Ling 151/Psych 156A: Acquisition of Language II

Lecture 11

Word meaning II

Announcements

Be working on HW4 (due 2/12/18)

Be working on the review questions for word meaning

Acquisition task

"I love my dax."

Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

What we know about the process of word learning

(1) Word meanings are learned from very few examples. Fast mapping is the extreme case of this, where one exposure is enough for children to infer the correct word-meaning mapping. However, cross-situational learning could work this way too, with a few very informative examples having a big impact.

kitty

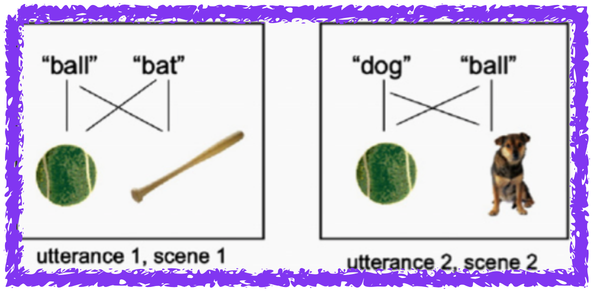
bear

ball

“Can I have the *zib*?”

[unknown]

20 months



12-14 months

What we know about the process of word learning

(2) Word meanings are often inferred from only positive examples. This means that children usually only see examples of what something is, rather than being explicitly told what something is not.

"What a cute dax!"

"I love my dax."

What we know about the process of word learning

(3) The target of word learning is a system of overlapping concepts. That is, words pick out different aspects of our world, and it's often the case that different words can refer to the same observable thing in the world.

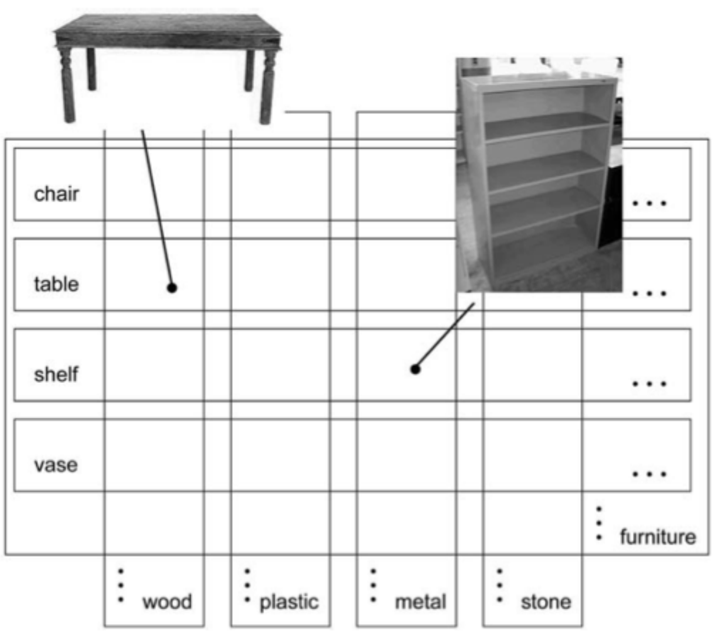
"I love my teddy."

"He's my favorite toy."

"He's brown and cuddly."

What we know about the process of word learning

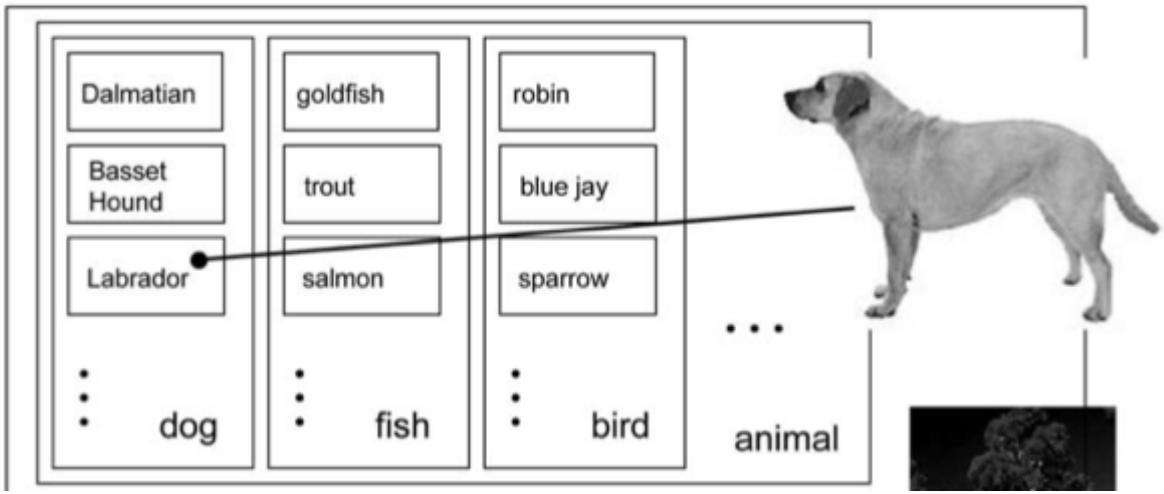
(3) The target of word learning is a system of overlapping concepts. That is, words pick out different aspects of our world, and it's often the case that different words can refer to the same observable thing in the world.



Shape vs. material labeling:
This is a desk.
It's made of wood.
This bookcase is also made of wood.

What we know about the process of word learning

(3) The target of word learning is a system of overlapping concepts. That is, words pick out different aspects of our world, and it's often the case that different words can refer to the same observable thing in the world.



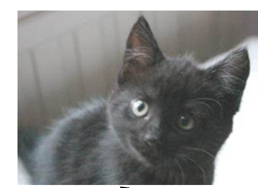
What level of specificity (object-kind labeling)?
"This is my labrador, who is a great dog, and a very friendly animal in general."

What we know about the process of word learning

(4) Inferences about word meaning based on examples should be graded, rather than absolute. That is, the child probably still has some uncertainty after learning from the input. This is particularly true if the input is ambiguous (as in cross-situational learning).

"I love my dax and my kleeg."

"There are my favorite dax and kleeg!"



Some uncertainty remains about whether "dax" is this or this.

Bayesian learning for word meaning mapping

Xu & Tenenbaum (2007: Psychological Review) hypothesize that a child using Bayesian learning would show these behaviors during word learning.

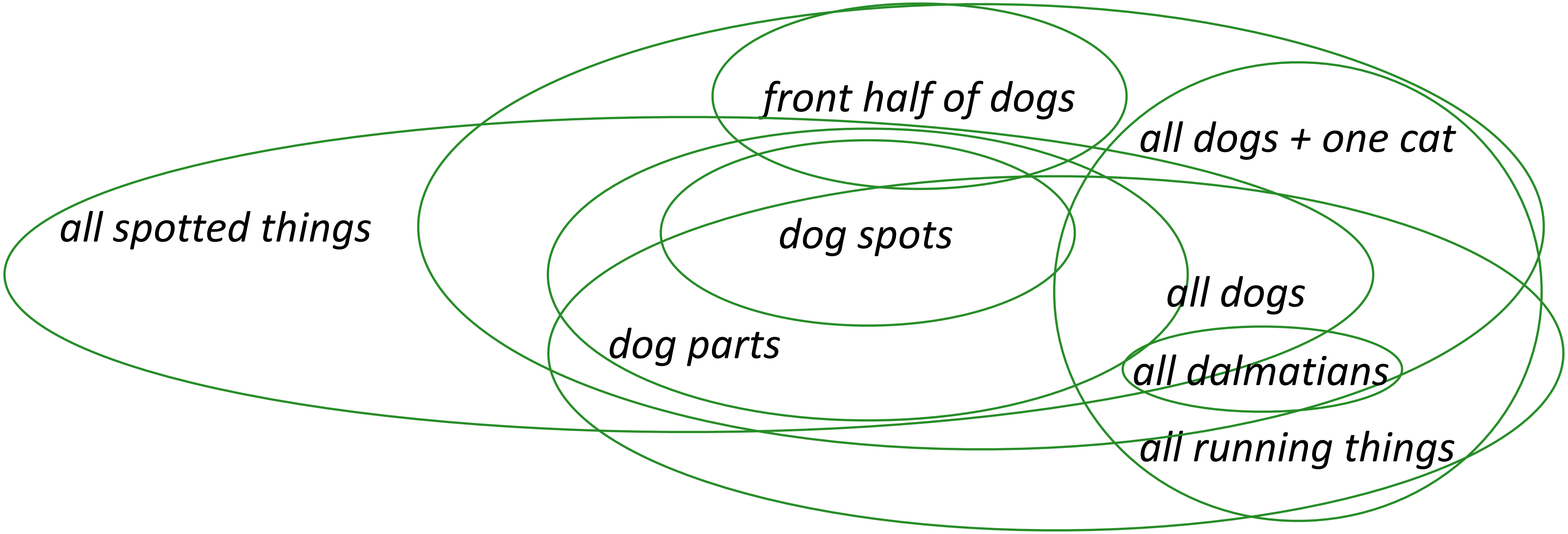
Claim: "Learners can rationally infer the meanings of words that label multiple overlapping concepts, from just a few positive examples. Inferences from more ambiguous patterns of data lead to more graded and uncertain patterns of generalization."

The importance of the hypothesis space

An important consideration: Bayesian learning can only operate over a defined hypothesis space.

Example of potential hypotheses for *dog*:

dog = …

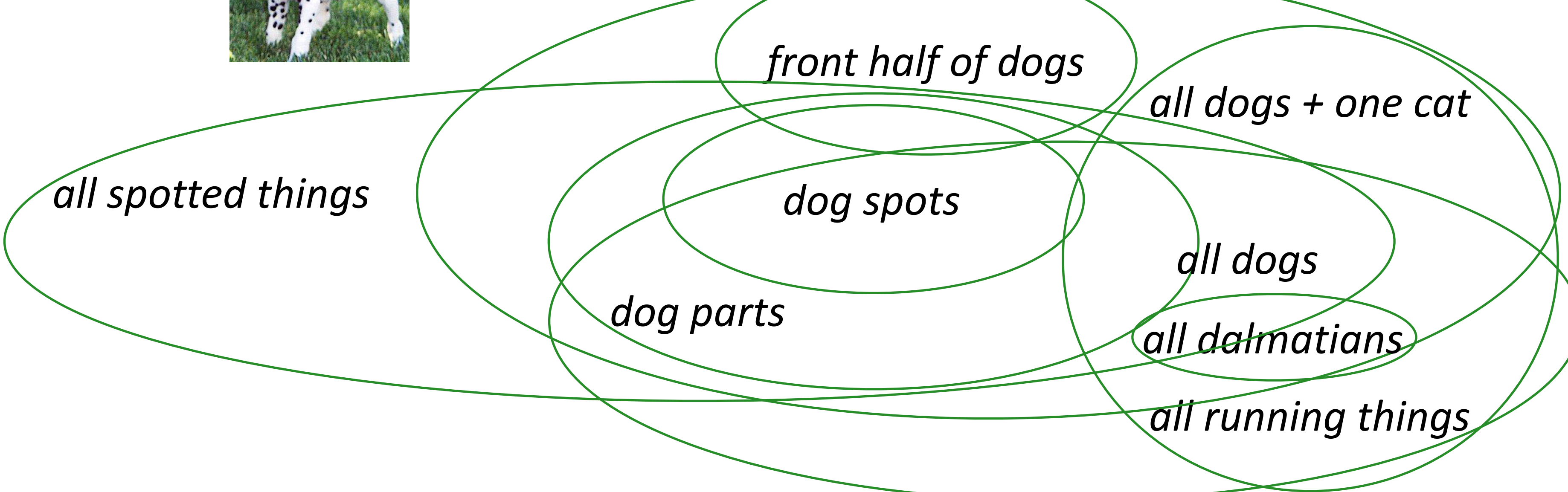


The importance of the hypothesis space

Two traditional constraints on children's hypothesis (learning biases):

Whole Object constraint: First guess is that a label refers to a whole object, rather than part of the object (*dog parts, front half of dog*) or an attribute of the object (*dog spots*)

dog = …

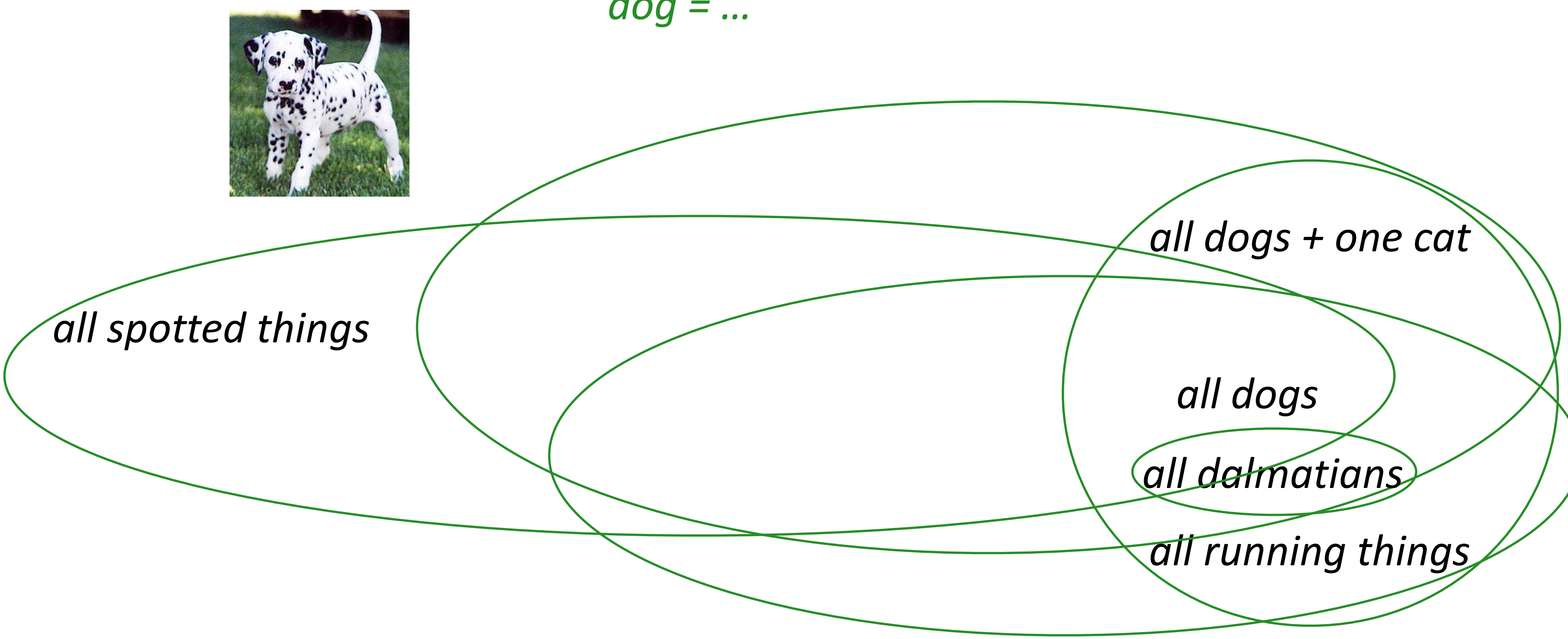


The importance of the hypothesis space

Two traditional constraints on children's hypothesis (learning biases):

Whole Object constraint: First guess is that a label refers to a whole object, rather than part of the object (*dog parts, front half of dog*) or an attribute of the object (*dog spots*)

dog = …

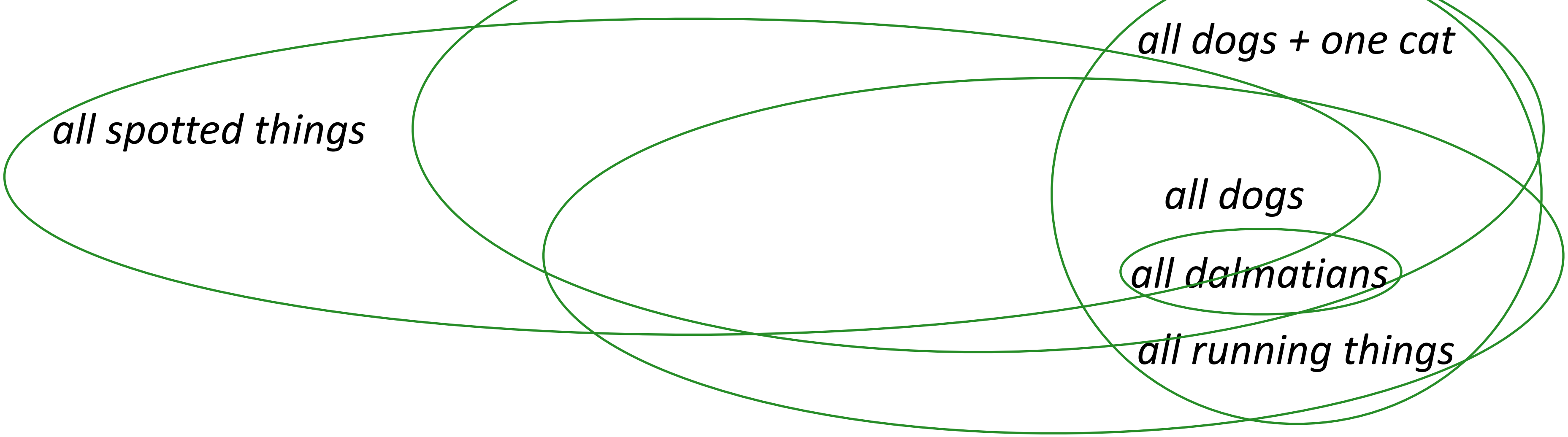


The importance of the hypothesis space

Two traditional constraints on children's hypothesis (learning biases):

