Psych 156A/ Ling 150:
Acquisition of Language II

Lecture 8

Word Meaning 1

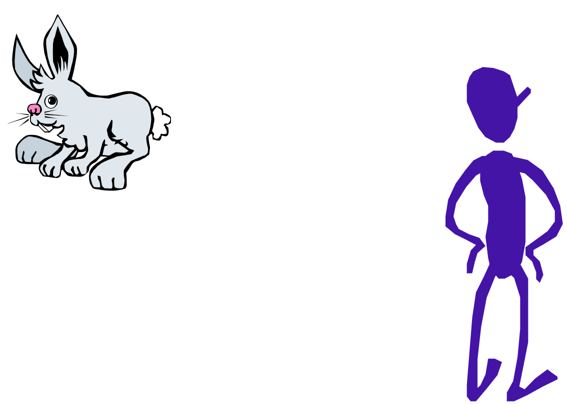
Announcements

Review questions available for word meaning

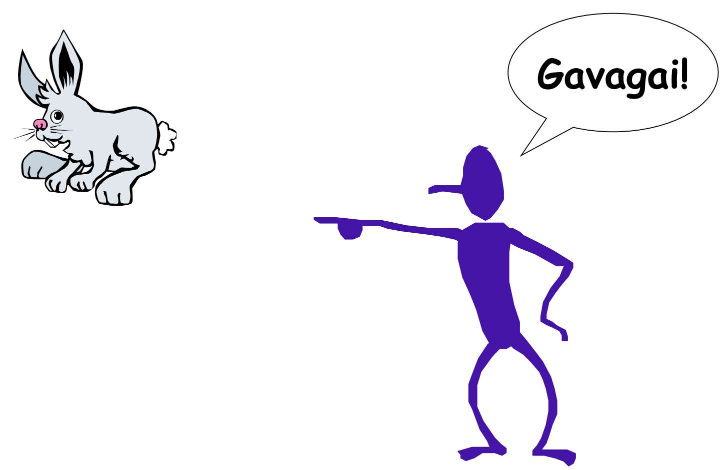
Be working on HW2 (due 5/8/14)
- Note: Remember that working in a group can be very beneficial.

Midterm review in class on 5/1/14

Midterm exam during class on 5/6/14



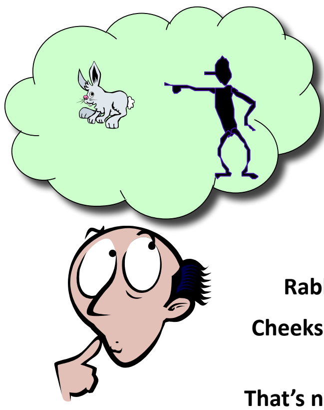
What does "gavagai" mean?



What does “gavagai” mean?

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

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



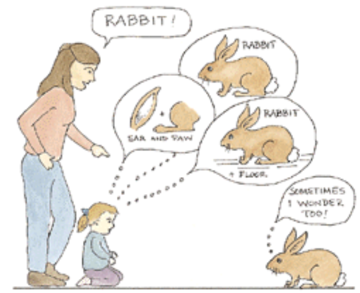
Thumping
Hopping
Scurrying

Stay!
Look!

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

That’s not a dog!

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.

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 (Bergelson & Swingley 2012).

eyes, mouth, hands, …

milk, spoon, juice, cookie, …

Computational problem

"I love my daxes."

