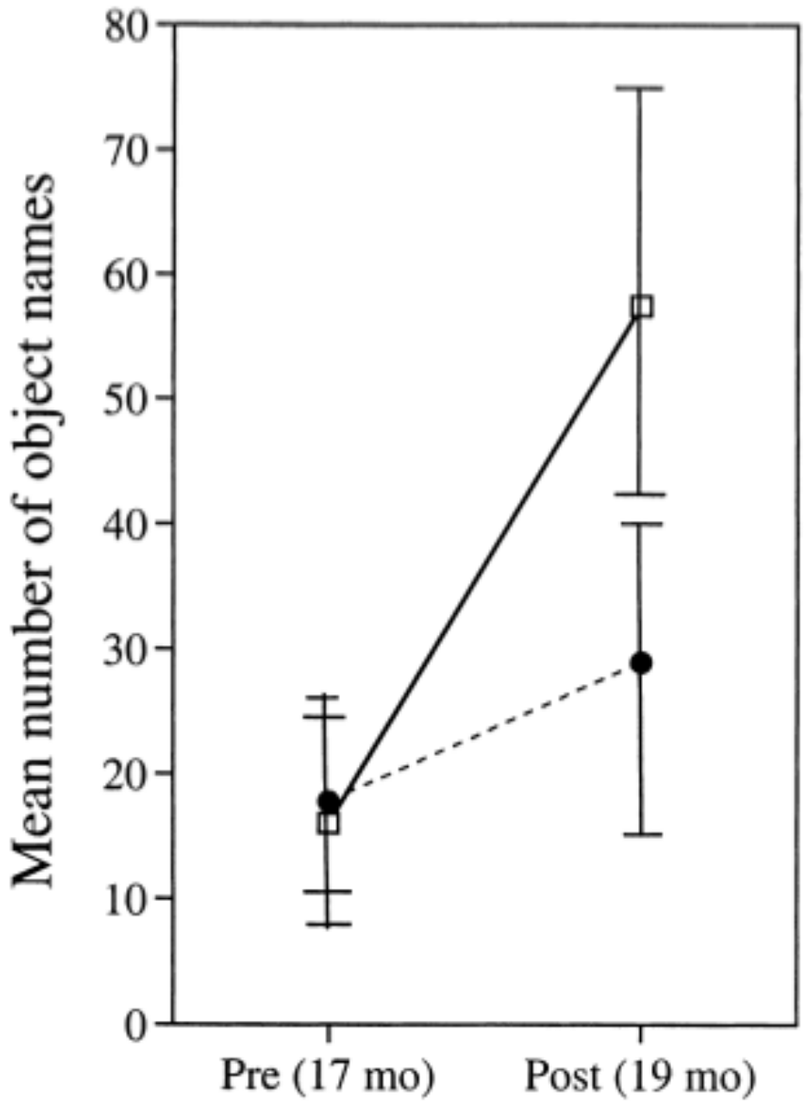
!

It’s a cup!