Taxonomic constraint (Markman 1989): First guess about an unknown label is that it applies to the taxonomic class (ex: *dog*, instead of *all running things* or *all dogs + one cat*)

dog = …



The importance of the hypothesis space

Two traditional constraints on children's hypothesis (learning biases):

Taxonomic constraint (Markman 1989): First guess about an unknown label is that it applies to the taxonomic class (ex: *dog*, instead of *all running things* or *all dogs + one cat*)

dog = …

Constraints on the hypothesis space

https://www.youtube.com/watch?v=Ci-5dVVvf0U

http://www.thelingspace.com/episode-35

2:33-4:14

Suspicious coincidences & Bayesian learning

Situation:

"*fep*" "*fep*" "*fep*" "*fep*"

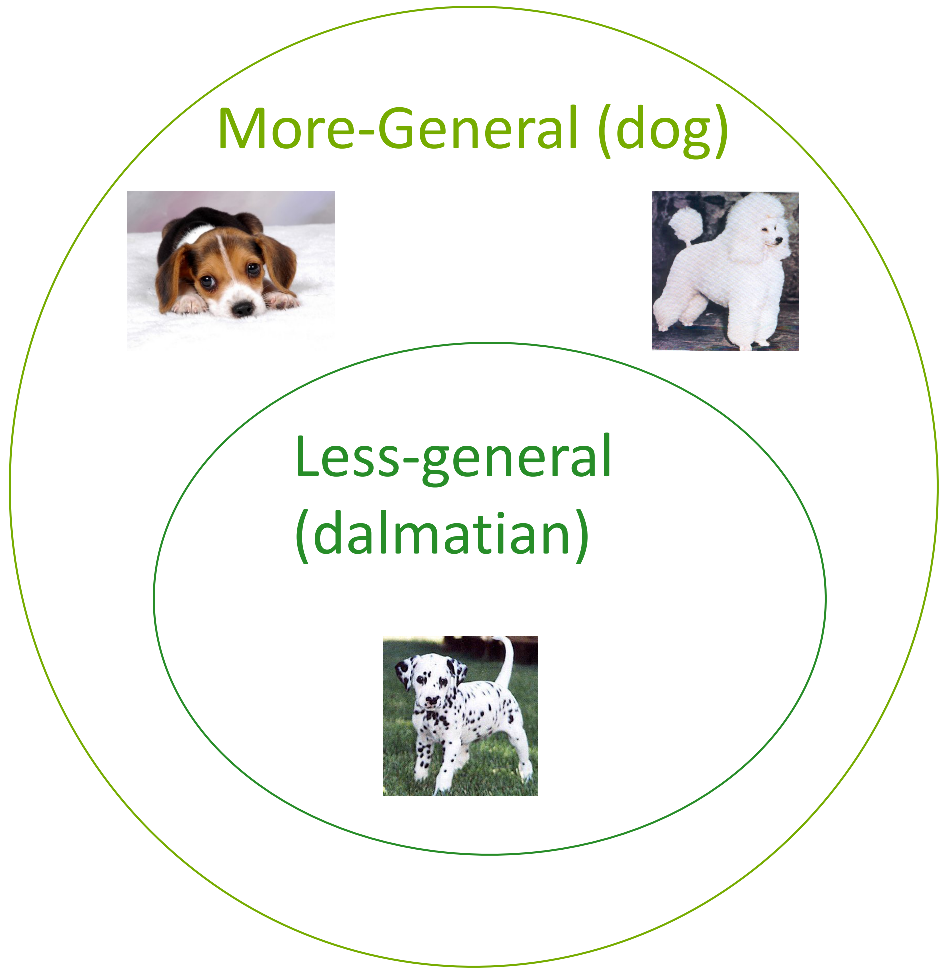
Suspicious: Why is no other animal or other kind of dog a *fep* if *fep* can really label any animal or any kind of dog?

Bayesian reasoning: Would expect to see other animals (or dogs) labeled as *fep* if *fep* really could mean those things. If *fep* continues not to be used this way, this is growing support that *fep* cannot mean those things.

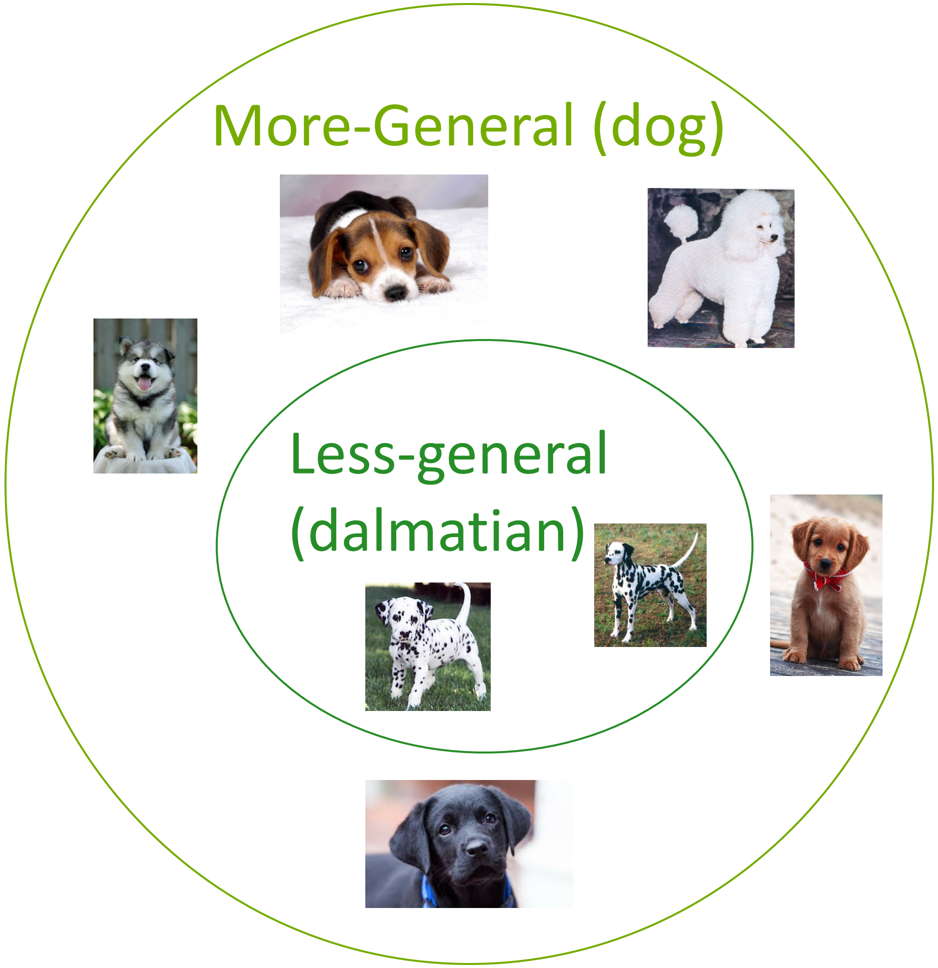
Formal instantiation of “suspicious coincidence”

Has to do with expectation of the data points that should be encountered in the input

If the more-general hypothesis (dog) is correct, the learner should encounter some data that can only be accounted for by the more-general hypothesis (like beagles or poodles). These data would be incompatible with the less-general hypothesis (dalmatian).

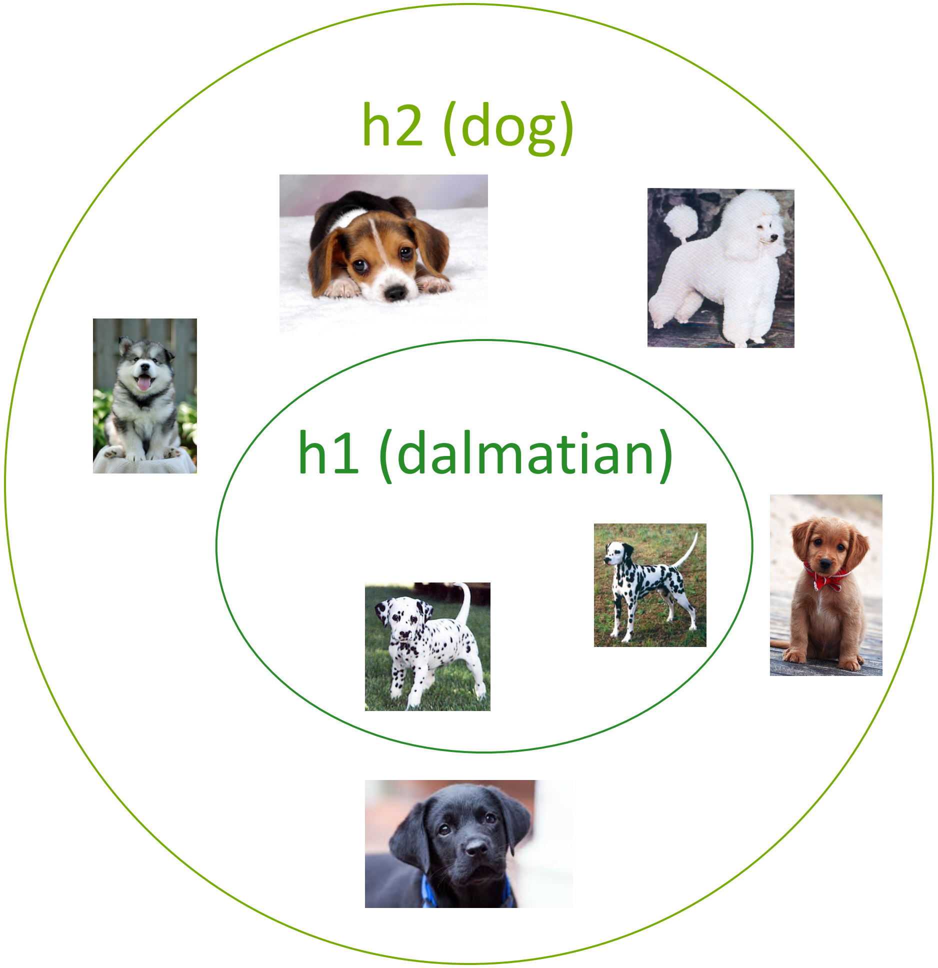


Using Bayesian inference to implement the sense of "suspicious coincidence"



Using Bayesian inference to implement the sense of "suspicious coincidence"

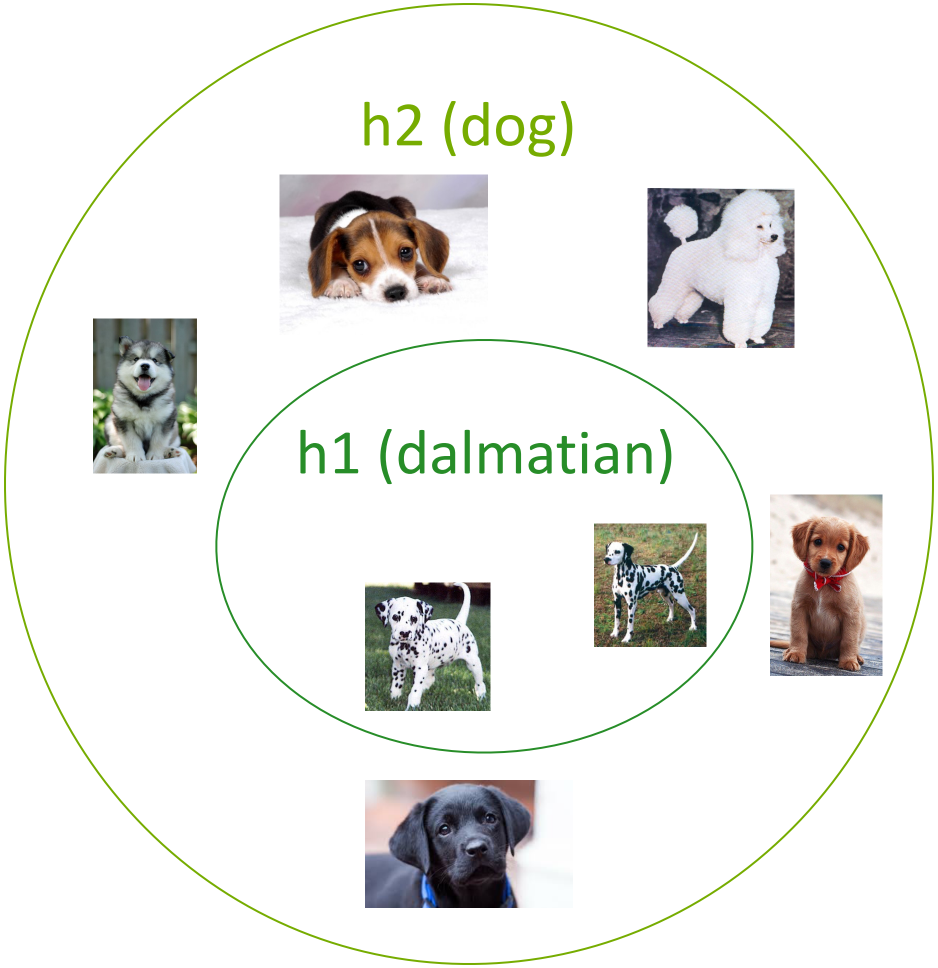
A Bayesian model assumes the learner has **some space of hypotheses H**, each of which represents a possible explanation for how **the data D** in the data intake were generated.



Using Bayesian inference to implement the sense of “suspicious coincidence”

Given **D**, the modeled child’s goal is to determine the probability of each possible hypothesis **h** $\in$ **H**. This is **P (h|D)**, the ***posterior*** for that hypothesis.

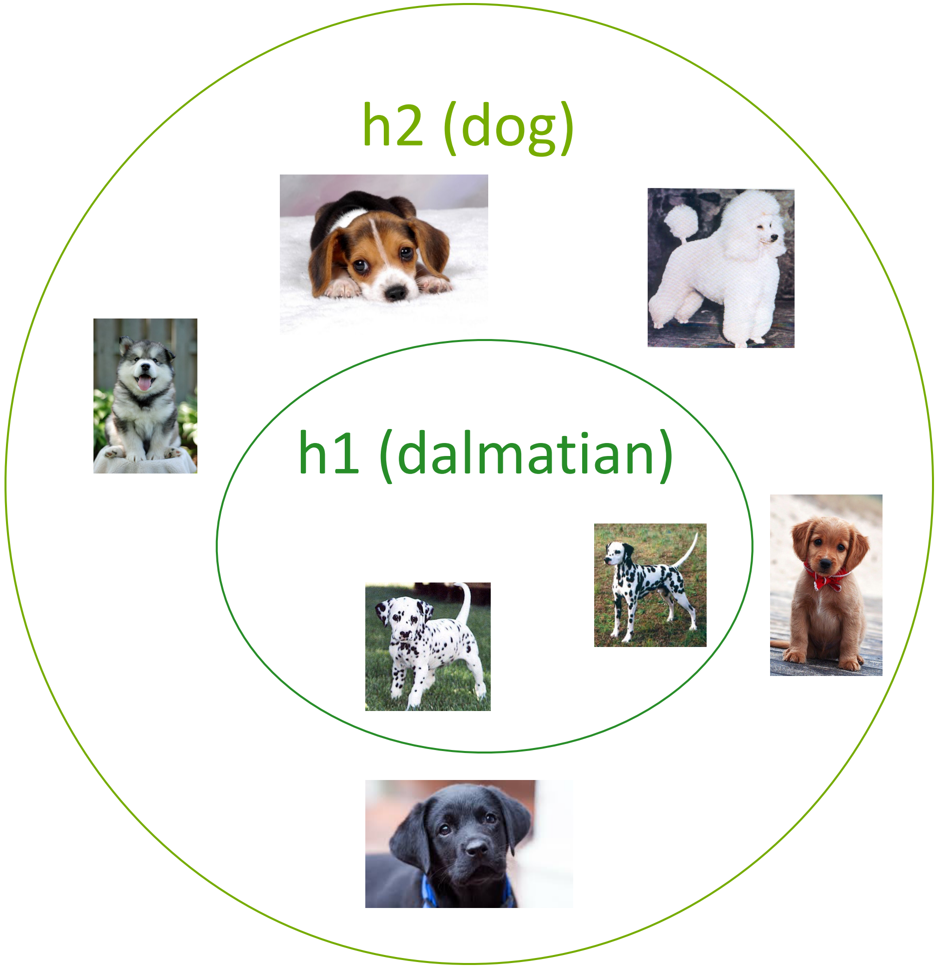
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Using Bayesian inference to implement the sense of “suspicious coincidence”