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

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

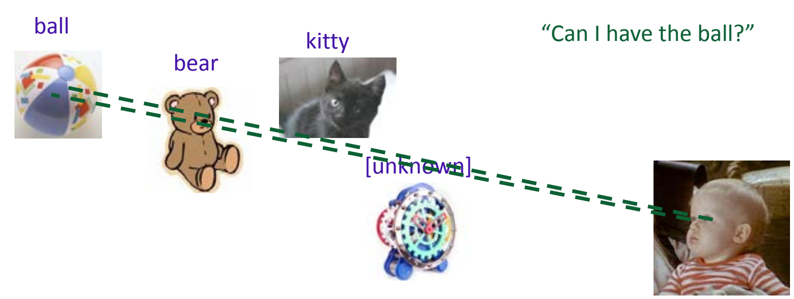
(Carey & Bartlett 1978, Dollaghan 1985, Mervis & Bertrand 1994, Medina,Snedecker, Trueswell, & Gleitman 2011)

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)



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)

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

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

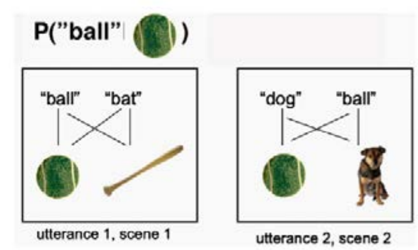


Fig. 1. Associations among words and referents across two individually ambiguous scenes. If a young learner calculates co-occurrences frequencies *across* these two trials, s/he can find the proper mapping of "Ball" to BALL.

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

Posterior probability of hypothesis H, given that data D have been observed

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

Posterior probability

Likelihood of seeing data D, given that H is true

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

Posterior probability

Prior probability of hypothesis H

Likelihood

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

Posterior probability

Likelihood

Prior

Probability of observing the data, no matter what hypothesis is true

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

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

Posterior probability

Likelihood

Prior

Probability of observing the data, no matter what hypothesis is true:
Calculate by summing over all hypotheses

How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

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

data

Posterior probability

Prior

Likelihood

Cross-situational learning

Let's apply Bayesian inference to this scenario.

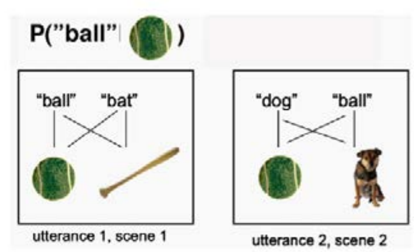


Fig. 1. Associations among words and referents across two individually ambiguous scenes. If a young learner calculates co-occurrences frequencies *across* these two trials, s/he can find the proper mapping of "Ball" to BALL.

Cross-situational learning

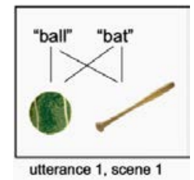
Let's apply Bayesian inference to this scenario.

Posterior probability that "ball" refers to 

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

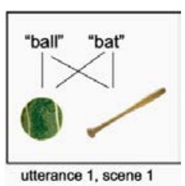
Hypothesis 2 (H2): "ball" = 

Since there are two hypotheses in the hypothesis space at this point
P(H1) = 1/2 = 0.5
P(H2) = 1/2 = 0.5

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

If this is the only data available,

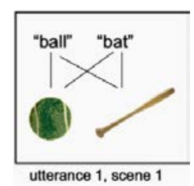
P(D | H1) = would this be observed if H1 were true? Yes. Therefore p(D | H1) = 1.0.

P(D | H2) = would this be observed if H2 were true? Yes. Therefore p(D | H2) = 1.0.

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

If this is the only data available,

$P(D) = \sum_h P(D \mid h) P(h) =$

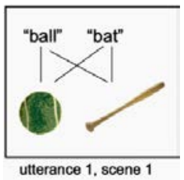
P(D | H1) * P(H1) = 1.0 * 0.5 = 0.5
P(D | H2) * P(H2) = 1.0 * 0.5 = 0.5

so
$\sum_h P(D \mid h) P(h) = 0.5 + 0.5 = 1.0$

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

If this is the only data available,

$$= \frac{P(D \mid H1) * P(H1)}{P(D)}$$

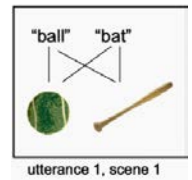
$$= \frac{1.0 * 0.5}{1.0} = 0.5$$

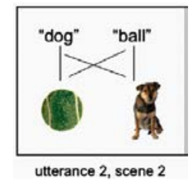
This feels intuitively right, since "ball" could refer to either object, given this data point.

