Ling 151/Psych 156A: Acquisition of Language II

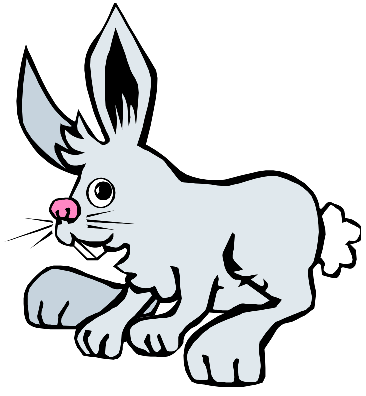
Lecture 10

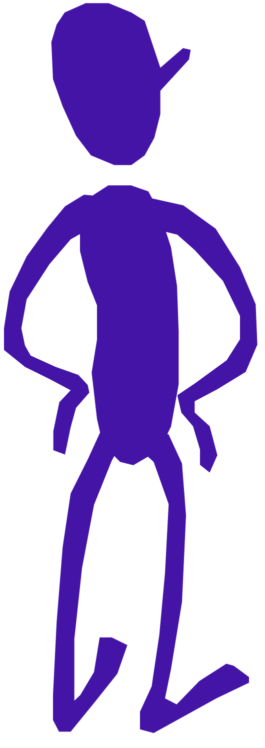
Word meaning I

Announcements

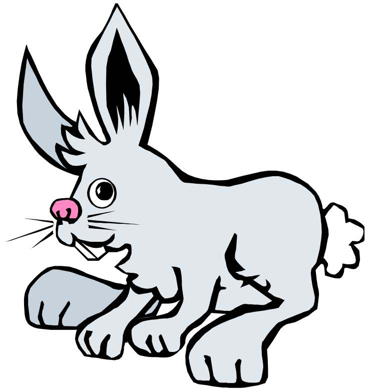
Review questions available for word meaning

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

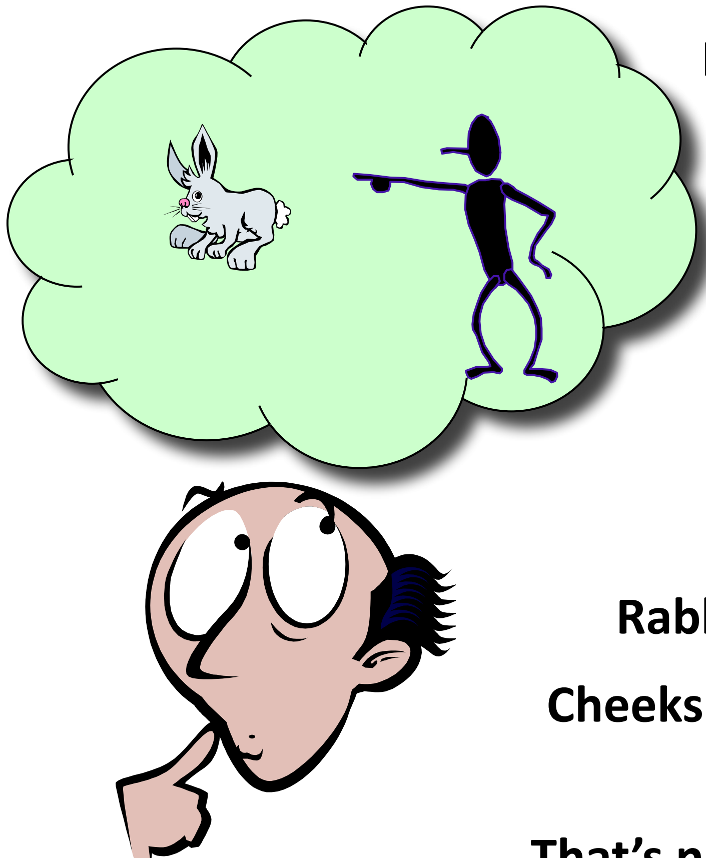




What does “gavagai” mean?



What does “gavagai” mean?



Rabbit?
Mammal?
gray rabbit?
Animal?
Carrot eater?
vegetarian?

Thumping
Hopping
Scurrying

Stay!
Look!

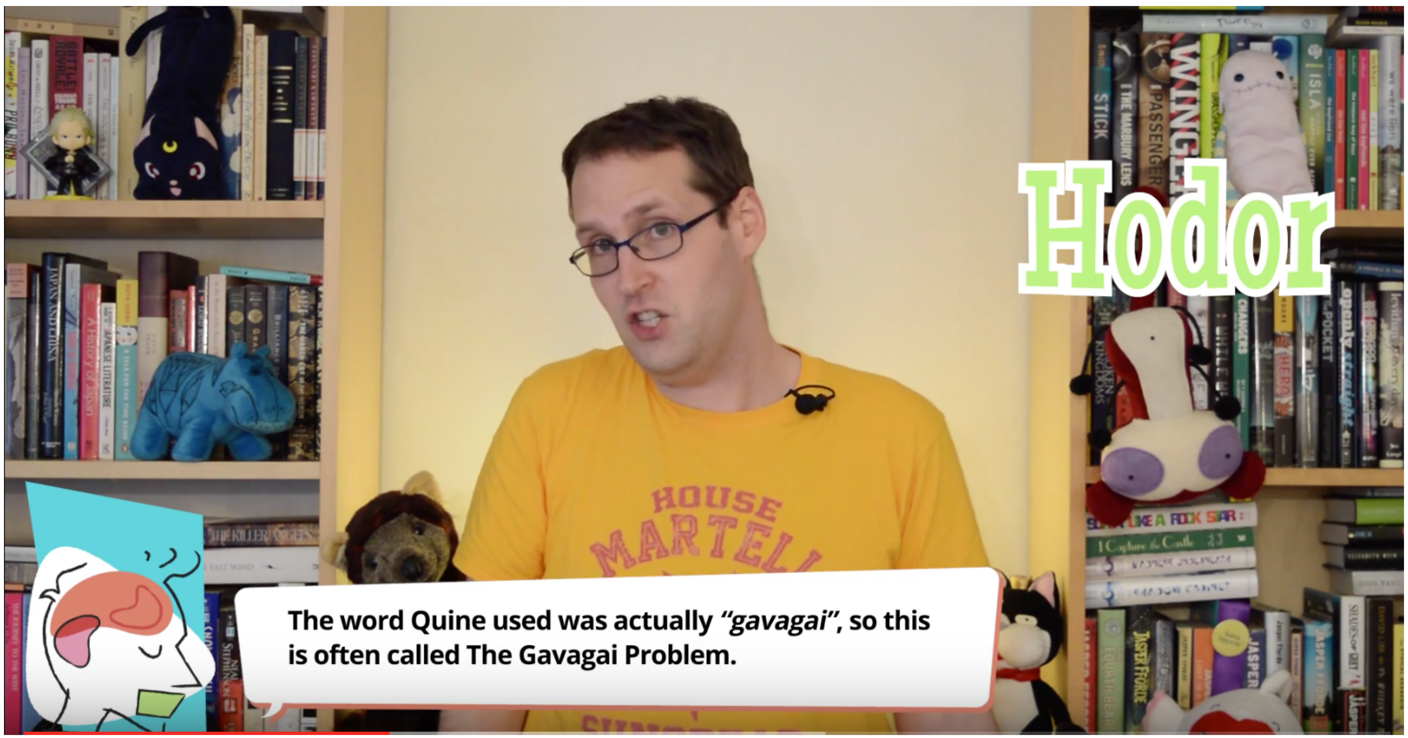
Ears?
Long ears?
Is it gray?
Fluffy?
What a cutie!

Meal!
Rabbit only until eaten!
Cheeks and left ear!

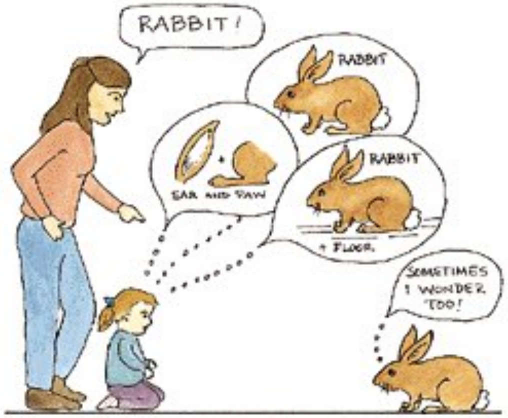
That’s not a dog!

What does “gavagai” mean?

http://www.thelingspace.com/episode-35
https://www.youtube.com/watch?v=Ci-5dVVvf0U
~2:03 - 2:32



Same problem the child faces



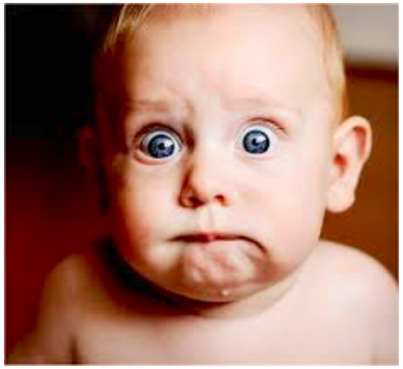
A little more context…

"Look! There's a goblin!"

