Psych 156A/ Ling 150: Acquisition of Language II

Lecture 9

Word meaning 2

Announcements

Be working on HW2 (due 5/5/16)

In-class midterm review 4/28/16 — Come with questions!

Midterm during class 5/3/16

Computational problem

"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.

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.

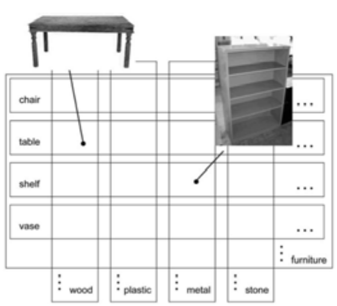
"I love my teddy."

"He's my favorite toy."

"He's brown and cuddly."

What we know about the process of word learning

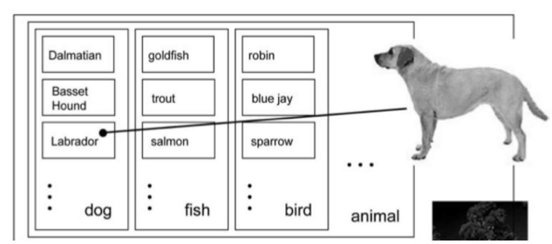
(3) The target of word-learning is a system of overlapping concepts. That is, words pick out different aspects of our world, and it's often the case that different words can refer to the same observable thing in the world.



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 abut 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 a potential hypothesis space for *dog*:

dog = dog parts, front half of dog, dog spots, all spotted things, all running things, all dogs + one cat

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*)

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*)

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

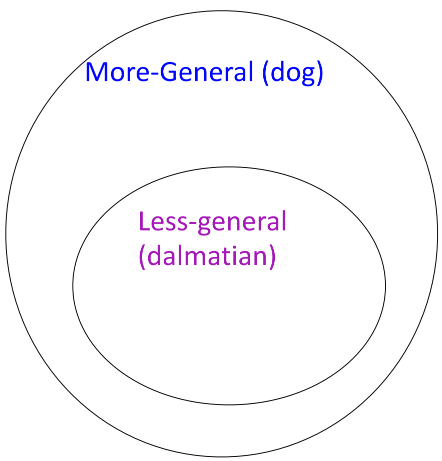
Suspicious: Why is no other animal or other kind of dog a *fep* if *fep* can really label any animal or any kind of dog?

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

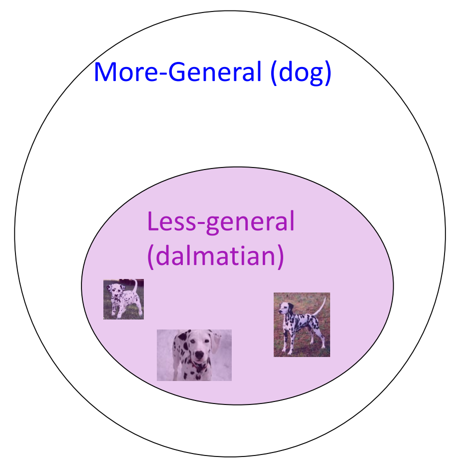
If the more-general generalization (dog) is correct, the learner should encounter some data that can only be accounted for by the more-general generalization (like beagles or poodles). These data would be incompatible with the less-general generalization (dalmatian).



Formal instantiation of "suspicious coincidence"

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

If the learner keeps *not* encountering data compatible only with the more-general generalization, the less-general generalization becomes more and more likely to be the generalization responsible for the language data encountered.



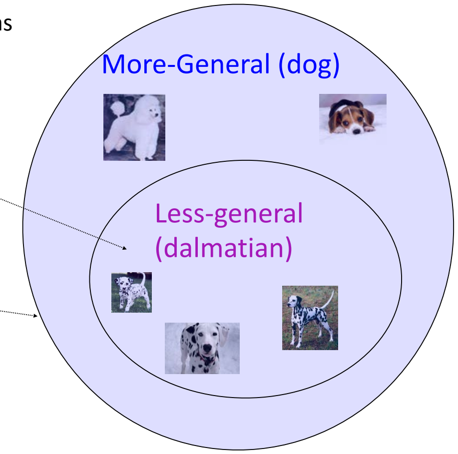
Formal instantiation of "suspicious coincidence"

Another way to think about it: probability of generating data points

Suppose there are only 5 dogs in the world that we know about, as shown in this diagram.

