Ling 151/Psych 156A: Acquisition of Language II

Lecture 19

Poverty of the stimulus II

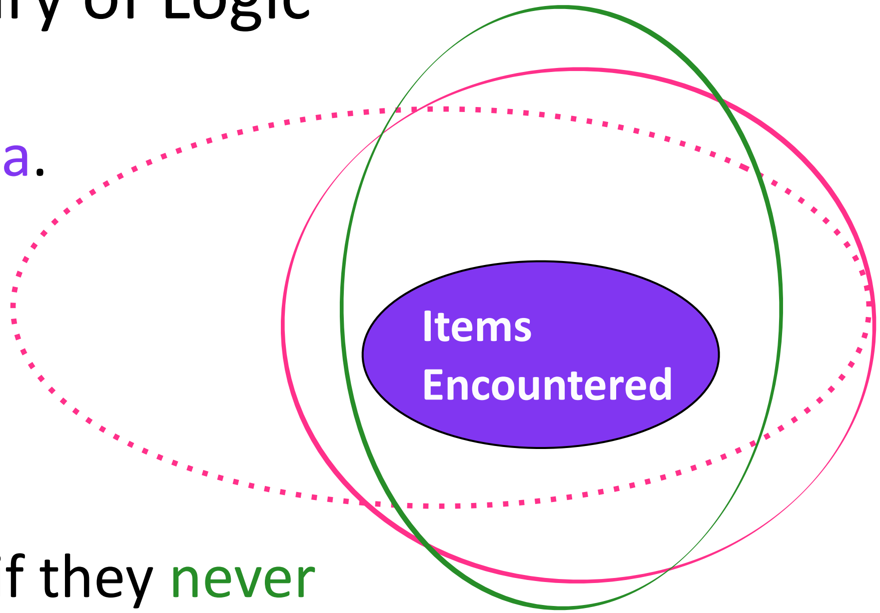
Announcements

Be working on HW7 (due: 3/7/18)

Be working on review questions

Please fill out course evaluation for this class - **in fact, let's take a few minutes and start/do it now in class.**

Poverty of the stimulus + constrained generalization leads to prior knowledge about language: Summary of Logic



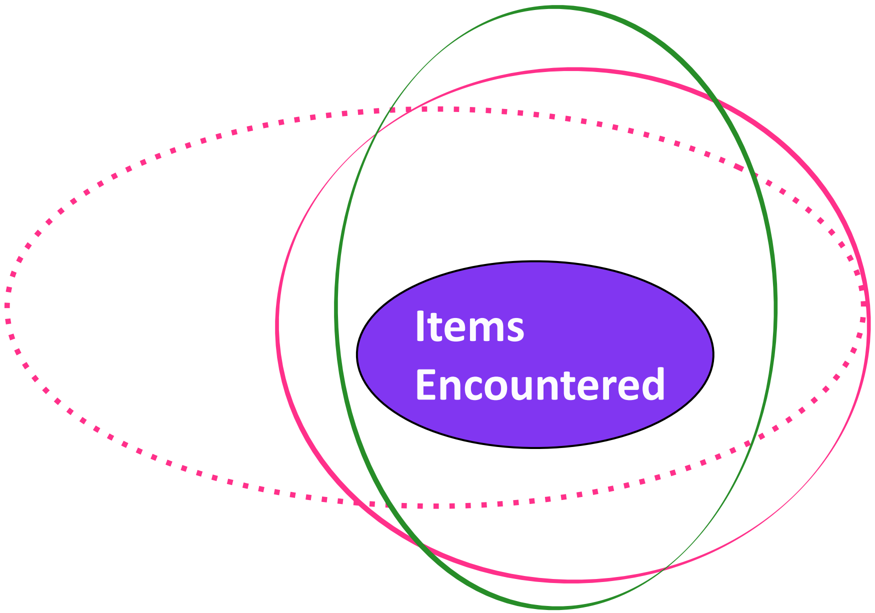
1) Suppose there are some data.

2) Suppose there are some incorrect hypotheses compatible with the data.

3) Suppose children behave as if they never entertain some of the incorrect hypotheses. That is, they make constrained generalizations.

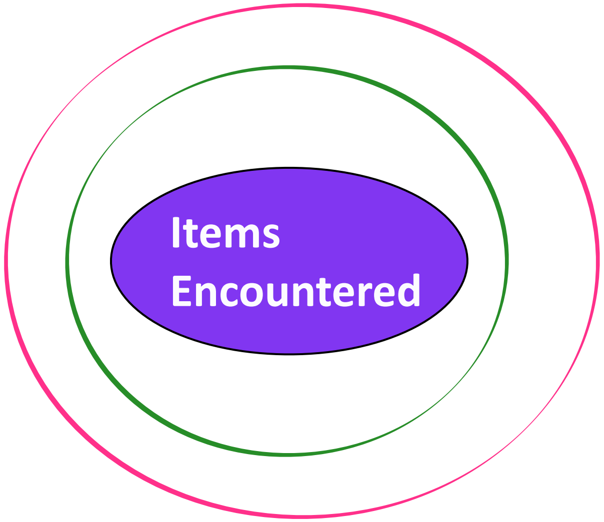
Conclusion: Children possess prior (possibly innate) knowledge ruling out those incorrect hypotheses from consideration.

Making generalizations that are underdetermined by the data



Children encounter a subset of the language's data, and have to decide how to generalize from that data

Bayesian inference is one way to do this, especially if the hypotheses are in a Subset-Superset relationship

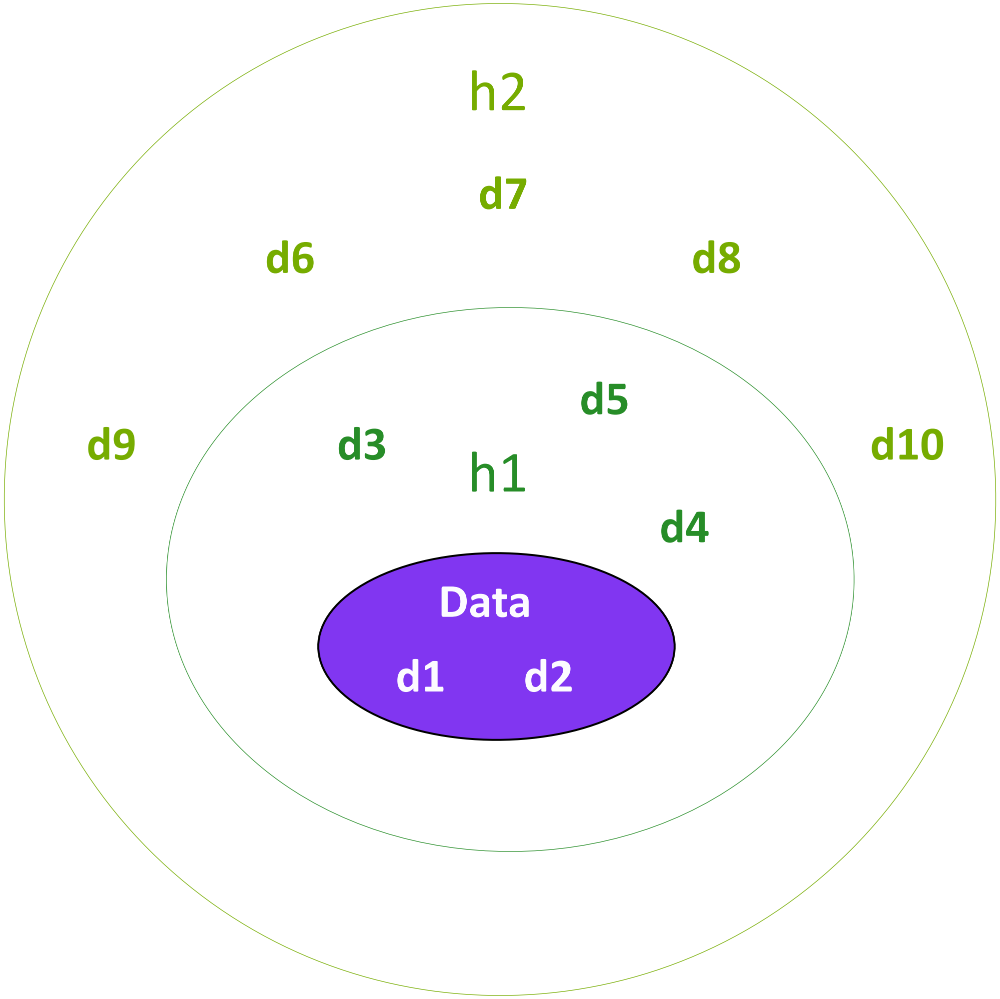


Children encounter a subset of the language's data, and have to decide how to generalize from that data

Bayesian reasoning

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.