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data





Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

Hypothesis 3 (H3): "ball" = 

Since there are three hypotheses in the hypothesis space at this point

P(H1) = 1/3 = 0.33

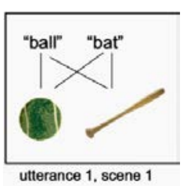
P(H2) = 1/3 = 0.33

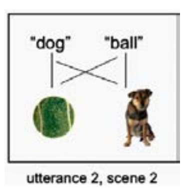
P(H3) = 1/3 = 0.33

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data





Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

Hypothesis 3 (H3): "ball" = 

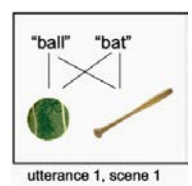
If this is the only data available,

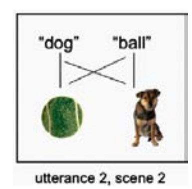
P(D | H1) = would this be observed if H1 were true? Yes. Therefore p(D | H1) = 1.0.

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data





Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

Hypothesis 3 (H3): "ball" = 

If this is the only data available,

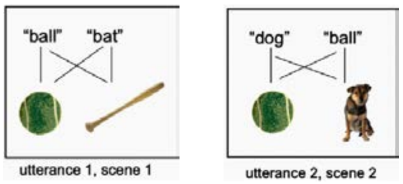
P(D | H2) = would this be observed if H2 were true? No. (Why would "ball" be said in the second scene?) Therefore p(D | H2) = 0.0.

P(D | H3) = would this be observed if H3 were true? No. (Why would "ball" be said in the first scene?) Therefore p(D | H3) = 0.0.

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

Hypothesis 3 (H3): "ball" = 

If this is the only data available,

$P(D) = \sum_h P(D \mid h)\, P(h) =$

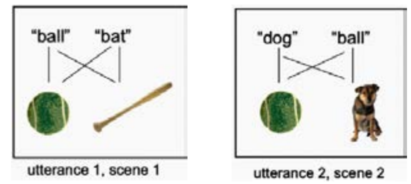
$P(D \mid H1) * P(H1) = 1.0 * 0.33 = 0.33$
$P(D \mid H2) * P(H2) = 0.0 * 0.33 = 0.0$
$P(D \mid H3) * P(H3) = 0.0 * 0.33 = 0.0$

so
$\sum_h P(D \mid h)\, P(h) = 0.33 + 0.0 + 0.0 = 0.33$

Cross-situational learning

Let's apply Bayesian inference to this scenario.

Observable data



Hypothesis 1 (H1): "ball" = 

Hypothesis 2 (H2): "ball" = 

Hypothesis 3 (H3): "ball" = 

If this is the only data available,

$= \frac{P(D \mid H1) * P(H1)}{P(D)}$

$= \frac{1.0 * 0.33}{0.33} = 1.0$

This feels intuitively right, since "ball" could only refer to the ball, when these two scenes are reconciled with each other.

Smith & Yu (2008)

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

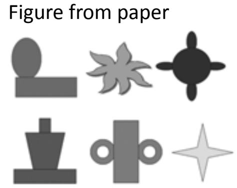
Smith & Yu (2008) ask: Can 12- and 14-month-old infants do this? (Relevant age for beginning word-learning.)

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

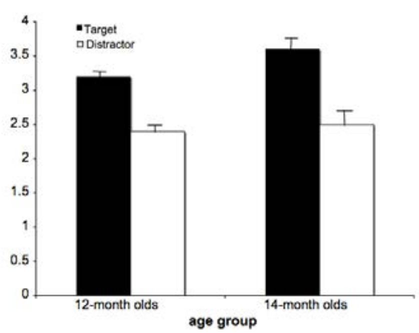
Testing: 12 trials with one word repeated 4 times and 2 objects (correct one and distracter) present

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

manu
manu
manu
manu

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

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 may work better because of this. There's some evidence that it matches infant behavioral results quite well (Stevens, Trueswell, Yang, & Gleitman 2013).