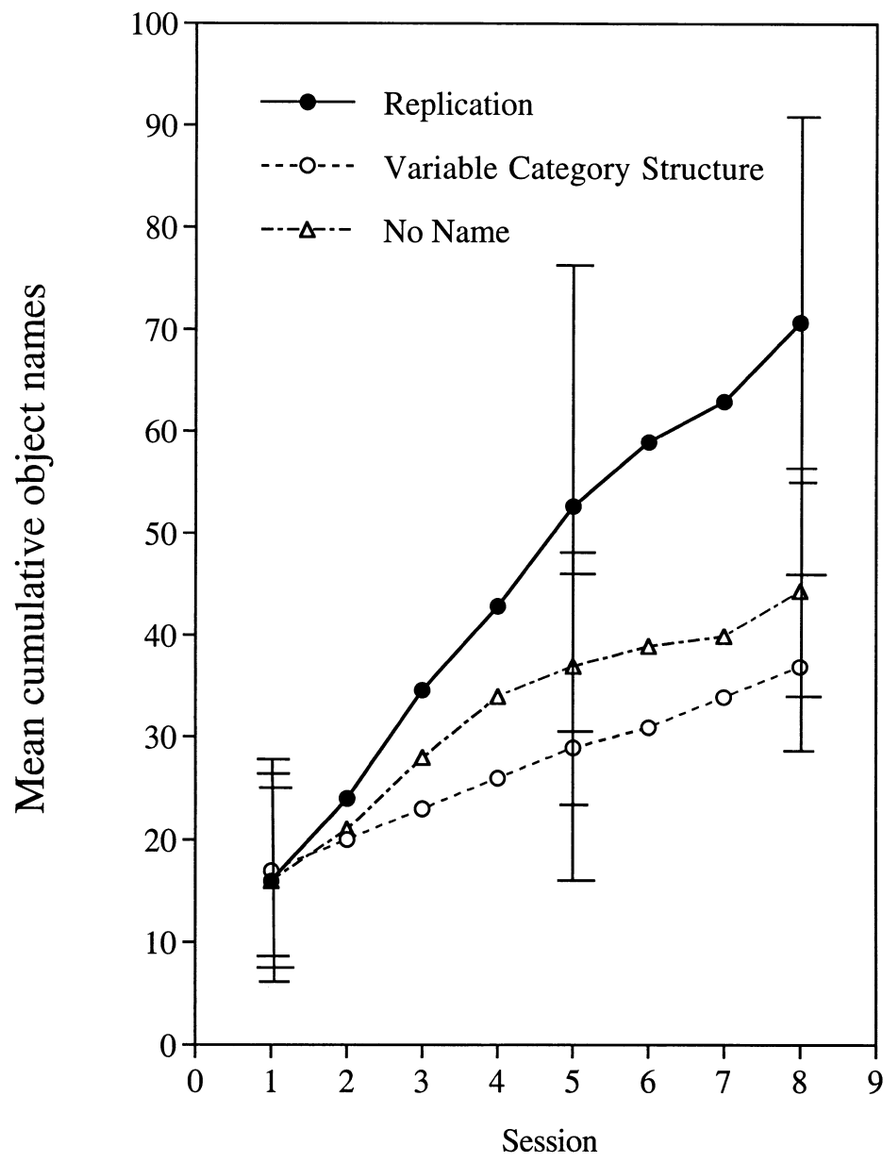
Shape bias training affects real word vocabulary learning

Result : teaching children names for only four artificial categories, each well organized by shape, accelerates word learning outside the laboratory