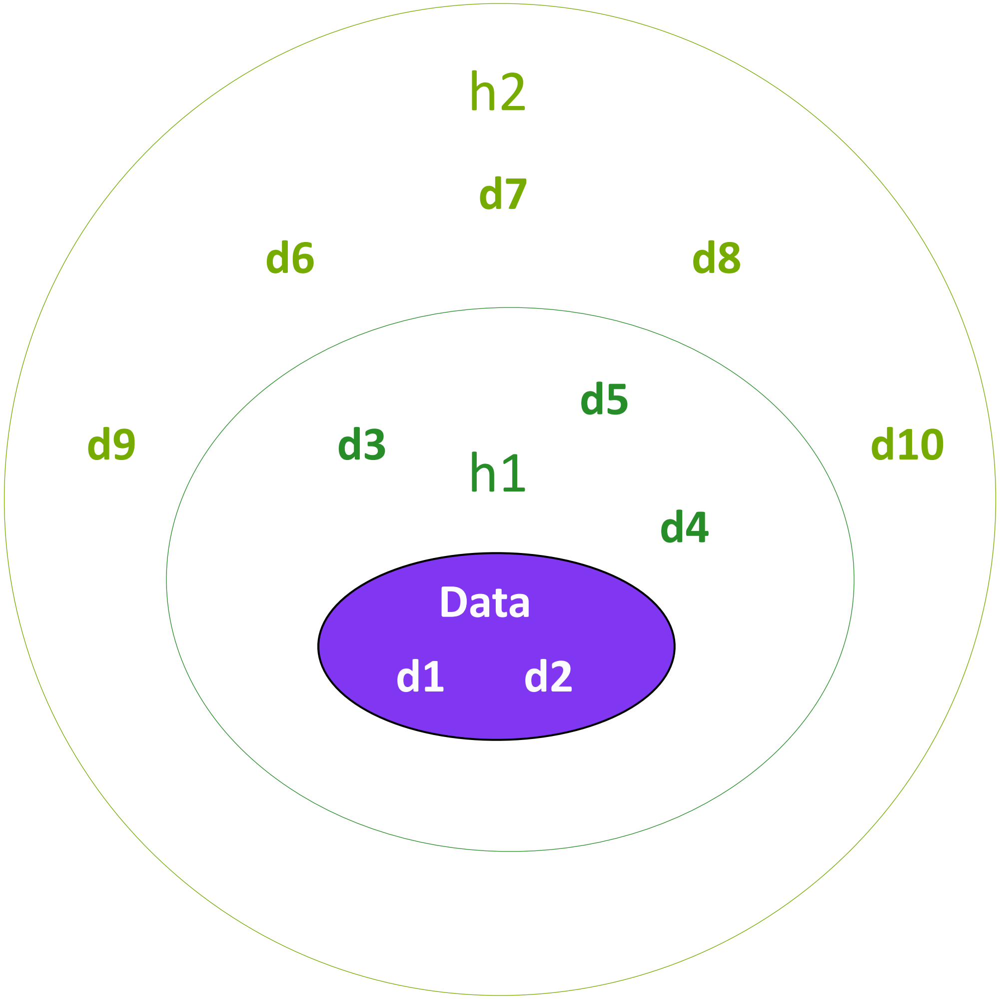
These hypotheses are also capable of generating other data points.



Bayesian reasoning

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

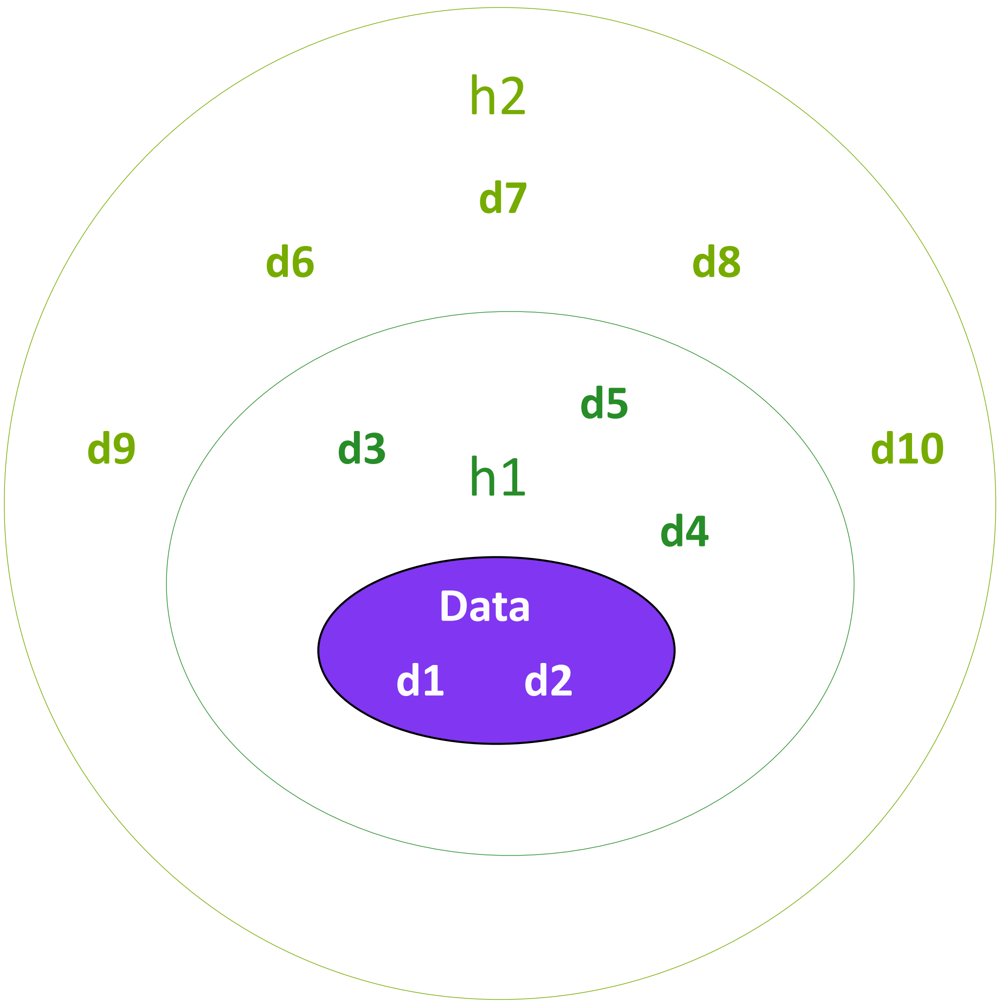
$$P(h|D)$$



Bayesian reasoning

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

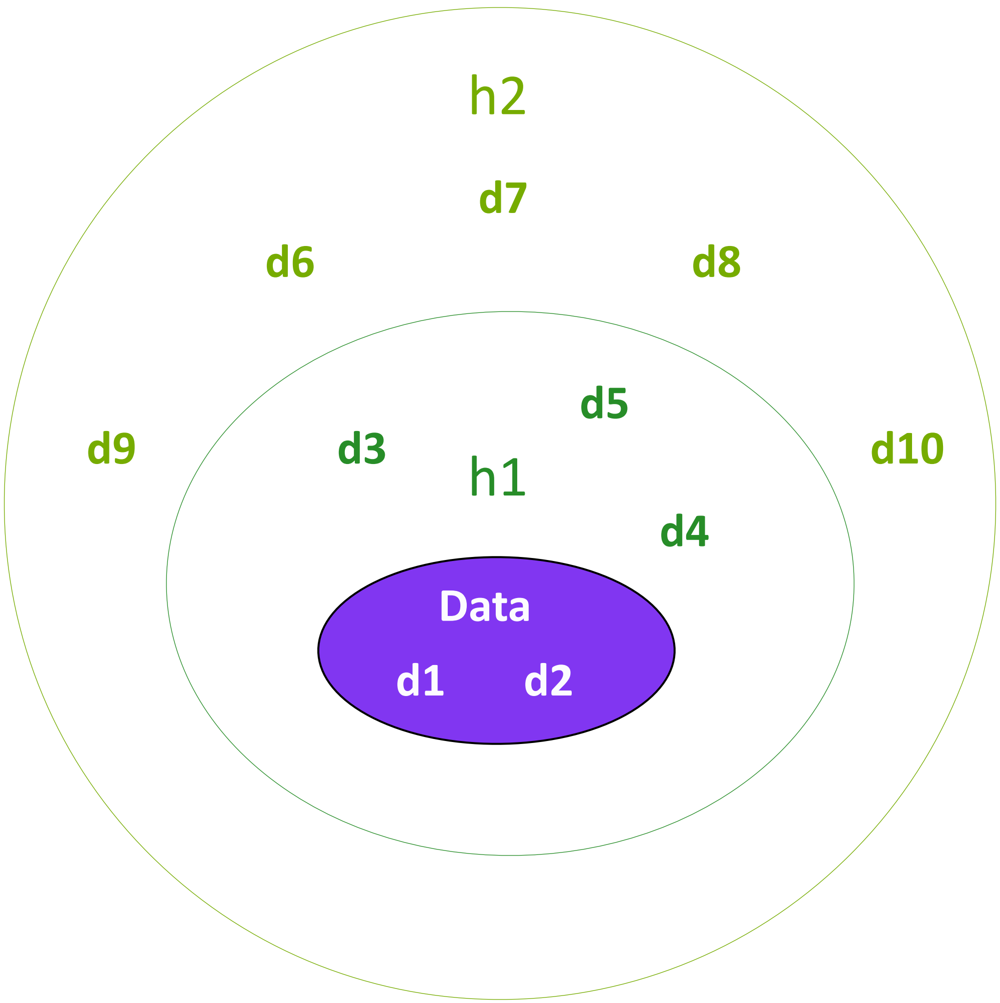
$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$



Bayesian reasoning

It also depends on P (h), which represents the *prior* of the hypothesis h. This encodes the probability of the hypothesis before any data have been encountered. Intuitively, this corresponds to how plausible the hypothesis is, irrespective of any data.