P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

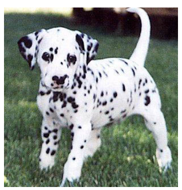
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Using Bayesian inference to implement the sense of "suspicious coincidence"

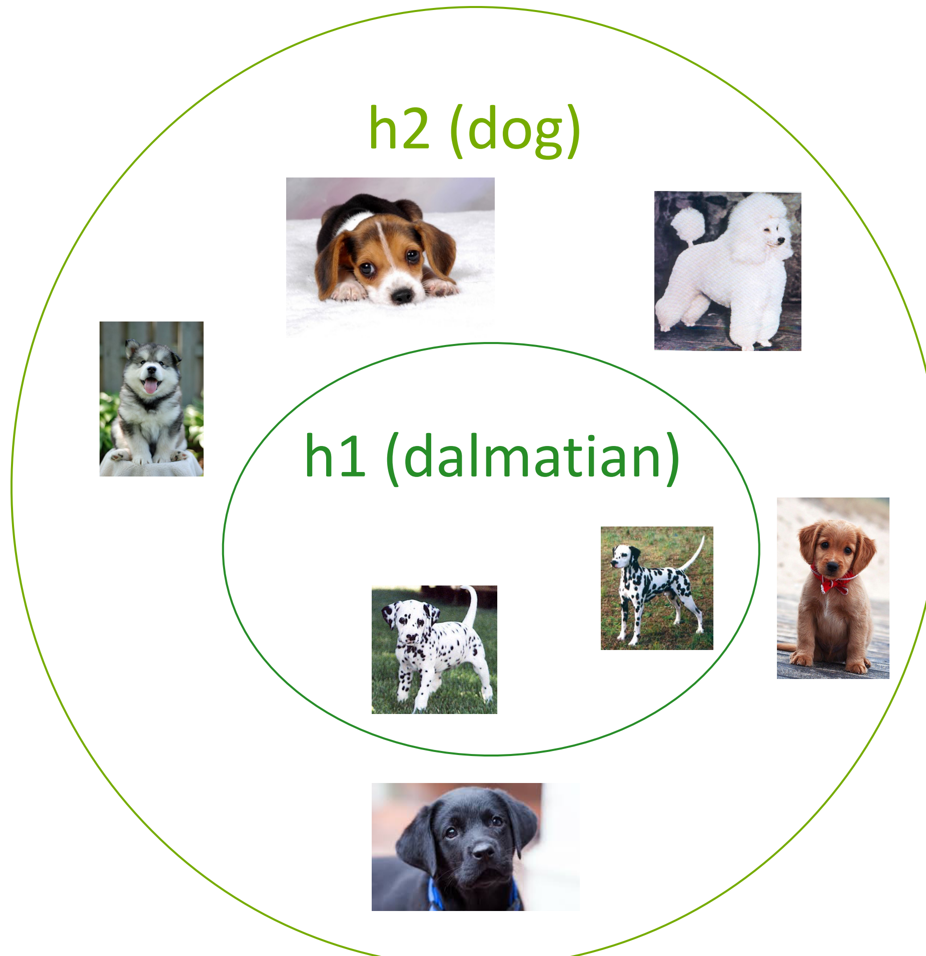
P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Data D

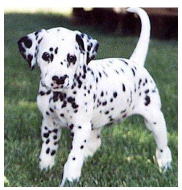
What if the data intake contained these data labeled "fep"?



Using Bayesian inference to implement the sense of "suspicious coincidence"

P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

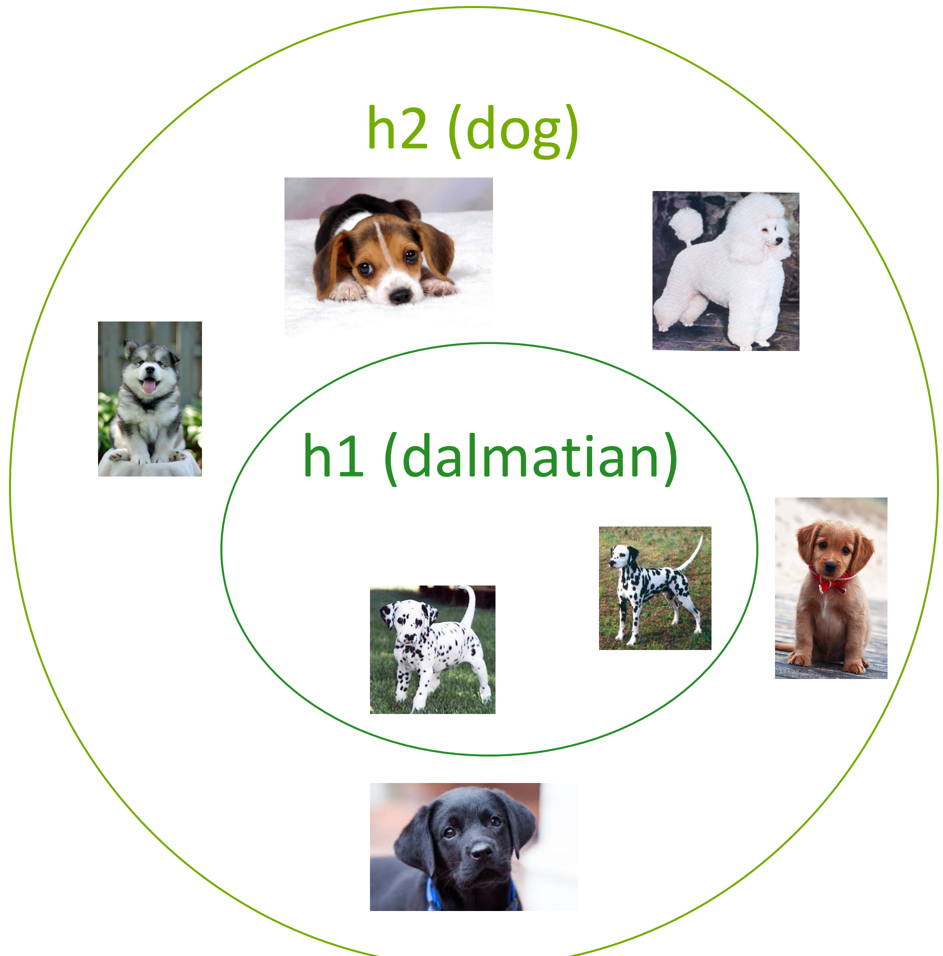
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Data D

P(D | h1) = 1/2 * 1/2 = 1/4

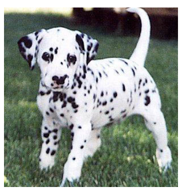
Since there are only two things in h1, the probability of either of them showing up when "fep" is used is 1/2.



Using Bayesian inference to implement the sense of "suspicious coincidence"

P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

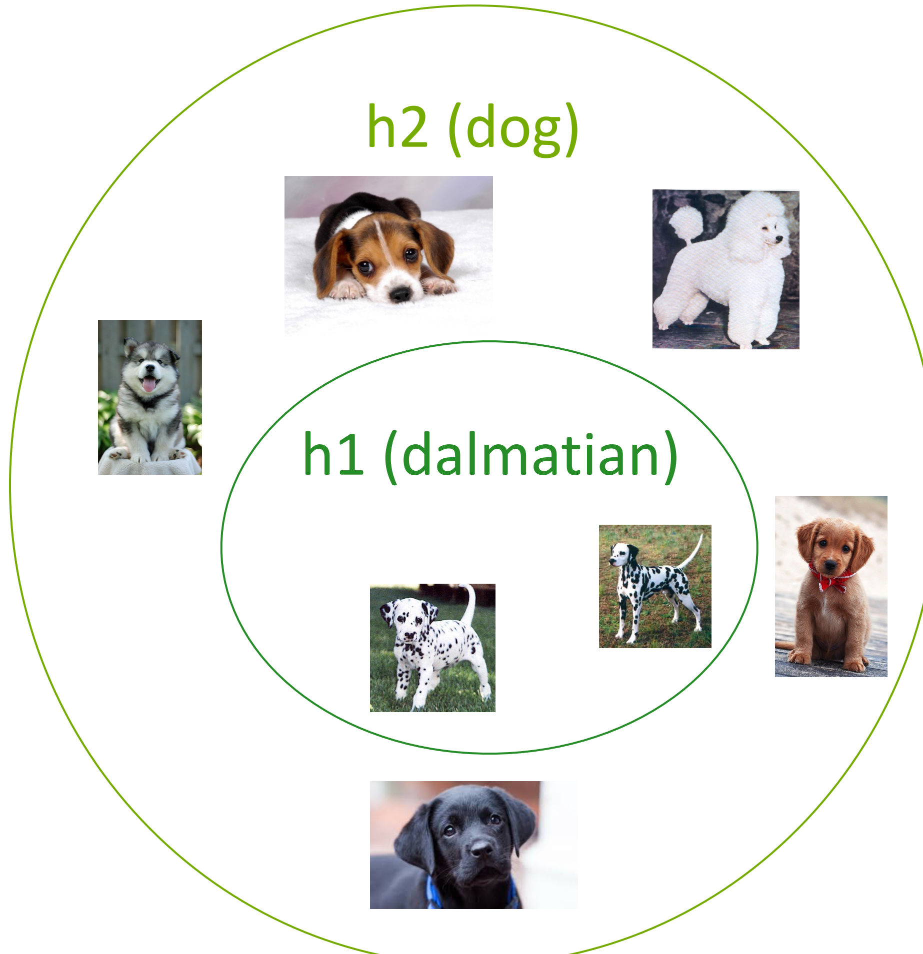


Data D

P(D | h1) = 1/4

P(D | h2) = 1/7 * 1/7 = 1/49

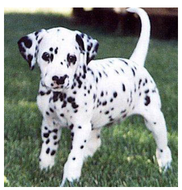
Since there are seven things in h2, the probability of any of them showing up when "fep" is used is 1/7.



Using Bayesian inference to implement the sense of "suspicious coincidence"

P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

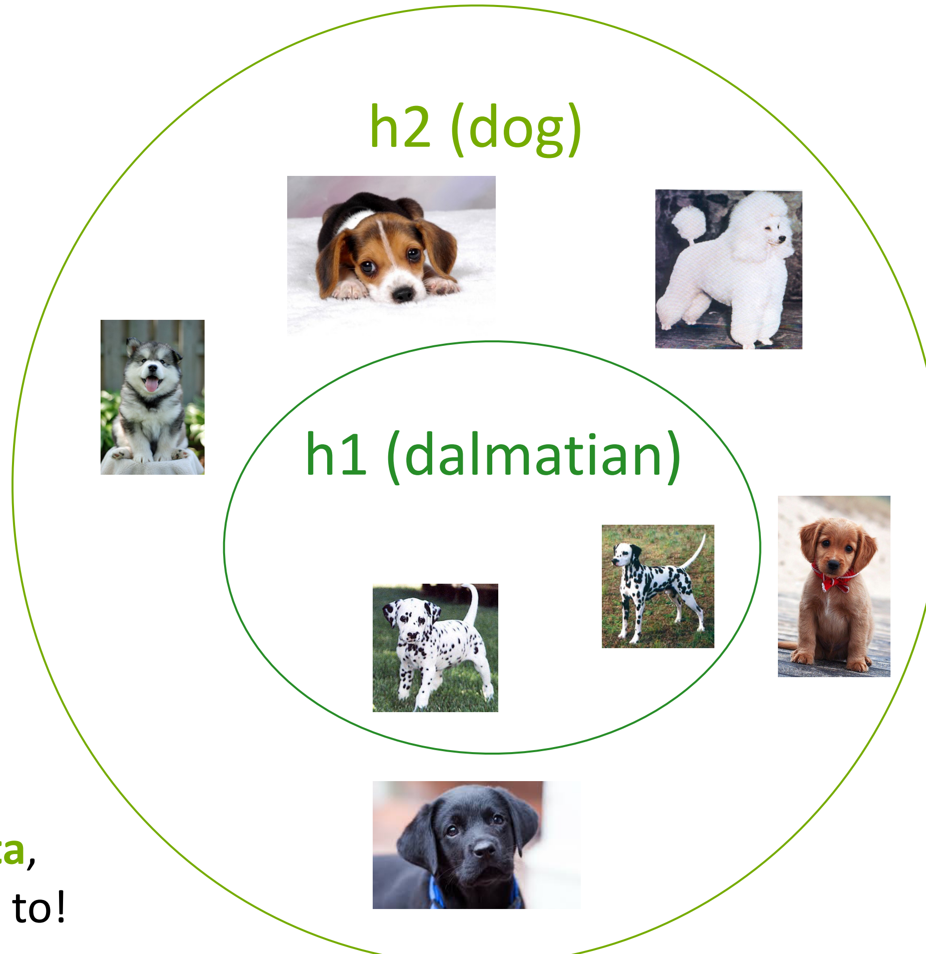
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Data D

P(D | h1) = 1/4

P(D | h2) = 1/49

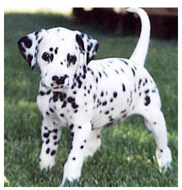


Here's where we can see the "suspicious coincidence" effect. It's a **suspicious coincidence for h2 to generate these data**, compared with h1 — h1 is far more likely to!

Using Bayesian inference to implement the sense of "suspicious coincidence"

P (D|h) represents the *likelihood* of the data D given hypothesis h, and describes how compatible that hypothesis is with the data.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

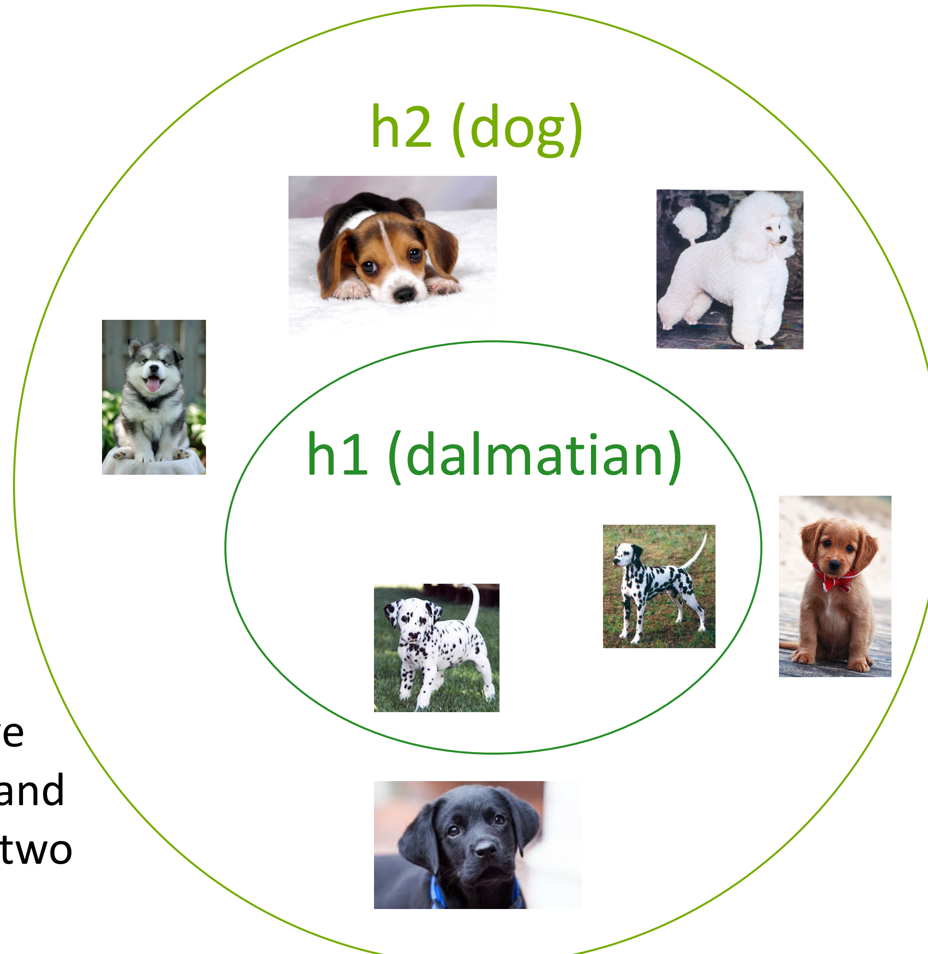


Data D

P(D | h1) = 1/4

P(D | h2) = 1/49

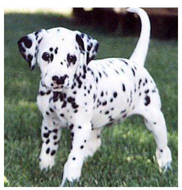
And we can see the result of this when we finish the Bayesian inference calculation and compare the relative probabilities of the two hypotheses.



Using Bayesian inference to implement the sense of "suspicious coincidence"

P (h) represents the *prior* of the hypothesis h, and represents the probability of the hypothesis before any data have been encountered. Intuitively, this corresponds to how plausible the hypothesis is, irrespective of any data.