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

Something else to think about…

Having more referents may not be a bad thing.

Why not?

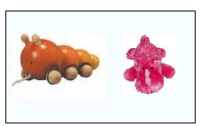
It's easier for the correct associations to emerge from spurious associations when there are more object-referent pairing opportunities. Let's see an example of this.

Why more may not always be harder…

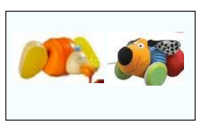
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

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"

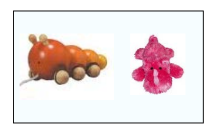
Why more may not always be harder…

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

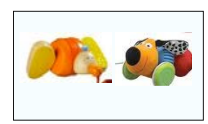
Can we tell whether "manu" refers to  or  ?

No - both hypotheses are equally compatible with these data.

"manu"
"colat"



"bosa"
"gasser"



"kaki"
"regli"

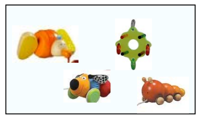
Why more may not always be harder…

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

Now, let's consider a more compex 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…

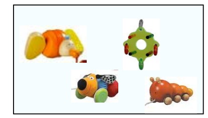
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

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

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

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

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



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

Why more may not always be harder…

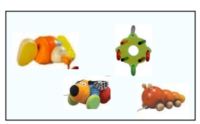
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

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

The second slide suggests "manu" can't be  - otherwise, that object would appear in the second slide.

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

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



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

Why more may not always be harder…

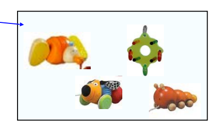
Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

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

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

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

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



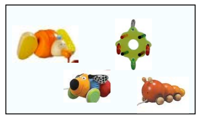
"manu"
"gasser"
"kaki"
"regli"

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"

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



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

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.

Why more may not always be harder…

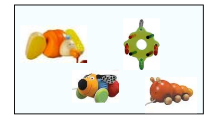
Let's walk through this scenario using Bayesian inference.

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

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

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

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



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

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"

Since there are four hypotheses in the hypothesis space at this point, the priors are:

P(H1) = 1/4 = 0.25
P(H2) = 1/4 = 0.25
P(H3) = 1/4 = 0.25
P(H4) = 1/4 = 0.25

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"

We can calculate the likelihoods, given this data point:

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

We can calculate the likelihood * prior for each hypothesis:

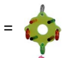
P(D | H1)*P(H1) = 1*0.25 = 0.25
P(D | H2)*P(H2) = 1*0.25 = 0.25
P(D | H3)*P(H3) = 1*0.25 = 0.25
P(D | H4)*P(H4) = 1*0.25 = 0.25

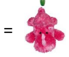
The sum (which we'll need for the denominator of the posterior) = 1

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

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"

We can now calculate the posterior for each hypothesis:

P(H1 | D) = 0.25/1 = 0.25
P(H2 | D) = 0.25/1 = 0.25
P(H3 | D) = 0.25/1 = 0.25
P(H4 | D) = 0.25/1 = 0.25

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"

These become the priors for the next data point.

P(H1) = 0.25
P(H2) = 0.25
P(H3) = 0.25
P(H4) = 0.25

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"

We can calculate the likelihoods, given this data point:

P(D | H1) = 1
P(D | H2) = 1
P(D | H3) = 1
P(D | H4) = 0 ( doesn't appear)

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"

We can calculate the likelihood * prior for each hypothesis:

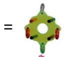
P(D | H1)*P(H1) = 1*0.25 = 0.25
P(D | H2)*P(H2) = 1*0.25 = 0.25
P(D | H3)*P(H3) = 1*0.25 = 0.25
P(D | H4)*P(H4) = 0*0.25 = 0

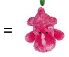
The sum (which we'll need for the denominator of the posterior) = 0.75

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

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"

We can now calculate the posterior for each hypothesis:

P(H1 | D) = 0.25/0.75 = 0.33
P(H2 | D) = 0.25/0.75 = 0.33
P(H3 | D) = 0.25/0.75 = 0.33
P(H4 | D) = 0/0.75 = 0

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

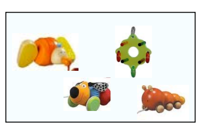
Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

These become the priors for the next data point.