Hypothesis 1 (H1): The less-general hypothesis is true, and *fep* means dalmatian.

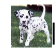
Hypothesis 2 (H2): The more-general hypothesis is true, and *fep* means dog.

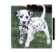


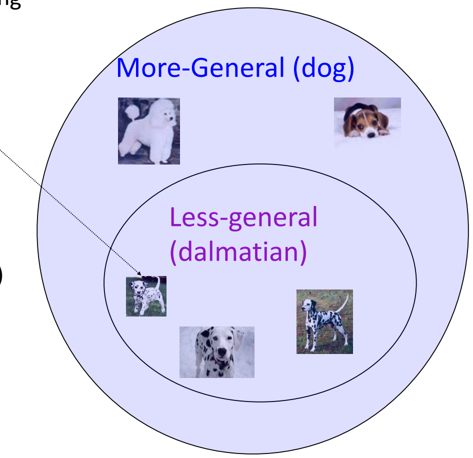
Formal instantiation of "suspicious coincidence"

Another way to think about it: probability of generating data points

What's the likelihood of selecting this dog for each hypothesis?

p(| H1) = 1/3
(since only three dogs are possible)

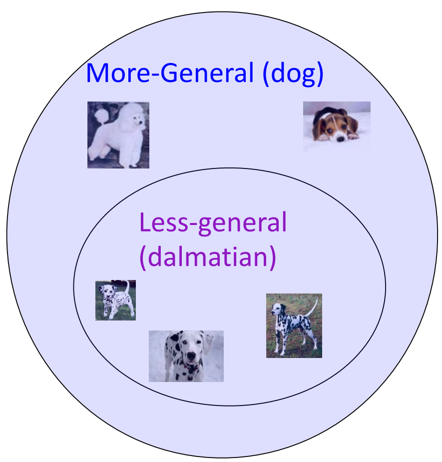
p(| H2) = 1/5
(since all five dogs are possible)



Formal instantiation of "suspicious coincidence"

Another way to think about it: probability of generating data points

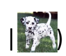
This means the likelihood for the less-general hypothesis is always going to be larger than the likelihood of the more-general hypothesis for data points that both hypotheses can account for.

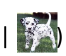


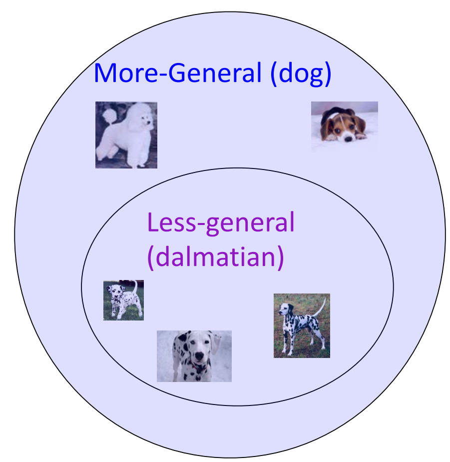
Formal instantiation of "suspicious coincidence"

Another way to think about it: probability of generating data points

If the prior is equal (ex: before any data, both hypotheses are equally likely), then the posterior probability will be greater for the less-general hypothesis.

p(H1 | ) ∝ p(| H1) * p(H1)
∝ 1/3 * p(H1)

p(H2 | ) ∝ p(| H2) * p(H2)
∝ 1/5 * p(H2)