The mapping problem

Even if something is explicitly labeled in the input ("Look! There's a goblin!"), how does the child know what *specifically* that word refers to? (Is it the head? The feet? The staff? The combination of eyes and hands? Attached goblin parts?...)

Quine (1960): An infinite number of hypotheses about word meaning are possible given the input the child has. That is, the input underspecifies the word's meaning.



The mapping problem

Wellwood, Gagliardi, & Lidz 2016

"Approaching the question first requires an appreciation of the kinds of word meanings that are the target of acquisition. Some words refer to object categories (dog, mammal) and others to event categories (run, watch): in acquiring such words, simply paying attention to the right aspects of the environment could in principle provide strong evidence that a novel word has a certain sort of meaning…."

The mapping problem

Wellwood, Gagliardi, & Lidz 2016

"However, this is only the very beginning of the story; many words refer to properties of objects or events (red, fluffy, fast, suddenly), and others refer to nothing at all (most, any, empty). Since any novel word could express innumerably many things, properties, or relations, understanding how children decide what a novel word means must be informed not only by a precise understanding of the kinds of data children have available to them, but also of the character of the biases and expectations they bring to the learning task."

So how do children figure it out? Obviously, they do….

Even by 6 to 9 months, infants recognize many familiar words in their language, like body parts and food items — that is, concrete objects (Bergelson & Swingley 2012, 2015).

eyes, mouth, hands, …

milk, spoon, juice, cookie, …

So how do children figure it out? Obviously, they do….

By 10 to 13 months old, infants understand words like “all gone”, “hug”, “bye”, and “wet” (Bergelson & Swingley 2013)

gone, hug, bye…

wet

Acquisition task

"I love my dax and my kleeg."

dax = *??*

kleeg = *??*

One solution: Fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping

(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina, Snedecker, Trueswell, & Gleitman 2011)

ball

bear

kitty

[unknown]

One solution: Fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping

(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina, Snedecker, Trueswell, & Gleitman 2011)

"Can I have the ball?"

ball

bear

kitty

[unknown]

One solution: Fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

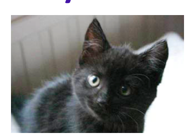
Experimental evidence of fast mapping

(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina, Snedecker, Trueswell, & Gleitman 2011)

"Can I have the ball?"

kitty

bear



ball

[unknown]

One solution: Fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping

(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina, Snedecker, Trueswell, & Gleitman 2011)

ball

bear

kitty

[unknown]

"Can I have the *zib*?"

One solution: Fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping

(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina, Snedecker, Trueswell, & Gleitman 2011)

"Can I have the *zib*?"

ball

bear

kitty

[unknown]

20 months

A slight problem…

"…not all opportunities for word learning are as uncluttered as the experimental settings in which fast-mapping has been demonstrated. In everyday contexts, there are typically many words, many potential referents, limited cues as to which words go with which referents, and rapid attentional shifts among the many entities in the scene." - Smith & Yu (2008)

A slight problem…

"…many studies find that children even as old as 18 months have difficulty in making the right inferences about the intended referents of novel words…infants as young as 13 or 14 months…can link a name to an object given repeated unambiguous pairings in a single session. Overall, however, these effects are fragile with small experimental variations often leading to no learning." - Smith & Yu (2008)

Cross-situational learning

Different approach: infants accrue statistical evidence across multiple trials that are individually ambiguous but can be disambiguated when the information from the trials is aggregated.

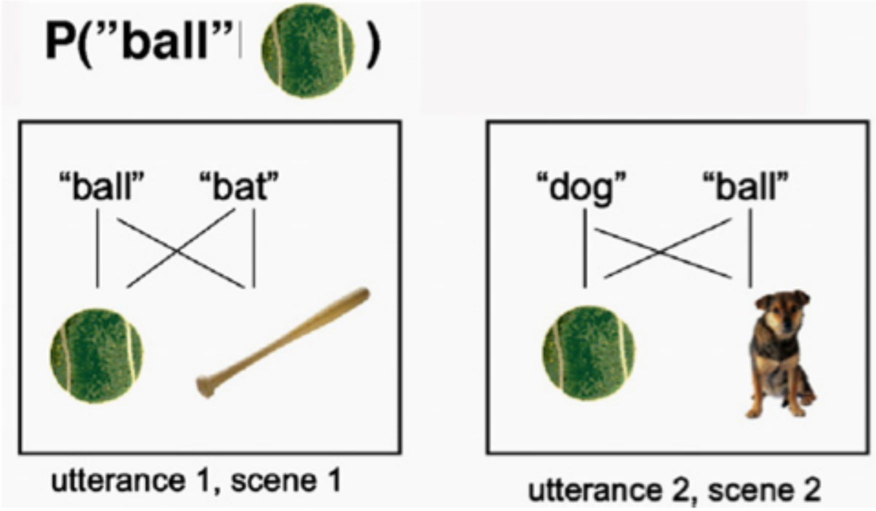


Fig. 1. Associations among words and referents across two individually ambiguous scenes. If a young learner calculates co-occurrences frequencies *across* these two trials, s/he can find the proper mapping of “Ball” to BALL.

Cross-situational learning

Accruing statistical evidence across multiple trials

This can be implemented with Bayesian inference.

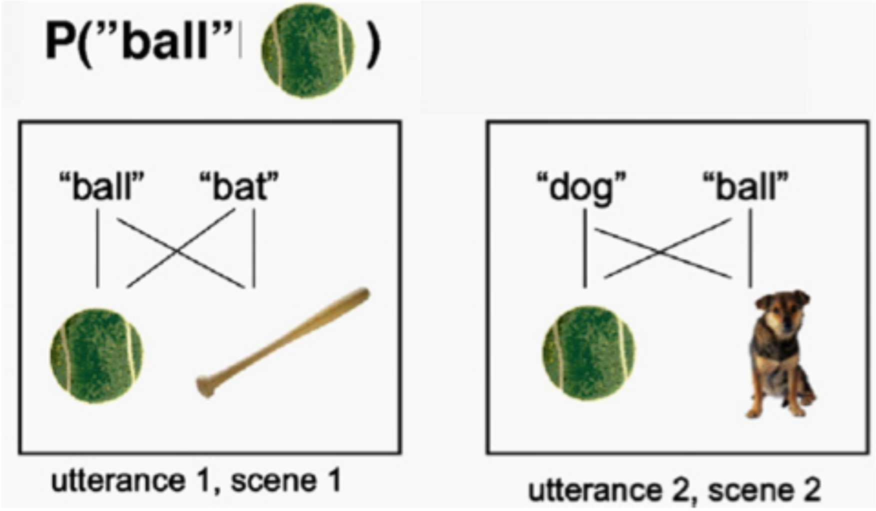


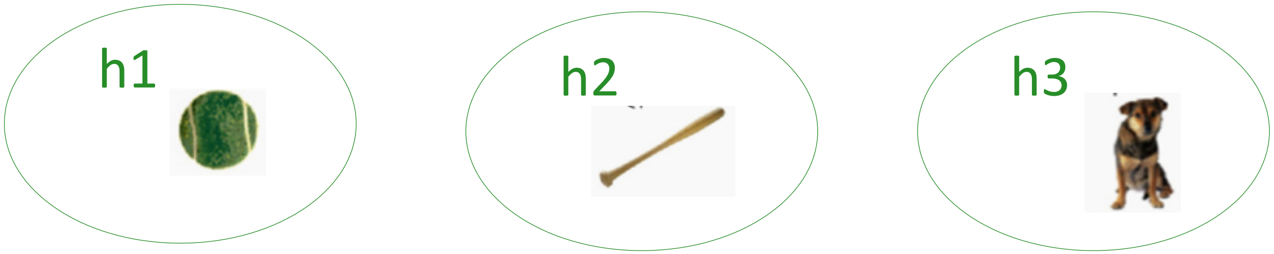
Fig. 1. Associations among words and referents across two individually ambiguous scenes. If a young learner calculates co-occurrences frequencies *across* these two trials, s/he can find the proper mapping of "Ball" to BALL.