P(H1) = 0.33
P(H2) = 0.33
P(H3) = 0.33
P(H4) = 0

data point 3



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

Sequential updating

Hypothesis 1 (H1): "manu" = 

Hypothesis 2 (H2): "manu" = 

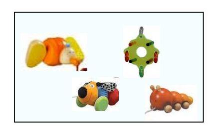
Hypothesis 3 (H3): "manu" = 

Hypothesis 4 (H4): "manu" = 

We can calculate the likelihoods, given this data point:

P(D | H1) = 0 ( doesn't appear)
P(D | H2) = 1
P(D | H3) = 0 ( doesn't appear)
P(D | H4) = 1

data point 3



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

Sequential updating

Hypothesis 1 (H1): “manu” =

Hypothesis 2 (H2): “manu” =

Hypothesis 3 (H3): “manu” =

Hypothesis 4 (H4): “manu” =

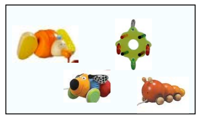
We can calculate the likelihood * prior for each hypothesis:

P(D | H1)*P(H1) = 0*0.33 = 0
P(D | H2)*P(H2) = 1*0.33 = 0.33
P(D | H3)*P(H3) = 0*0.33 = 0
P(D | H4)*P(H4) = 1*0 = 0

The sum (which we’ll need for the denominator of the posterior) = 0.33

Σ P(D|h)*P(h)

data point 3



“manu”
“gasser”
“kaki”
“regli”

Sequential updating

Hypothesis 1 (H1): “manu” =

Hypothesis 2 (H2): “manu” =

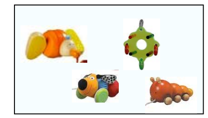
Hypothesis 3 (H3): “manu” =

Hypothesis 4 (H4): “manu” =

We can now calculate the posterior for each hypothesis:

P(H1 | D) = 0/0.33 = 0
P(H2 | D) = 0.33/0.33 = 1
P(H3 | D) = 0/0.33 = 0
P(H4 | D) = 0/0.33 = 0

data point 3



“manu”
“gasser”
“kaki”
“regli”

Sequential updating

Hypothesis 1 (H1): “manu” =

Hypothesis 2 (H2): “manu” =

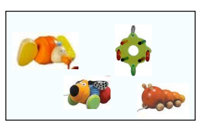
Hypothesis 3 (H3): “manu” =

Hypothesis 4 (H4): “manu” =

We can now calculate the posterior for each hypothesis:

P(H1 | D) = 0/0.33 = 0
P(H2 | D) = 0.33/0.33 = 1
P(H3 | D) = 0/0.33 = 0
P(H4 | D) = 0/0.33 = 0

data point 3



“manu”
“gasser”
“kaki”
“regli”

The utility of probabilities

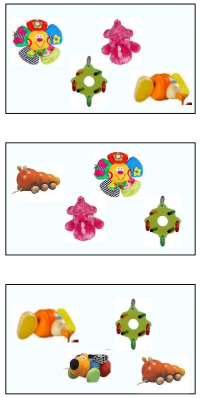
Partial knowledge of some words appears to be very helpful for helping learners figure out the meaning of words they don’t know yet (Yurovsky, Fricker, & Yu 2013).

“bosa”
“gasser”
“manu”
“colat”

P(H1 | D) = 0.25/0.75 = 0.33
P(H2 | D) = 0.25/0.75 = 0.33
P(H3 | D) = 0.25/0.75 = 0.33
P(H4 | D) = 0/0.75 = 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).

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.

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.

Questions?

You should be able to do up through question 7 on HW2 and up through question 5 on the word meaning review questions.