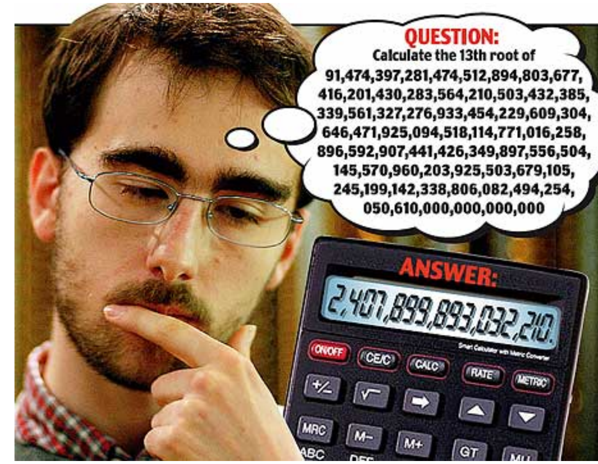
Constructing a hypothesis space from the Web

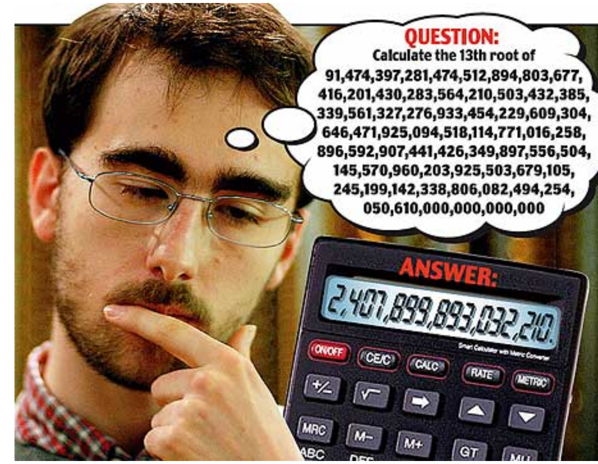
for large-scale Bayesian word learning

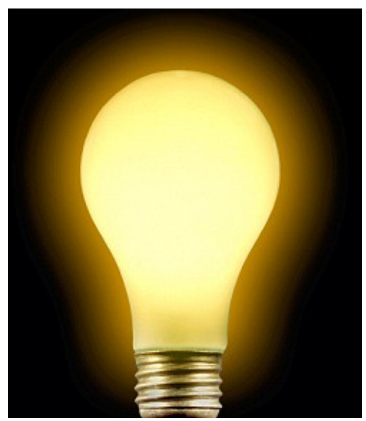
Joshua T. Abbott
Joseph L. Austerweil
Thomas L. Griffiths

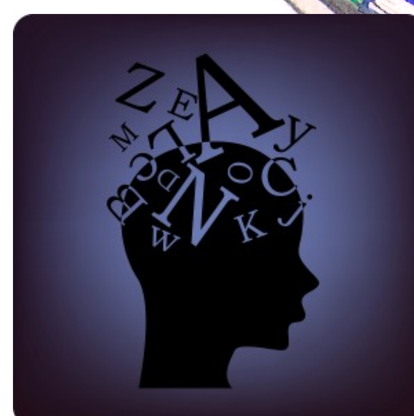

Garry Kasparov
VS
DEEP BLUE


QUESTION:
Calculate the 13th root of
91,474,397,281,474,512,894,803,677,
416,201,430,283,564,210,503,432,385,
339,561,327,276,933,454,229,609,304,
646,471,925,094,518,114,771,016,258,
896,592,907,441,426,349,897,556,504,
145,570,960,203,925,503,679,105,
245,199,142,338,806,082,494,254,
050,610,000,000,000,000
ANSWER:
2,407,899,893,032,210.
ON/OFF
CE/C
CALC
RATE
METRIC
MRC
M−
M+
GT

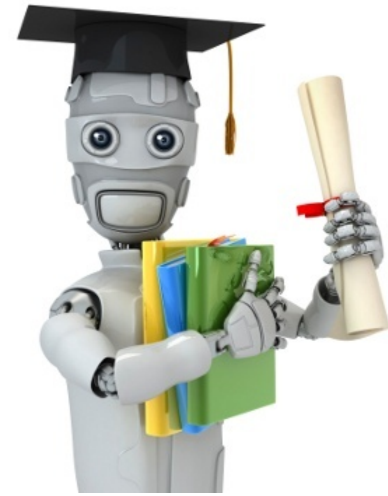

Garry Kasparov
VS


QUESTION:
Calculate the 13th root of
91,474,397,281,474,512,894,803,677,
416,201,430,283,564,210,503,432,385,
339,561,327,276,933,454,229,609,304,
646,471,925,094,518,114,771,016,258,
896,592,907,441,426,349,897,556,504,
145,570,960,203,925,503,679,105,
245,199,142,338,806,082,494,254,
050,610,000,000,000,000
ANSWER:
2,407,899,893,032,210.
ON/OFF
CE/C
CALC
RATE
METRIC
MRC
M–
M+
GT



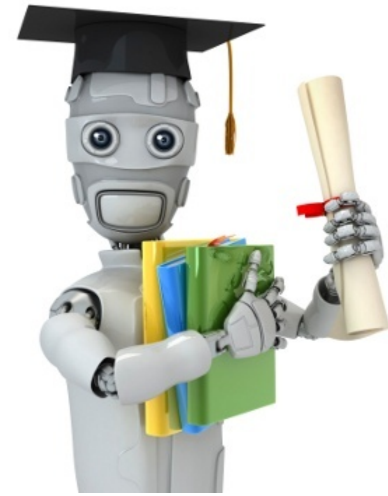



SPANISH
GERMAN
GRAMMER
HINDI
ITALIAN 1



Cognitive Science

Machine Learning



Cognitive Science

developing high-quality models of human cognition

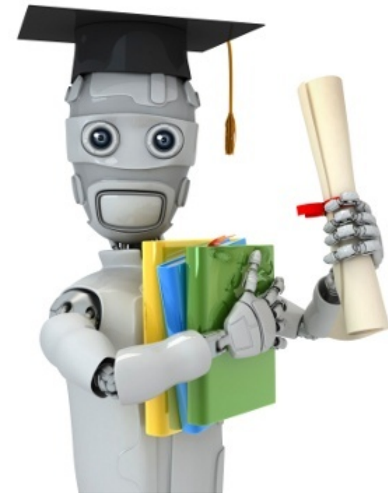
small-scale experiments with toy/artificial stimuli

Machine Learning

Cognitive Science

developing high-quality models of human cognition

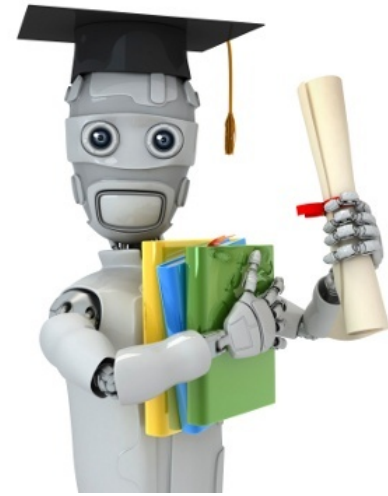
small-scale experiments with toy/artificial stimuli



Machine Learning

solving the problem - not exploring how people do it

large-scale experiments with online data sources



Cognitive Science

developing high-quality models of human cognition

small-scale experiments with toy/artificial stimuli

Machine Learning

solving the problem - not exploring how people do it

large-scale experiments with online data sources

Case study: Word learning

Xu & Tenenbaum (2007) showed that a handcrafted Bayesian model can explain word learning in 3 domains.

Case study: Word learning

Xu & Tenenbaum (2007) showed that a handcrafted Bayesian model can explain word learning in 3 domains.

2 issues we explore:

- Can we automatically construct a word learning model from online resources?
- Can we generalize to multiple domains?

Here is a DAK

(Xu & Tenenbaum, 2007)

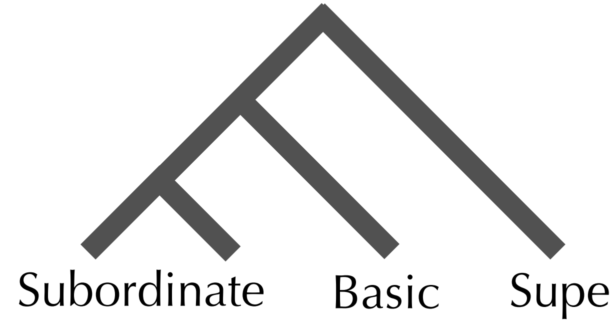
Here is a DAK

(Xu & Tenenbaum, 2007)

Can you help Mr. Frog find the other "DAKS"?

(Xu & Tenenbaum, 2007)

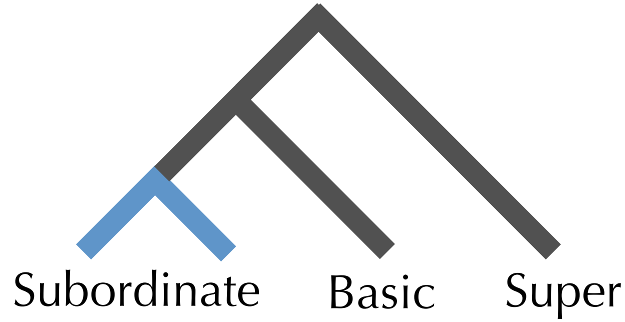
Here is a DAK



Can you help Mr. Frog find the other "DAKS"?

(Xu & Tenenbaum, 2007)

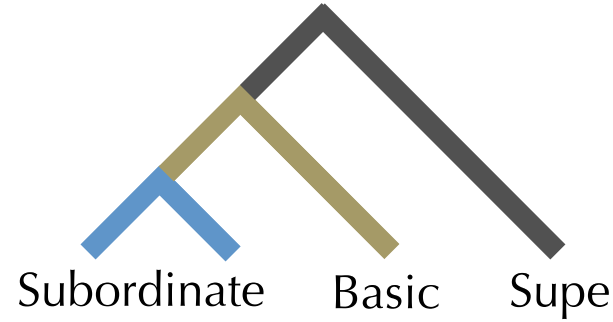
Here is a DAK



Can you help Mr. Frog find the other "DAKS"?

(Xu & Tenenbaum, 2007)

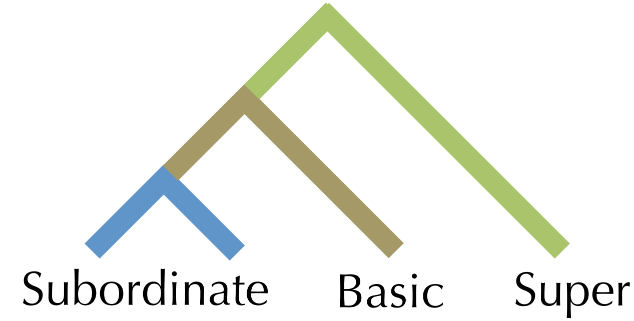
Here is a DAK



Can you help Mr. Frog find the other "DAKS"?