Cross-situational learning

Accruing statistical evidence across multiple trials

A Bayesian model assumes the learner has **some space of hypotheses H**…

“ball” refers to…

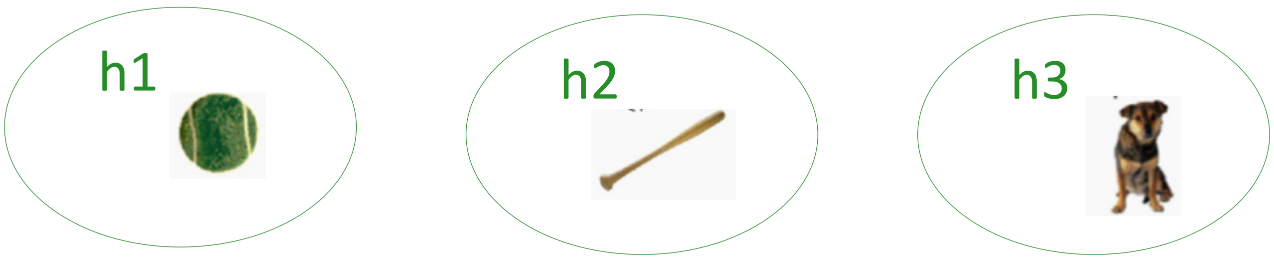


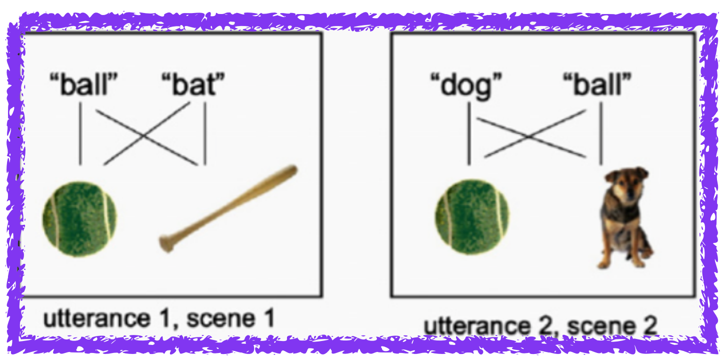
Cross-situational learning

Accruing statistical evidence across multiple trials

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.

"ball" refers to...

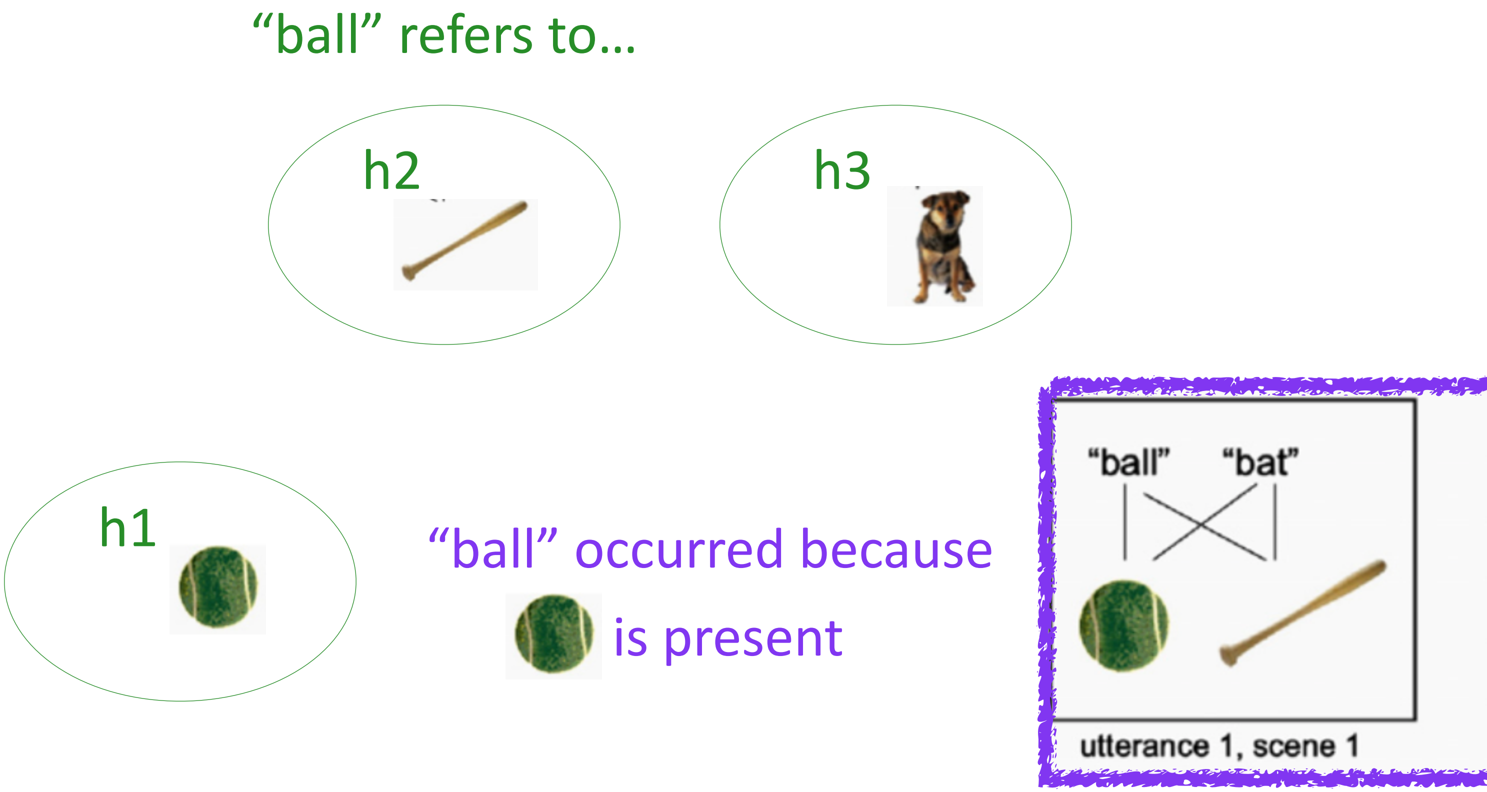




Cross-situational learning

Accruing statistical evidence across multiple trials

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.



Cross-situational learning

Accruing statistical evidence across multiple trials

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.

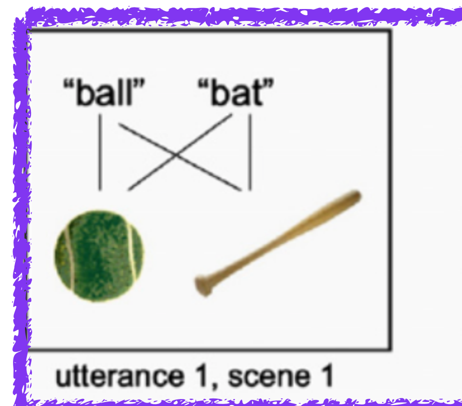
"ball" refers to…

h1

h3

h2

"ball" occurred because  is present



Cross-situational learning

Accruing statistical evidence across multiple trials

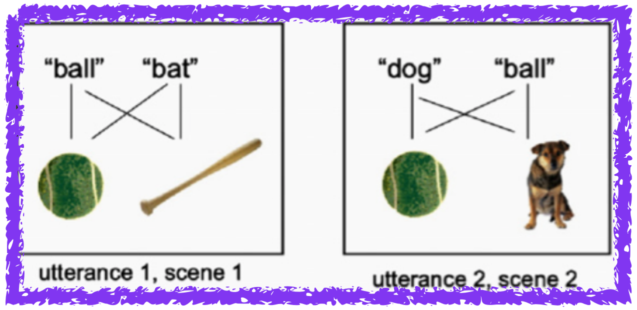
Given **D**, the modeled child's goal is to determine the probability of each possible hypothesis **h** ∈ **H**, written as **P (h|D)** - the ***posterior*** for that hypothesis.

"ball" refers to...

h1

h2

h3



$$P(h|D)$$

Cross-situational learning

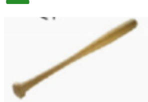
Accruing statistical evidence across multiple trials

This depends on a few different aspects (which have their own probabilities).

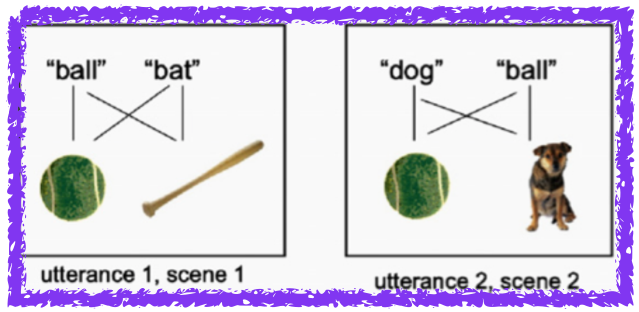
“ball” refers to…

h1

h2



h3



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

Cross-situational learning

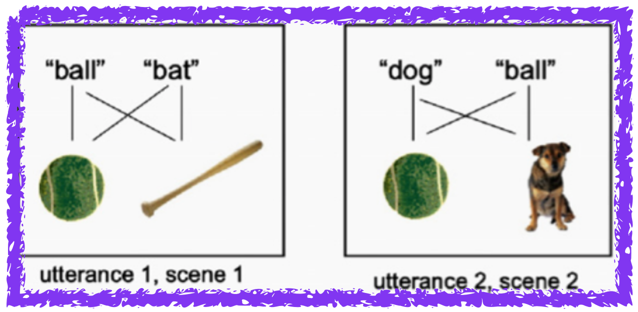
Accruing statistical evidence across multiple trials