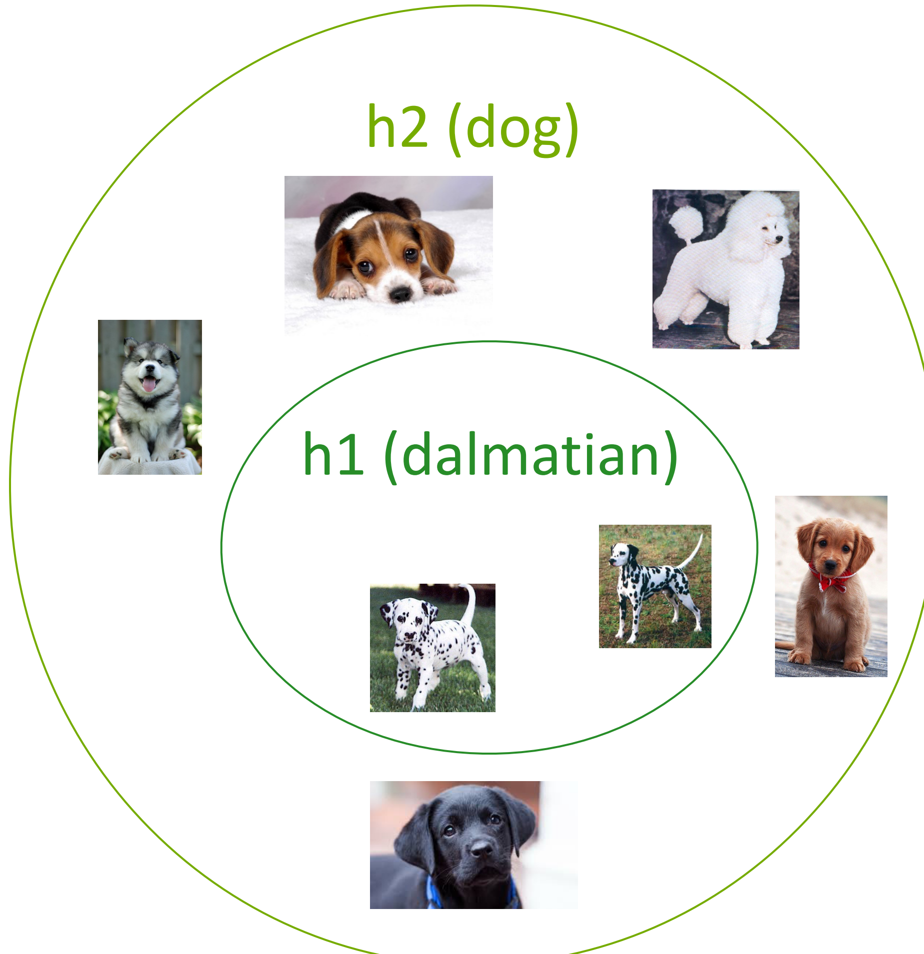
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Data D

P(D | h1) = 1/4

P(D | h2) = 1/49

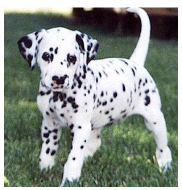


Suppose we let both hypotheses be equally likely before any data have been seen.

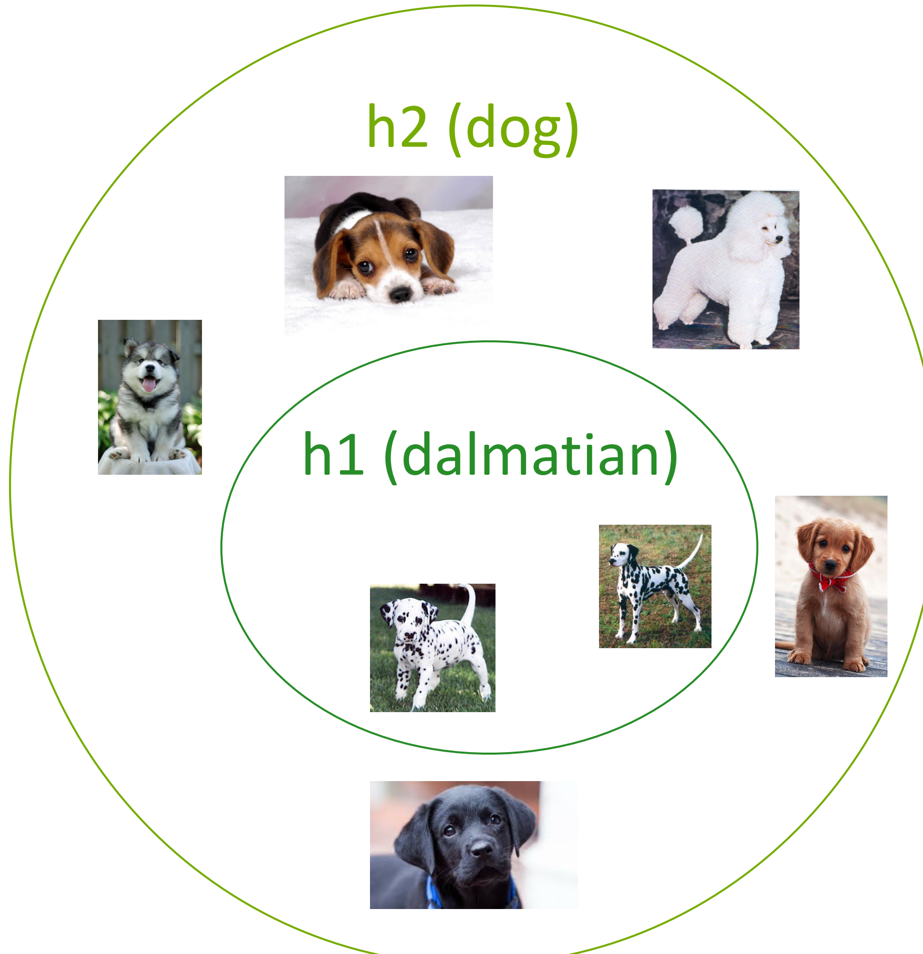
Using Bayesian inference to implement the sense of "suspicious coincidence"

P (h) represents the *prior* of the hypothesis h, and represents the probability of the hypothesis before any data have been encountered. Intuitively, this corresponds to how plausible the hypothesis is, irrespective of any data.

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)}$$



Data D



P(D | h1) = 1/4

P(D | h2) = 1/49

P(h1) = 1/2

P(h2) = 1/2

uniform probability

Using Bayesian inference to implement the sense of “suspicious coincidence”

The posterior P(h|D) is proportional to

the likelihood of the hypothesis * the prior of the hypothesis.

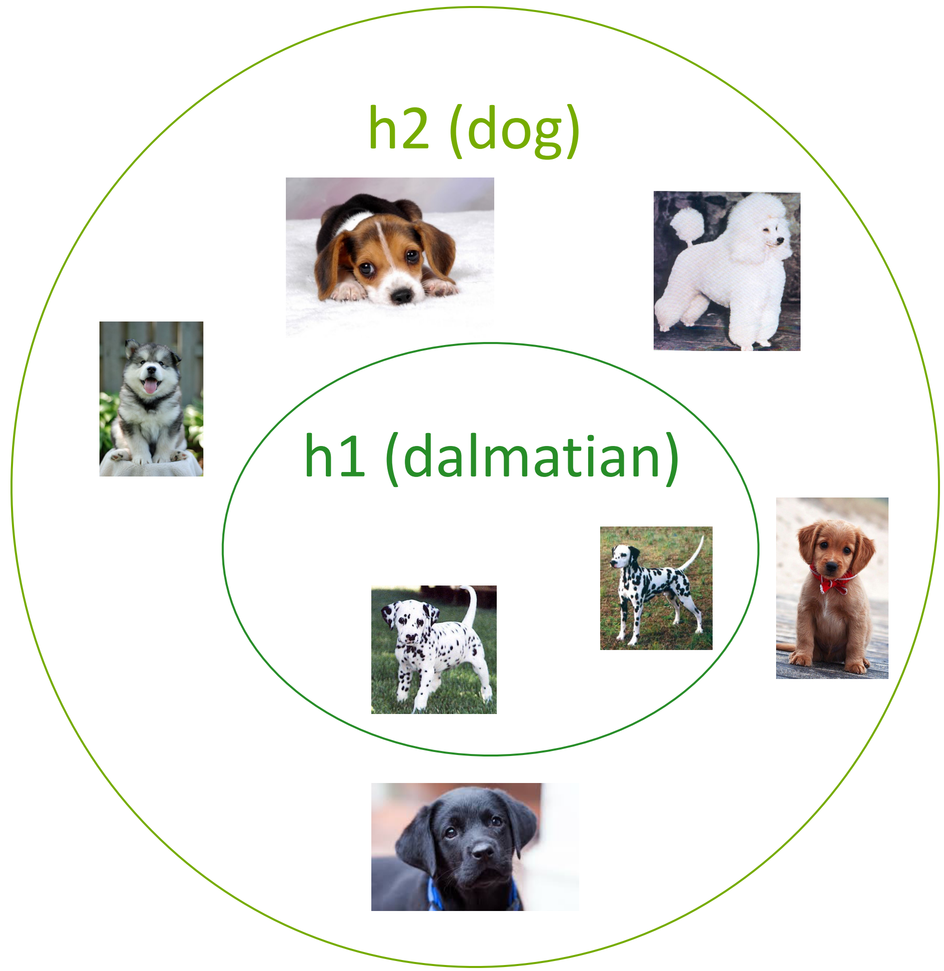
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

$$\propto P(D|h) * P(h)$$

Data D

P(D | h1) = 1/4 P(h1) = 1/2

P(D | h2) = 1/49 P(h2) = 1/2



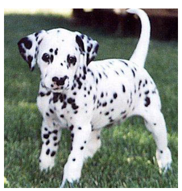
Using Bayesian inference to implement the sense of "suspicious coincidence"

The posterior P(h|D) is proportional to

the likelihood of the hypothesis * the prior of the hypothesis.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

$$\propto P(D|h) * P(h)$$



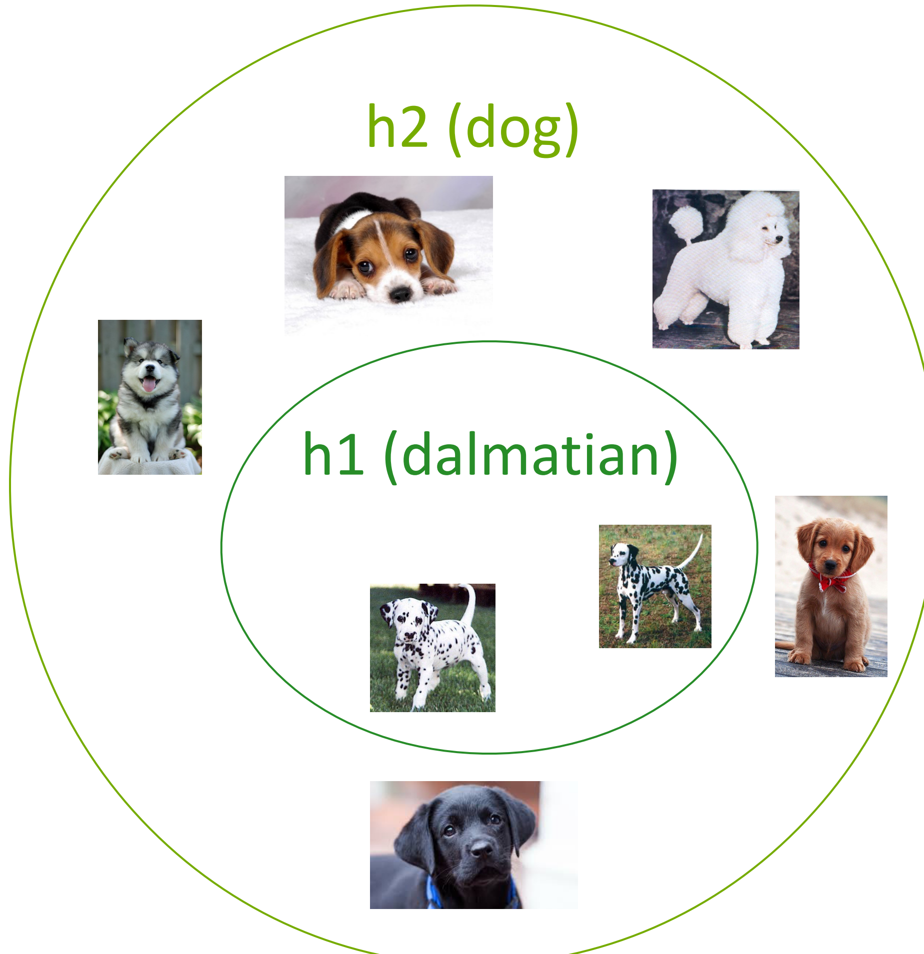
Data D

P(D | h1) = 1/4　　P(h1) = 1/2

P(D | h2) = 1/49　　P(h2) = 1/2

P(h1 | D) ∝ 1/4 * 1/2 = **1/8**

P(h2 | D) ∝ 1/49 * 1/2 = **1/98**



Using Bayesian inference to implement the sense of "suspicious coincidence"

The posterior P(h|D) is proportional to

the likelihood of the hypothesis * the prior of the hypothesis.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

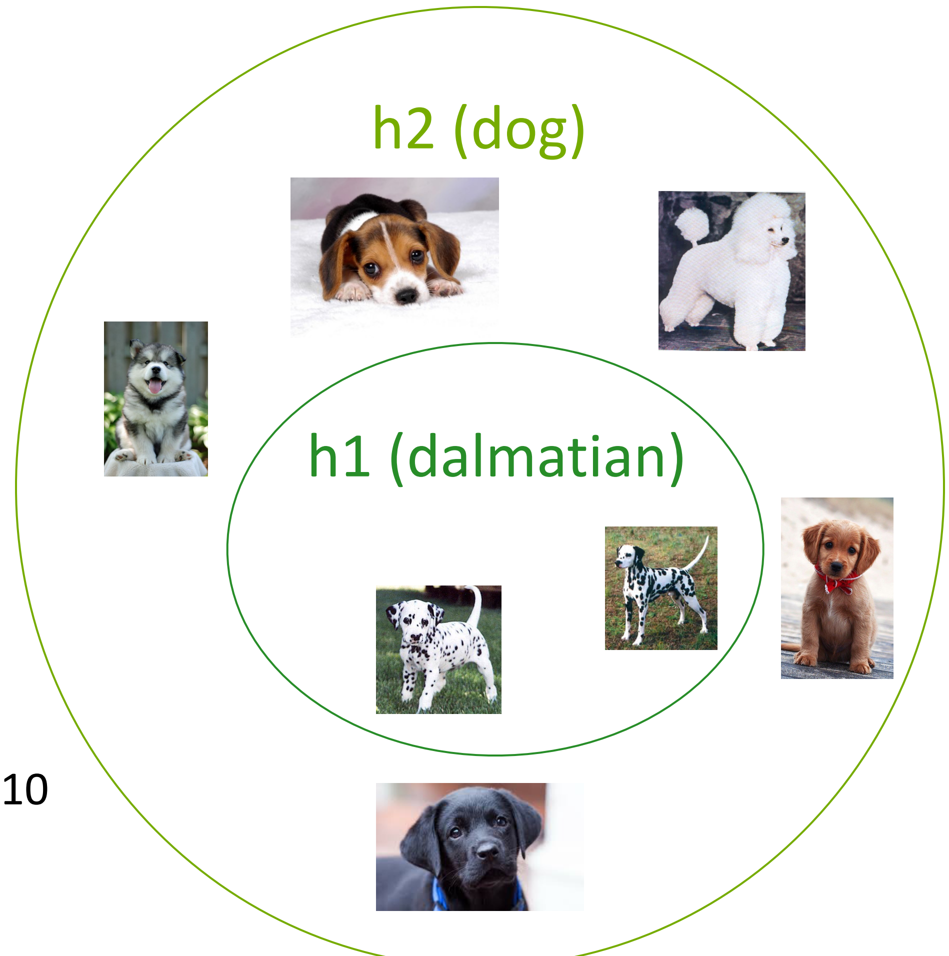
$$\propto P(D|h) * P(h)$$

Data D

P(h1 | D) ∝ **1/8**

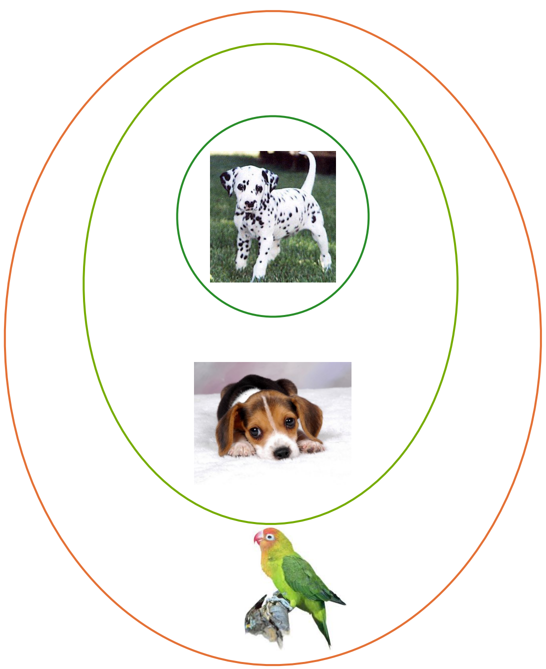
P(h2 | D) ∝ **1/98**

The more specific hypothesis h1 is over 10 times as probable as the more-general hypothesis h2, given these data!