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



Bayesian reasoning

The posterior probability is proportional to the likelihood * the prior for each hypothesis.

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

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

Data D = d1 d2

likelihoods

P(D | h1) = 1/5 * 1/5 = 1/25

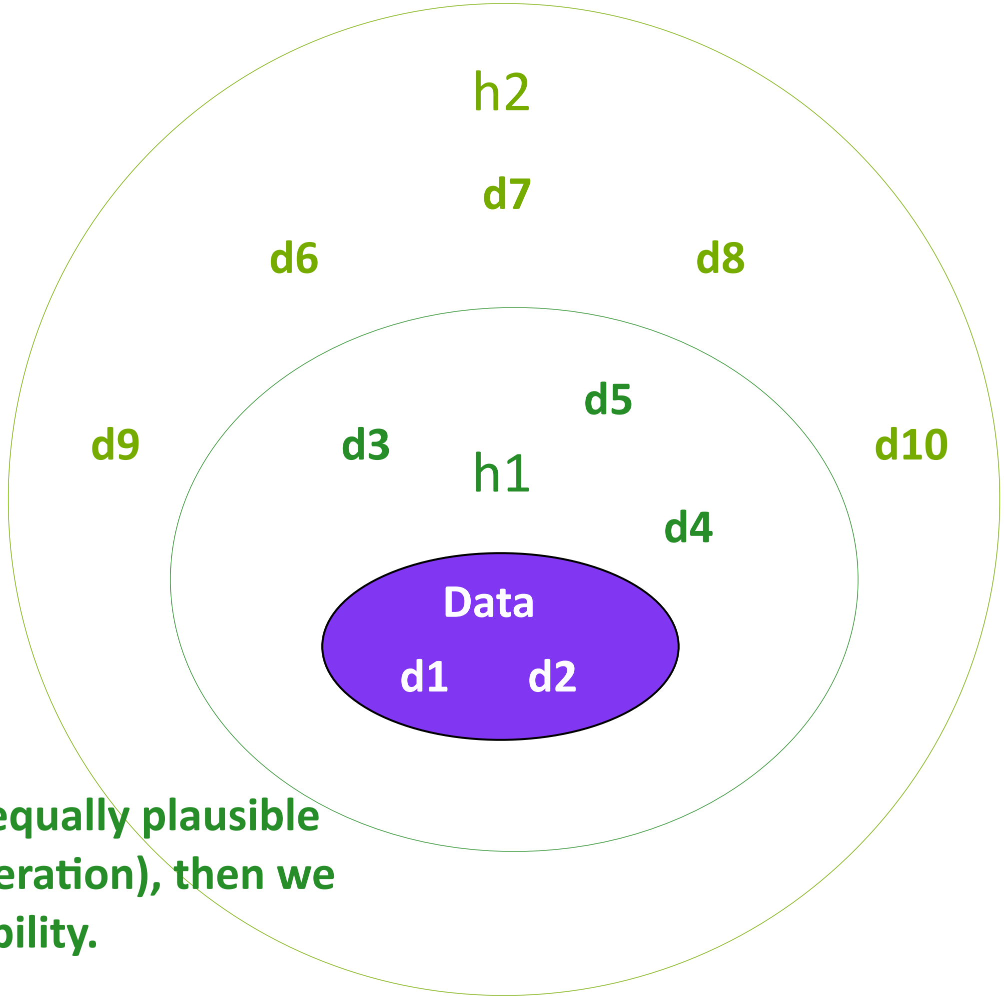
P(D | h2) = 1/10 * 1/10 = 1/100

priors

P(h1) = 1/2

P(h2) = 1/2

If we assume they're equally plausible (no other biases in operation), then we have a uniform probability.



Bayesian reasoning

The posterior probability is proportional to the likelihood * the prior for each hypothesis.

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

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

Data D = d1 d2

likelihoods

P(D | h1) = 1/25

P(D | h2) = 1/100

priors

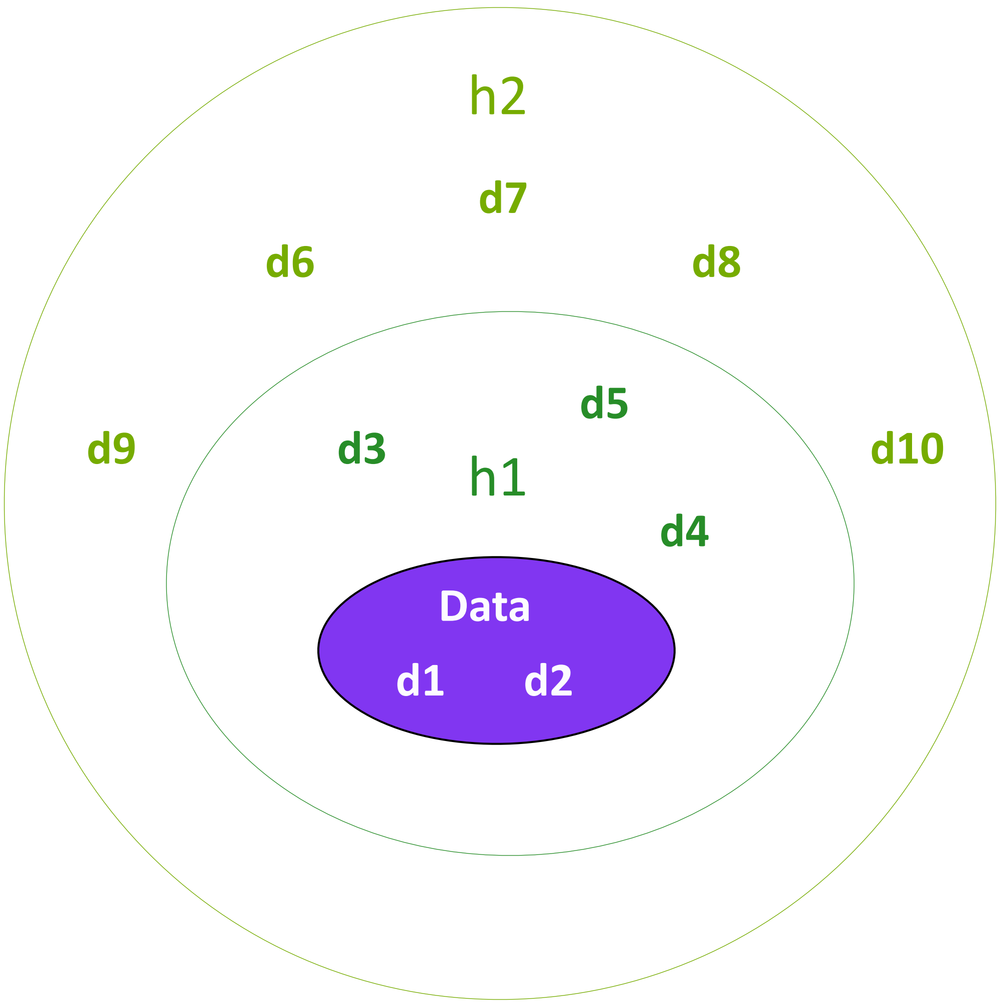
P(h1) = 1/2

P(h2) = 1/2

posteriors

P(h1 | D) = 1/25 * 1/2 = 1/50

P(h2 | D) = 1/100 * 1/2 = 1/200



Bayesian reasoning

The posterior probability is proportional to the likelihood * the prior for each hypothesis.