“ball” refers to…

h1

h2

h3



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

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

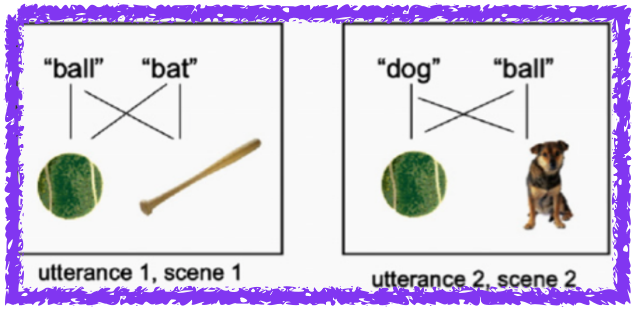
Cross-situational learning

Accruing statistical evidence across multiple trials

h1

✔ P(D | h1) = 1 * 1 = 1

“ball” refers to…



Given these data, h1 would predict that “ball” should be said in both scene 1 and scene 2 — which it is.

h2

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

Moreover, because it’s the only object in h1, that object occurring when “ball” is said is 1/1 = 1.

h3

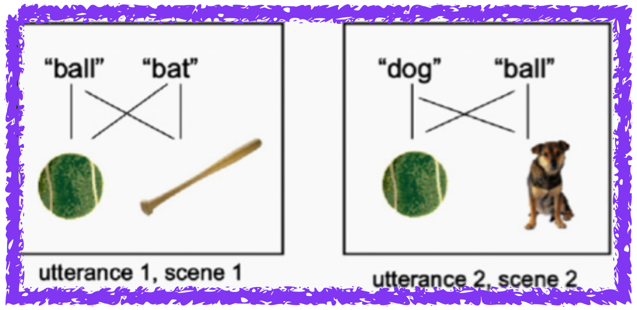
Cross-situational learning

Accruing statistical evidence across multiple trials

P(D | h2) = 1 * 0 = 0

h2

"ball" refers to...



h1

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

P(D | h1) = 1

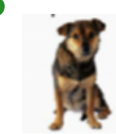
h3

Given these data, h2 would predict that "ball" should be said only in scene 1 but *not* in scene 2. This makes the likelihood of generating the data in the second scene 0 for this hypothesis.

Cross-situational learning

Accruing statistical evidence across multiple trials

h3

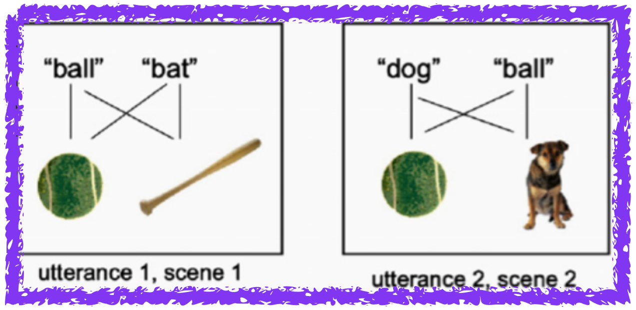


✗ P(D | h3) = 0 * 1 = 0

"ball" refers to…

h1

h2



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

✔ P(D | h1) = 1

✗ P(D | h2) = 0

Similarly, given these data, h3 would predict that "ball" should be said only in scene 2 but *not* in scene 1. This makes the likelihood of generating the data in the first scene 0 for this hypothesis.

Cross-situational learning

Accruing statistical evidence across multiple trials

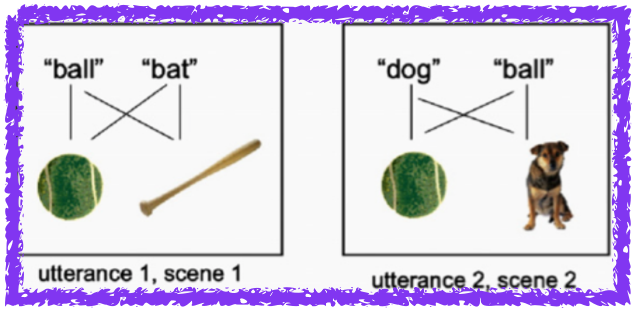
P (h) represents the *prior* of the hypothesis h, and represents the probability of the hypothesis before any data have been encountered. Intuitively, this corresponds to how plausible the hypothesis is, irrespective of any data.

"ball" refers to…

h1

h2

h3



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

$P(D \mid h1) = 1$

$P(D \mid h2) = 0$

$P(D \mid h3) = 0$

If there's no reason to consider one hypothesis more complex than another, the hypotheses will typically receive **uniform** probability (all of them have the same probability).

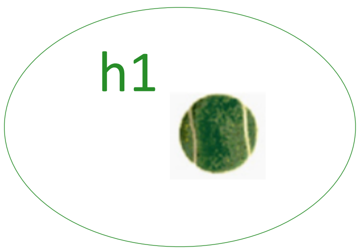
This is typically 1 over the total hypotheses available.

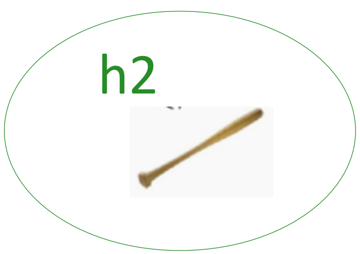
Cross-situational learning

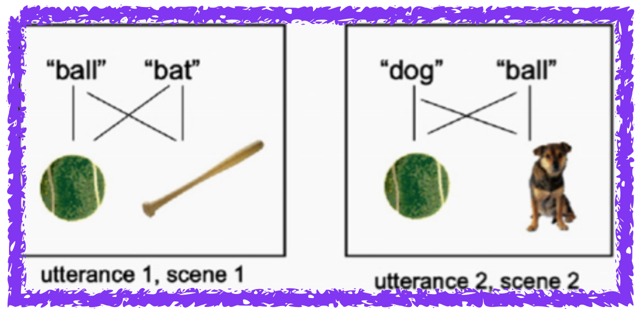
Accruing statistical evidence across multiple trials

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.

"ball" refers to…







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

This is typically 1 over the total hypotheses available.

uniform probability

✔ $P(D \mid h1) = 1$	$P(h1) = 1/3$
✗ $P(D \mid h2) = 0$	$P(h2) = 1/3$
✗ $P(D \mid h3) = 0$	$P(h3) = 1/3$

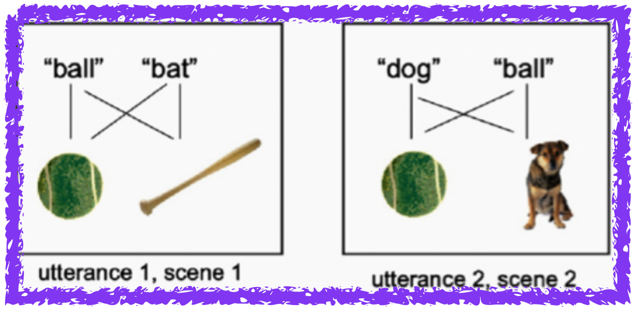
Cross-situational learning

Accruing statistical evidence across multiple trials

P(D) represents the probability of the data irrespective of any hypothesis. It serves as a normalizing factor so that the posterior probabilities sum to 1.

"ball" refers to…

h1



h2

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

h3

P(D | h1) = 1   P(h1) = 1/3

P(D | h2) = 0   P(h2) = 1/3

P(D | h3) = 0   P(h3) = 1/3

Cross-situational learning

Accruing statistical evidence across multiple trials

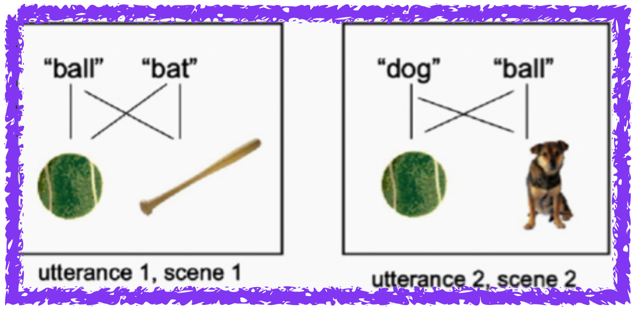
P(D) is calculated by summing over all possible hypotheses the following:

"ball" refers to...

h1

h2

h3



$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} = \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(D | h1) = 1   P(h1) = 1/3

P(D | h2) = 0   P(h2) = 1/3

P(D | h3) = 0   P(h3) = 1/3

Cross-situational learning

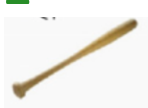
Accruing statistical evidence across multiple trials

P(D) is calculated by summing over all possible hypotheses the following:

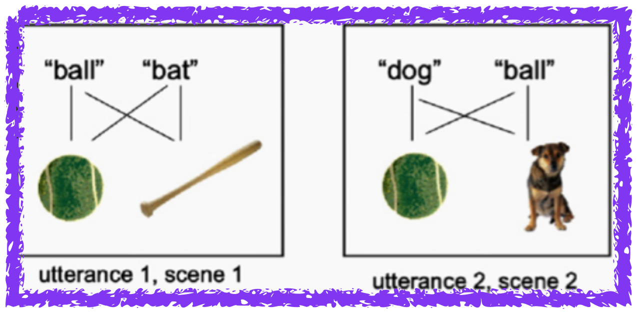
"ball" refers to…

h1

h2



h3



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

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} = \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(D | h1) = 1 P(h1) = 1/3