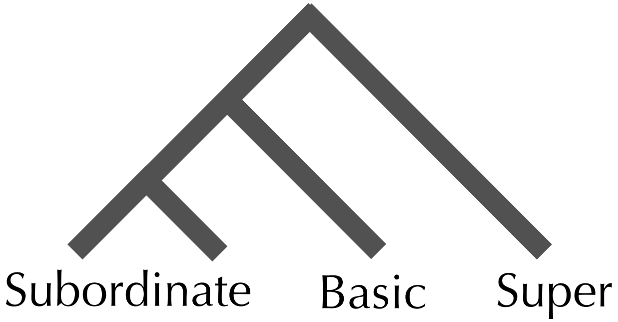
(Xu & Tenenbaum, 2007)

Here is a DAK

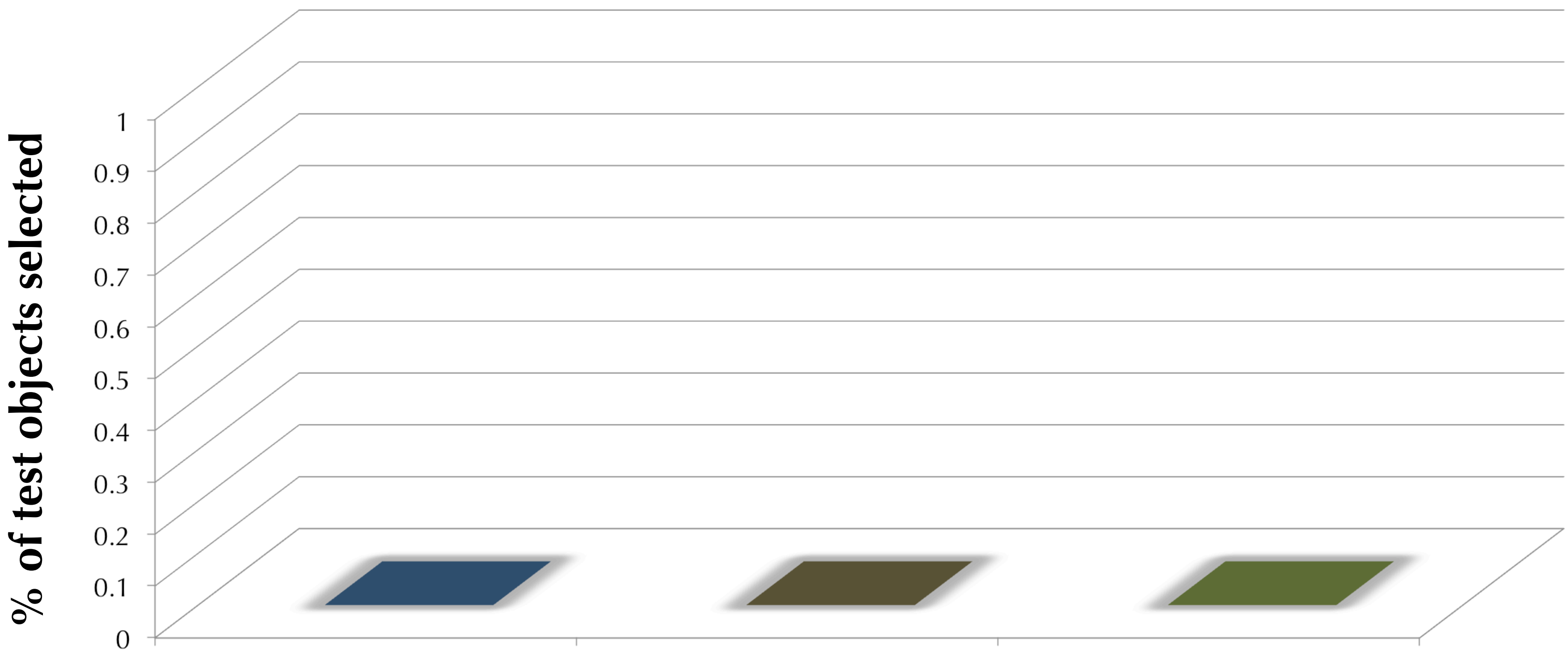


Can you help Mr. Frog find the other "DAKS"?

Here is a DAK

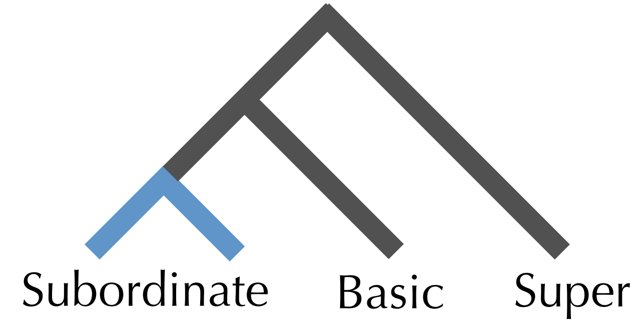


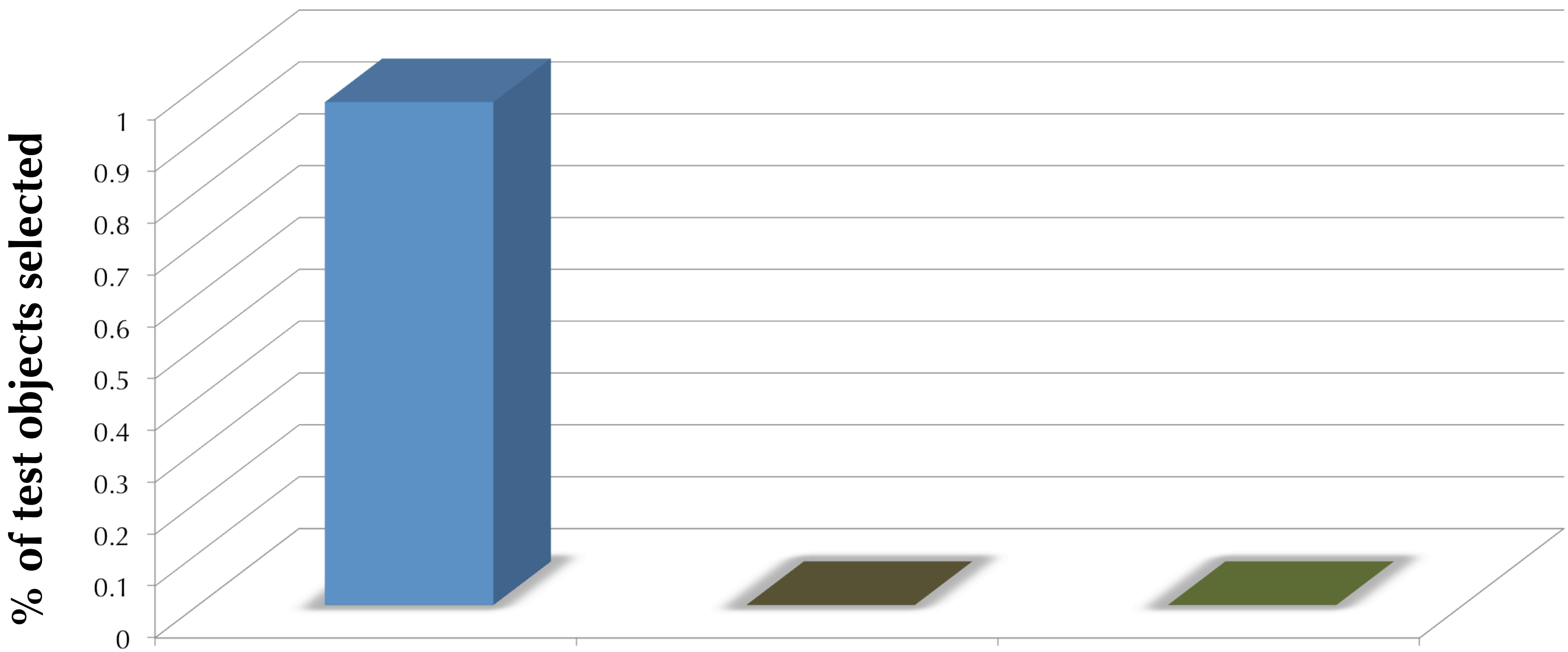
(Xu & Tenenbaum, 2007)



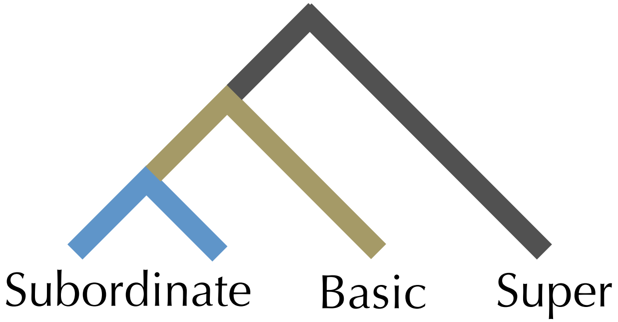
(Xu & Tenenbaum, 2007)