Suspicious coincidences and children

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

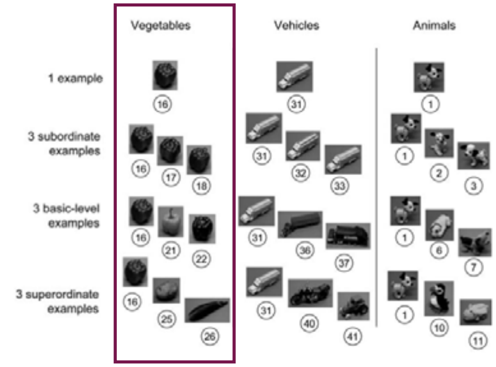
Testing children

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

The vegetable class had these levels:

subordinate: green pepper
basic: pepper
superordinate: vegetable



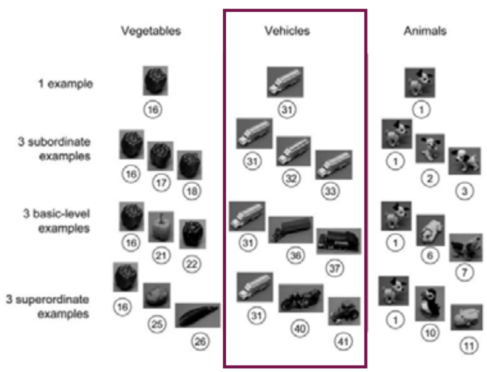
Testing children

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

The vehicle class had these levels:

subordinate: yellow truck
basic: truck
superordinate: vehicle



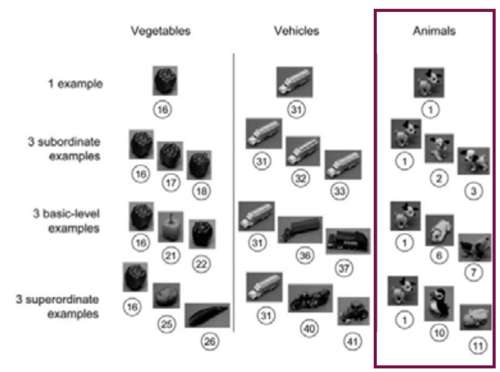
Testing children

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

The animal class had these levels:

subordinate: terrier
basic: dog
superordinate: animal



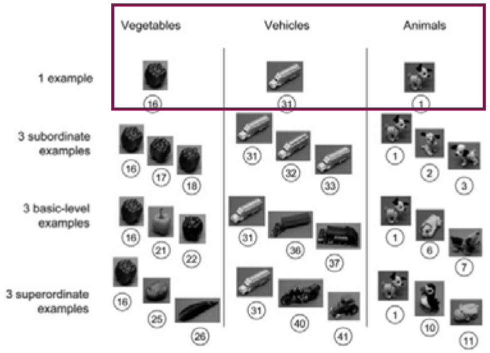
Testing children

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

There were four conditions:

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



Testing children

Subjects: 3- and 4-year-old children

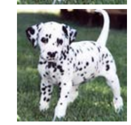
Task, part 1: Children were presented with three examples of a novel word ("blick", "fep", or "dax") during training. ("This is a blick/fep/dax") There were three classes of stimuli: vegetables, vehicles, and animals.

There were four conditions:

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

"This is a fep."

"This is a fep."



"This is a fep."

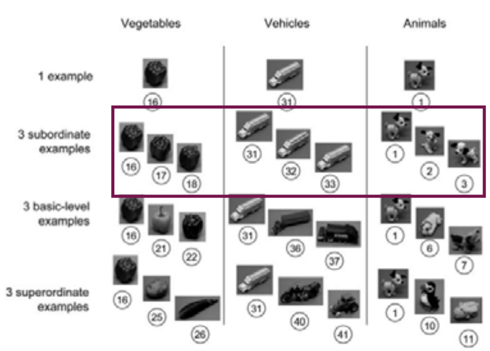
Testing children

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

There were four conditions:

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



Testing children

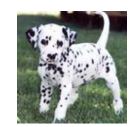
Subjects: 3- and 4-year-old children

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

There were four conditions:

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

"This is a fep."



"This is a fep."

"This is a fep."

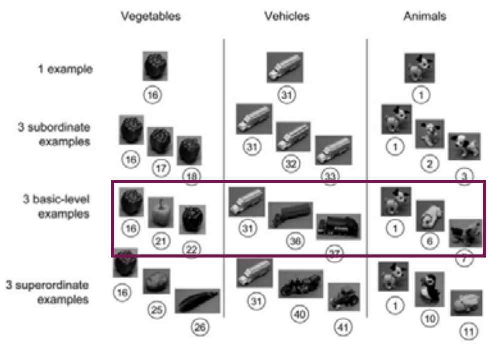
Testing children

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

There were four conditions:

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



Testing children

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

There were four conditions:

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

“This is a fep.”