Suspicious coincidences and children $P(h|D) \propto P(D|h) * P(h)$

Xu & Tenenbaum (2007) wanted to see if children have this kind of response to suspicious coincidences. If so, that means that they make specific generalizations when they encounter data that are compatible with multiple hypotheses about word meaning, in particular:



subordinate (least-general), ex: dalmatian

basic, ex: dog

superordinate (most-general), ex: animal

Testing children

Xu & Tenenbaum (2007)

Subjects: 3- and 4-year-old children

Task, part 1: Children were presented with three examples of a novel word ("blick", "fep", or "dax") during training.

("**This is a blick/fep/dax**")

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

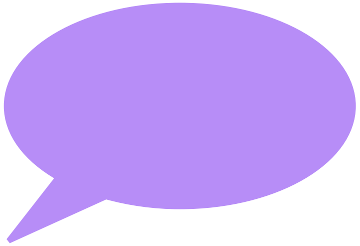
Task, part 1: "**This is a blick/fep/dax**"

There were three classes of stimuli: vegetables, vehicles, and animals.

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children



Learning

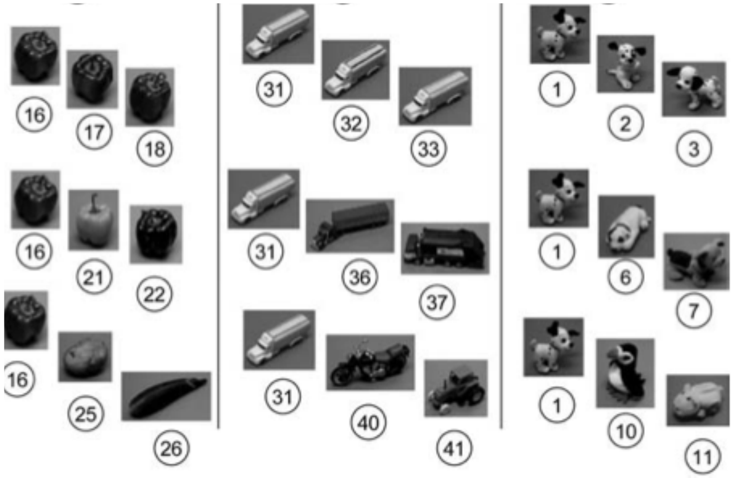
Task, part 1: "**This is a blick/fep/dax**"

Each class had these levels:

subordinate

basic

superordinate



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: “**This is a blick/fep/dax**”

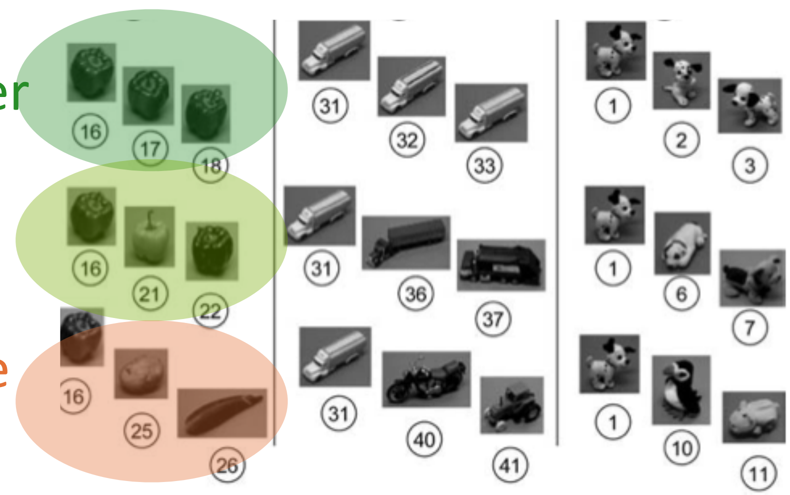
Each class had these levels:

vegetables vehicles animals

subordinate: green pepper

basic: pepper

superordinate: vegetable



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: “**This is a blick/fep/dax**”

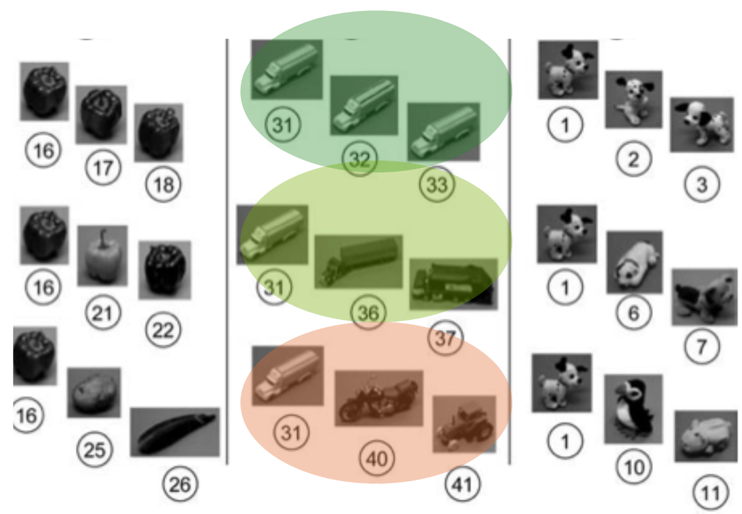
Each class had these levels:

subordinate: yellow truck

basic: truck

superordinate: vehicle

vegetables **vehicles** animals



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: “**This is a blick/fep/dax**”

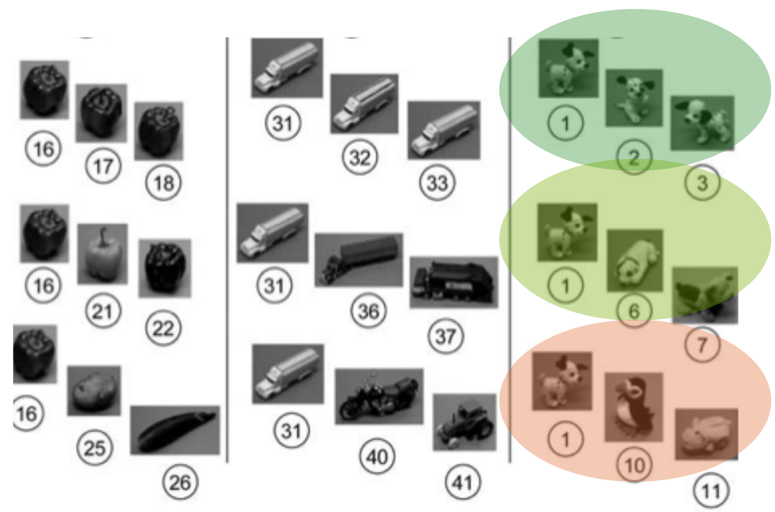
Each class had these levels:

subordinate: terrier

basic: dog

superordinate: animal

vegetables vehicles **animals**



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

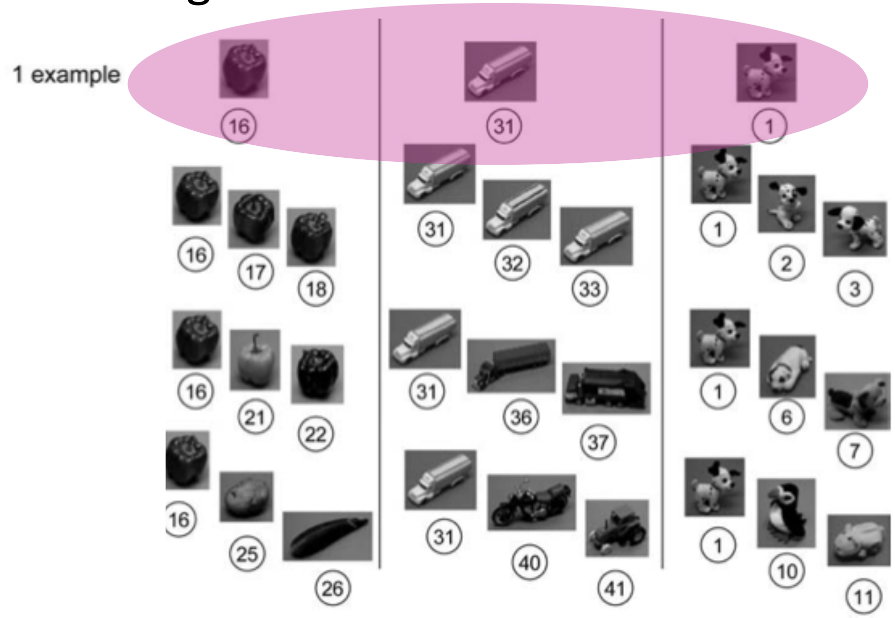
Learning

Task, part 1: “**This is a blick/fep/dax**”

There were 4 conditions

The **1-example** condition presented the same object & label three times.

vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

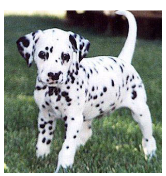
3- and 4-year-old children

Learning

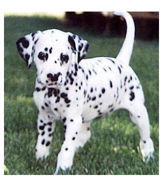
Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

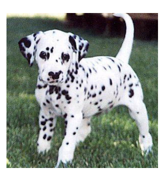
This is a fep



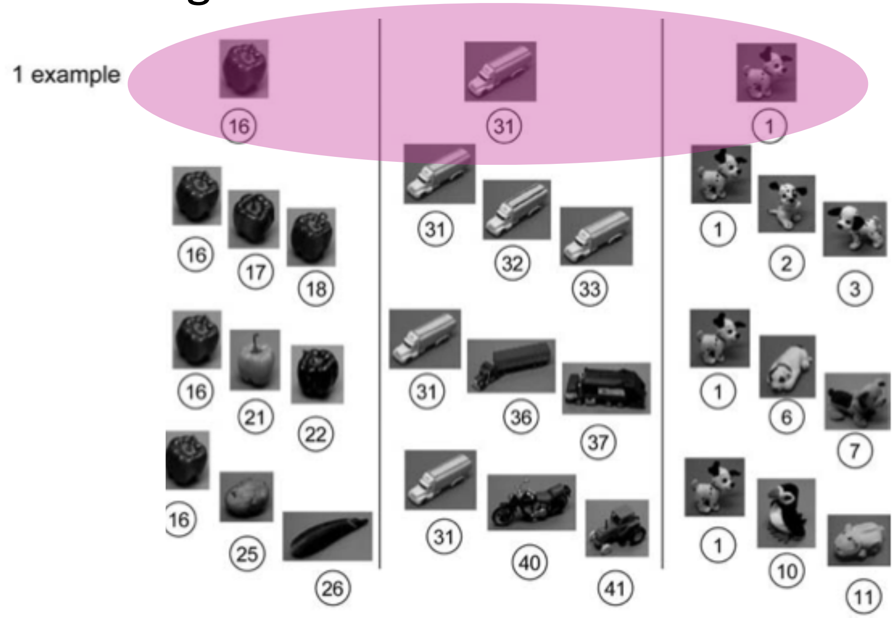
This is a fep



This is a fep



vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

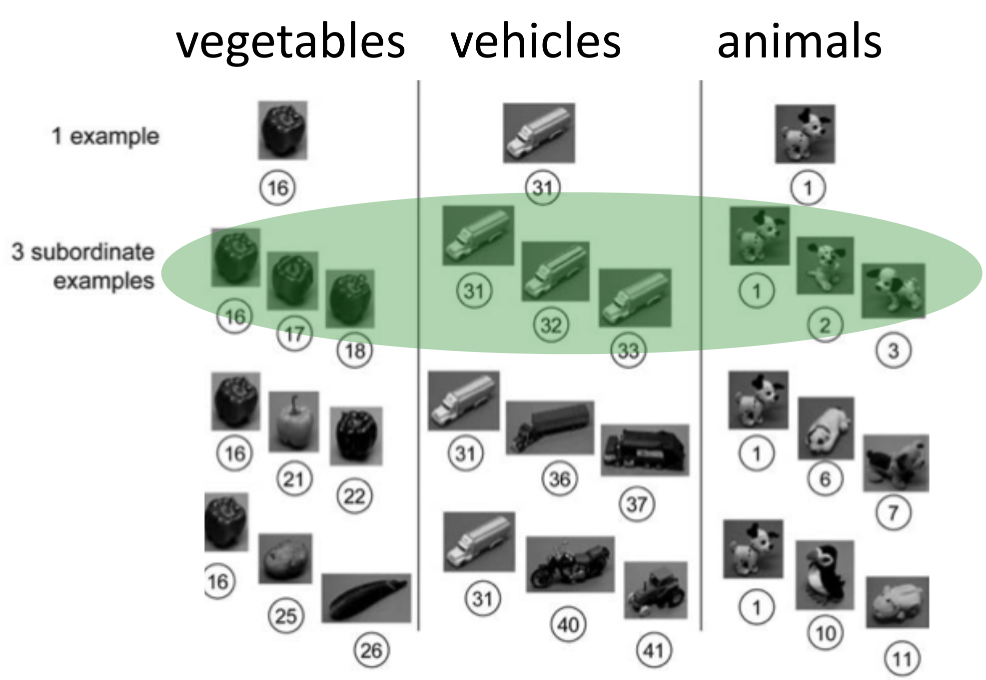
3- and 4-year-old children

Learning

Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

The **3-subordinate** example condition presented a subordinate object & label three times.



Testing children

Xu & Tenenbaum (2007)

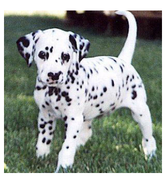
3- and 4-year-old children

Learning

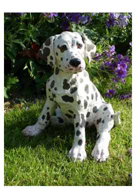
Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

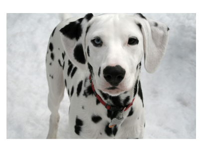
This is a fep



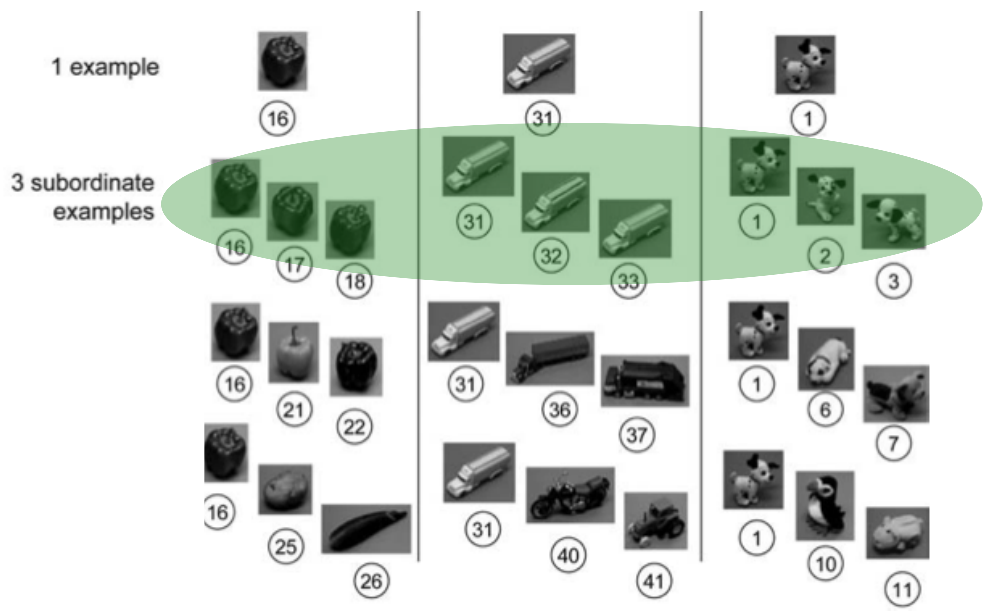
This is a fep



This is a fep



vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

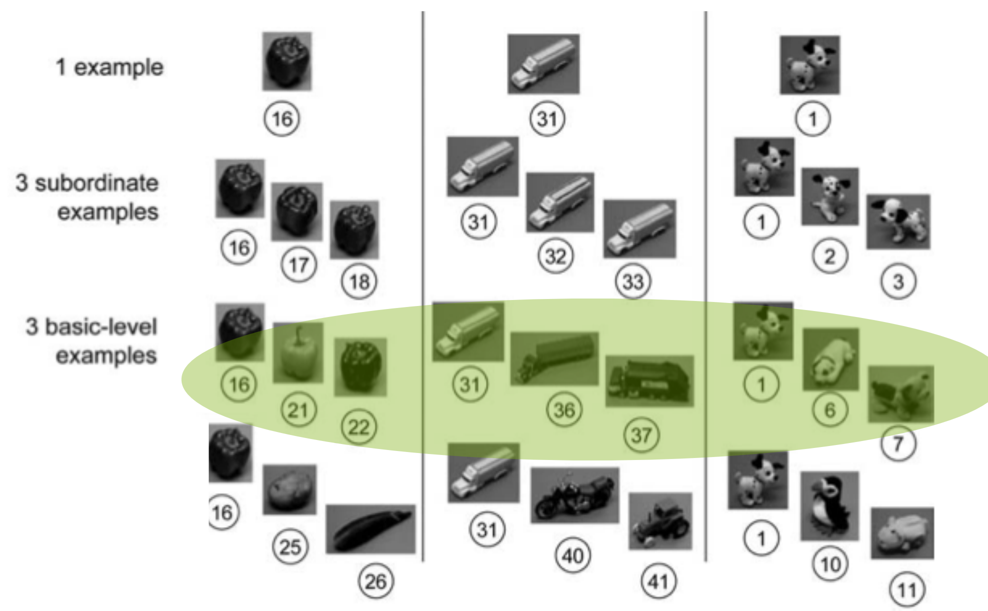
Learning

Task, part 1: “**This is a blick/fep/dax**”