Here is a DAK

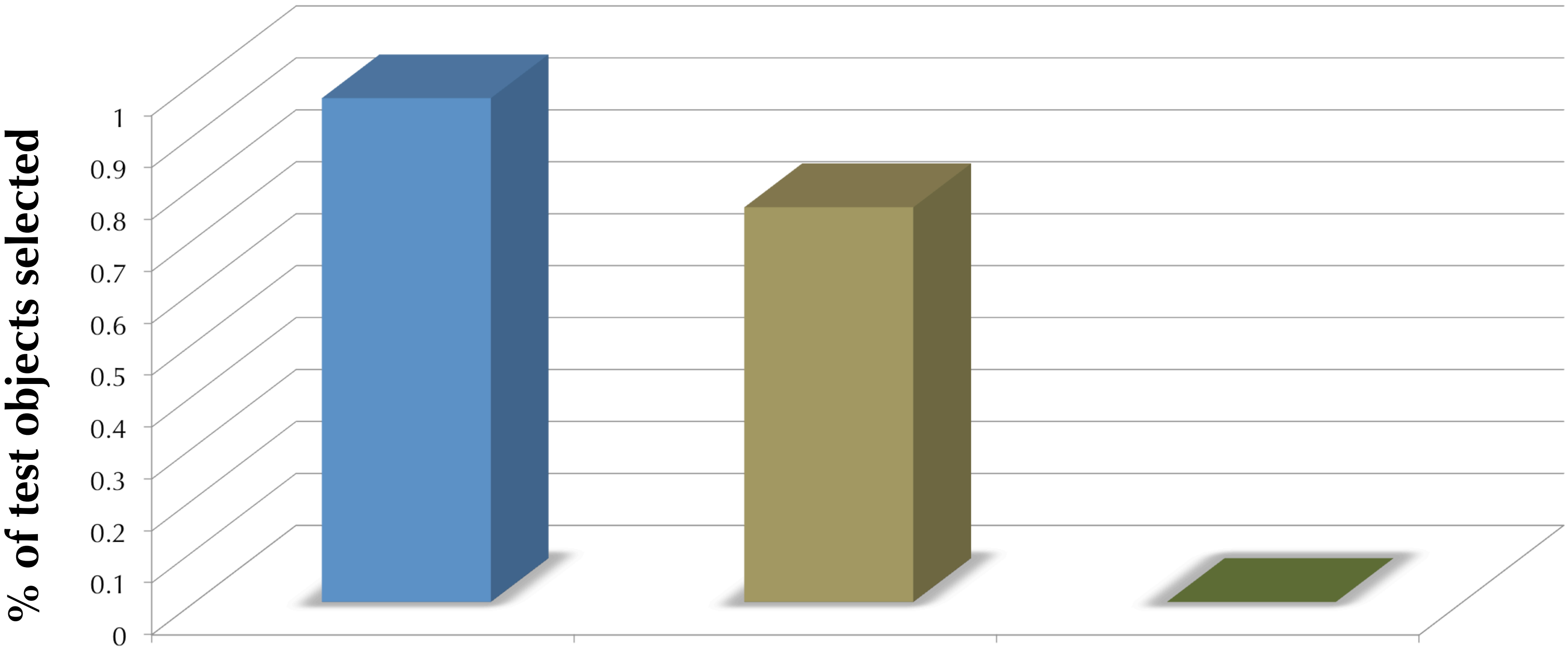




Here is a DAK

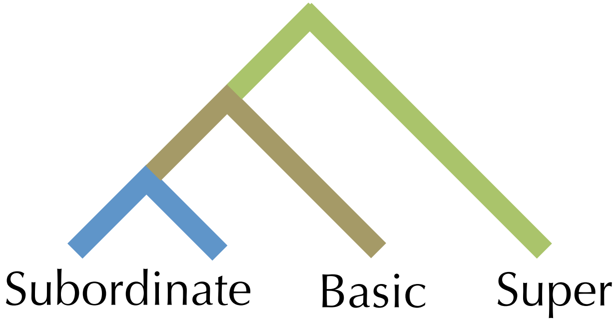


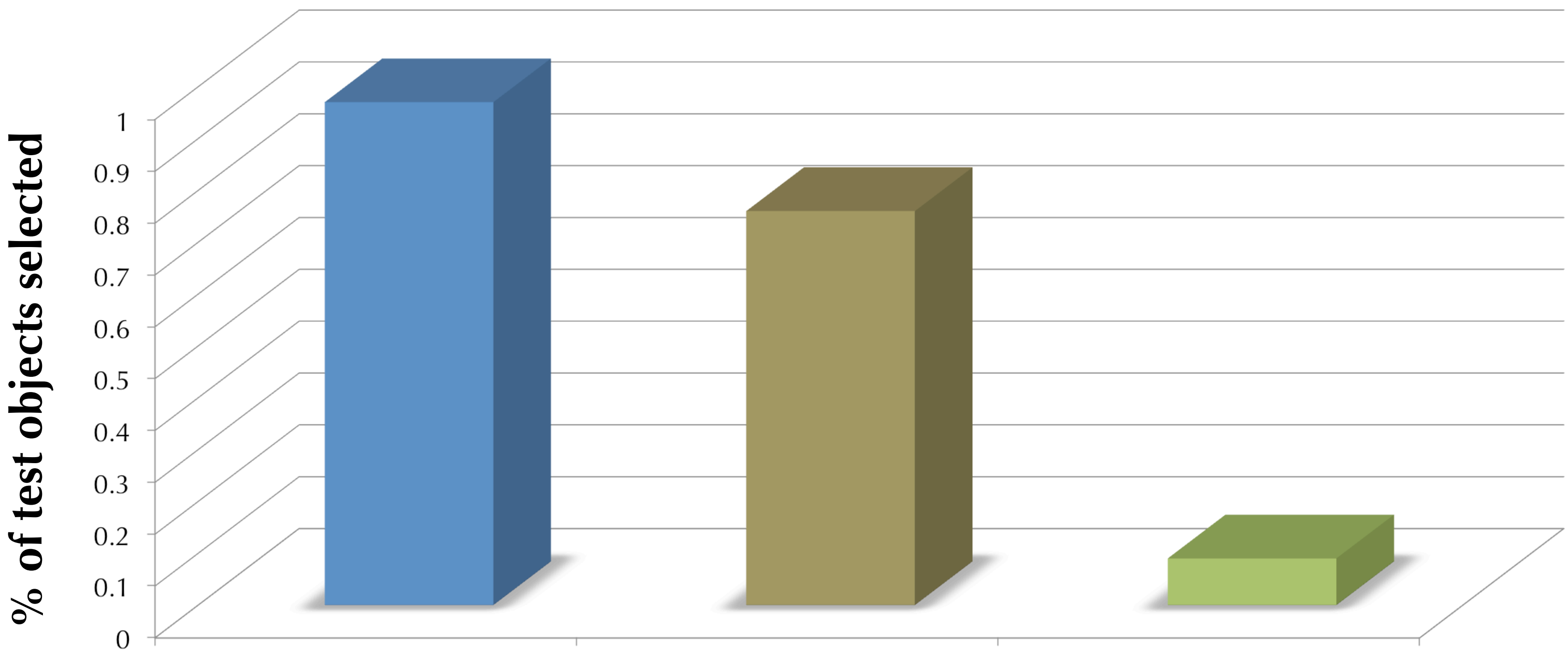
(Xu & Tenenbaum, 2007)



(Xu & Tenenbaum, 2007)

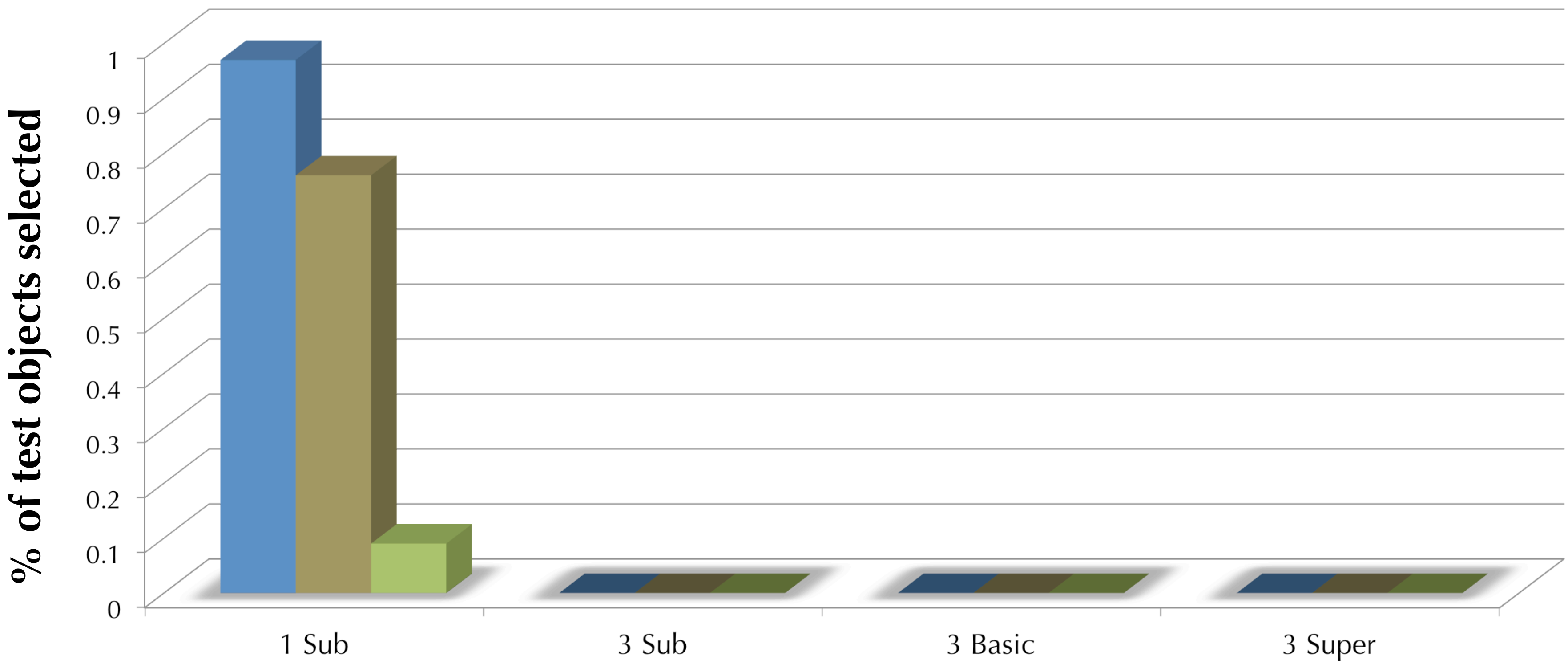
Here is a DAK





(Xu & Tenenbaum, 2007)