“This is a fep.”

“This is a fep.”

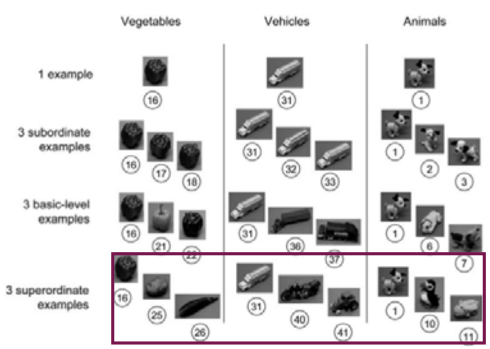
Testing children

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

There were four conditions:

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



Testing children

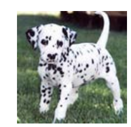
Subjects: 3- and 4-year-old children

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

There were four conditions:

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

“This is a fep.”



“This is a fep.”

“This is a fep.”

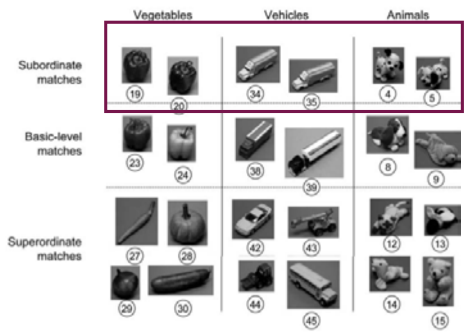
Testing children

Subjects: 3- and 4-year-old children

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

There were three kinds of matches available:

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



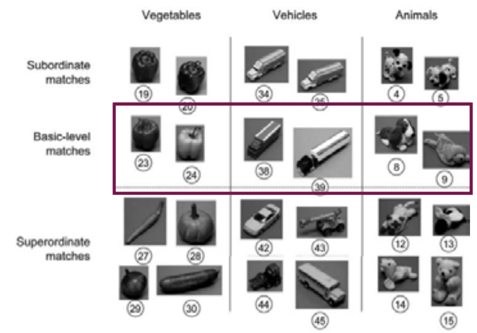
Testing children

Subjects: 3- and 4-year-old children

Task, part 2: generalization (asked to help Mr.Frog identify only things that are "blicks"/ "feps"/ "daxes" 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)



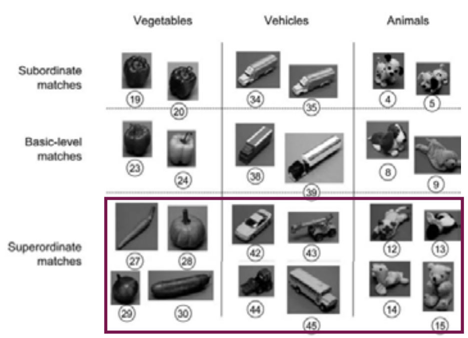
Testing children

Subjects: 3- and 4-year-old children

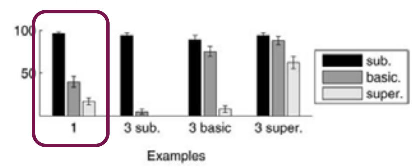
Task, part 2: generalization (asked to help Mr.Frog identify only things that are "blicks"/ "feps"/ "daxes" from a set of new objects)

There were three kinds of matches available:

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

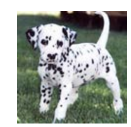


Children's generalizations

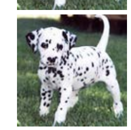


When children heard a single example three times, they readily generalized to the subordinate class, but were less likely to generalize to the basic-level, and even less likely to generalize to the superordinate level. This shows that young children are fairly conservative in their generalization behavior.

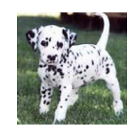
"This is a fep."



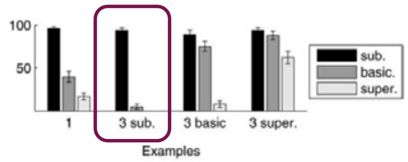
"This is a fep."