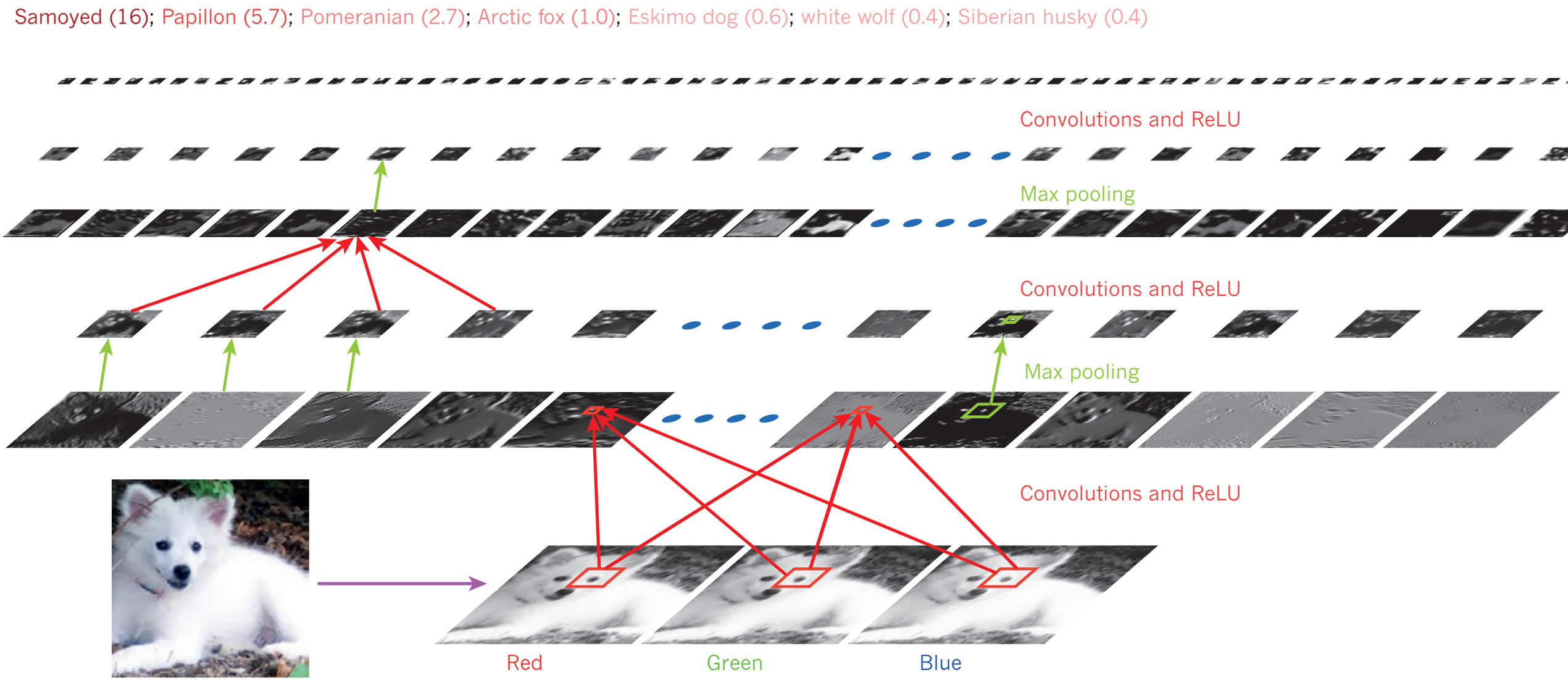
Study 1



Study 2 (replication)



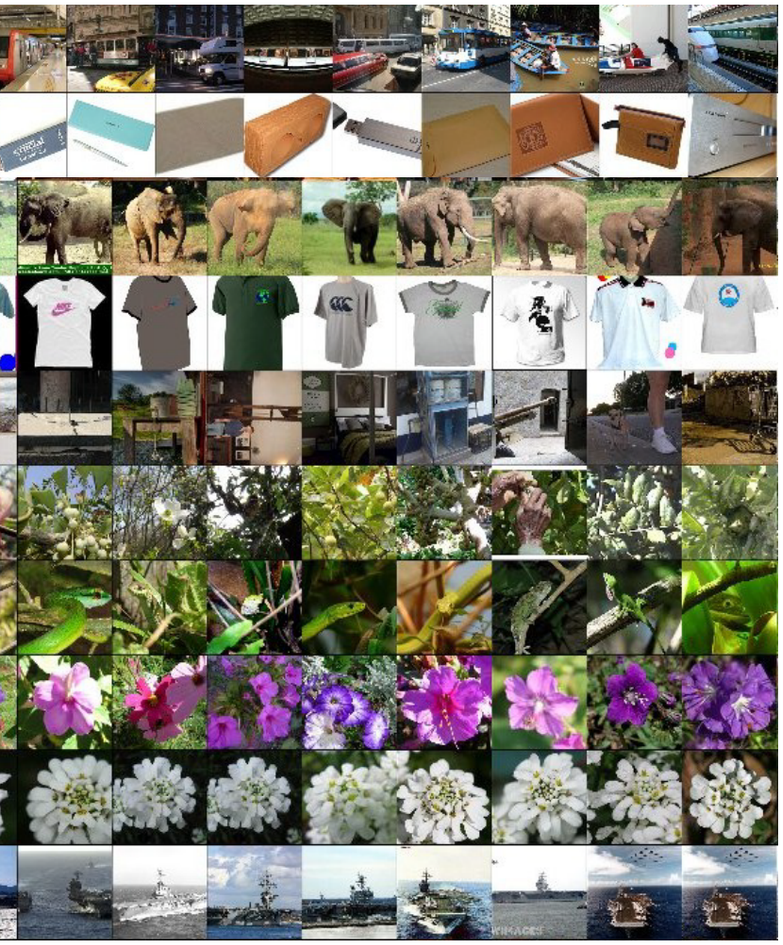
Does the shape bias emerge from large-scale category learning? Evaluation with a deep convolutional neural network



From LeCun, Bengio, & Hinton (2015).

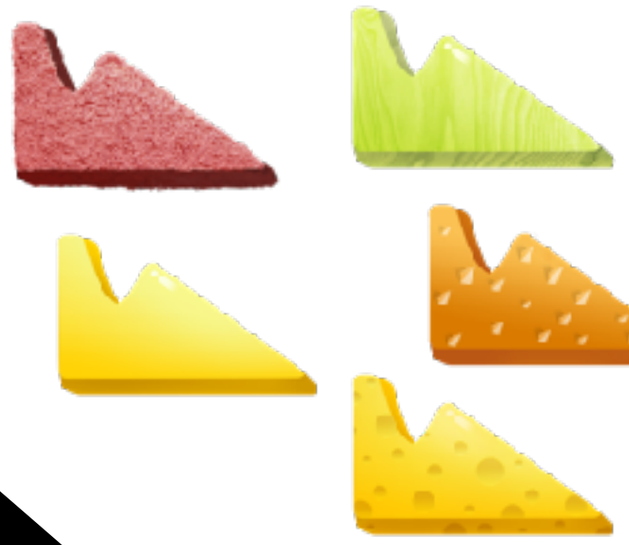
Does the shape bias emerge from large-scale category learning? Evaluation with a deep convolutional neural network

large-scale, imperfect prior experience (1.2 million images)



novel categories

“dax”