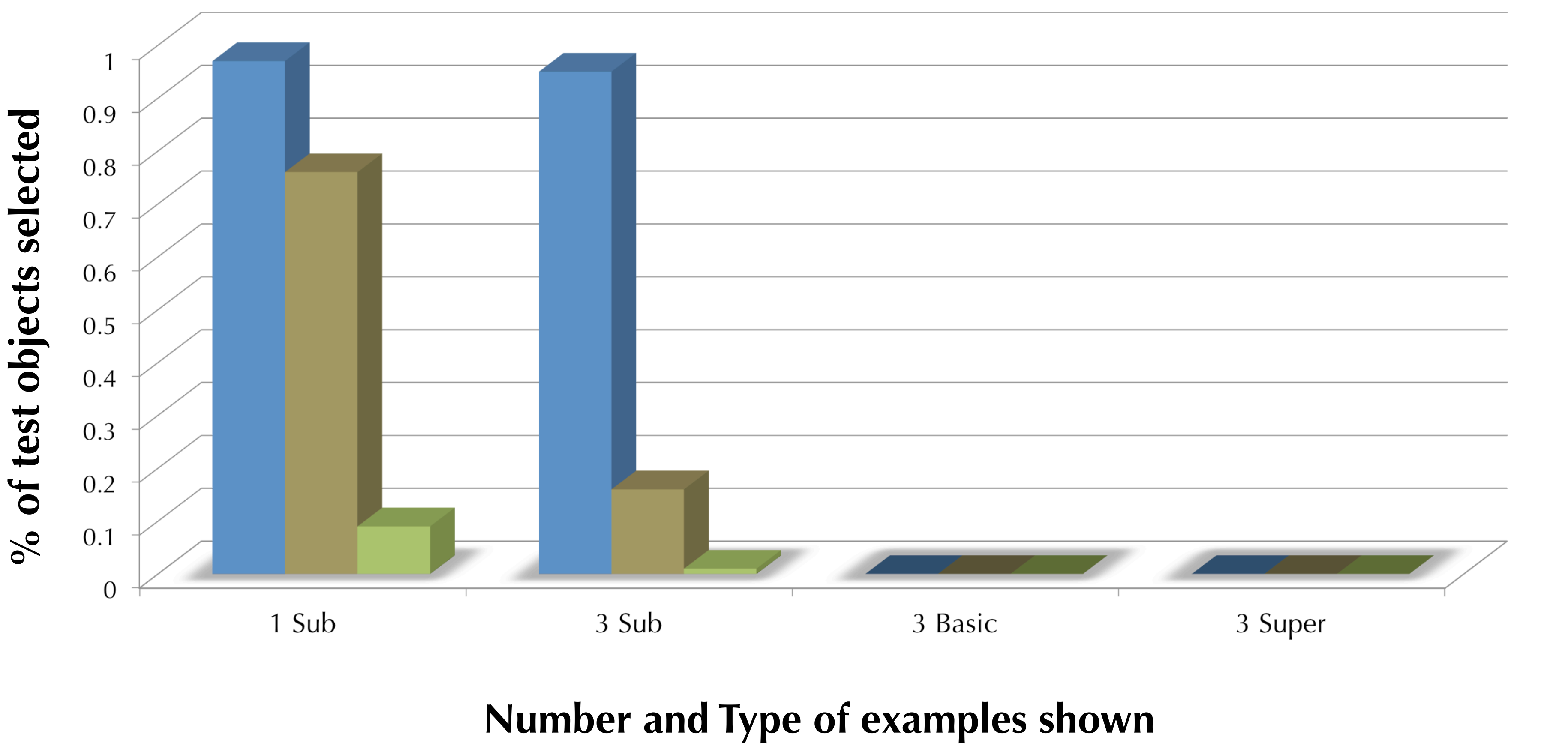
Here is a DAK



Number and Type of examples shown

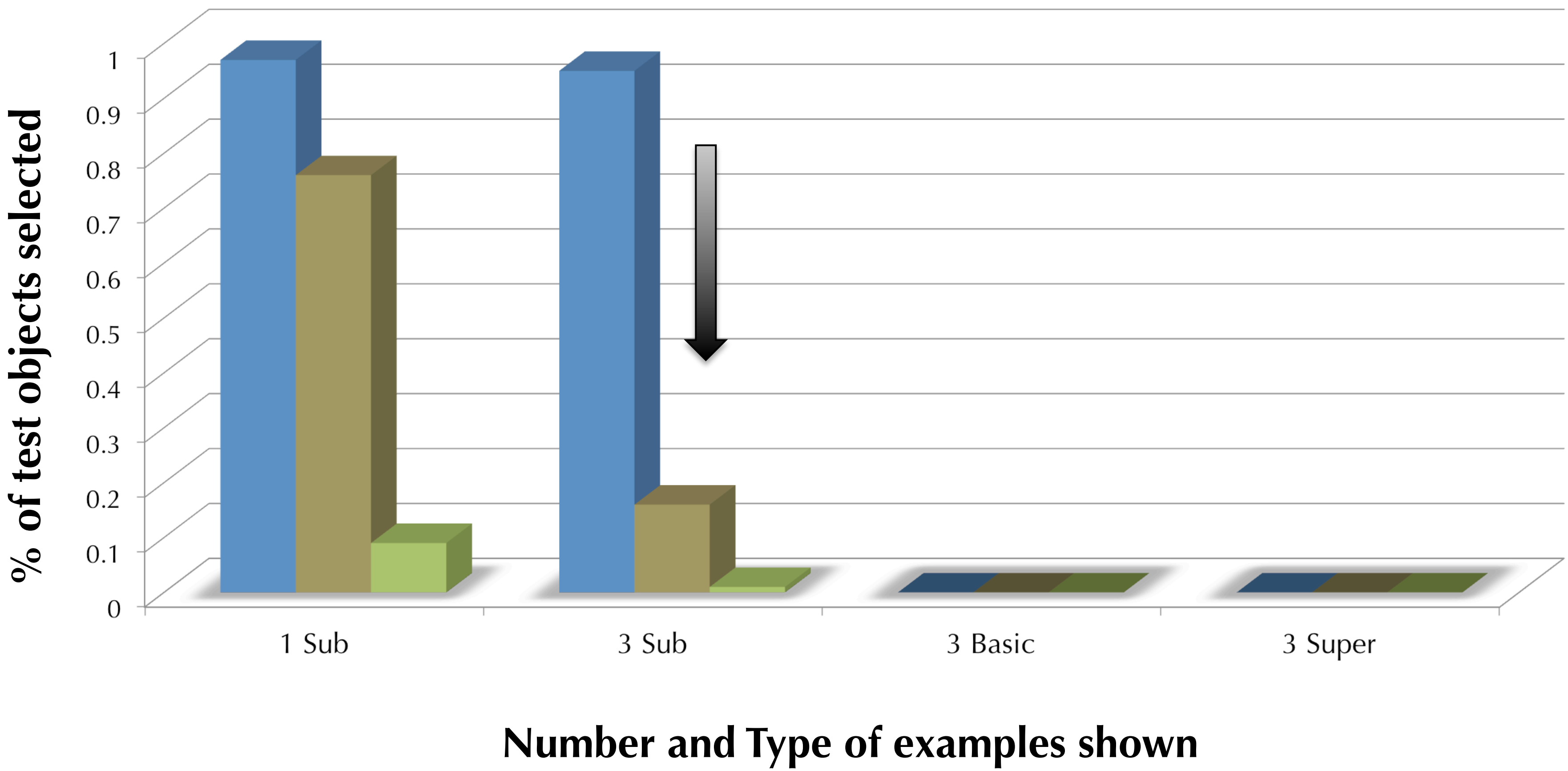
Here are three FEPS

(Xu & Tenenbaum, 2007)



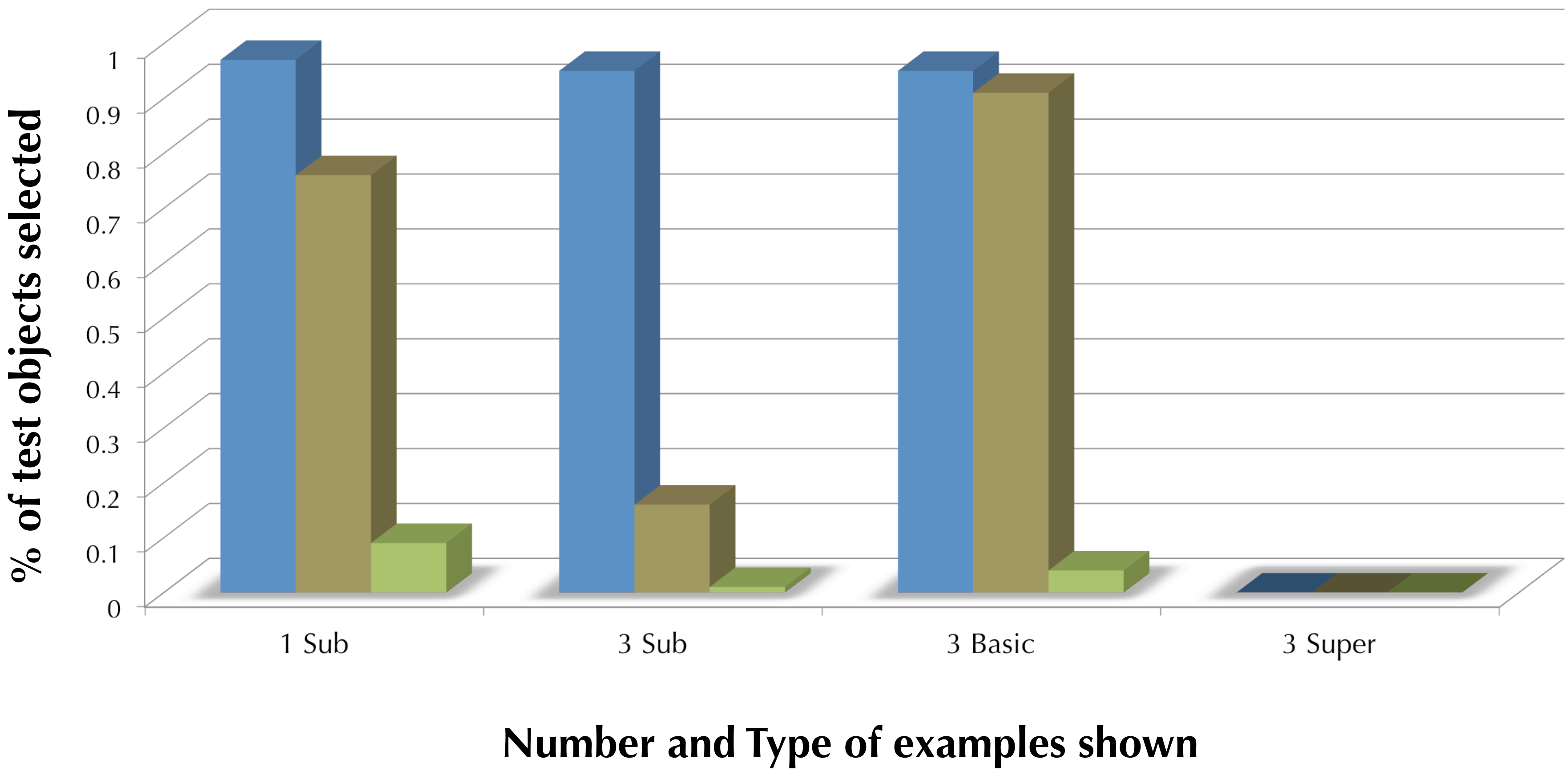
Here are three FEPS

(Xu & Tenenbaum, 2007)