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

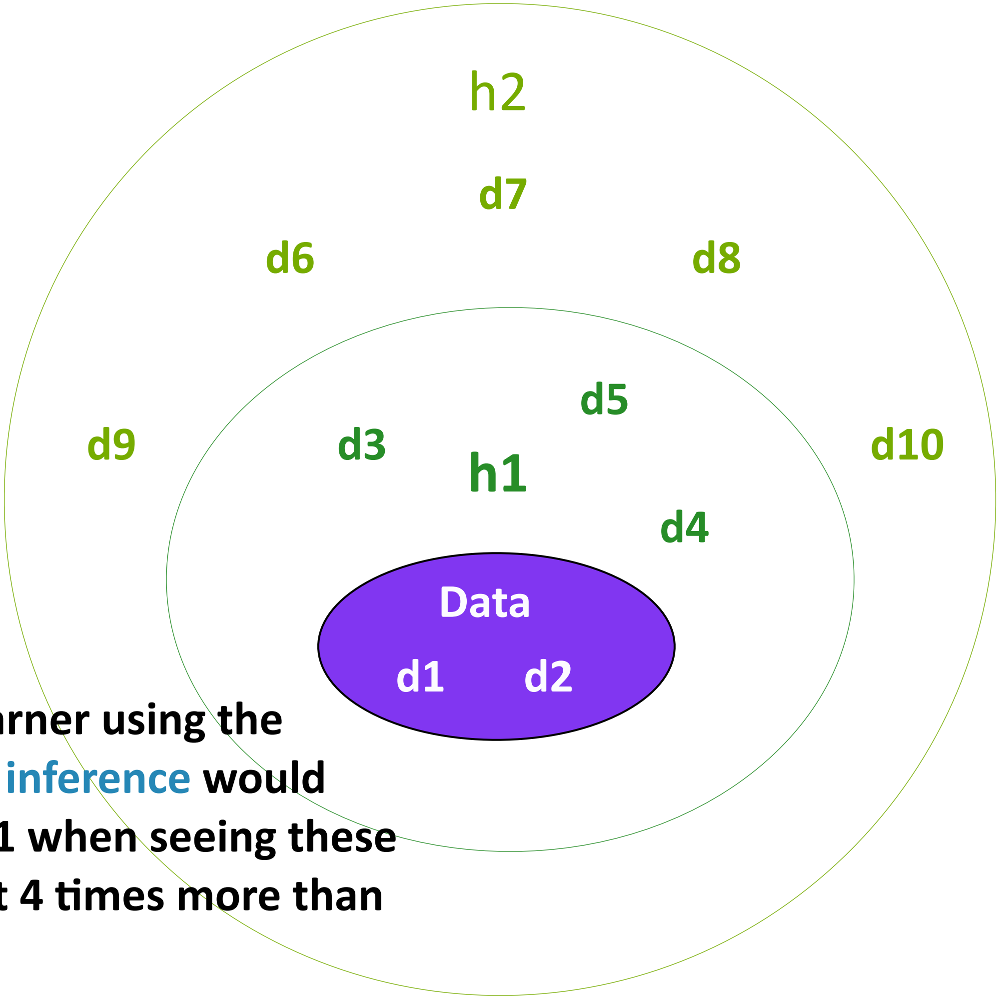
$$\propto P(D|h) * P(h)$$

Data D = d1 d2

posteriors

P(h1 | D) = 1/25 * 1/2 = 1/50

P(h2 | D) = 1/100 * 1/2 = 1/200

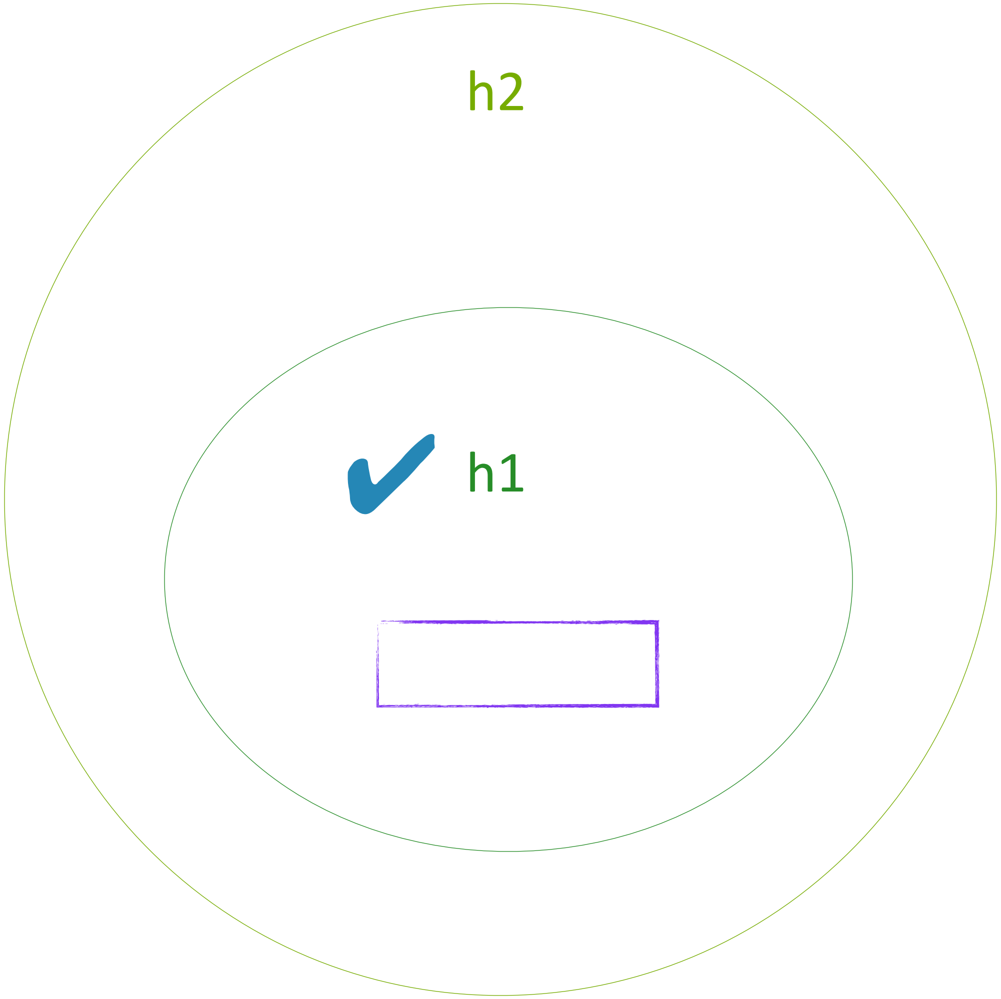


Even if no other biases are at work, a learner using the domain-general mechanism of Bayesian inference would prefer the smaller (subset) hypothesis h1 when seeing these ambiguous data. Here, it would prefer it 4 times more than h2. This is all due to the likelihood.

Bayesian reasoning

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

We have behavioral evidence that infants reason in a way that leads to similar conclusions when given this kind of scenario.



Bayesian reasoning

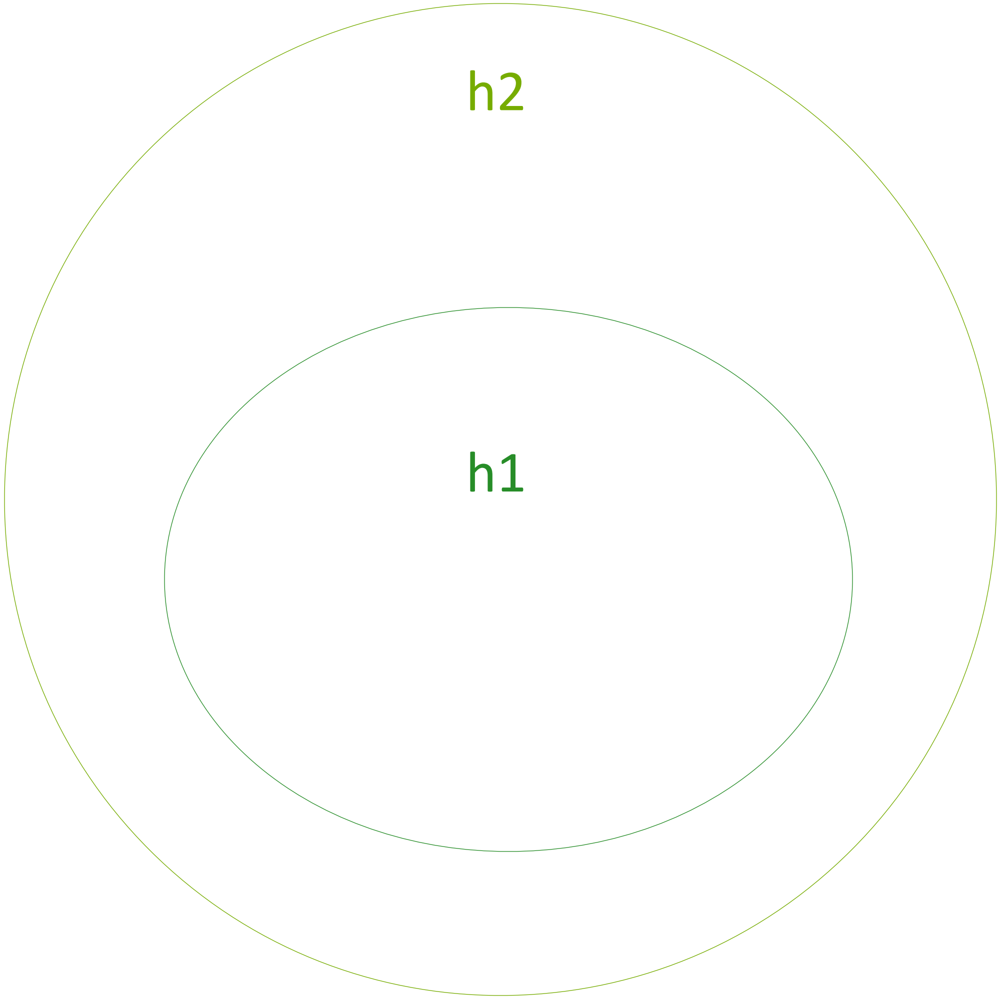
$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)}$$

Gerken 2006, 2010

artificial language study

Infants were trained on data from an artificial language, which consisted of words following a certain pattern.