“wif"

“zup"

“lug”

Shape bias could be emergent property of large-scale category learning

Here is a "dax"

Slide courtesy of Subhankar Ghosh

Where is the other "dax"?

1	2	3
		
color match	shape match	texture match
convnet choices: **6%**	**84%**	**10%**

Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study

Samuel Ritter [*1] **David G.T. Barrett** [*1] **Adam Santoro** [1] **Matt M. Botvinick** [1]

Abstract

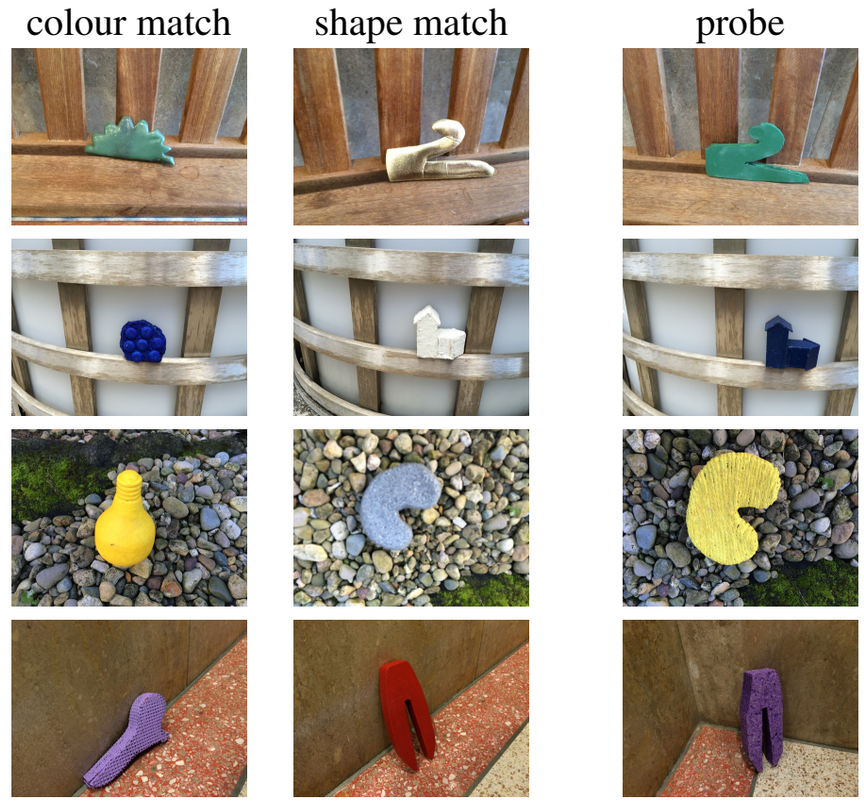
Deep neural networks (DNNs) have achieved unprecedented performance on a wide range of complex tasks, rapidly outpacing our understanding of the nature of their solutions. This has caused a recent surge of interest in methods for rendering modern neural systems more interpretable. In this work, we propose to address the interpretability problem in modern DNNs using the rich history of problem descriptions, theories and experimental methods developed by cognitive psychologists to study the human mind. To explore the potential value of these tools, we chose a well-established analysis from developmental psychology that explains how children learn word labels for objects, and applied that analysis to DNNs. Using datasets of stimuli inspired by the original cognitive psychology experiments, we find that state-of-the-art one shot learning models trained on ImageNet exhibit a similar bias to that observed in humans: they prefer to categorize objects according to shape rather than color. The magnitude of this shape bias varies greatly among archi-