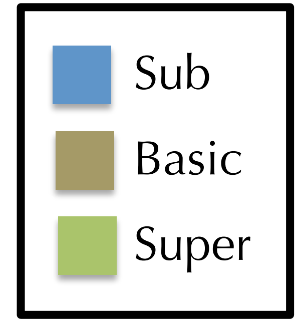
Here are three BLICKS

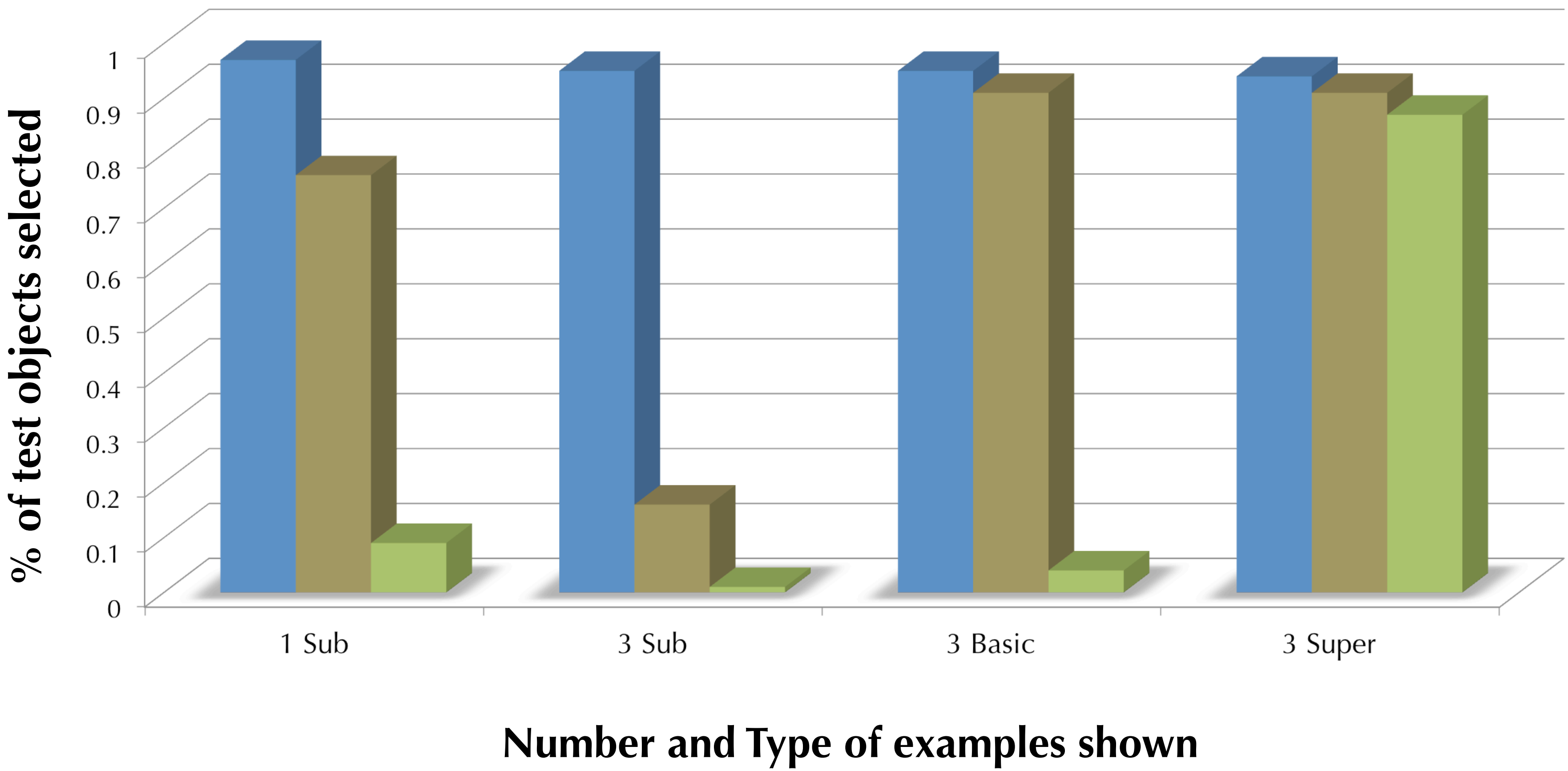
(Xu & Tenenbaum, 2007)



Here are three ZIVS

(Xu & Tenenbaum, 2007)





Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)

Here are three ZIVS

$X = \{$

x_1

x_2

x_3

$\}$

We want to compute:

Prob( is a ZIV | X)

Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)

$$P(h \mid X) \propto P(X \mid h)P(h)$$

h: hypothesis

X: data

Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)

Posterior

$$P(h \mid X) \propto P(X \mid h)P(h)$$

h: hypothesis

X: data

Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)

Posterior

Prior

$$P(h \mid X) \propto P(X \mid h)P(h)$$

h: hypothesis

X: data

Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)

Posterior Likelihood Prior

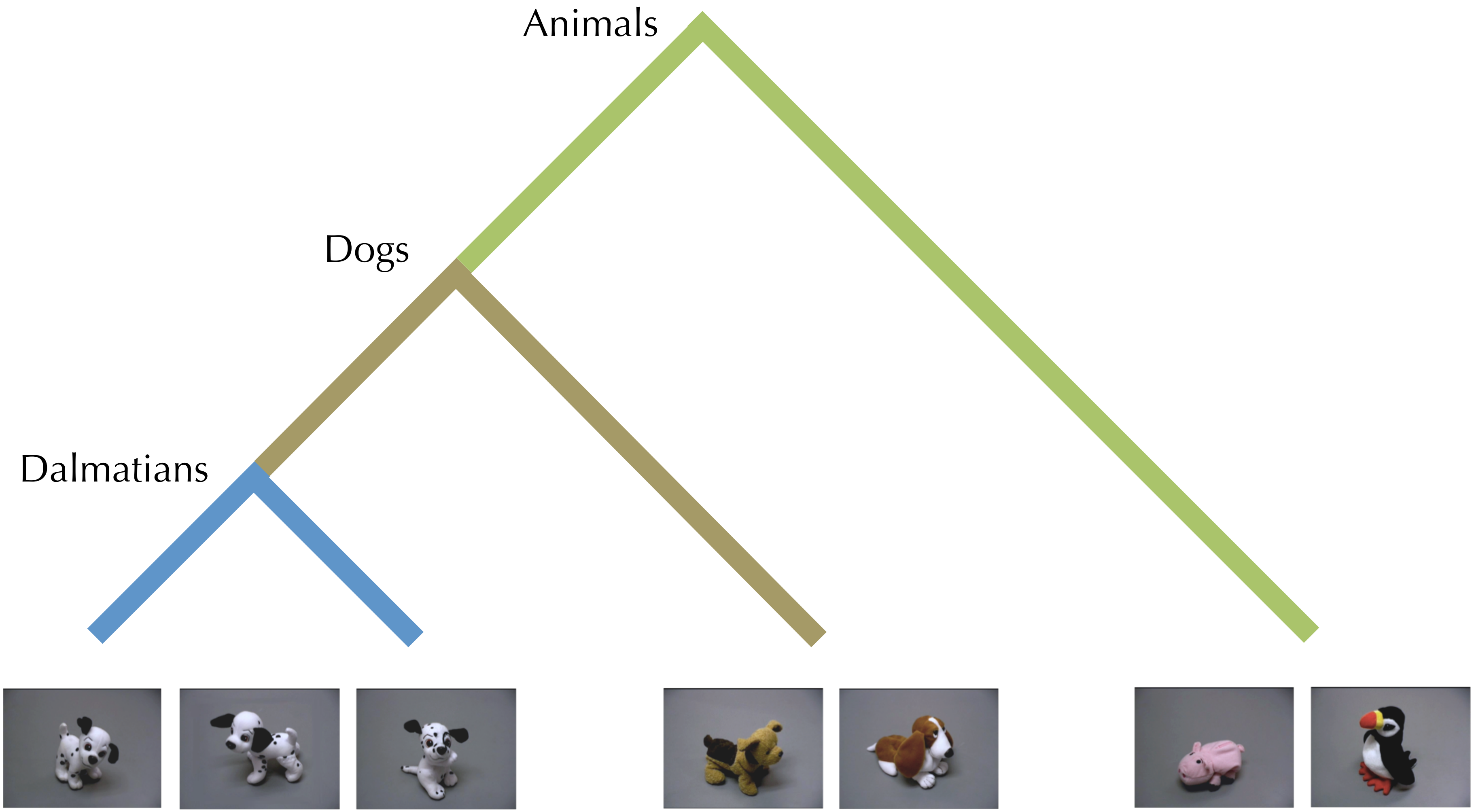
$$P(h \mid X) \propto P(X \mid h) P(h)$$

h: hypothesis

X: data

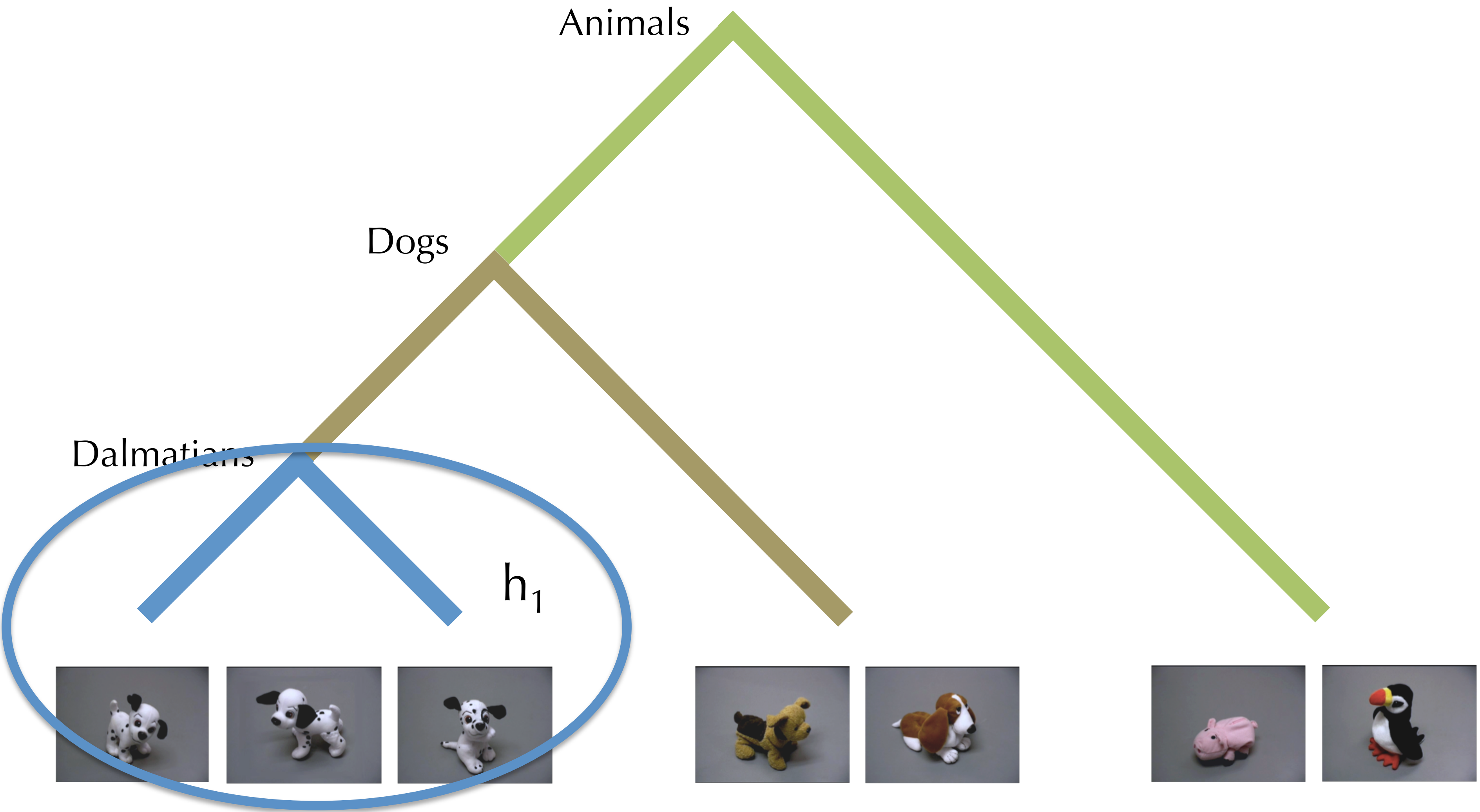
Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)