"This is a fep."



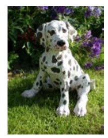
Children's generalizations



When children had only subordinate examples as input, they readily generalized to the subordinate class, but almost never generalized beyond that. They were sensitive to the suspicious coincidence, and chose the least-general hypothesis compatible with the data.

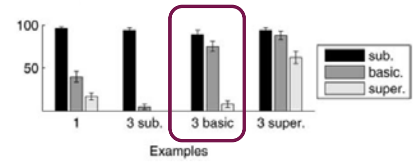
"This is a fep."

"This is a fep."



"This is a fep."

Children's generalizations



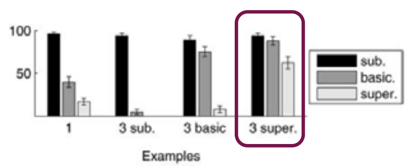
When children had basic-level examples as input, they readily generalized to the subordinate class and the basic-level class, but almost never generalized beyond that. They were again sensitive to the suspicious coincidence, and chose the least-general hypothesis compatible with the data.

"This is a fep."

"This is a fep."

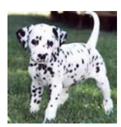
"This is a fep."

Children's generalizations



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

"This is a fep."

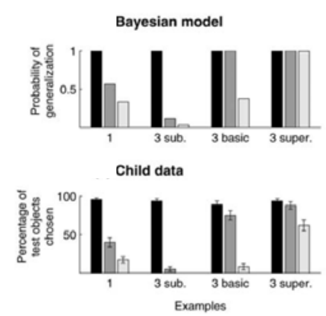


"This is a fep."

"This is a fep."

Modeling children's responses

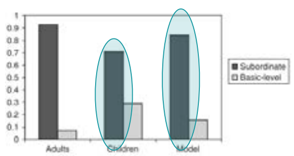
Xu & Tenenbaum (2007) found that children's responses were best captured by a learning model that used Bayesian inference (and so was sensitive to suspicious coincidences).



Children are sensitive to how the data are selected

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

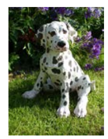
If the child believes the data are randomly sampled from the all the available data out there, it's a very strong suspicious coincidence that only subordinate-level items are selected. Subordinate-level is hypothesis.



Picked at random…

"This is a fep."

"This is a fep."

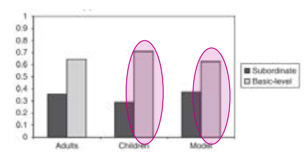


"This is a fep."

Children are sensitive to how the data are selected

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 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

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]

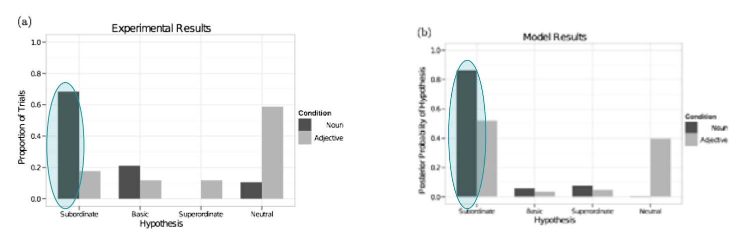
"This is a *blick*." [Noun use]

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

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)

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

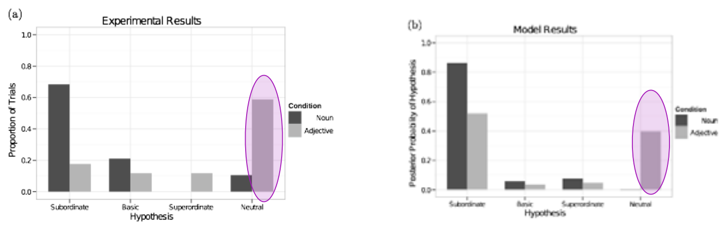


Children's adjective and noun learning are consistent with Bayesian inference

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)

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…

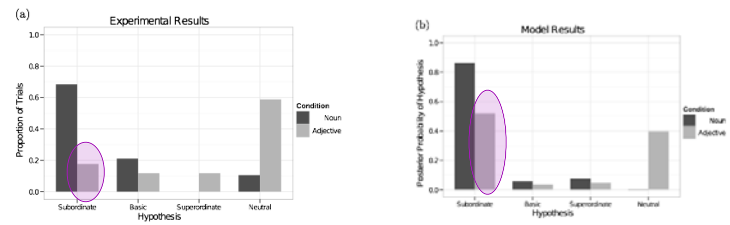


Children's adjective and noun learning are consistent with Bayesian inference

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)

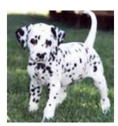
…though the model still likes to pick up on the suspicious coincidence of the subordinate level, moreso than children do.