P(D | h2) = 0 P(h2) = 1/3

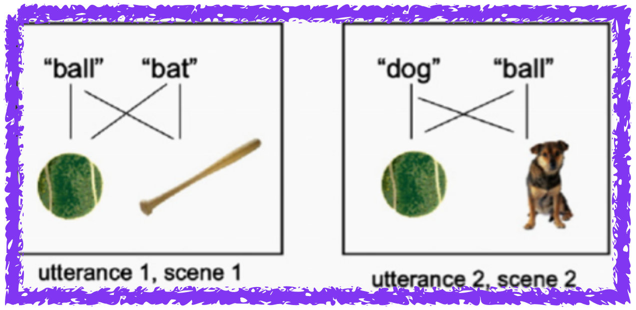
P(D | h3) = 0 P(h3) = 1/3

Cross-situational learning

Accruing statistical evidence across multiple trials

P(D) is calculated by summing over all possible hypotheses the following:

“ball” refers to…



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

h1

h2

h3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h'\in H} P(D|h')*P(h')}$$

P(D | h1) = 1 P(h1) = 1/3

P(D | h2) = 0 P(h2) = 1/3

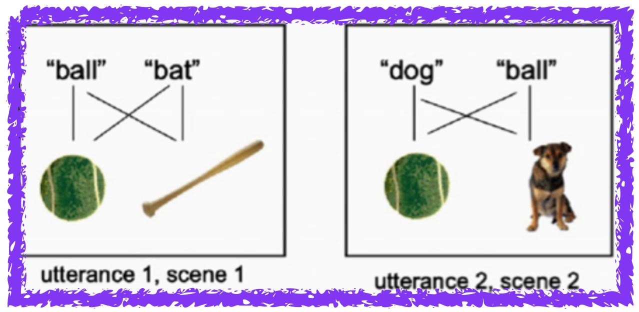
P(D | h3) = 0 P(h3) = 1/3

Cross-situational learning

Accruing statistical evidence across multiple trials

When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

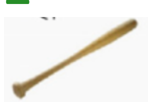
"ball" refers to…



Why is this so?

h1

h2



h3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h'\in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

P(D | h1) = 1 — P(h1) = 1/3

P(D | h2) = 0 — P(h2) = 1/3

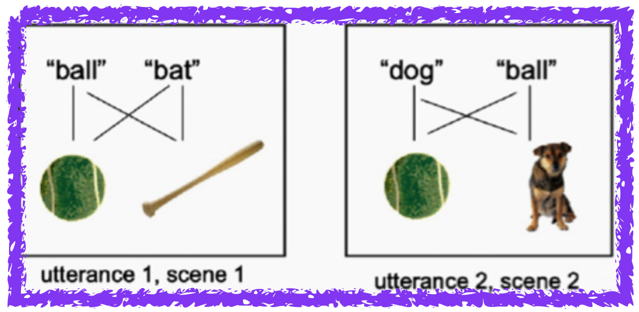
P(D | h3) = 0 — P(h3) = 1/3

Cross-situational learning

Accruing statistical evidence across multiple trials

When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

“ball” refers to…



h1

h2

h3

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} = \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')} \propto P(D|h) * P(h)$$

		likelihood * prior
✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

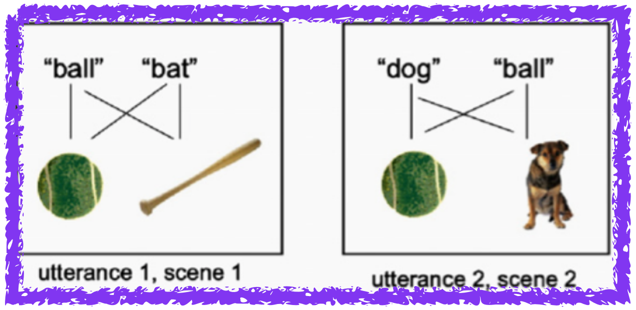
When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to...

h1

h2

h3



sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} = \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')} \propto P(D|h) * P(h)$$

likelihood * prior

✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

Cross-situational learning

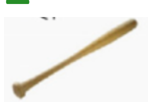
Accruing statistical evidence across multiple trials

When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

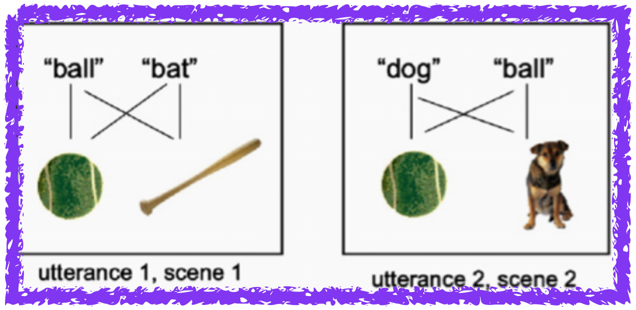
"ball" refers to…

h1

h2



h3



$P(h1 \mid D) = \frac{1/3}{\mathbf{1/3}} = 1$

sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h'\in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

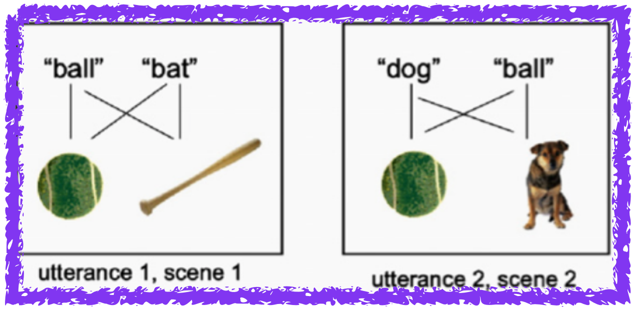
Cross-situational learning

Accruing statistical evidence across multiple trials

When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to…

h1



$P(h1 \mid D) = \mathbf{1}$

$P(h2 \mid D) = \frac{0}{\mathbf{1/3}} = 0$

sum: 1/3 + 0 + 0 = 1/3

h2

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

h3

✔ $P(D \mid h1) = 1$	$P(h1) = 1/3$	$1 * 1/3 = \mathbf{1/3}$
✗ $P(D \mid h2) = 0$	$P(h2) = 1/3$	$0 * 1/3 = \mathbf{0}$
✗ $P(D \mid h3) = 0$	$P(h3) = 1/3$	$0 * 1/3 = \mathbf{0}$

Cross-situational learning

Accruing statistical evidence across multiple trials

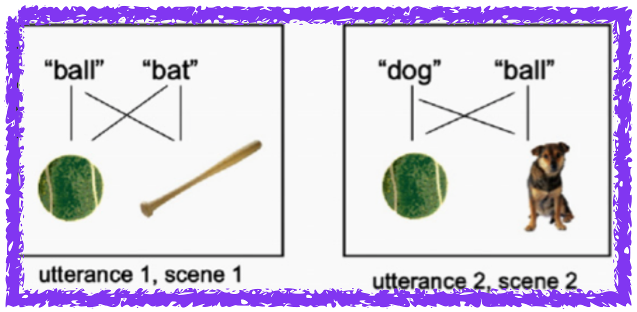
When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to…

h1

h2

h3



P(h1 | D) = **1**

P(h2 | D) = **0**

P(h3 | D) = $\frac{0}{\mathbf{1/3}}$ = 0

sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

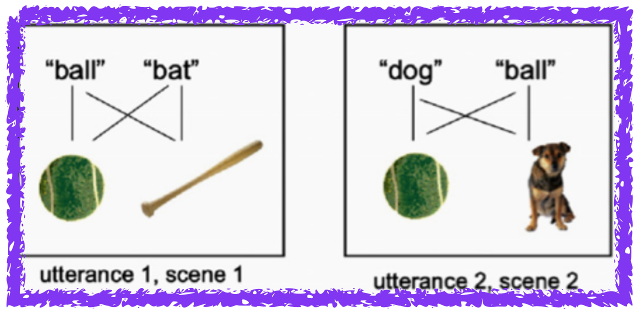
When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to…

h1

h2

h3



P(h1 | D) = **1**

P(h2 | D) = **0**

P(h3 | D) = **0**

sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

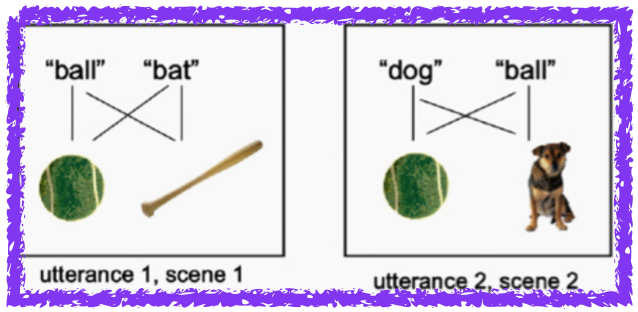
When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to…

h1

h2

h3



P(h1 | D) = **1**

P(h2 | D) = **0**

P(h3 | D) = **0**

Conclusion:

h1 is the only one left with any probability

sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ P(D | h1) = 1	P(h1) = 1/3	1 * 1/3 = **1/3**
✗ P(D | h2) = 0	P(h2) = 1/3	0 * 1/3 = **0**
✗ P(D | h3) = 0	P(h3) = 1/3	0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

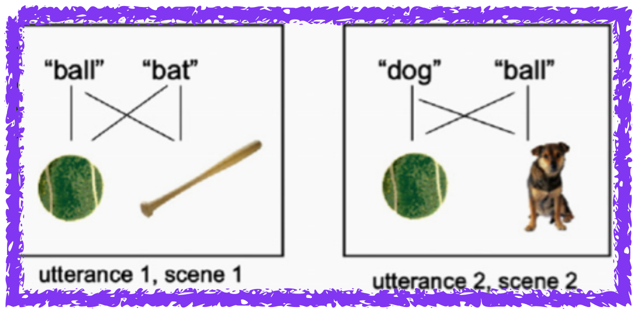
When we only care about how one hypothesis compares to another (as we do here), calculating P(D) can be skipped over.