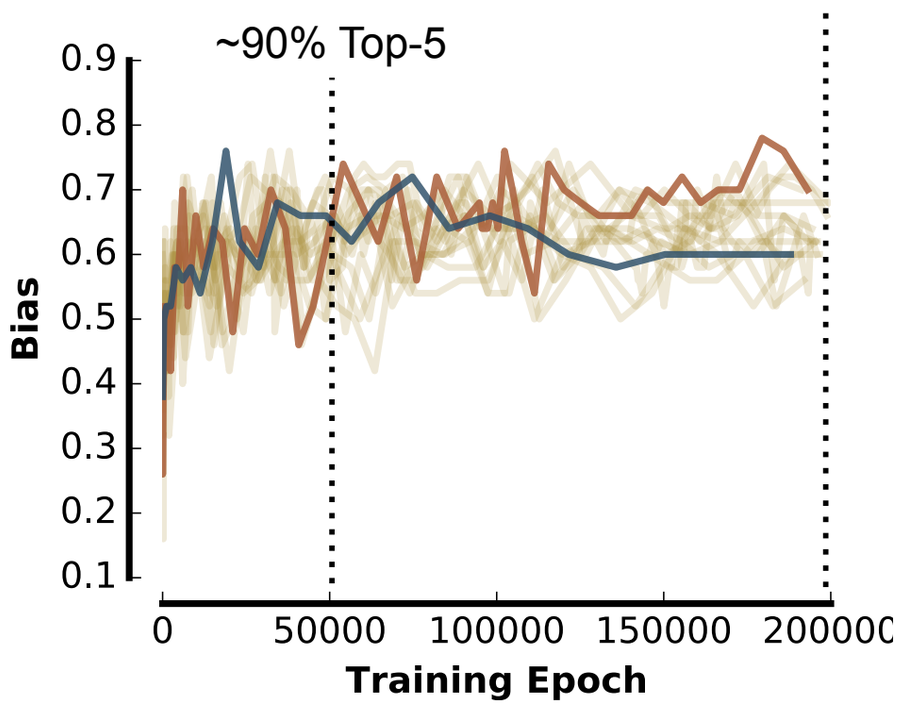
1. Introduction

During the last half-decade deep learning has significantly improved performance on a variety of tasks (for a review, see LeCun et al. (2015)). However, deep neural network (DNN) solutions remain poorly understood, leaving many to think of these models as black boxes, and to question whether they can be understood at all (Bornstein, 2016; Lipton, 2016). This opacity obstructs both basic research seeking to improve these models, and applications of these models to real world problems (Caruana et al., 2015).

Recent pushes have aimed to better understand DNNs: tailor-made loss functions and architectures produce more interpretable features (Higgins et al., 2016; Raposo et al., 2017) while output-behavior analyses unveil previously opaque operations of these networks (Karpathy et al., 2015). Parallel to this work, neuroscience-inspired methods such as activation visualization (Li et al., 2015), ablation analysis (Zeiler & Fergus, 2014) and activation maximization (Yosinski et al., 2015) have also been applied.

Altogether, this line of research developed a set of promising tools for understanding DNNs, each paper producing a glimmer of insight. Here, we propose another tool for the kit, leveraging methods inspired not by neuroscience, but instead by psychology. Cognitive psychologists have long wrestled with the problem of understanding another





Learning Inductive Biases with Simple Neural Networks

Reuben Feinman (reuben.feinman@nyu.edu)
Center for Neural Science
New York University

Brenden M. Lake (brenden@nyu.edu)
Department of Psychology and Center for Data Science
New York University

Abstract

People use rich prior knowledge about the world in order to efficiently learn new concepts. These priors–also known as "inductive biases"–pertain to the space of internal models considered by a learner, and they help the learner make inferences that go beyond the observed data. A recent study found that deep neural networks optimized for object recognition develop the shape bias (Ritter et al., 2017), an inductive bias possessed by children that plays an important role in early word learning. However, these networks use unrealistically large quantities of training data, and the conditions required for these biases to develop are not well understood. Moreover, it is unclear how the learning dynamics of these networks relate to developmental processes in childhood. We investigate the development and influence of the shape bias in neural networks using controlled datasets of abstract patterns and synthetic images, allowing us to systematically vary the quantity and form of the experience provided to the learning algorithms. We find that simple neural networks develop a shape bias after seeing as few as 3 examples of 4 object categories. The development of these biases predicts the onset of vocabulary acceleration in our networks, consistent with the developmental process in children.