There were 4 conditions

The **3-basic-level** example condition presented a basic-level object & label three times.

vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

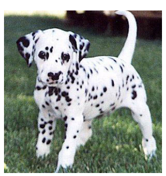
3- and 4-year-old children

Learning

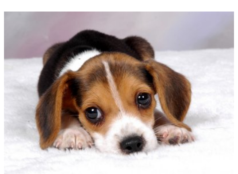
Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

This is a fep

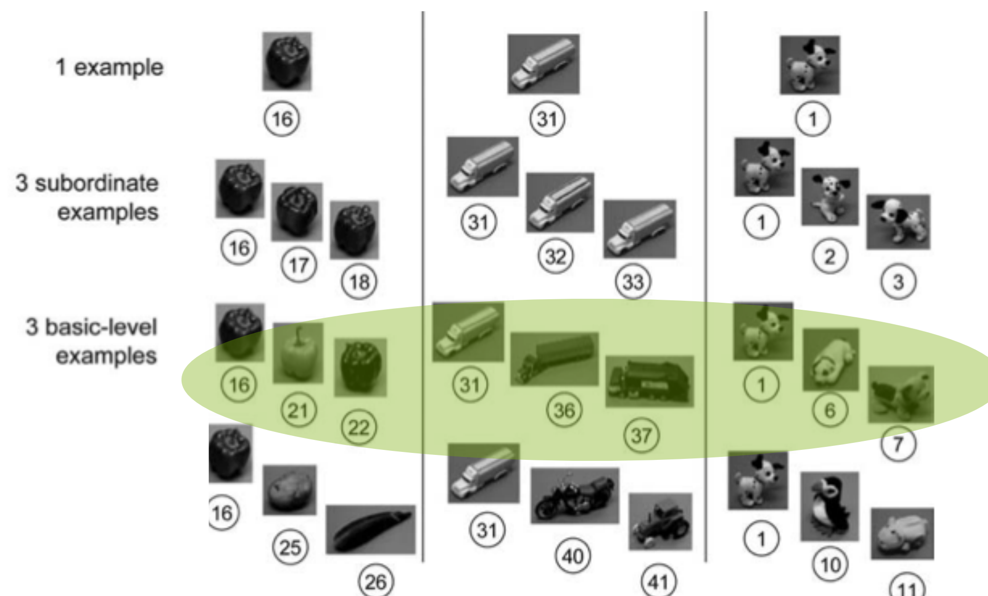


This is a fep



This is a fep

vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

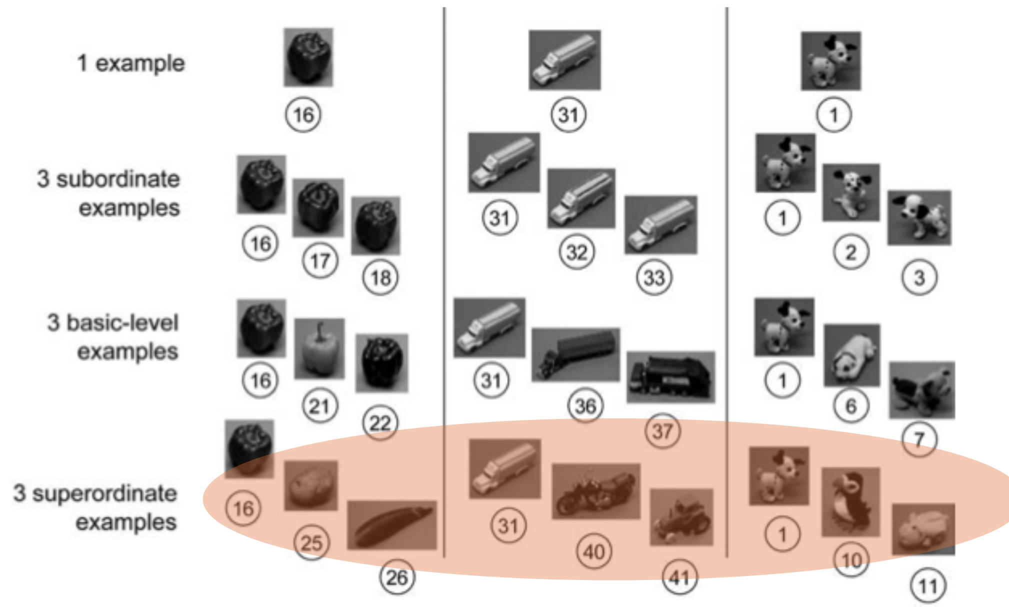
Learning

Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

The **3-superordinate** example condition presented a superordinate object & label three times.

vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

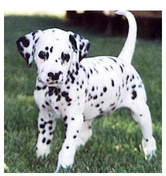
3- and 4-year-old children

Learning

Task, part 1: "**This is a blick/fep/dax**"

There were 4 conditions

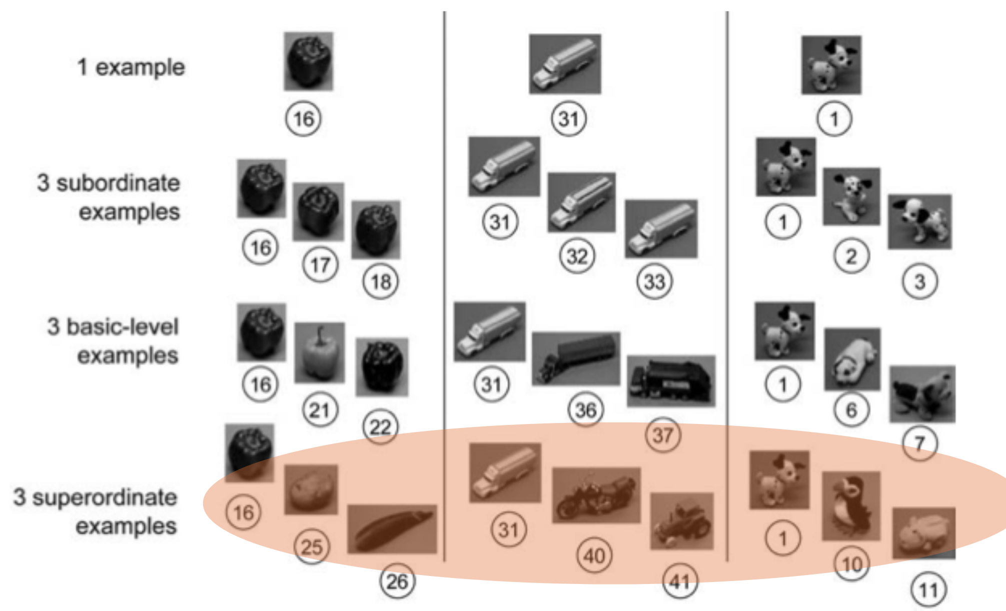
This is a fep



This is a fep

This is a fep

vegetables vehicles animals



Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify only things that are "blicks"/ "feps"/ "daxes" from a set of new objects

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

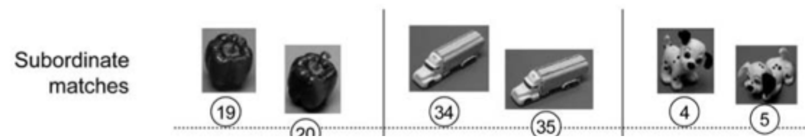
Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

There were three kinds of matches available:

vegetables vehicles animals



Subordinate matches (which were the least general, given the examples the children were trained on)

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

There were three kinds of matches available:

Basic-level matches (which were more general, given the examples the children were trained on)

vegetables vehicles animals

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

Learning

Task, part 1: “**This is a blick/fep/dax**”

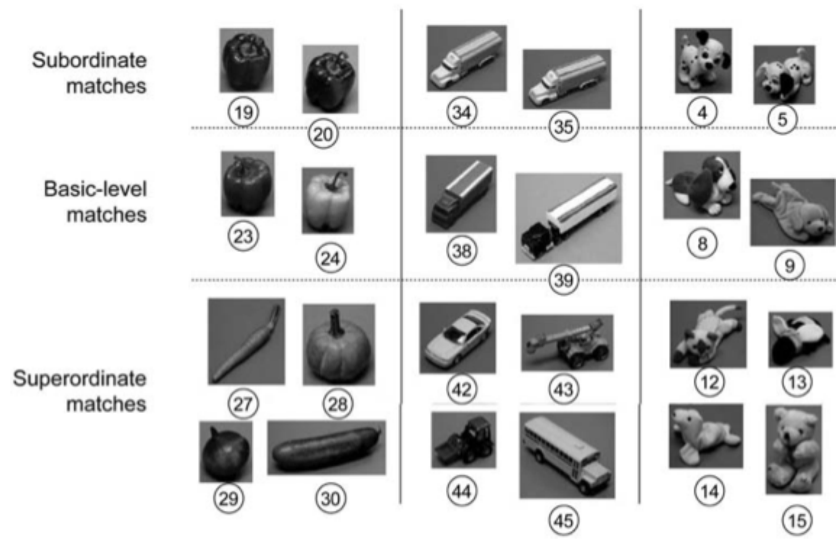
Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

There were three kinds of matches available:

Superordinate-level matches (which were the most general, given the examples the children were trained on)

vegetables vehicles animals

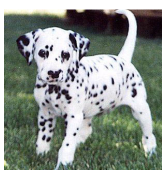


Testing children

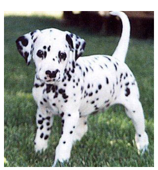
Xu & Tenenbaum (2007)

3- and 4-year-old children

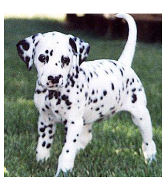
This is a fep



This is a fep



This is a fep



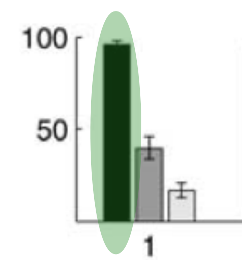
Learning

Task, part 1: "**This is a blick/fep/dax**"

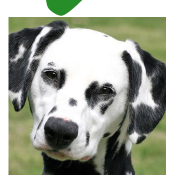
Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

When children heard a single example three times, they readily generalized to the subordinate class, but were less likely to generalize to the basic-level, and even less likely to generalize to the superordinate level.



✔



X

X

Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep

This is a fep

This is a fep

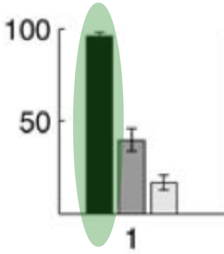
Learning

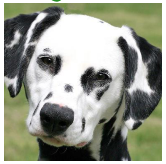
Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

This shows that young children are fairly conservative in their generalization behavior.



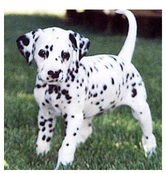


Testing children

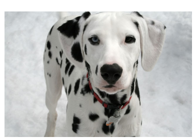
Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep



This is a fep



This is a fep

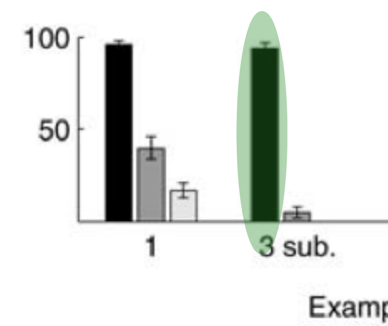
Learning

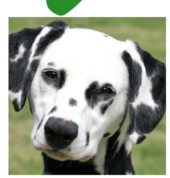
Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

When children had only subordinate examples as input, they readily generalized to the subordinate class, but almost never generalized beyond that.



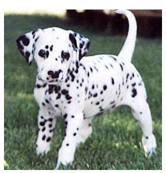


Testing children

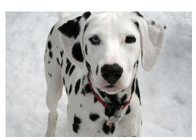
Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep



This is a fep



This is a fep

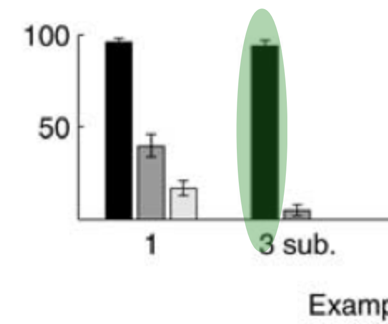
Learning

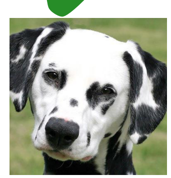
Task, part 1: “**This is a blick/fep/dax**”

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

They were sensitive to the suspicious coincidence, and chose the least-general hypothesis compatible with the data.





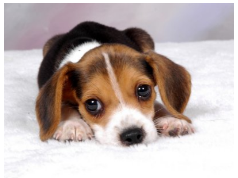
Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep

This is a fep



This is a fep

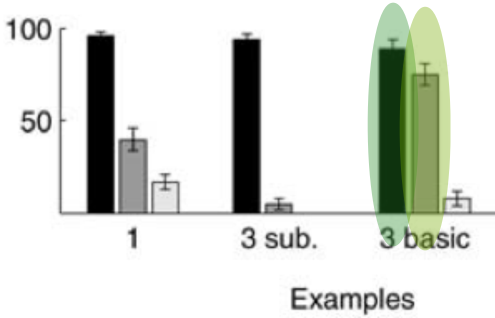
Learning

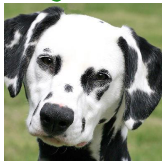
Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

When children had basic-level examples as input, they readily generalized to the subordinate class and the basic-level class, but almost never generalized beyond that.



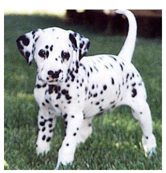


Testing children

Xu & Tenenbaum (2007)

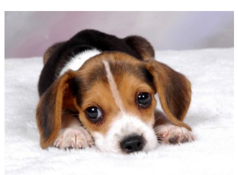
3- and 4-year-old children

This is a fep



This is a fep

This is a fep



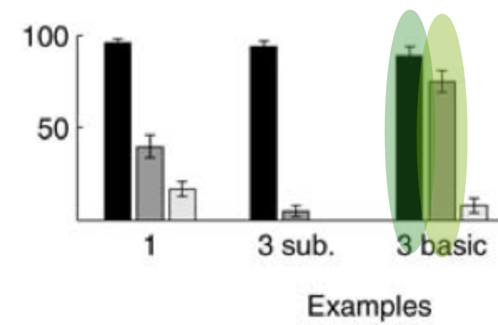
Learning

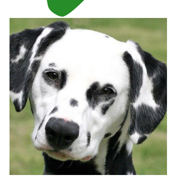
Task, part 1: “**This is a blick/fep/dax**”

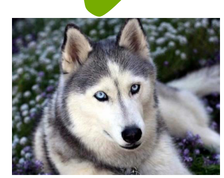
Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

They were again sensitive to the suspicious coincidence, and chose the least-general hypothesis compatible with the data.





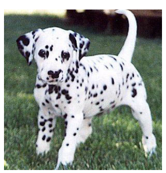


Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep



This is a fep

This is a fep

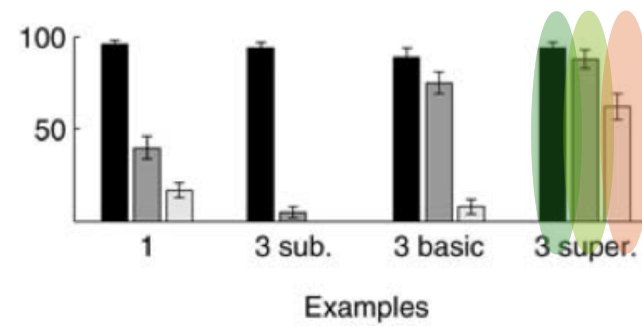
Learning

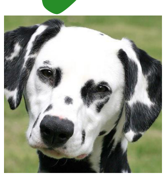
Task, part 1: "**This is a blick/fep/dax**"

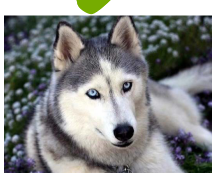
Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

When children had superordinate-level examples as input, they readily generalized to the subordinate class and the basic-level class, and often generalized to the superordinate class.