"ball" refers to...

h1

h2

h3



$P(h1 \mid D) =$ **1**

$P(h2 \mid D) =$ **0**

$P(h3 \mid D) =$ **0**

...which is exactly what we knew before we normalized.

sum: 1/3 + 0 + 0 = 1/3

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ $P(D \mid h1) = 1$	$P(h1) = 1/3$	1 * 1/3 = **1/3**
✗ $P(D \mid h2) = 0$	$P(h2) = 1/3$	0 * 1/3 = **0**
✗ $P(D \mid h3) = 0$	$P(h3) = 1/3$	0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

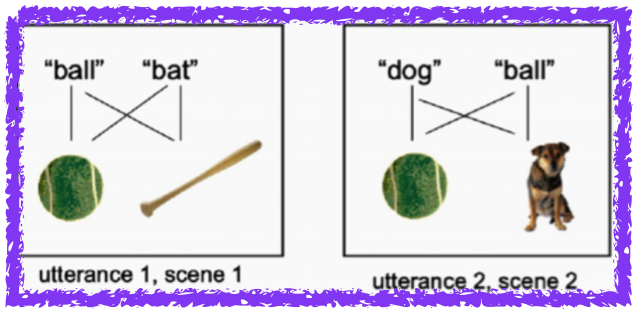
The determining factor here is data coverage — that is, the likelihood. Can the hypothesis account for the data or not?

“ball” refers to…

h1

h2

h3



Only hypothesis 1 can account for the data in both these scenes.

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

likelihood * prior

✔ P(D | h1) = 1 — 1 * 1/3 = **1/3**

✗ P(D | h2) = 0 — 0 * 1/3 = **0**

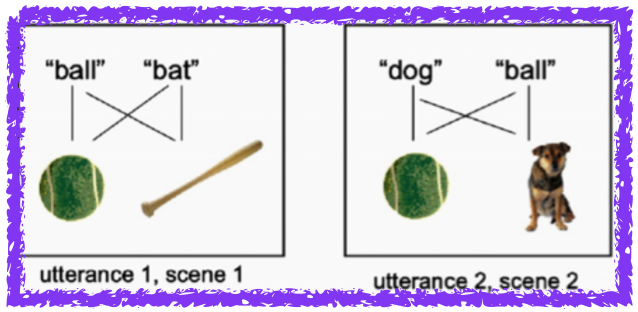
✗ P(D | h3) = 0 — 0 * 1/3 = **0**

Cross-situational learning

Accruing statistical evidence across multiple trials

So, can very young children reason like this?

“ball” refers to…



h1

h2

$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)} = \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')} \propto P(D|h)*P(h)$$

h3

Smith & Yu (2008)

Yu & Smith (2007): Adults seem able to reason like this in cross-situational learning (in experimental setups).

Smith & Yu (2008) ask: Can 12- and 14-month-old infants do this? (Relevant age for beginning word-learning.)




Smith & Yu (2008): Experiment

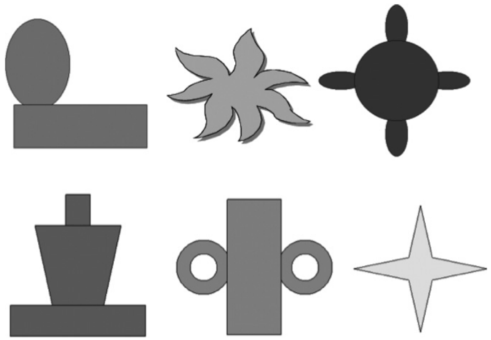
Infants were trained on six novel words obeying phonotactic probabilities of English:

bosa, gasser, manu, colat, kaki, regli

These words were associated with six brightly colored shapes

(sadly greyscale in the paper)

Figure from paper



What the shapes are probably more like

Smith & Yu (2008): Experiment

Training: 30 slides with 2 objects named with two words (total time: 4 min)

Example training slides

manu
colat

bosa
manu

Smith & Yu (2008): Experiment

Testing: 12 trials with one word repeated 4 times and 2 objects (correct one and distracter) present

manu

manu

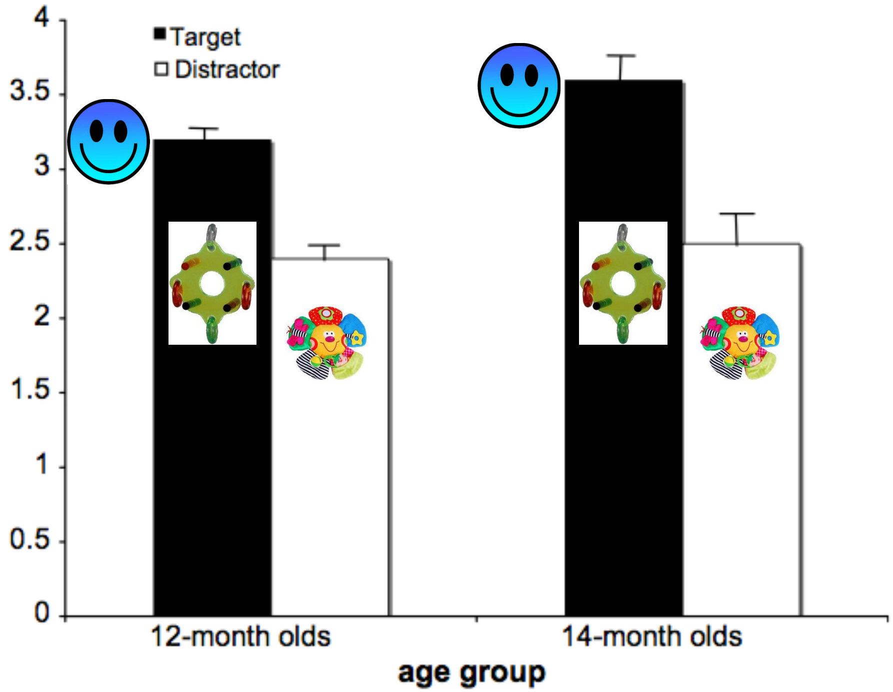
manu

manu

Which one does the infant think is *manu*? That should be the one the infant prefers to look at.

Smith & Yu (2008): Experiment

Results: Infants preferentially look at target over distracter, and 14-month-olds looked longer than 12-month-olds. This means they were able to tabulate distributional information across situations.



Implication: 12 and 14-month-old infants can do cross-situational learning that relies on a reasoning process like Bayesian inference

Something to think about…

The real world isn't necessarily as simple as these experimental setups - often times, there will be many potential referents.

(A similar issue to the one fast-mapping has.)

Fig. 1. (*A*) A plausible word learning environment for the word shoe. (*B*) The simulated word-learning environment for shoe found in most cross-situational word-learning experiments.

Something to think about…

A strategy where learners hang on to one hypothesis at a time until it's proven incorrect and only then switch to a different one (called "Propose But Verify") may work better because of this. There's some evidence that it matches infant and toddler behavioral results quite well (Stevens, Trueswell, Yang, & Gleitman 2013, Woodard, Gleitman, & Trueswell 2016) and may be more effective for navigating the hypothesis space (Romberg & Yu 2014).

Fig. 1. (*A*) A plausible word learning environment for the word shoe. (*B*) The simulated word-learning environment for shoe found in most cross-situational word-learning experiments.

Something to think about…

A strategy where learners hang on to one hypothesis at a time until it's proven incorrect and only then switch to a different one (called "Propose But Verify") may work better.

Some more discussion about this: http://facultyoflanguage.blogspot.com/2013/03/learning-fast-and-slow-i-how-children.html

Fig. 1. (*A*) A plausible word learning environment for the word shoe. (*B*) The simulated word-learning environment for shoe found in most cross-situational word-learning experiments.