Keywords: neural networks; inductive biases; learning-to-learn; word learning

Humans possess the remarkable ability to learn a new concept from seeing just a few examples. A child can learn the meaning of a new word such as "fork" after observing only one or a handful of different forks (Bloom, 2000). In con-

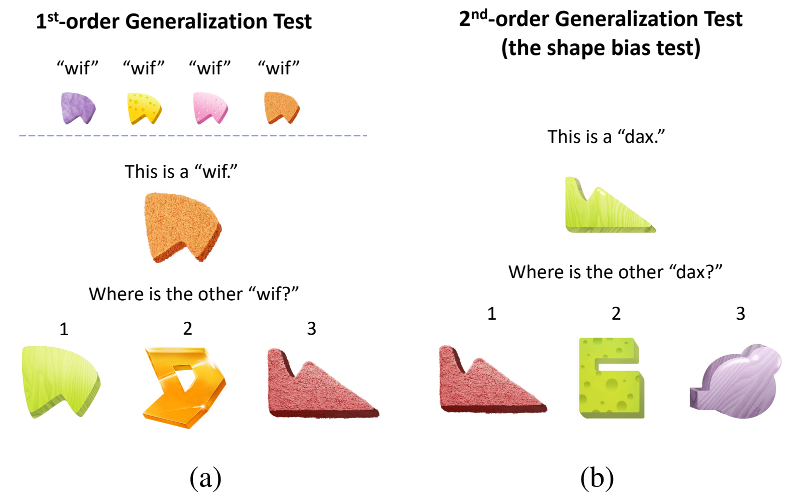
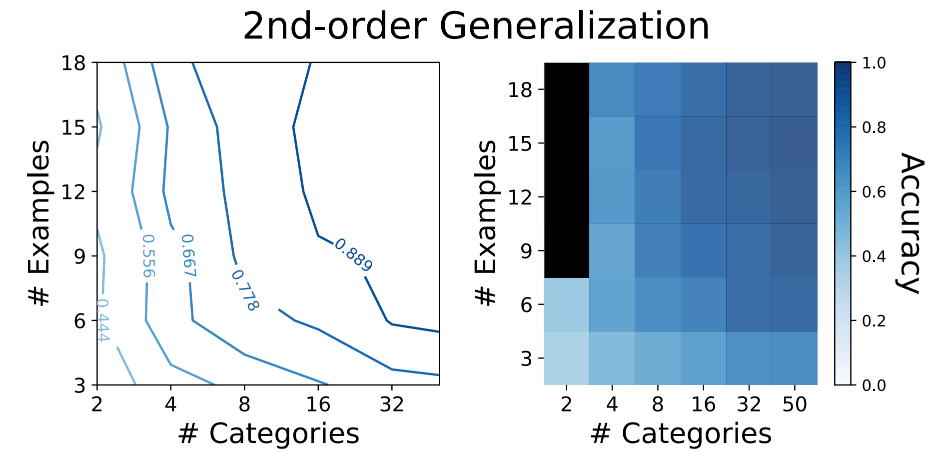


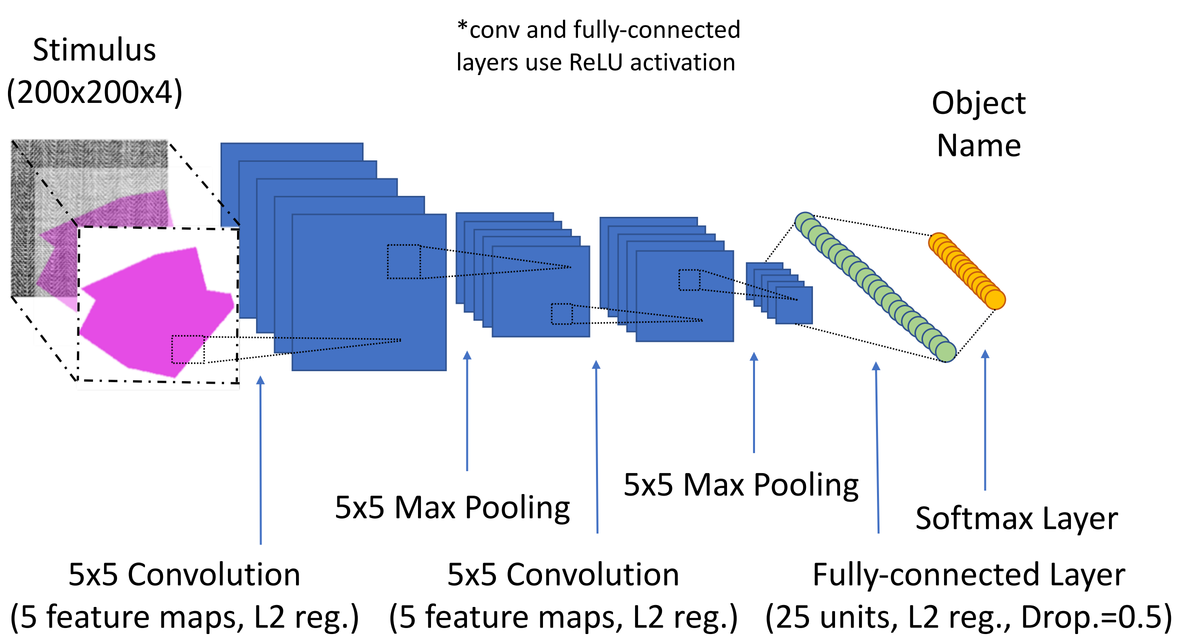
Figure 1: Shape bias generalization tests. The 1st-order test, shown in (a), assesses if a child has learned to generalize a familiar object name to a novel exemplar according to shape. This is the first step of shape bias development. The 2nd-order test, shown in (b), assesses if the child has learned to generalize a novel name to a novel exemplar by shape, the second and final step of shape bias development.