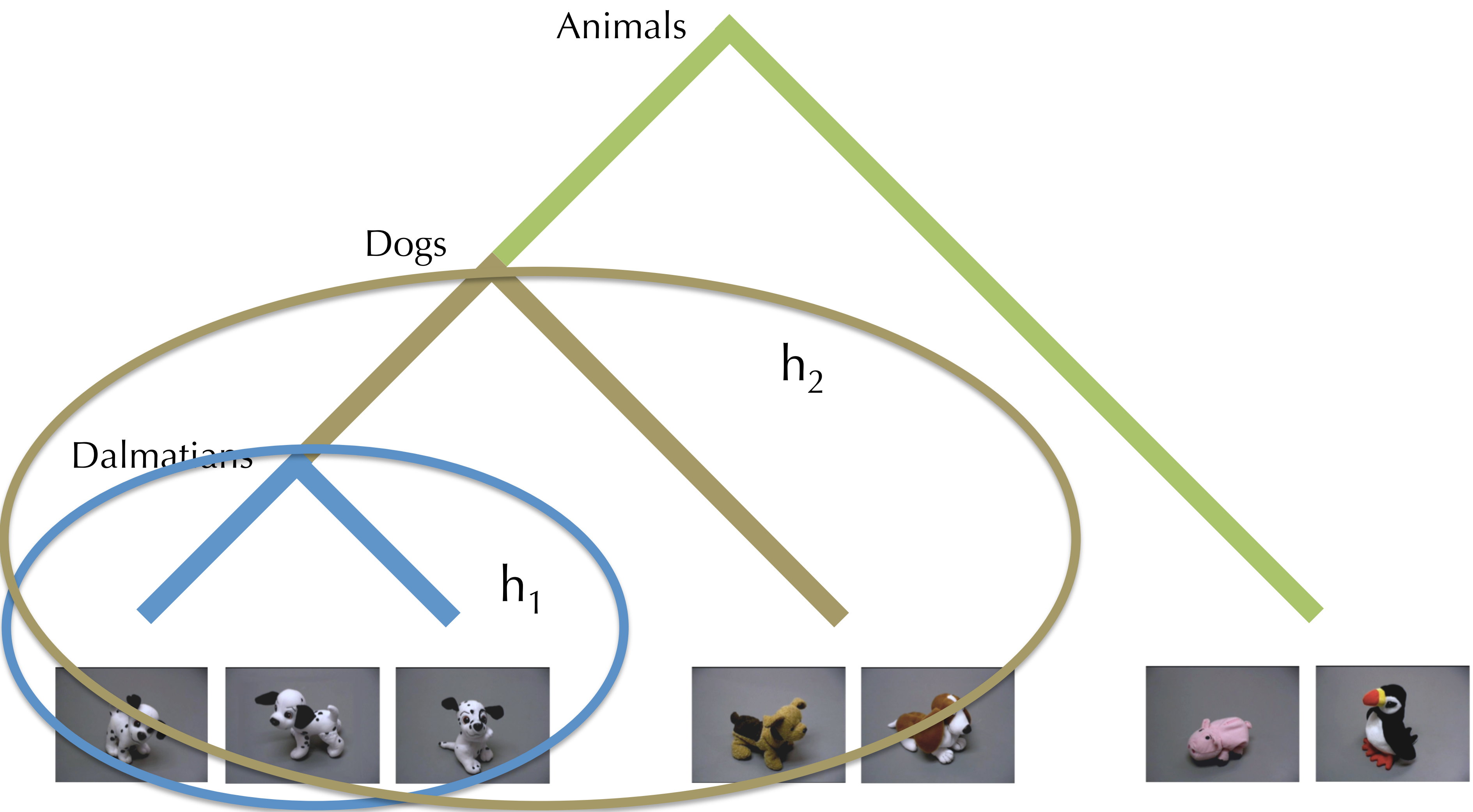
Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)



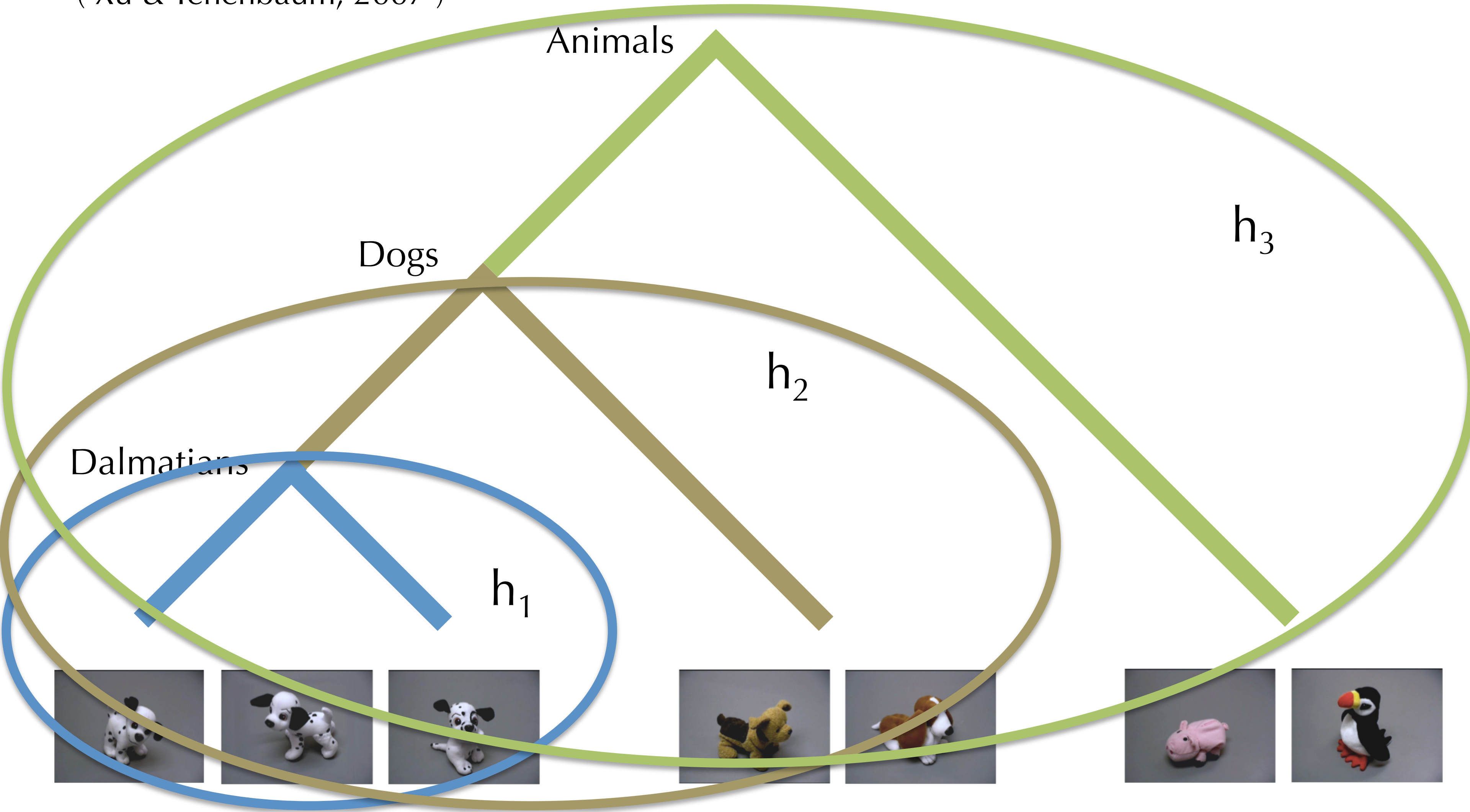
Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)



Word Learning as Bayesian Inference

(Xu & Tenenbaum, 2007)



Challenges for scaling

challenges

- small, hand-constructed domains
- toy stimuli
- constructing hypothesis space based on pairwise similarity judgments requires $O(n^2)$ judgments

Challenges for scaling

challenges

- small, hand-constructed domains
- toy stimuli
- constructing hypothesis space based on pairwise similarity judgments requires $O(n^2)$ judgments

solutions

- WordNet
  A lexical database for English
- IMAGENET
- automatically derived from WordNet structure

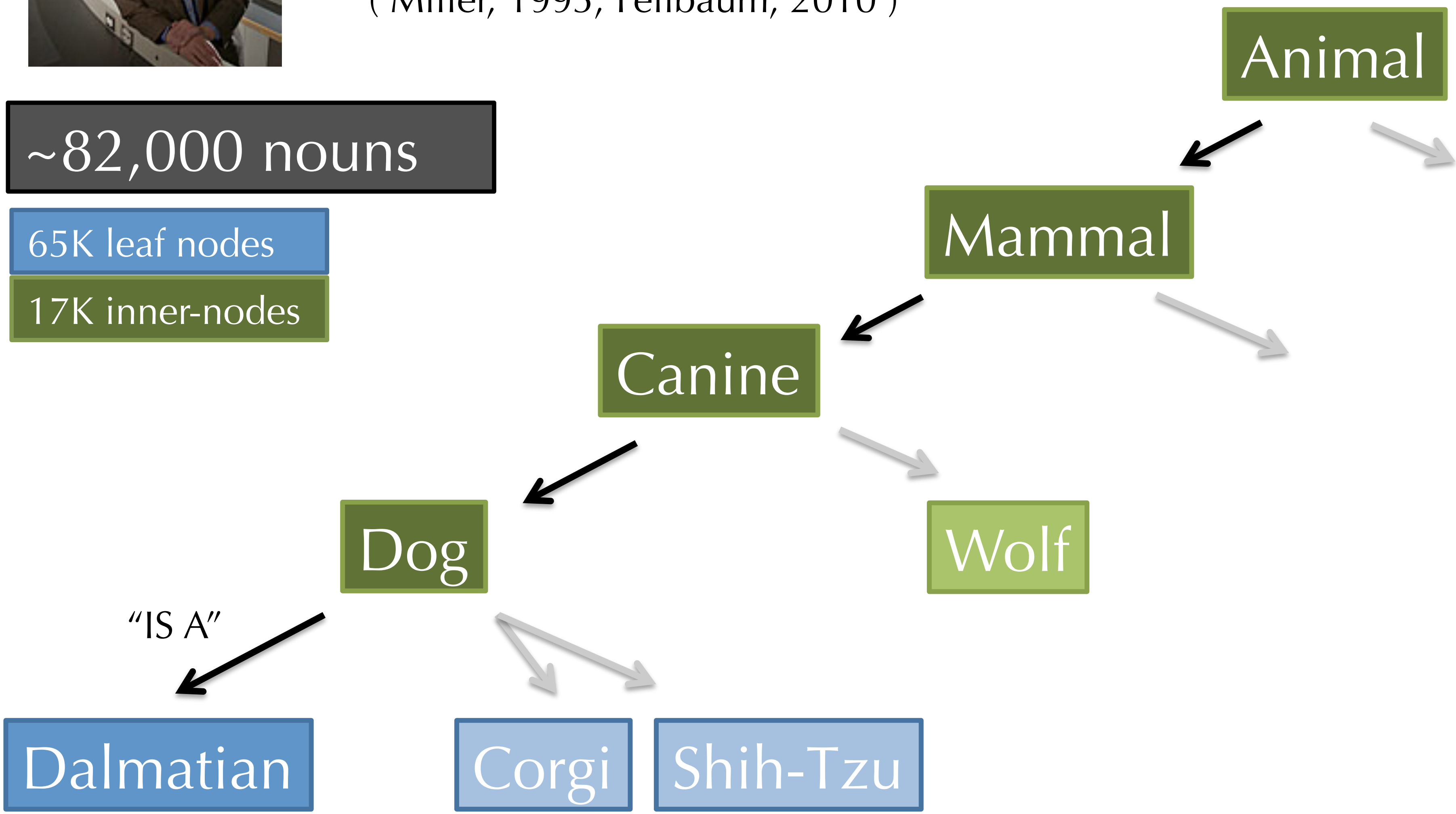
WordNet

(Miller, 1995; Fellbaum, 2010)