Data D

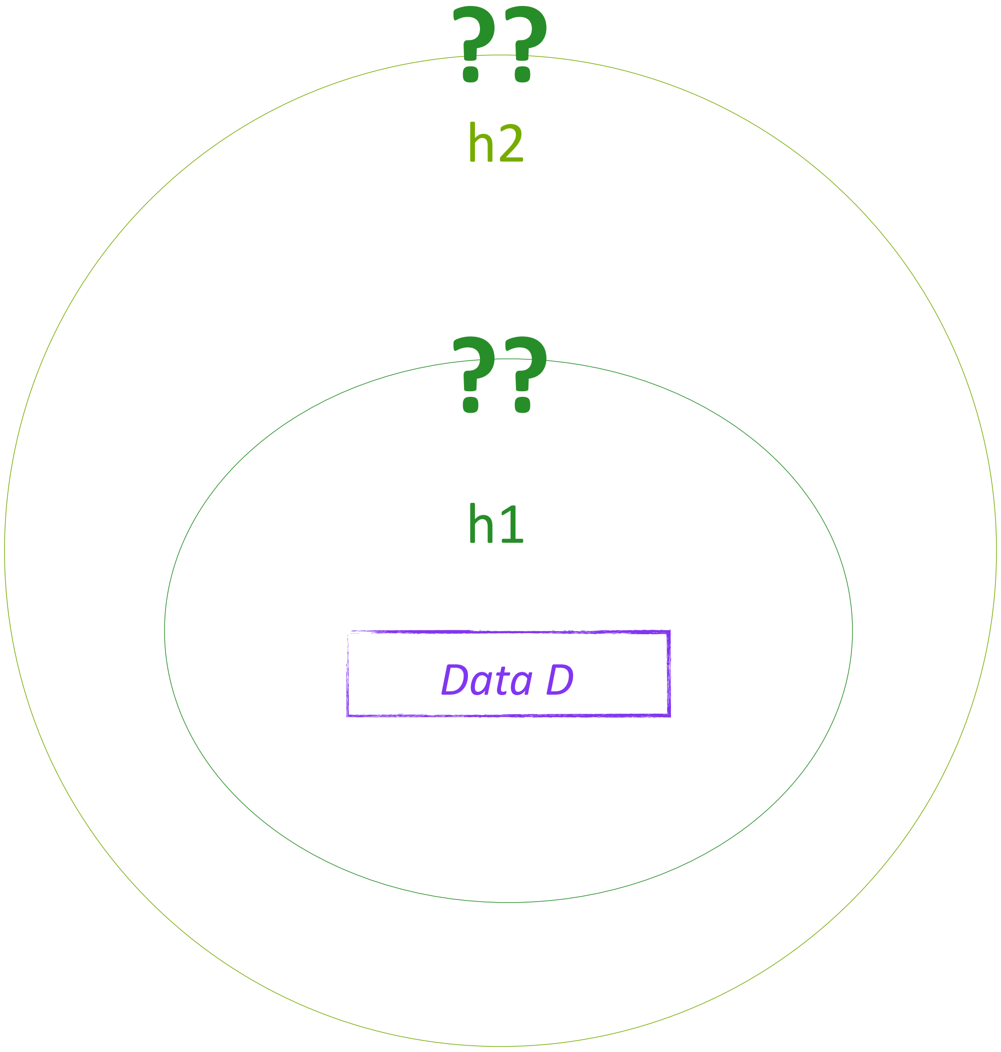


Bayesian reasoning

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

Gerken 2006, 2010
artificial language study

The infant's job: determine the **generalization** that describes the pattern for words of the artificial language.

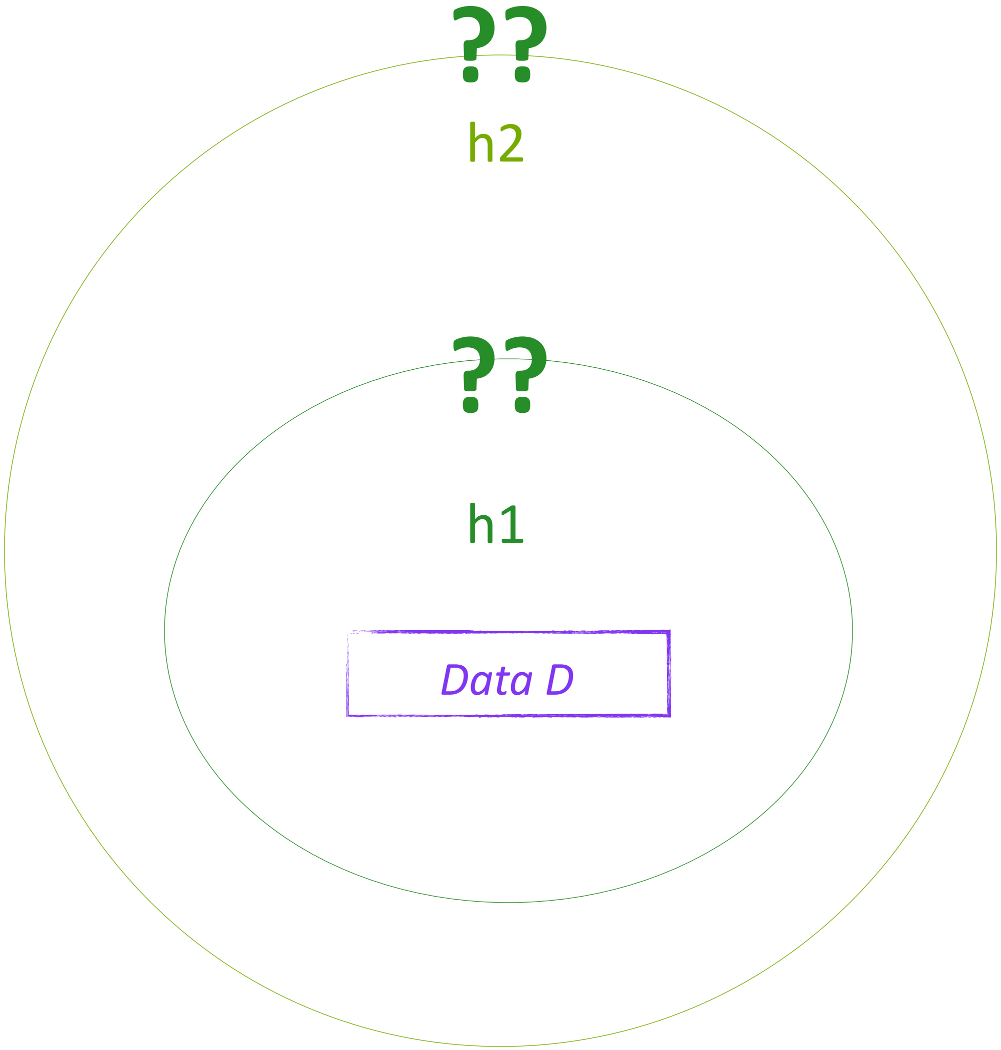


Bayesian reasoning

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

Gerken 2006, 2010
artificial language study

Marcus et al. (1999) found that very young infants will notice that words made up of 3 syllables follow a pattern that can be represented as AAB or ABA.