not always clear, results show that children, adults and primates can "learn-to-learn" or form higher-order generalizations that improve the efficiency of future learning (Harlow, 1949; Smith et al., 2002; Dewar & Xu, 2010).

Researchers have proposed a number of computational



simple convolutional net can learn a shape bias with as few as 6 examples of 8 object categories



Count noun framing

Here is a “zif”.

Which is the other zif?

A

B

C

Mass noun framing (e.g., Roger Brown, 1957)

Here is some “zif”.

Can you find some more zif?

A

B

C

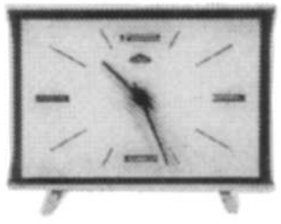
Taxonomic vs. shape bias: Cimpian & Markman (2005)

- 3-5 year olds
- triads with pitting shape bias vs. taxonomic bias

"See this? This is a dax”

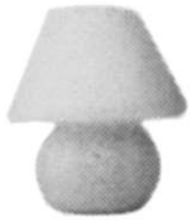
“find another dax”

shape choice (28%)



taxonomic choice (72%)

"See this? This is a dax”

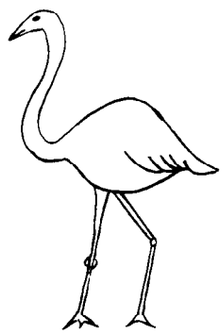


“find another dax”

- Cimpian and Markman : pre-school children learn “words as kinds”, not simply as classes of shapes (although shape can be a cue to kind)
- Linda Smith: “3-5 yo already know these words, and it's not genuine word learning. For real word learning, attention to shape matters”

Essentialism in categorization and category-based induction (Gelman & Markman, 1986)

Provided



"This bird's heart has a right aortic arch only"

"This bat's heart has a left aortic arch only"

Query

"What does this bird's heart have?"

Results: 4 year olds generalize based on category membership ~68% of time, overriding a distractor chosen for strong perceptual similarity

Review: Keil's (1989) transformation study of essentialism

Participants were kids in grades K, 2, and 4 (ages approx. 5, 7, and 10 years old)

Examples of two descriptions used in the first transformations study

Natural kind: Raccoon/skunk

The doctors took a raccoon (show picture of raccoon) and shaved away some of its fur. They dyed what was left all black. Then they bleached a single stripe all white down the center of its back. Then, with surgery (explained to child in preamble), they put in its body a sac of super smelly odor, just like a skunk has (with younger children "odor" was replaced with "super smelly yucky stuff"). When they were all done, the animal looked like this (show picture of skunk). After the operation was this a skunk or a raccoon? (Both pictures were present at the time of the final question.)

Artifact: Coffeepot/birdfeeder

The doctors took a coffeepot that looked like this (show picture of coffeepot). They sawed off the handle, sealed the top, took off the top knob, sealed closed the spout, and sawed it off. They also sawed off the base and attached a flat piece of metal. They attached a little stick, cut a window in it, and filled the metal container with birdfood. When they were done, it looked like this (show picture of birdfeeder). After the operation was this a coffeepot or a birdfeeder? (Both pictures were present at the time of the final question.)