Accounting for other observed behavior

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

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

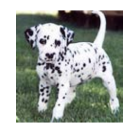


This explicitly tells children that this object can be labeled as both "dalmatian" and "dog", and moreover that "dog" is a more general term than "dalmatian".

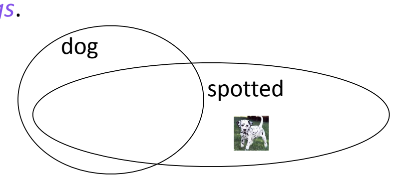
Accounting for other observed behavior

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*.



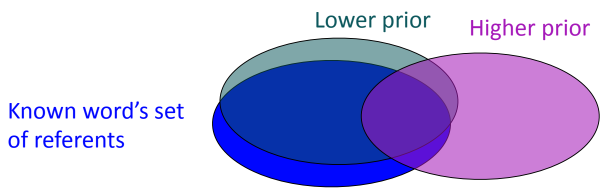
This hypothesis now has 0 probability.

Accounting for other observed behavior

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

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?

Potential explanations:

(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

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?

Potential explanations:

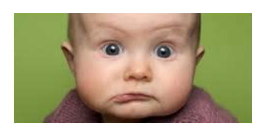
(2) The hypothesis spaces of young children may not be sufficiently constrained to make strong inferences. For example, even though adults know that the set of dogs is much larger than the set of dalmatians, young children may not know this - especially if their family dog is a dalmatian, and they don’t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

An open question

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?

Potential explanations:

(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 input. (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.



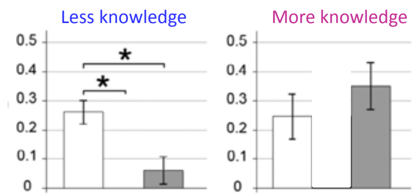
Changes over time

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!



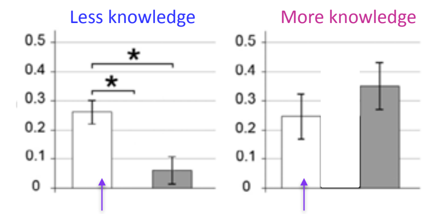
Changes over time

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.

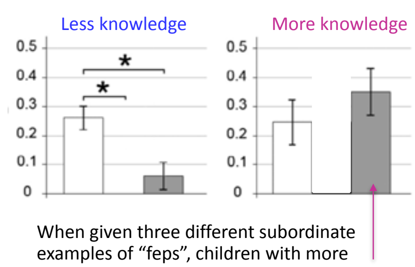
Changes over time

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.

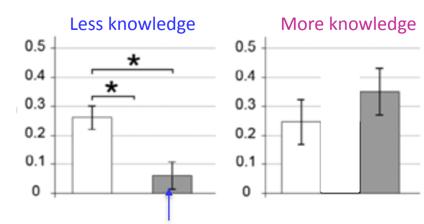
Changes over time

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.

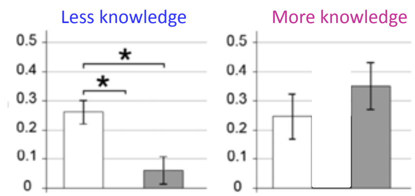
Changes over time

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?

Jenkins et al. 2015: What this means

"…the Bayesian model in isolation and in its current form cannot capture the U-shaped trend."

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

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 the U-shaped trend."

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

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 the U-shaped trend."

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.

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?

Use the remaining time to work on HW2 and the review questions for word meaning. You should be able to do all the questions on HW2 and all the review questions.