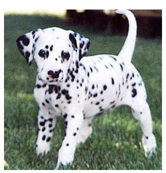


Testing children

Xu & Tenenbaum (2007)

3- and 4-year-old children

This is a fep



This is a fep

This is a fep

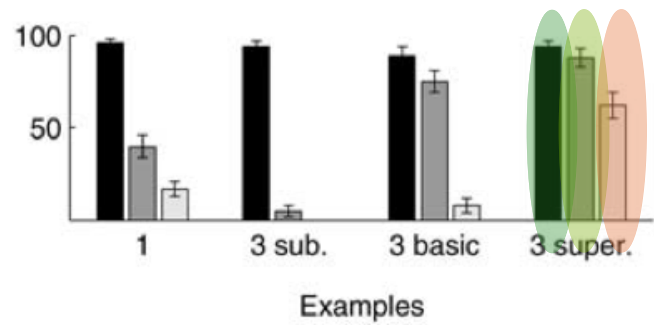
Learning

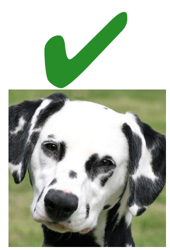
Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

They were again sensitive to the suspicious coincidence, though they were still a little uncertain how far to extend the generalization.





Testing children

Xu & Tenenbaum (2007)

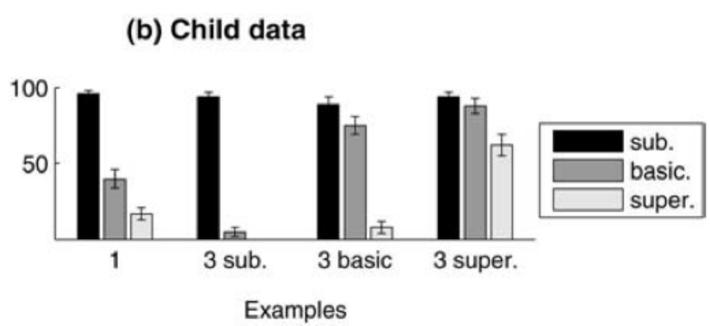
3- and 4-year-old children

Learning

Task, part 1: "**This is a blick/fep/dax**"

Generalization

Task, part 2: help Mr. Frog identify things from a set of new objects

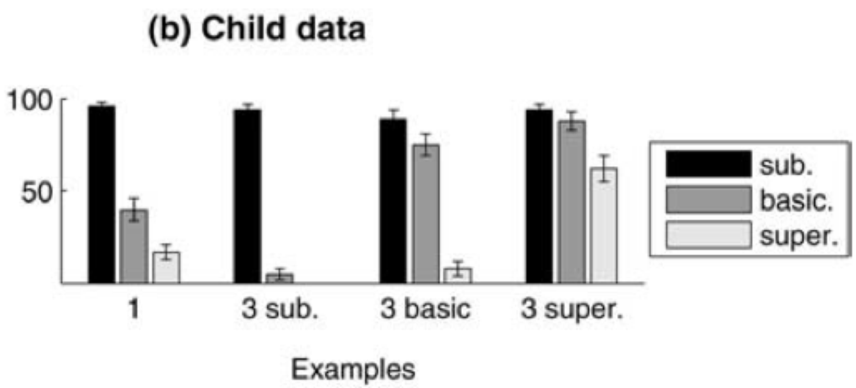


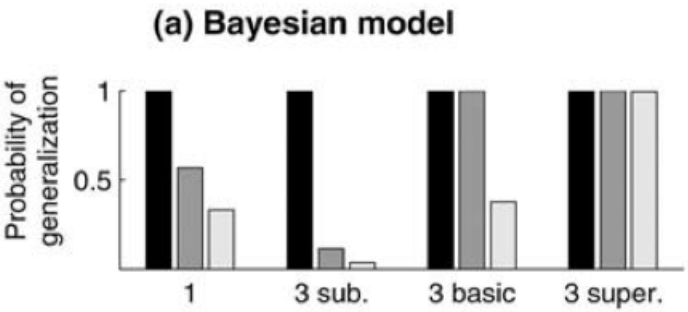
Matching children

$$P(h|D) \propto P(D|h) * P(h)$$

3- and 4-year-old children

Xu & Tenenbaum (2007) found that children's responses were best captured by a learning model that used Bayesian inference (and so was sensitive to suspicious coincidences).



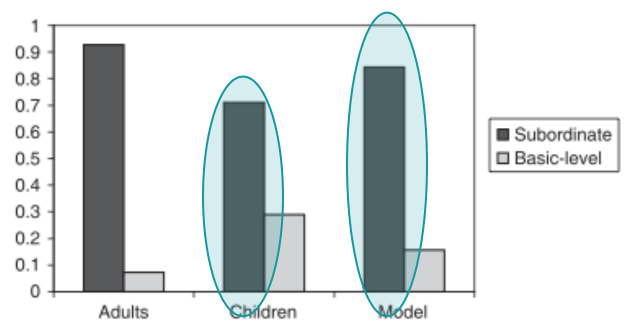


Children are sensitive to how the data are selected

$$P(h|D) \propto P(D|h) * P(h)$$

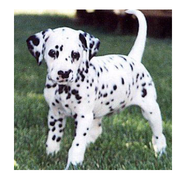
Like a Bayesian learner, children are also sensitive to how the data are selected (Xu & Tenenbaum 2007, *Developmental Science*).

If the child believes the data are randomly sampled from all the available data out there, it's a very strong suspicious coincidence that only subordinate-level items are selected. Subordinate-level is favored hypothesis.

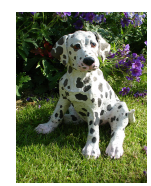


Picked at random…

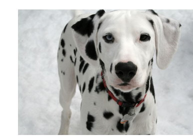
This is a fep

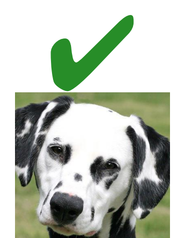


This is a fep



This is a fep



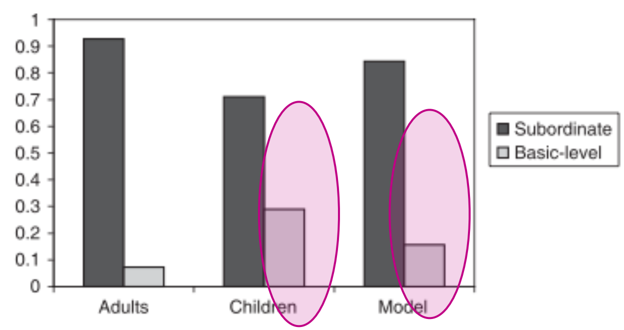


Children are sensitive to how the data are selected

$$P(h|D) \propto P(D|h) * P(h)$$

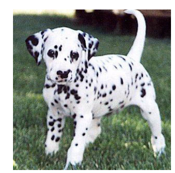
Like a Bayesian learner, children are also sensitive to how the data are selected (Xu & Tenenbaum 2007, *Developmental Science*).

If the child instead believes the data are selected because they're similar to each other, it's not a very suspicious coincidence that only subordinate-level items are selected. Basic-level is favored hypothesis.

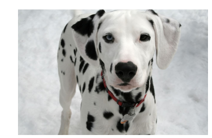


Picked not at random…

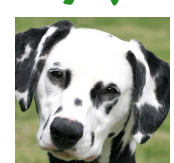
This is a fep



This is a fep



This is a fep



Children's adjective and noun learning are consistent with Bayesian inference

[Extra]

$$P(h|D) \propto P(D|h) * P(h)$$

Children can also use syntactic category information (like whether something is used as an adjective or a noun) to help make inferences about what the word means, in addition to the suspicious coincidences associated with the data selection.

(Gagliardi, Bennett, Lidz, & Feldman 2012)

"This is a *blicky* one." [Adjective use]

"This is a *blick*." [Noun use]

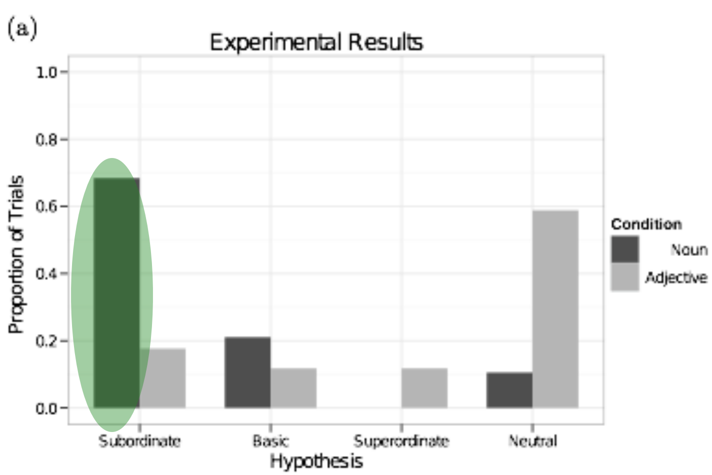
Children's adjective and noun learning are consistent with Bayesian inference

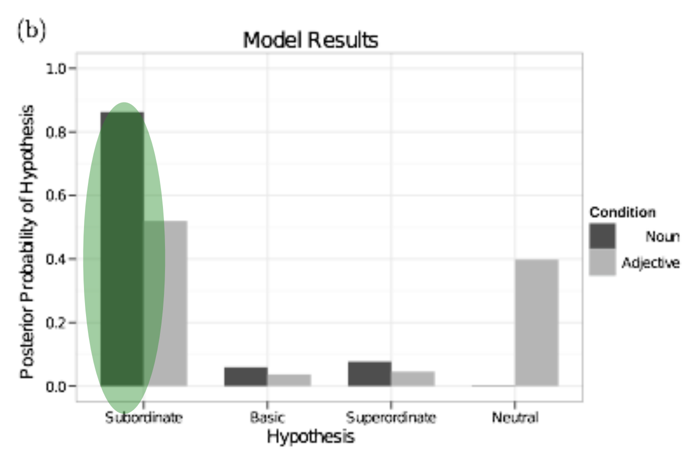
[Extra]

$$P(h|D) \propto P(D|h) * P(h)$$

Gagliardi, Bennett, Lidz, & Feldman 2012

Given 3 subordinate examples of a *blick*, children and the Bayesian model prefer *blick* to refer to the subordinate class only.





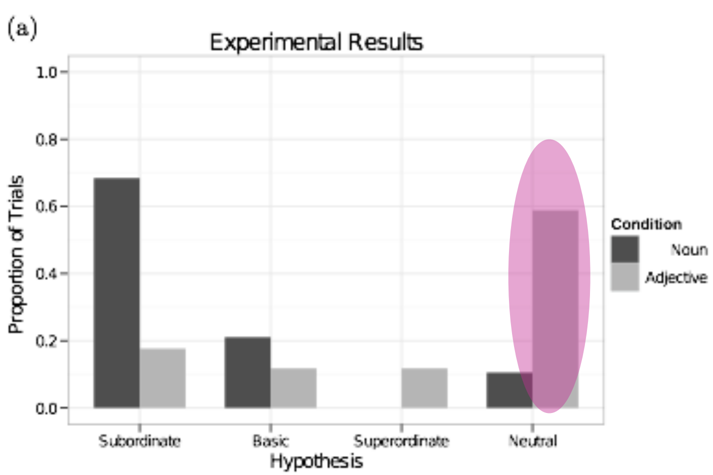
Children's adjective and noun learning are consistent with Bayesian inference

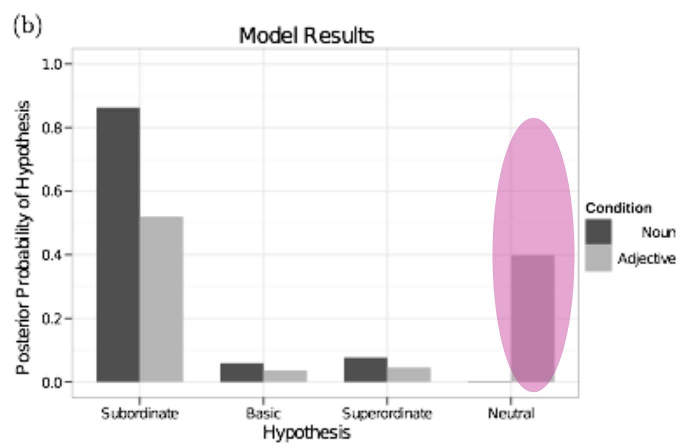
[Extra]

$$P(h|D) \propto P(D|h) * P(h)$$

Gagliardi, Bennett, Lidz, & Feldman 2012

Given 3 subordinate examples of a *blicky* one, children and the Bayesian model have considerable belief that *blicky* is neutral with respect to level, and simply represents the property…





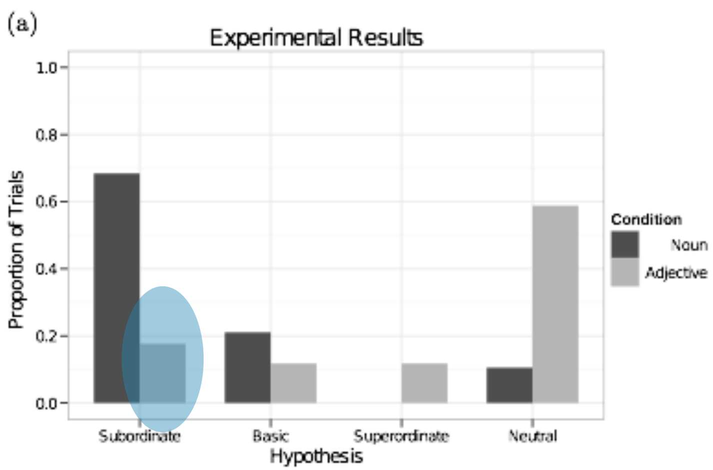
Children's adjective and noun learning are consistent with Bayesian inference

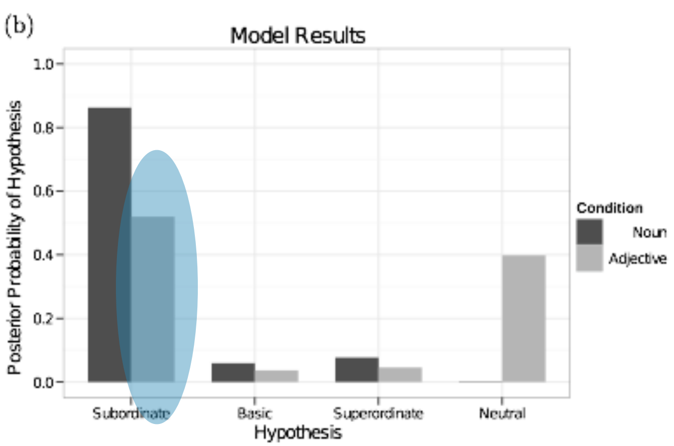
[Extra]

$$P(h|D) \propto P(D|h) * P(h)$$

Gagliardi, Bennett, Lidz, & Feldman 2012

…though the model still likes to pick up on the suspicious coincidence of the subordinate level, moreso than children do.





Accounting for other observed behavior

$$P(h|D) \propto P(D|h) * P(h)$$

How could a child using Bayesian inference make use of evidence like the following:

"That's a dalmatian. It's a kind of dog."

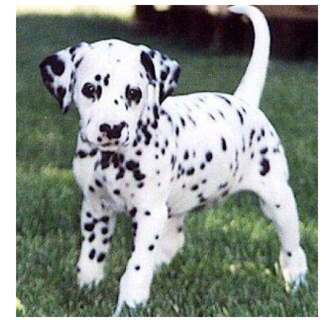
This explicitly tells children that this object can be labeled as both "dalmatian" and "dog", and moreover that "dog" is a more general term than "dalmatian".

Accounting for other observed behavior

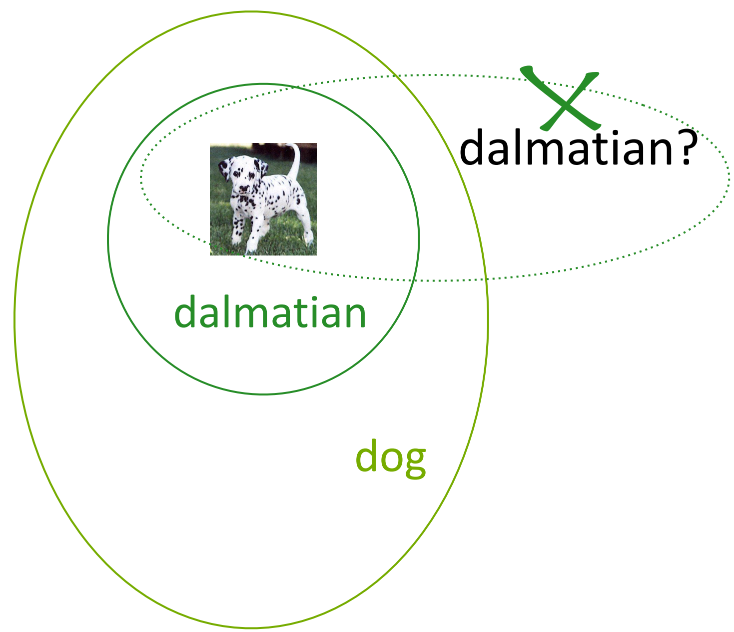
$$P(h|D) \propto P(D|h) * P(h)$$

How could a child using Bayesian inference make use of evidence like the following:

"That's a dalmatian. It's a kind of dog."