Keil's (1989) transformation study of essentialism and the "perceptual to conceptual shift"

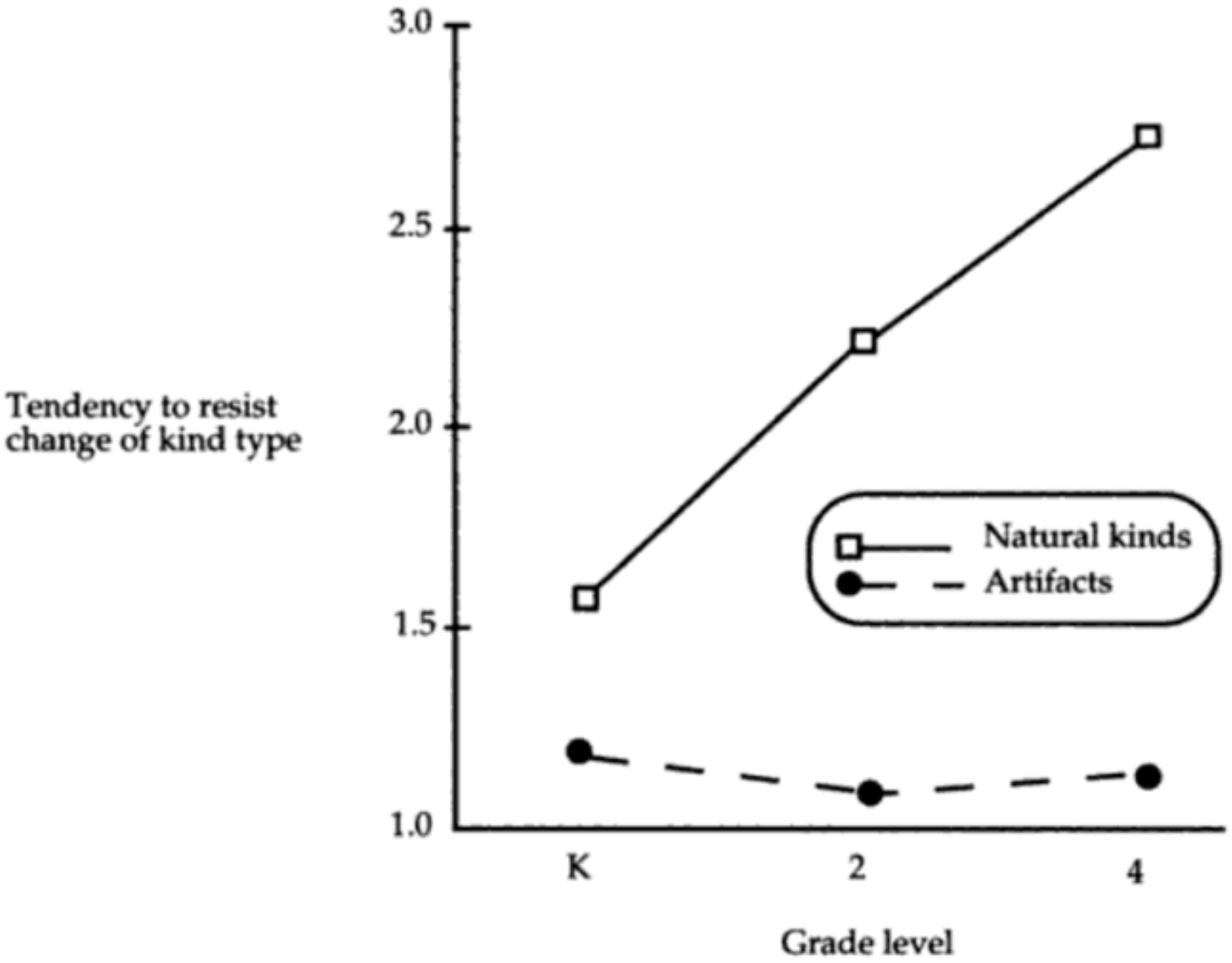


Figure 9.1
Mean scores for natural kind and artifact terms in the first transformations study. 1 = judgment that operation changed kind type, 2 = judgment indicating indecision as to whether operation changed kind type, 3 = judgment that operation did not change kind type.

Conclusion

Word learning as a window into conceptual development

- Word learning one of the most heavily researched and controversial topics in cognitive development
  - ✳ literature often does not make distinction between word learning and concept learning

Children learn new concepts very quickly, aided by key biases and constraints:

- basic level bias
- taxonomic bias
- whole object bias
- mutual exclusivity bias
- shape bias

Perceptually or conceptually driven?
The degree to which early word learning is driven by perceptual and attentional learning — versus reflecting kinds embedded in folks theories — is still an active debate