Something else to think about…

Having more referents may not be a bad thing.

Why not?

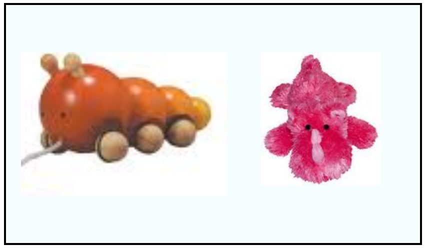
It's easier for the correct associations to separate from spurious associations when there are more object-referent pairing opportunities. Let's see an example of this.

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

First, let's consider their condition, where two objects are shown at a time. Let's say we get three slides/ scenes of data.

"manu"
"colat"



"bosa"
"gasser"

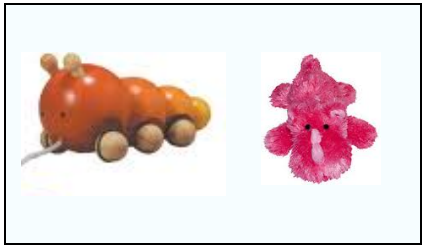
"kaki"
"regli"

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

"manu"
"colat"

Can we tell whether "manu" refers to



"bosa"
"gasser"

No - both hypotheses are equally compatible with these data.

"kaki"
"regli"

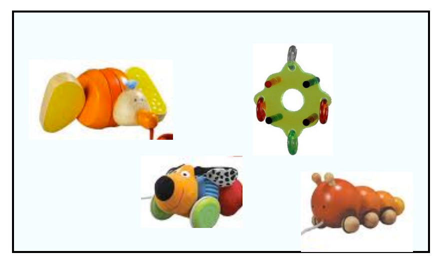
Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Now, let's consider a more complex condition, where four objects are shown at a time. Let's say we get three slides/scenes of data.

"manu"
"colat"
"bosa"
"regli"

"bosa"
"gasser"
"manu"
"colat"



"manu"
"gasser"
"kaki"
"regli"

Why more may not always be harder…

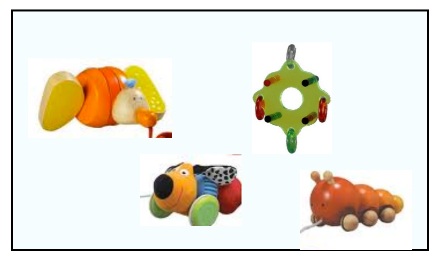
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

"manu"
"colat"
"bosa"
"regli"

Can we tell whether "manu" refers to or  or  or  ? 

"bosa"
"gasser"
"manu"
"colat"

Well, the first slide isn't helpful in distinguishing between these four hypotheses…



"manu"
"gasser"
"kaki"
"regli"

Why more may not always be harder…

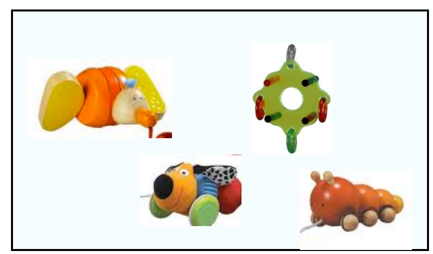
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

“manu”
“colat”
“bosa”
“regli”

Can we tell whether “manu” refers to or  or  or  ? 

“bosa”
“gasser”
“manu”
“colat”

The second slide suggests “manu” can’t be  - otherwise, that object would appear in the second slide.



“manu”
“gasser”
“kaki”
“regli”

Why more may not always be harder…

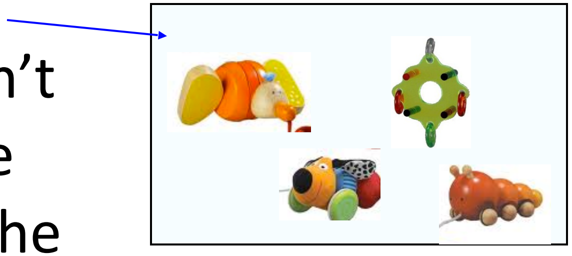
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

"manu"
"colat"
"bosa"
"regli"

Can we tell whether "manu" refers to or  or  

"bosa"
"gasser"
"manu"
"colat"

The third slide suggests "manu" can't be  or  - otherwise, those objects would would appear in the third slide.



"manu"
"gasser"
"kaki"
"regli"

Why more may not always be harder…

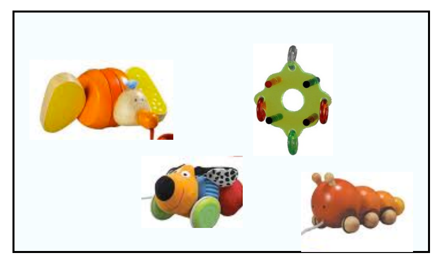
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Therefore, "manu" is 

This shows us that having more things appear (and be named) at once actually offers more opportunities for the correct associations to emerge.

"manu"
"colat"
"bosa"
"regli"

"bosa"
"gasser"
"manu"
"colat"



"manu"
"gasser"
"kaki"
"regli"

Why more may not always be harder…

Let's walk through this scenario using Bayesian inference.

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

We'll see an example of how sequential updating would work (instead of calculating the posterior just once, based on all of the data).

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 1

"manu"
"colat"
"bosa"
"regli"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

Since there are four hypotheses in the hypothesis space at this point, if we assume uniform probability for them, the priors are:

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

We can also calculate the likelihood of each hypothesis generating this data point — specifically, the probability of “manu” being said if that hypothesis was correct.

Sequential updating

data point 1

“manu”
“colat”
“bosa”
“regli”

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

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

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

likelihood of each hypothesis generating this data point

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

“manu” would be said in the scene, and “manu” being said when this object is present is 1/1.

Hypothesis 1 (H1): “manu” = 

P(D | H1) = 1/1 = 1

Sequential updating

Hypothesis 1 (H1): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

likelihood of each hypothesis generating this data point

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

And the same is true for all the other hypotheses.

P(D | H1) = 1

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

P(D | H2) = 1/1 = 1
P(D | H3) = 1/1 = 1
P(D | H4) = 1/1 = 1

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 1

"manu"
"colat"
"bosa"
"regli"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

Because we'll be using the posterior probabilities for subsequent updating, we need to actually do the normalization.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 1 * 1/4 = 1/4

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 1 * 1/4 = 1/4

sum 1

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 1 * 1/4 = 1/4

sum 1

Posterior probability

P(H1 | D) = 1/4 / 1 = 1/4
P(H2 | D) = 1/4 / 1 = 1/4
P(H3 | D) = 1/4 / 1 = 1/4
P(H4 | D) = 1/4 / 1 = 1/4

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h)*P(h)}{\sum_{h'\in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 1 * 1/4 = 1/4

sum 1

Posterior probability

P(H1 | D) = 1/4
P(H2 | D) = 1/4
P(H3 | D) = 1/4
P(H4 | D) = 1/4

Interpretation: After this data point, all hypotheses are equally likely still.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 1

“manu”
“colat”
“bosa”
“regli”

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 1 * 1/4 = 1/4

sum 1

Posterior probability

P(H1 | D) = 1/4
P(H2 | D) = 1/4
P(H3 | D) = 1/4
P(H4 | D) = 1/4

Sequential updating

These posterior probabilities for data point 1 become the prior probabilities for data point 2.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) =
P(H2) =
P(H3) =
P(H4) =

Posterior probability

P(H1 | D) = 1/4
P(H2 | D) = 1/4
P(H3 | D) = 1/4
P(H4 | D) = 1/4

Sequential updating

These posterior probabilities for data point 1 become the prior probabilities for data point 2.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

We can now calculate the likelihoods for data point 2.

Sequential updating

data point 2

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

"bosa"
"gasser"
"manu"
"colat"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

We can now calculate the likelihoods for data point 2.

"manu" would be said in the scene, and "manu" being said when this object is present is 1/1.

Hypothesis 1 (H1): "manu" = 

P(D | H1) = 1/1 = 1

Sequential updating

Hypothesis 1 (H1): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

Hypothesis 4 (H4): “manu” = 

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

We can now calculate the likelihoods for data point 2.

The same is true for H2 and H3.

P(D | H1) = 1

Hypothesis 2 (H2): “manu” = 

P(D | H2) = 1/1 = 1
P(D | H3) = 1/1 = 1

Hypothesis 3 (H3): “manu” = 

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

data point 2

"bosa"
"gasser"
"manu"
"colat"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

We can now calculate the likelihoods for data point 2.

However, H4 would not account for this data point. It would not predict "manu" should be said because the object isn't present.