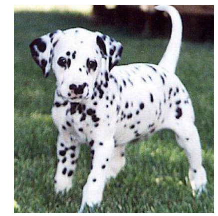


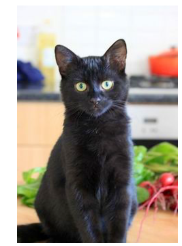
A Bayesian learner can treat this as conclusive evidence that *dalmatian* is a subset of *dog* and give 0 probability to any hypothesis where *dalmatian* is not contained within the set of *dogs*.



Accounting for other observed behavior

$$P(h|D) \propto P(D|h) * P(h)$$



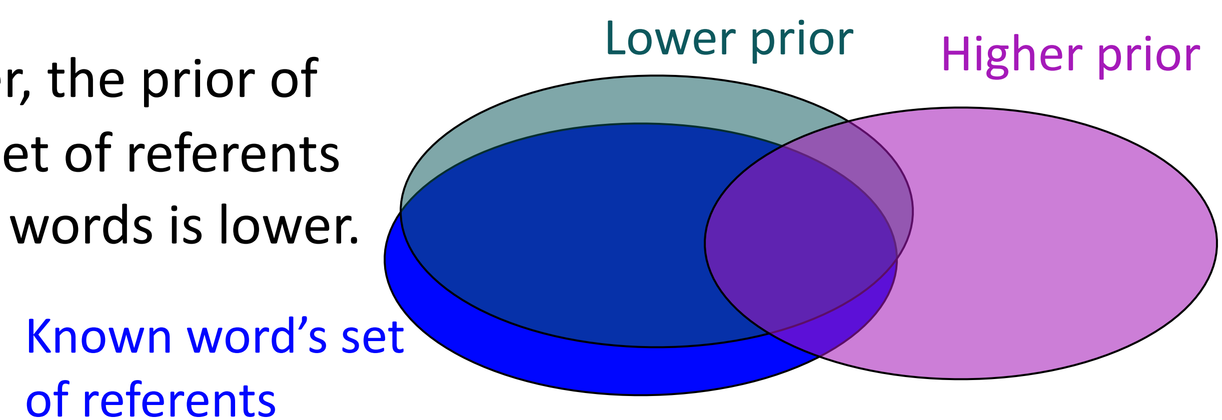


How could a child using Bayesian inference incorporate lexical contrast, where the meaning of all words must somehow differ?

This is particularly important when the child already knows some words like “dog”

(ex: “cat”, “puppy”, “pet”)

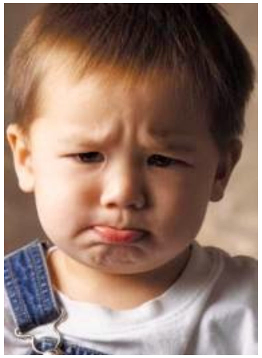
In a Bayesian learner, the prior of hypotheses whose set of referents overlap with known words is lower.



An open question

$$P(h|D) \propto P(D|h) * P(h)$$

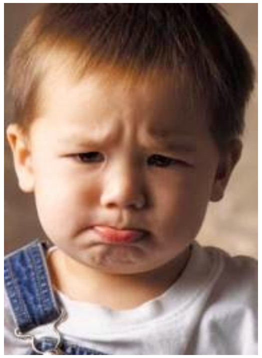
Early word-learning (younger than 3-years-old) appears to be slow & laborious – if children are using Bayesian inference, this shouldn't be the case. Why would this occur?



An open question

$$P(h|D) \propto P(D|h) * P(h)$$

Early word-learning (younger than 3-years-old) appears to be slow & laborious - why?



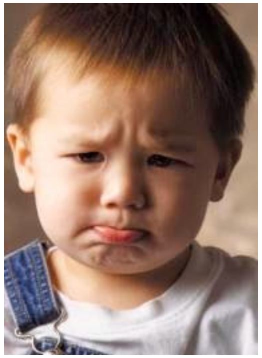
Potential explanation:

(1) Bayesian inference capacity isn't yet active in early word-learners. Even though older children (such as the ones tested in Xu & Tenenbaum (2007)) can use this ability, younger children cannot.

An open question

$$P(h|D) \propto P(D|h) * P(h)$$

Early word-learning (younger than 3-years-old) appears to be slow & laborious - why?



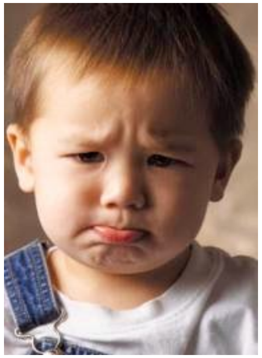
Potential explanation:

(2) The hypothesis spaces of young children may not be sufficiently constrained to make strong inferences. For example, even though adults know that the set of dogs is much larger than the set of dalmatians, young children may not know this - especially if their family dog is a dalmatian, and they don't know many other dogs.

An open question

$$P(h|D) \propto P(D|h) * P(h)$$

Early word-learning (younger than 3-years-old) appears to be slow & laborious - why?



Potential explanation: [related idea]

(2) The hypothesis spaces of young children may have different constraints on them than the hypothesis spaces of adults, older children, or even children further along in linguistic development.

Differences between early & late-talkers

$$P(h|D) \propto P(D|h) * P(h)$$

Colunga & Sims 2017:

There are different word-learning biases at play between "early-talkers" with larger vocabularies and "late-talkers" with smaller vocabularies. This is because the words they already know constrain their hypotheses about what other words could refer to.

18- to 22-month-olds

early talkers:

vocab = 151 to 526 words

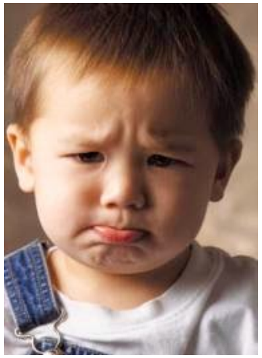
late talkers:

vocab = 9 to 82 words

An open question

$$P(h|D) \propto P(D|h) * P(h)$$

Early word-learning (younger than 3-years-old) appears to be slow & laborious - why?



Potential explanation:

(3) Young children's ability to remember words and/or their referents isn't stable. That is, even if someone points out a dalmatian to a child, the child can't remember the word form or the referent long enough to use that word-meaning mapping as intake. (Remember - there's a lot going on in children's worlds, and they have limited cognitive resources!) This makes the child's input much less informative than that same input would be to an adult.

An open question

$$P(h|D) \propto P(D|h) * P(h)$$

Frank, Lewis, & MacDonald 2016

An explicit model of children's very early word-learning suggests the changes in children's word-learning behavior can be "attributable to more gradual changes in processing abilities".

"The conclusion of our analysis is that even if young infants were trying to learn in precisely the same way as older toddlers, they would be too slow and too fallible to extract much signal from their input data."

Changes over time

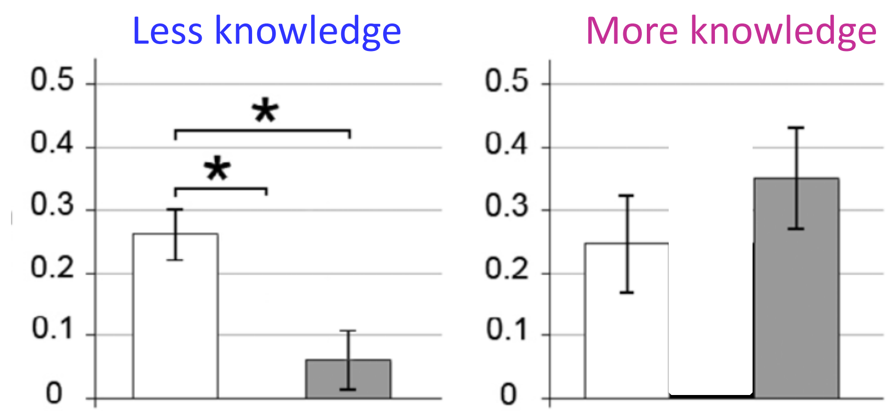
$$P(h|D) \propto P(D|h) * P(h)$$

As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015:

The Bayesian model from Xu & Tenenbaum (2007) predicts that the suspicious coincidence effect should get stronger as more subordinate (ex: dalmatian) and basic-level (ex: dog) members are learned.

But they found that children with more knowledge of category members demonstrated less sensitivity to suspicious coincidences!



Changes over time

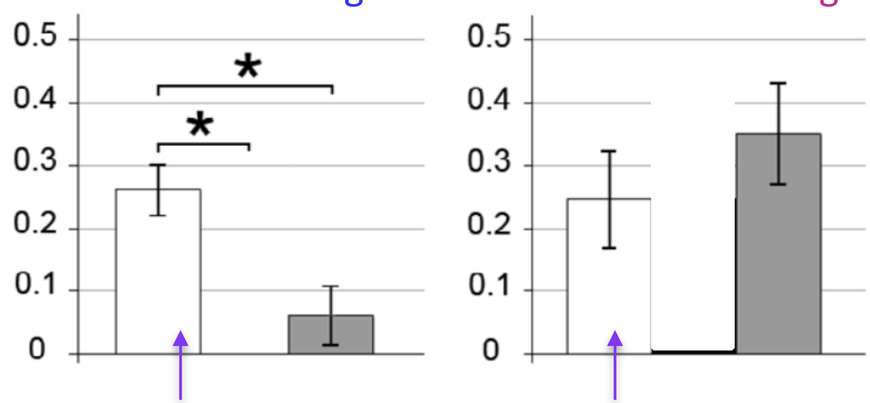
$$P(h|D) \propto P(D|h) * P(h)$$

As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015:

The Bayesian model from Xu & Tenenbaum (2007) predicts that the suspicious coincidence effect should get stronger as more subordinate (ex: dalmatian) and basic-level (ex: dog) members are learned.

But they found that children with more knowledge of category members demonstrated less sensitivity to suspicious coincidences!



When given one example of a "fep", both kinds of children generalize to the basic-level category about the same amount. This is their basic-level bias.

Changes over time

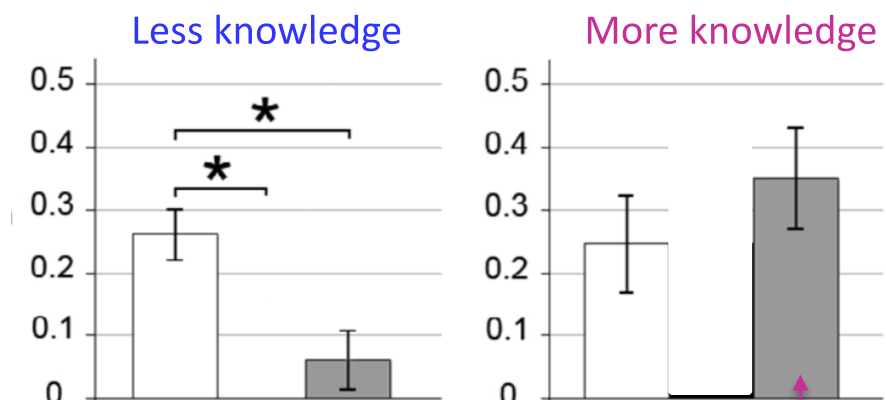
$$P(h|D) \propto P(D|h) * P(h)$$

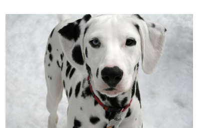
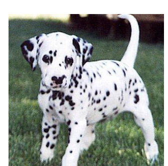
As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015:

The Bayesian model from Xu & Tenenbaum (2007) predicts that the suspicious coincidence effect should get stronger as more subordinate (ex: dalmatian) and basic-level (ex: dog) members are learned.

But they found that children with more knowledge of category members demonstrated less sensitivity to suspicious coincidences!



When given three different subordinate examples of "feps", children with more category member knowledge *still* generalized to the basic-level.

Changes over time

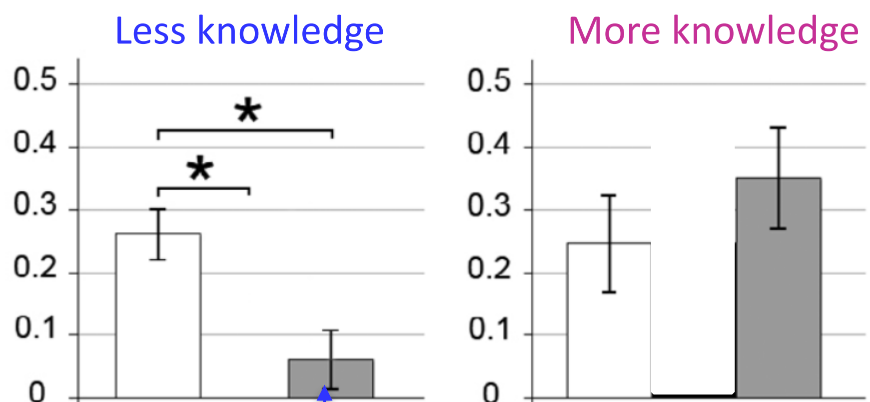
$$P(h|D) \propto P(D|h) * P(h)$$

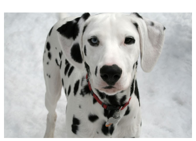
As children acquire more knowledge, does their word-learning behavior change over time?

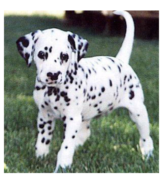
Jenkins et al. 2015:

The Bayesian model from Xu & Tenenbaum (2007) predicts that the suspicious coincidence effect should get stronger as more subordinate (ex: dalmatian) and basic-level (ex: dog) members are learned.

But they found that children with more knowledge of category members demonstrated less sensitivity to suspicious coincidences!







Meanwhile, children with less category member knowledge were sensitive to the suspicious coincidence and didn't generalize.

Changes over time

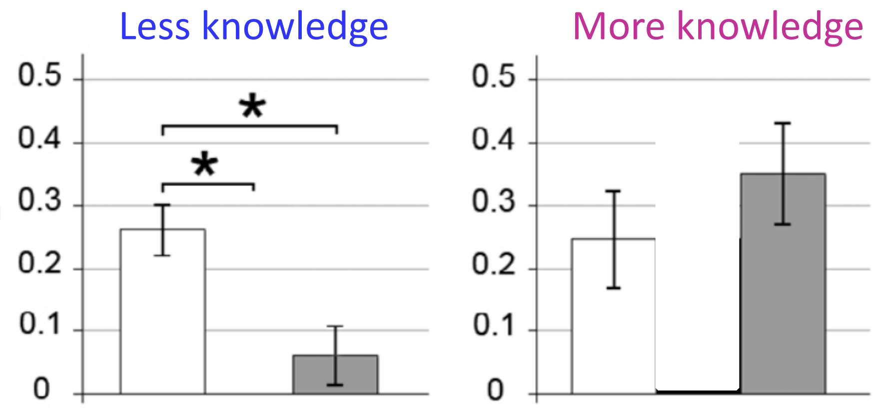
$$P(h|D) \propto P(D|h) * P(h)$$

As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015:

The Bayesian model from Xu & Tenenbaum (2007) predicts that the suspicious coincidence effect should get stronger as more subordinate (ex: dalmatian) and basic-level (ex: dog) members are learned.

What's going on?



Changes over time

As children acquire more knowledge, does their word-learning behavior change over time?

$$P(h|D) \propto P(D|h) * P(h)$$

Jenkins et al. 2015: What this means

"…the Bayesian model in isolation and in its current form cannot capture" this behavior.

One idea: The influence of language experience

"One possibility is that children with greater category knowledge might have learned that, in general, subordinate level categories are labeled with compound labels, like "sheepdog," "delivery truck" or "Bell pepper." Basic-level categories, on the other hand, tend to have single morpheme labels like "dog," "truck," and "pepper."

Changes over time

$$P(h|D) \propto P(D|h) * P(h)$$

As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015: What this means

"…the Bayesian model in isolation and in its current form cannot capture" this behavior.

One idea: The influence of language experience

In child-directed speech, Jenkins et al. found that compound nouns are subordinate-level categories nearly 3 times out of 4, while single morpheme labels are basic-level categories nearly 95 times out of 100.

Changes over time

$$P(h|D) \propto P(D|h) * P(h)$$

As children acquire more knowledge, does their word-learning behavior change over time?

Jenkins et al. 2015: What this means

"…the Bayesian model in isolation and in its current form cannot capture" this behavior.

One idea: The influence of language experience

Therefore, when the more experienced child hears "fep", she assumes it's a basic-level item.

Recap

Word learning is difficult because many words refer to concepts that can overlap in the real world. This means that there isn't just one word for every thing in the world - there are many words, each picking out a different aspect of that thing.

Bayesian learning may be a strategy that can help children overcome this difficulty, and experimental evidence suggests that their behavior is consistent with a Bayesian learning strategy.

However, Bayesian learning may not be active or help sufficiently at the very earliest stages of word-learning, given other constraints children have.

Also, children's sensitivity to suspicious coincidences changes over time, and may be impacted by other linguistic cues they can use to figure out what a word means.

Questions?

You should be able to do up through 2 on HW2
and up through 15 on the word meaning review questions.