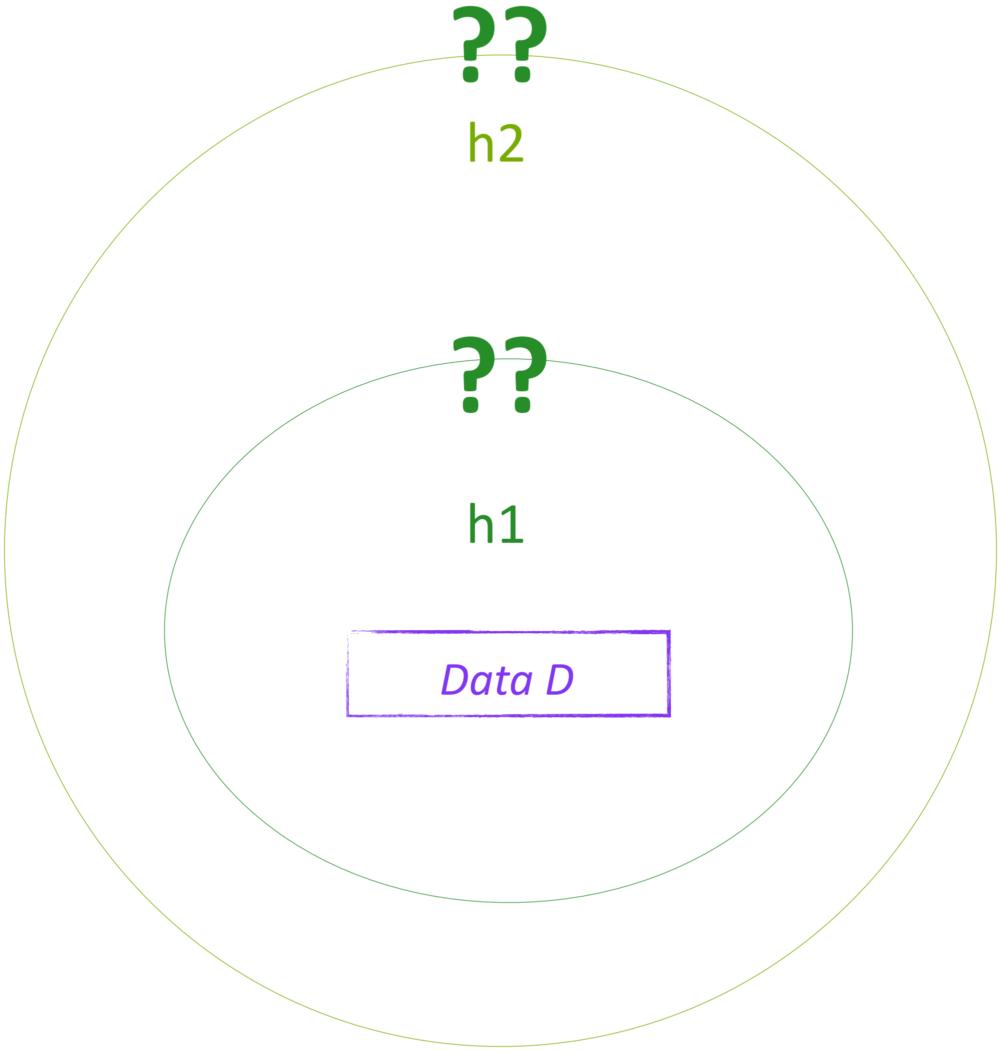
Bayesian reasoning

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

Gerken 2006, 2010
artificial language study

Marcus et al. (1999) found that very young infants will notice that words made up of 3 syllables follow a pattern that can be represented as AAB or ABA.

Example:
A syllables = le, wi
B syllables = di, je



Bayesian reasoning

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

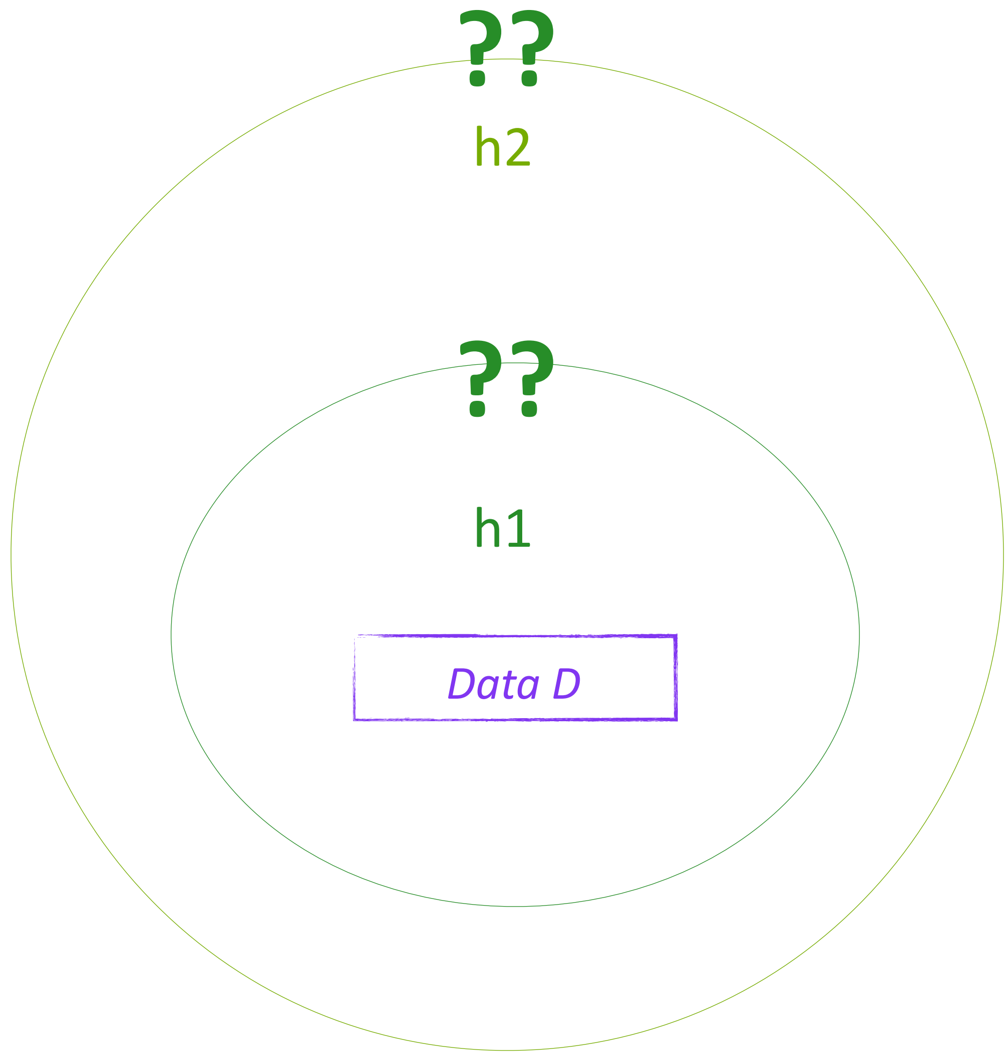
Gerken 2006, 2010
artificial language study

AAB or ABA

A syllables = le, wi
B syllables = di, je

AAB language words:
leledi, leleje, wiwidi, wiwije

ABA language words:
ledile, lejele, widiwi, wijewi



Bayesian reasoning

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

Gerken 2006, 2010
artificial language study

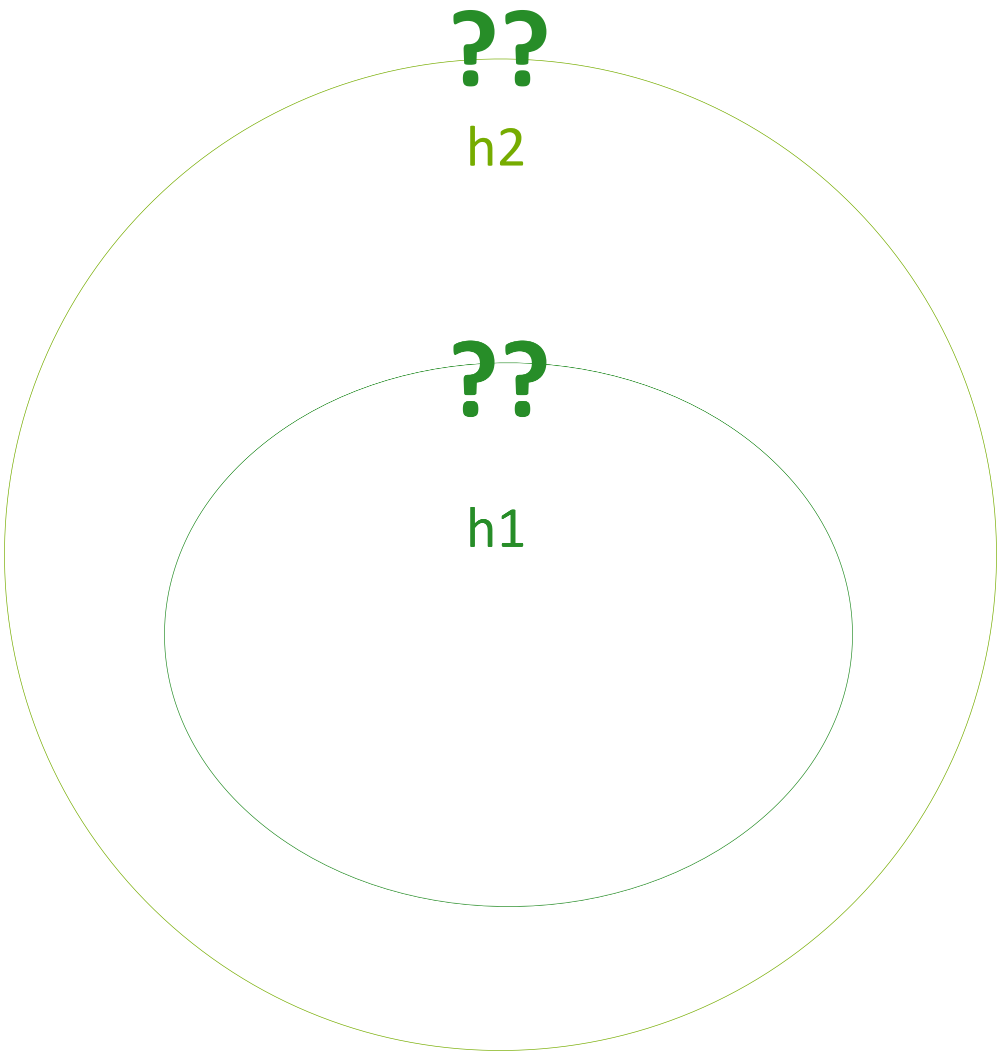
AAB or ABA

AAB language words:
leledi, leleje, wiwidi, wiwije

ABA language words:
ledile, lejele, widiwi, wijewi

What kind of generalization would children make if they were given particular kinds of data from these same artificial languages?