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1

Hypothesis 4 (H4): "manu" = 

P(D | H4) = 0/1 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 2

"bosa"
"gasser"
"manu"
"colat"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

Because again we'll be using the posterior probability for subsequent updating, we need to do the normalization.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 0 * 1/4 = 0

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 0 * 1/4 = 0

sum 3/4

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 2

“bosa”
“gasser”
“manu”
“colat”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 0 * 1/4 = 0

sum 3/4

Posterior probability

P(H1 | D) = 1/4 / 3/4 = 1/3
P(H2 | D) = 1/4 / 3/4 = 1/3
P(H3 | D) = 1/4 / 3/4 = 1/3
P(H4 | D) = 0/ 3/4 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 2

"bosa"
"gasser"
"manu"
"colat"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 0 * 1/4 = 0

sum 3/4

Posterior probability

P(H1 | D) = 1/3
P(H2 | D) = 1/3
P(H3 | D) = 1/3
P(H4 | D) = 0

H4 has been ruled out, but the other three are equally possible.

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 2

"bosa"
"gasser"
"manu"
"colat"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h'\in H} P(D|h')*P(h')}$$

P(H1) = 1/4
P(H2) = 1/4
P(H3) = 1/4
P(H4) = 1/4

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0

likelihood * prior

(D | H1) * P(H1) = 1 * 1/4 = 1/4
(D | H2) * P(H2) = 1 * 1/4 = 1/4
(D | H3) * P(H3) = 1 * 1/4 = 1/4
(D | H4) * P(H4) = 0 * 1/4 = 0

sum 3/4

Posterior probability

P(H1 | D) = 1/3
P(H2 | D) = 1/3
P(H3 | D) = 1/3
P(H4 | D) = 0

These posteriors become the priors for data point 3.

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

These posteriors become the priors for data point 3.

Sequential updating

data point 3

“manu”
“gasser”
“kaki”
“regli”

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

We can now calculate the likelihoods for data point 3.

“manu” would not be said in the scene if H1 were correct.

Hypothesis 1 (H1): “manu” = 

P(D | H1) = 0/1 = 0

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

data point 3

“manu”
“gasser”
“kaki”
“regli”

Hypothesis 4 (H4): “manu” = 

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

We can now calculate the likelihoods for data point 3.

The same is true for H3.

P(D | H1) = 0

Hypothesis 3 (H3): “manu” = 

P(D | H3) = 0/1 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 3 (H3): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

We can now calculate the likelihoods for data point 3.

However, "manu" would be said if either H2 or H4 were true.

P(D | H1) = 0
P(D | H3) = 0

Hypothesis 2 (H2): "manu" = 

Hypothesis 4 (H4): "manu" = 

P(D | H2) = 1/1 = 1
P(D | H4) = 1/1 = 1

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

P(D | H1) = 0
P(D | H2) = 1
P(D | H3) = 0
P(D | H4) = 1

Since this is the last data point, we don't actually need to do the normalization step unless we want to get a probability rather than a relative sense of how much more likely one hypothesis is than another.

But we can, just for practice.

Sequential updating

Hypothesis 1 (H1): “manu” = 

Hypothesis 2 (H2): “manu” = 

Hypothesis 3 (H3): “manu” = 

Hypothesis 4 (H4): “manu” = 

data point 3

“manu”
“gasser”
“kaki”
“regli”

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

P(D | H1) = 0
P(D | H2) = 1
P(D | H3) = 0
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 0 * 1/3 = 0
(D | H2) * P(H2) = 1 * 1/3 = 1/3
(D | H3) * P(H3) = 0 * 1/3 = 0
(D | H4) * P(H4) = 1 * 0 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h) * P(h)}{\sum_{h' \in H} P(D|h') * P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

P(D | H1) = 0
P(D | H2) = 1
P(D | H3) = 0
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 0 * 1/3 = 0
(D | H2) * P(H2) = 1 * 1/3 = 1/3
(D | H3) * P(H3) = 0 * 1/3 = 0
(D | H4) * P(H4) = 1 * 0 = 0

sum 1/3

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

P(D | H1) = 0
P(D | H2) = 1
P(D | H3) = 0
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 0 * 1/3 = 0
(D | H2) * P(H2) = 1 * 1/3 = 1/3
(D | H3) * P(H3) = 0 * 1/3 = 0
(D | H4) * P(H4) = 1 * 0 = 0

sum 1/3

Posterior probability

P(H1 | D) = 0 / 1/3 = 0
P(H2 | D) = 1/3 / 1/3 = 1
P(H3 | D) = 0 / 1/3 = 0
P(H4 | D) = 0 / 1/3 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

data point 3

"manu"
"gasser"
"kaki"
"regli"

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

$$= \frac{P(D|h)*P(h)}{\sum_{h' \in H} P(D|h')*P(h')}$$

P(H1) = 1/3
P(H2) = 1/3
P(H3) = 1/3
P(H4) = 0

P(D | H1) = 0
P(D | H2) = 1
P(D | H3) = 0
P(D | H4) = 1

likelihood * prior

(D | H1) * P(H1) = 0 * 1/3 = 0
(D | H2) * P(H2) = 1 * 1/3 = 1/3
(D | H3) * P(H3) = 0 * 1/3 = 0
(D | H4) * P(H4) = 1 * 0 = 0

sum 1/3

Posterior probability

P(H1 | D) = 0
P(H2 | D) = 1
P(H3 | D) = 0
P(H4 | D) = 0

Only hypothesis 2 is left!

The utility of probabilities

Partial knowledge of some words appears to be very helpful for learners figuring out the meaning of words they don't know yet (Yurovsky, Fricker, & Yu 2013).

"bosa"
"gasser"
"manu"
"colat"

This may relate to the priors they give some hypotheses. For example, if they know "manu" = , then they would set the prior for other words referring to  as 0.

Some other factors in cross-situational learning

Even if there are more referents, cross-situational learning is more successful when some referents are immediately repeated from situation to situation (Kachergis, Yu, & Shiffrin 2012).

"manu"
"colat"
"bosa"
"regli"

"bosa"
"gasser"
"manu"
"colat"

"manu"
"gasser"
"kaki"
"regli"

Some other factors in cross-situational learning

The child's perspective of real world events may make cross-situational learning more feasible, as compared to a neutral third party (the way a photograph represents the world). This is likely because certain things are more salient from a child's perspective due to object foregrounding and degree of clutter in line of sight (Yurovsky, Smith, & Yu 2013).

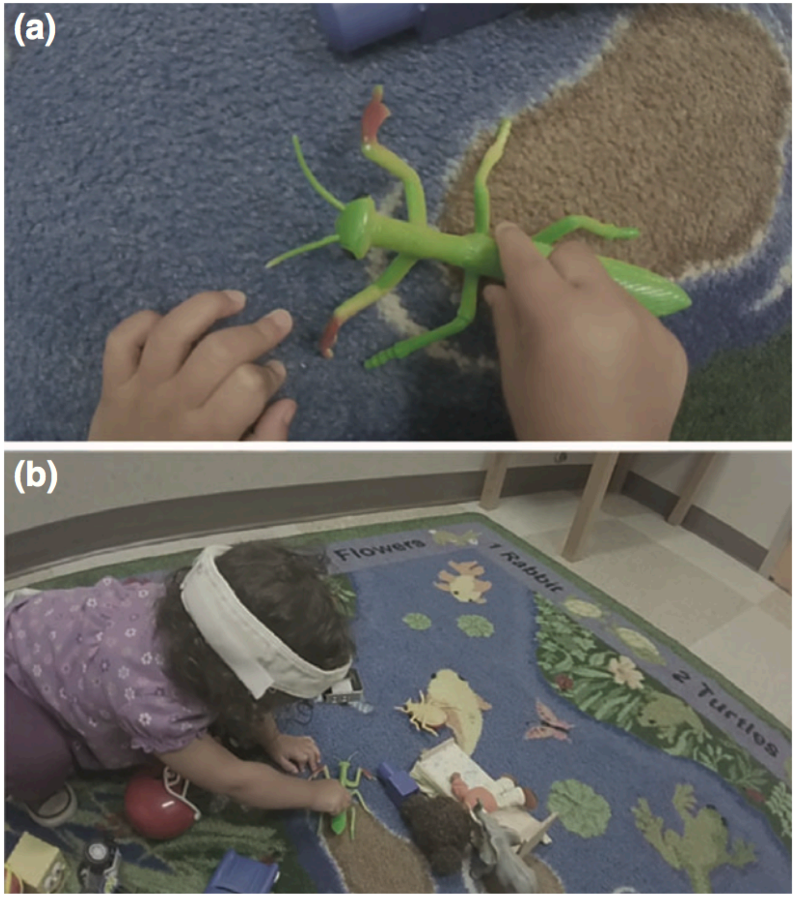


FIGURE 2 | Differences in the number of namable objects in view from the child's (a) and parent's (b) perspective.

Samuelson & McMurray 2017

Recap: Word-meaning mapping

Cross-situational learning, which relies on distributional information across situations, can help children learn which words refer to which things in the world.

One way to implement the reasoning process behind cross-situational learning is Bayesian inference. It can be done in a batch over all the data observed, or sequentially as the data are observed one by one.

Experimental evidence suggests that infants are capable of this kind of reasoning in controlled experimental setups, which may or may not resemble real life contexts with respect to how many referents are available.

Questions?

You should be able to do up through question 1 on HW4 and up through question 6 on the word meaning review questions.