~82,000 nouns

65K leaf nodes

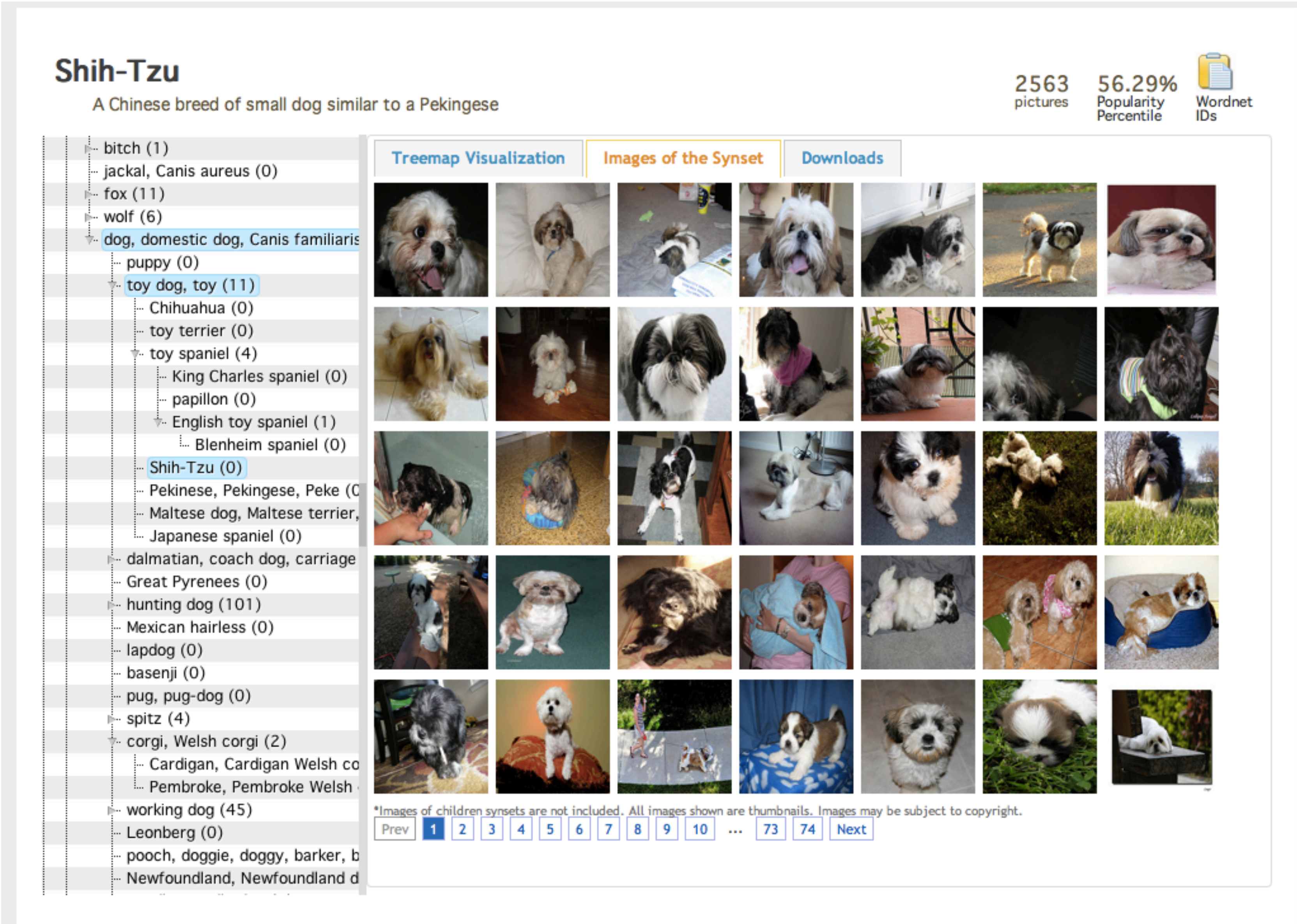
17K inner-nodes



IMAGENET

(Deng et al., 2009)

14.2 million images



Large-scale Word Learning

Here are four ZIVS

Here are three BLICKS

Here are five FEPS

Large-scale Word Learning

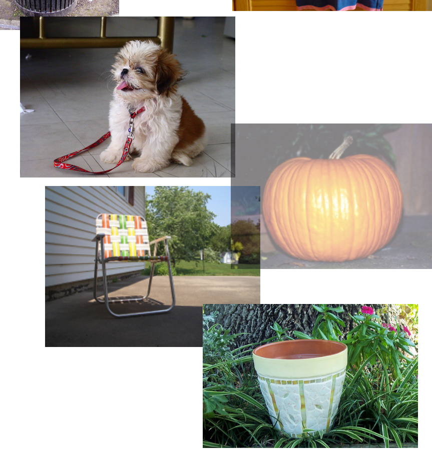
Here are four ZIVS

Here are three BLICKS

Here are five FEPS

Is this a BLICK?

Is this a ZIV?



Is this a FEP?

Experimental Validation

- Replicate Xu & Tenenbaum (2007) using naturalistic images as stimuli and a word learning model based on WordNet
- Test how this approach generalizes to a new set of domains

TRAINING SET

1 subordinate

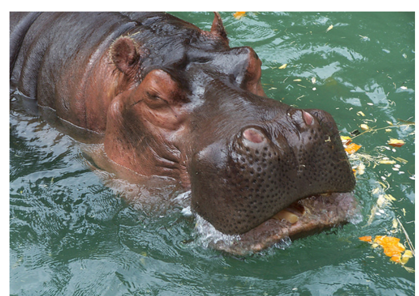
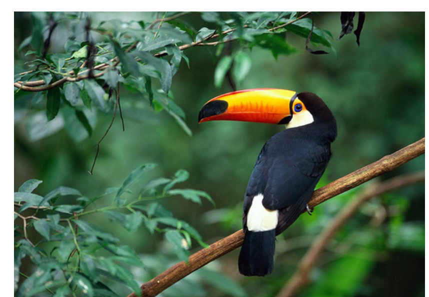
3 subordinate