Data D



Bayesian reasoning

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

Gerken 2006, 2010

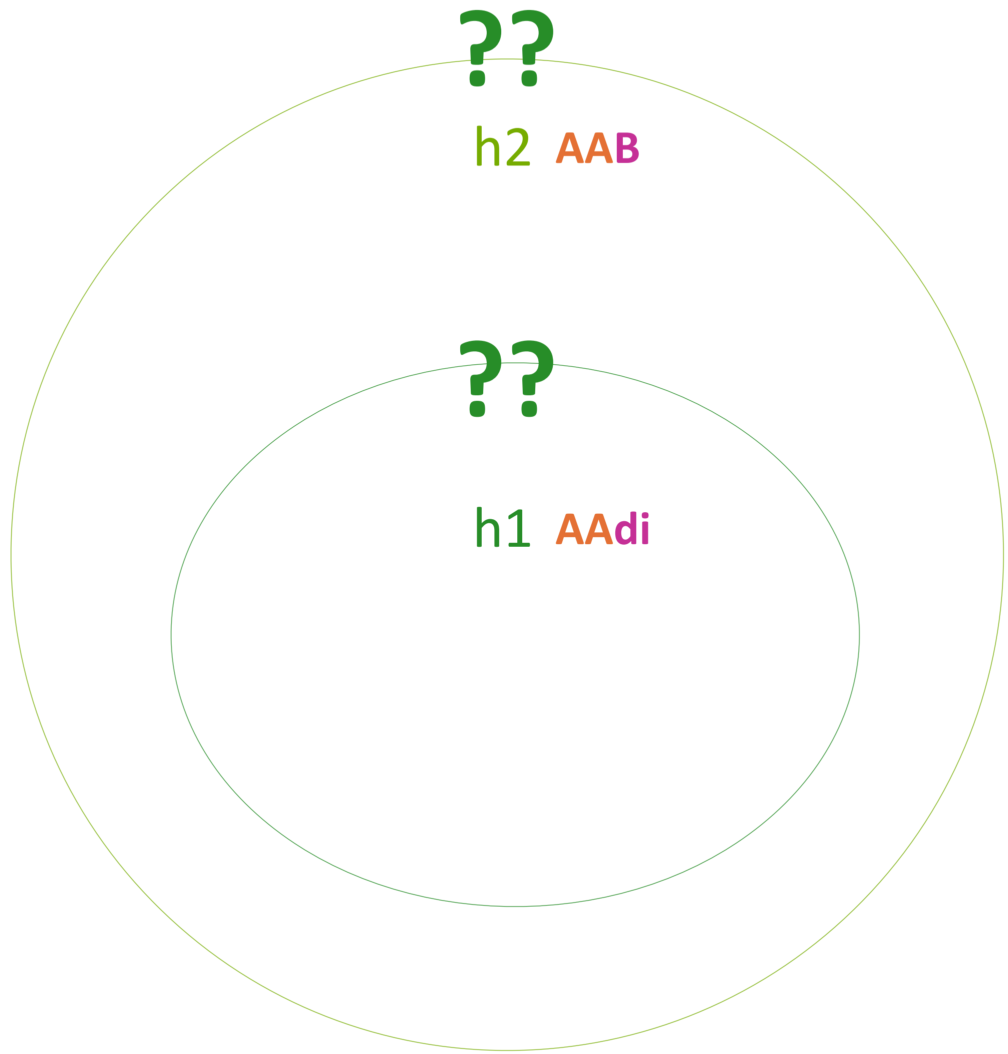
artificial language study

AAB

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Infants only see a subset of the language

Data D



Bayesian reasoning

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

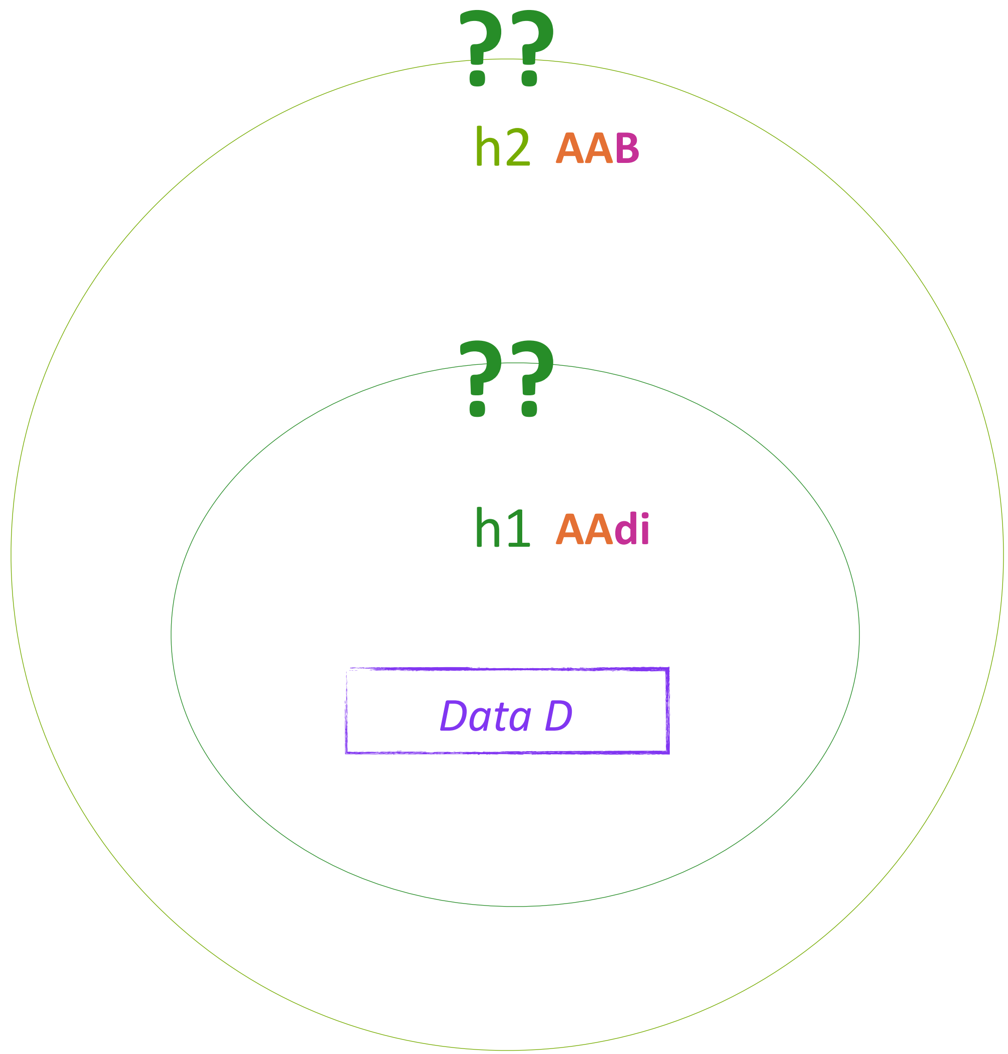
Gerken 2006, 2010
artificial language study

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Experimental condition

Training on four word types:
leledi, wiwidi, jijidi, dededi

Consistent with both a less-general hypothesis (h1) and a more-general hypothesis (h2).