3 basic-level



3 superordinate

TEST SET

2 subordinate

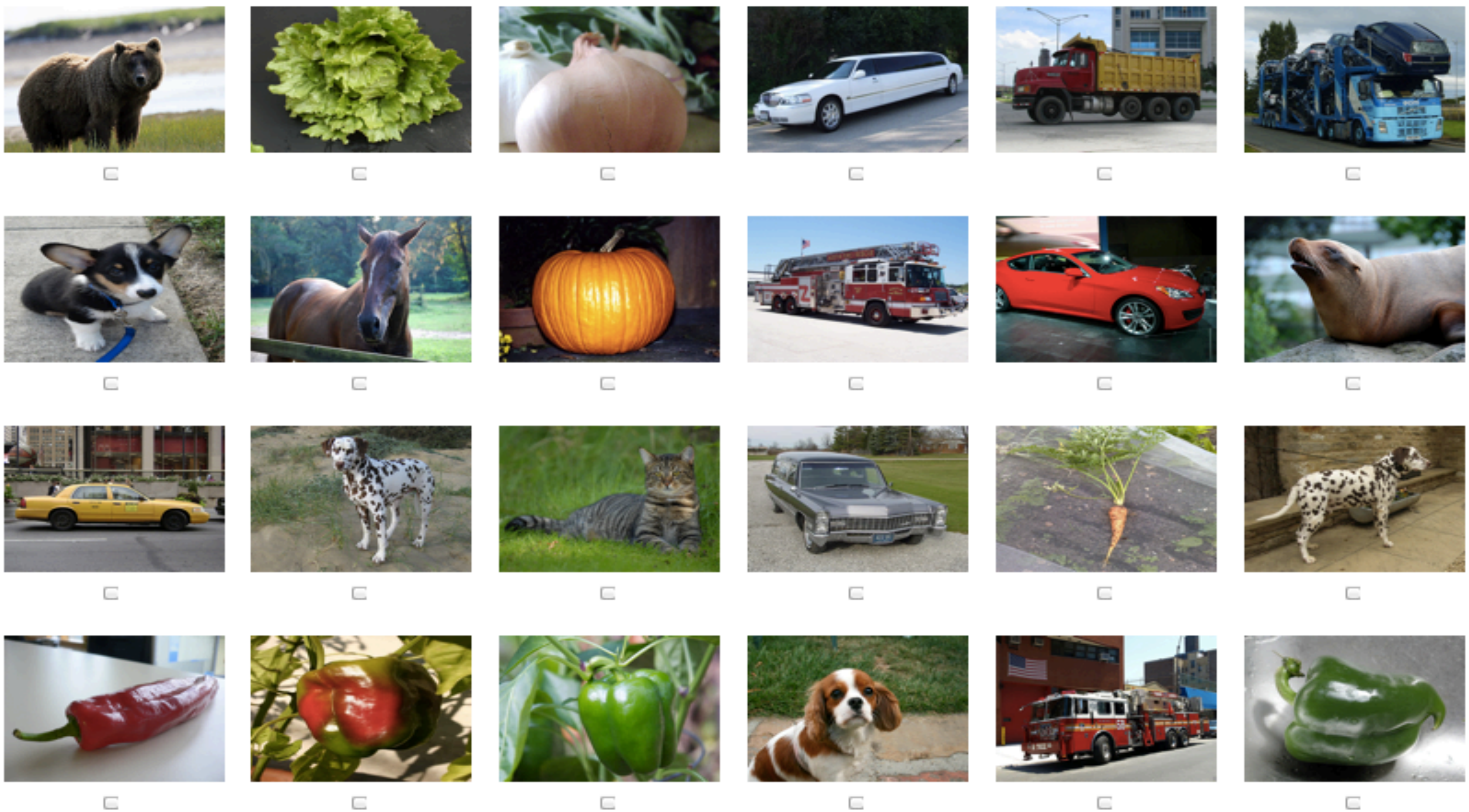

2 basic-level


4 superordinate

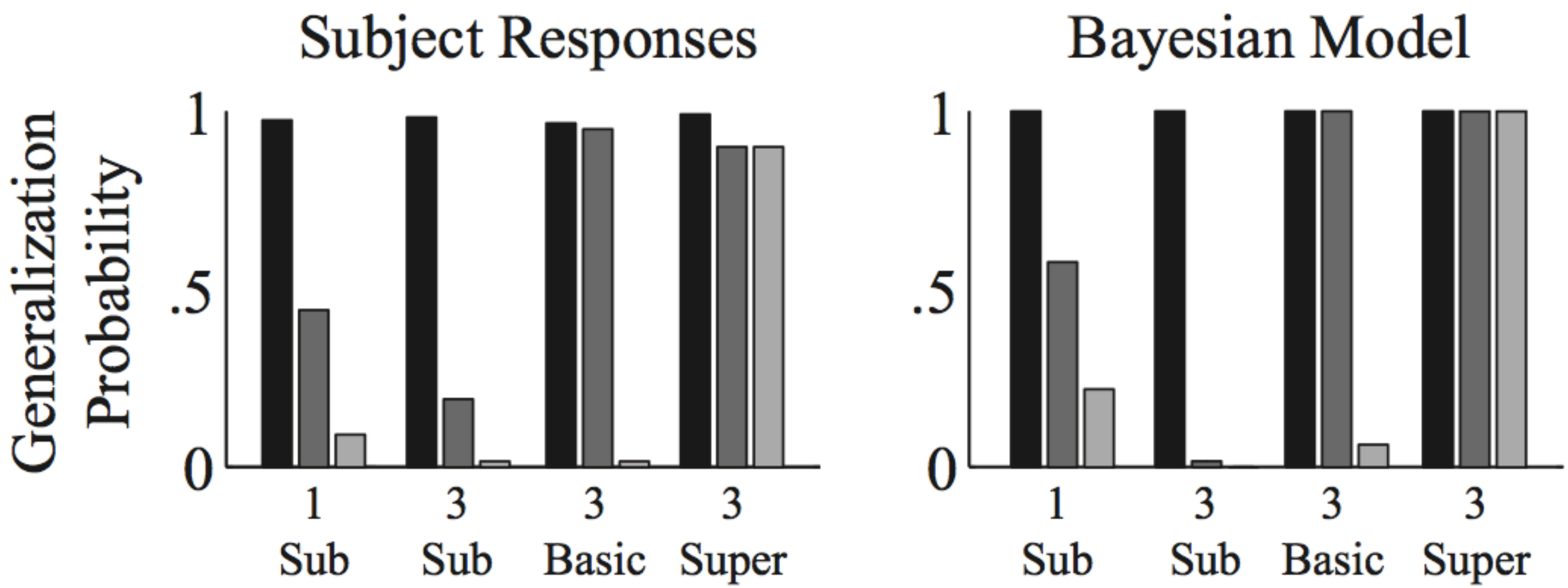



Here are three FEPS

Can you help Mr. Frog find the other "FEPS"?



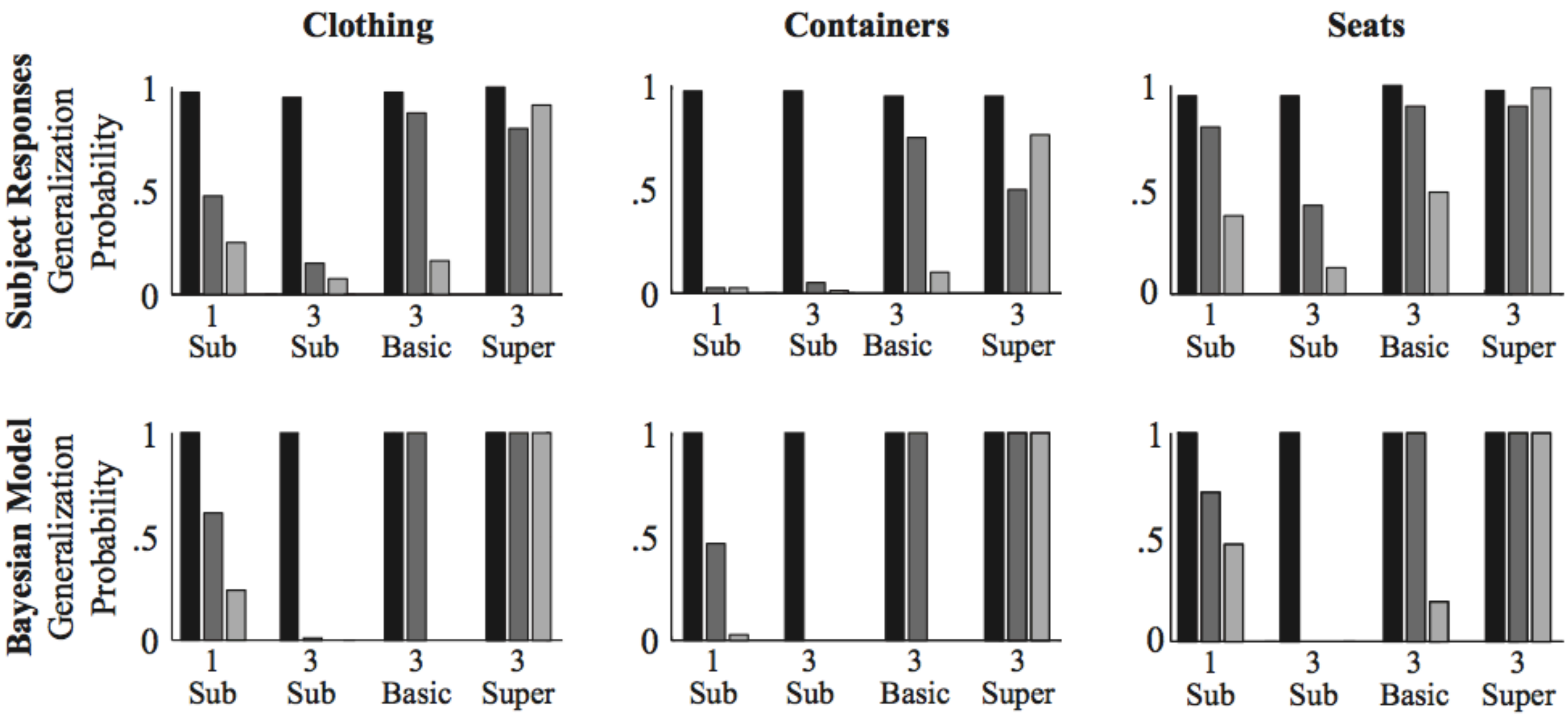
Replicating Xu & Tenenbaum (2007)



Extending to Novel Domains

	Clothing		Containers		Seats	
Object level	1	2	1	2	1	2
Subordinate						
Basic						
Superordinate						

Extending to Novel Domains



Conclusions

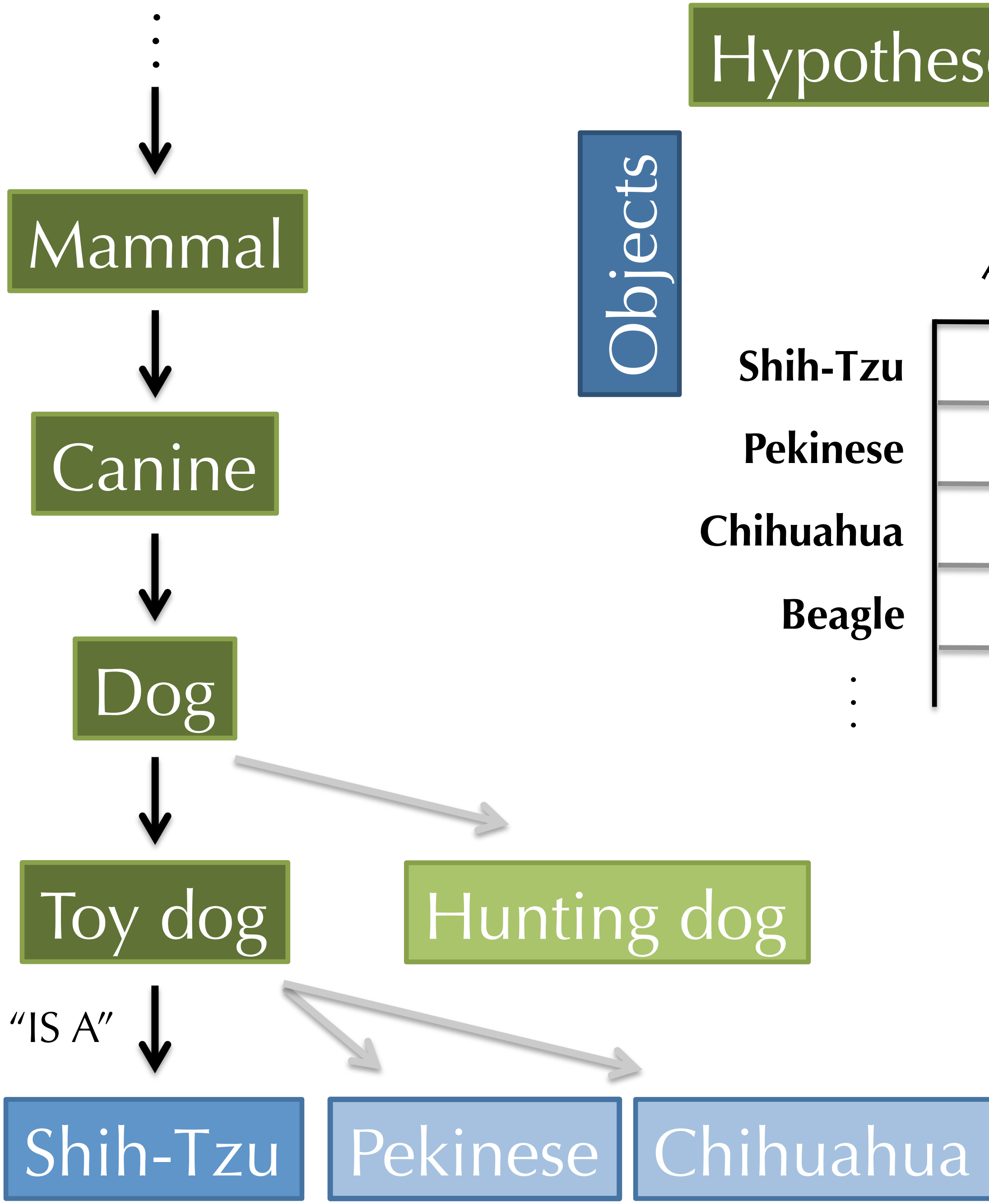
Integrating methods from Cognitive Science and Machine Learning has the potential to benefit both fields

Conclusions

Integrating methods from Cognitive Science and Machine Learning has the potential to benefit both fields

- Extended a cognitive model of word learning to operate on large-scale data – evaluation with naturalistic stimuli and over more domains
- Step towards bringing machines closer to human performance in word learning

Questions?



Hypotheses

Objects	Toy dog	Hunting dog	Dog	Canine	Mammal	. . .
Shih-Tzu	1	0	1	1	1	
Pekinese	1	0	1	1	1	
Chihuahua	1	0	1	1	1	
Beagle	0	1	1	1	1	
⋮						