Bayesian reasoning

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

Gerken 2006, 2010

artificial language study

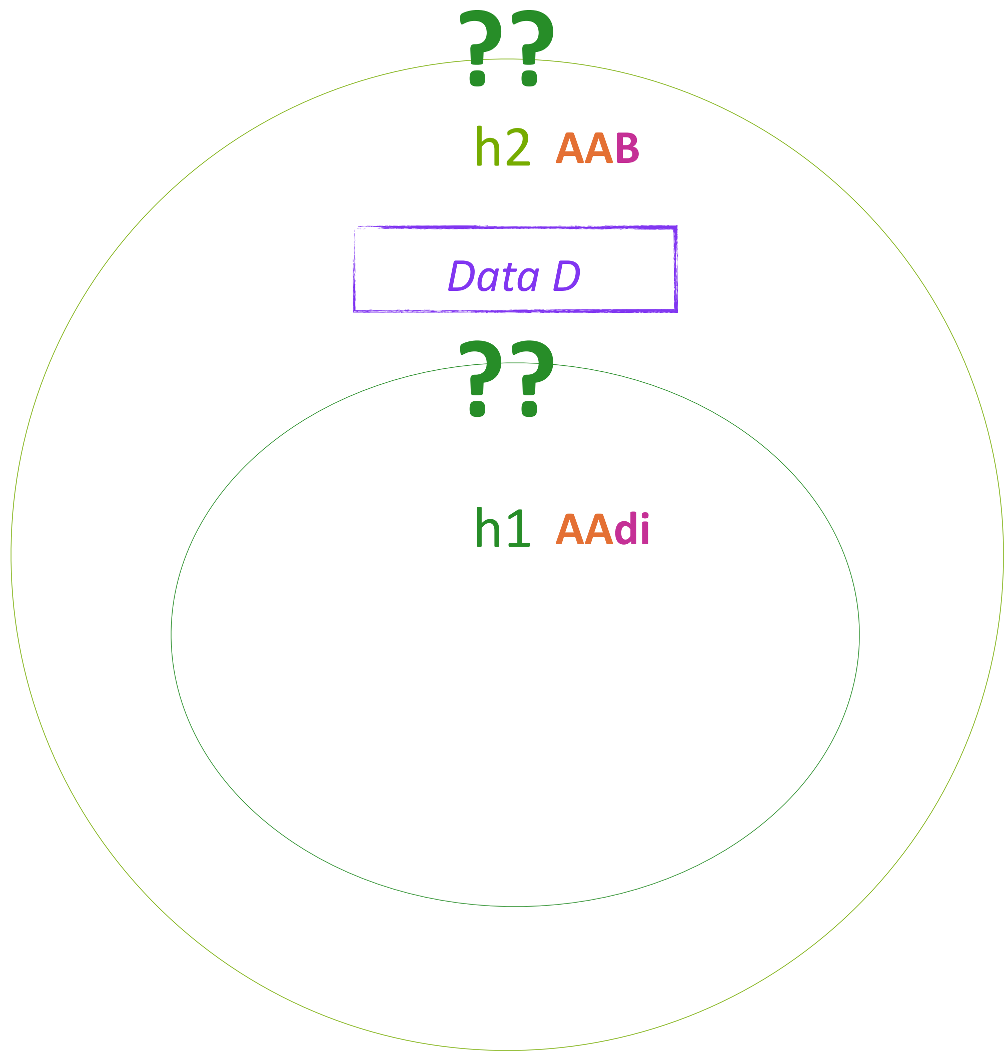
AAB

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Control condition

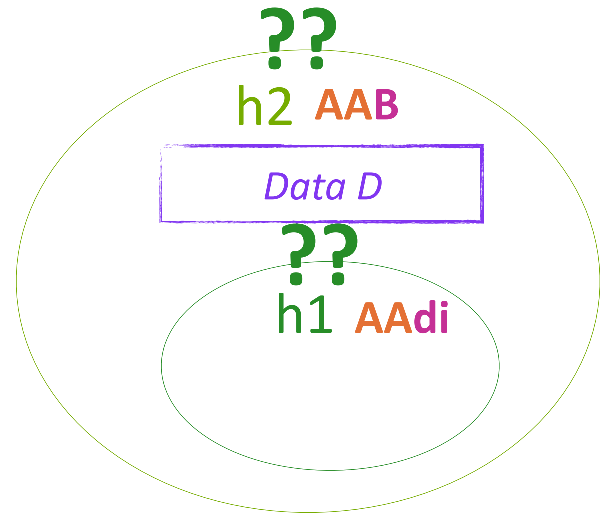
Training on four word types:
leledi, wiwije, jijili, dedewe

Consistent only with the more-general hypothesis (h2).



Bayesian reasoning

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



Gerken 2006, 2010

artificial language study

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	ededi	dedeje	dedeli	dedewe

Control condition

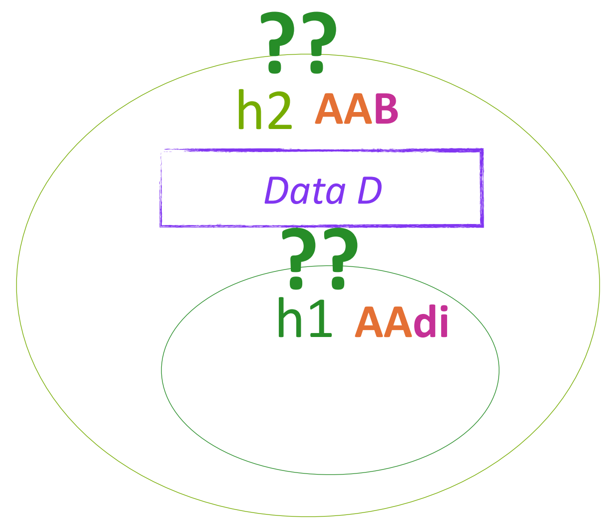
Training on four word types:
leledi, wiwije, jijili, dedewe

Consistent only with the more-general hypothesis (h2).

This control condition is used to see what children's behavior is when the data are only consistent with one of the generalizations.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Control condition

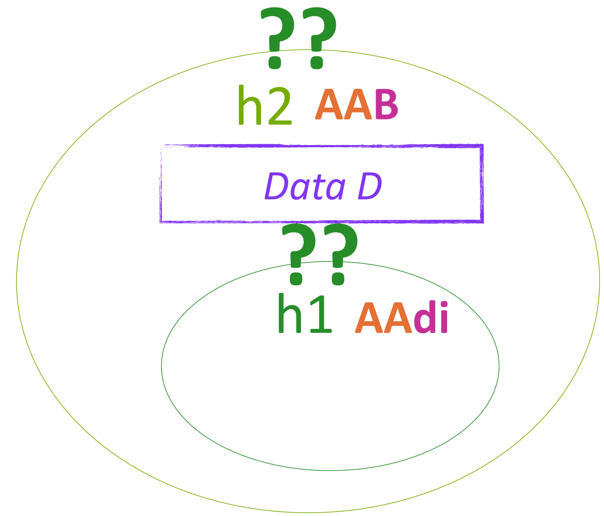
Training on four word types:
leledi, wiwije, jijili, dedewe

Consistent only with the more-general hypothesis (h2).

If children fail to make the generalization in the control condition, then the results in the experimental condition will not be informative. (Perhaps the task was too hard for children.)

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure

with 9-month-olds

Control condition

leledi, wiwije, jijili, dedewe

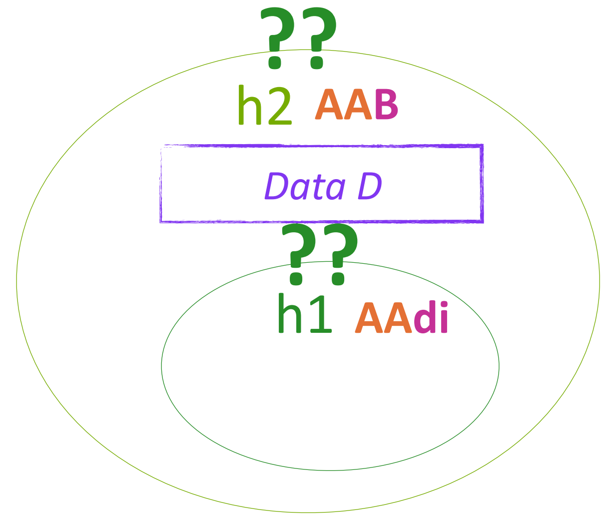
Training: 2 minutes hearing artificial language words

Test: AAB pattern words using novel syllables vs.
ABA pattern words using novel syllables

Ex: novel syllables: ko, ba
kokoba vs.
kobako

Bayesian reasoning

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



Gerken 2006, 2010

artificial language study

Task type: Head Turn Preference Procedure

with 9-month-olds

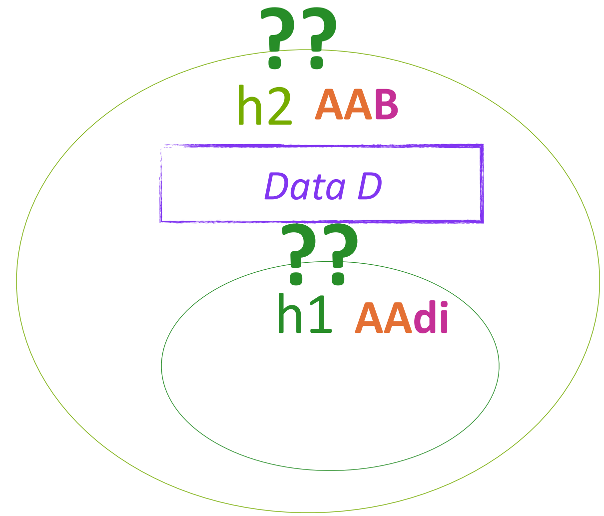
Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako

Behavior: If children learn the more-general pattern (AAB), they will prefer to listen to an AAB pattern word like kokoba, over a word that does not follow the AAB pattern, like kobako.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

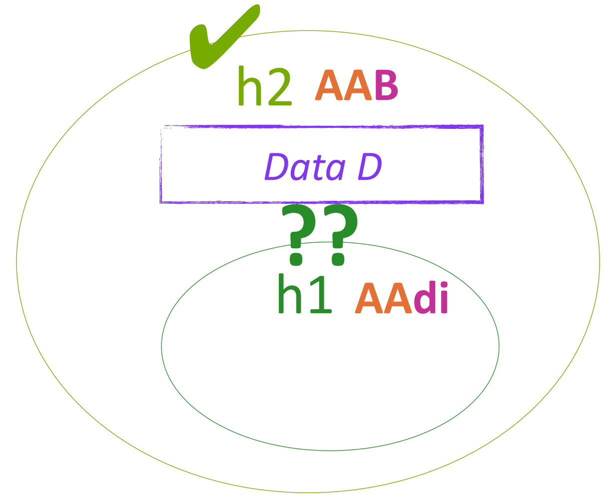
Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako

Behavior: Children listened longer on average to test items consistent with the AAB pattern [13.51 sec], as opposed to items inconsistent with it [10.14 sec].

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

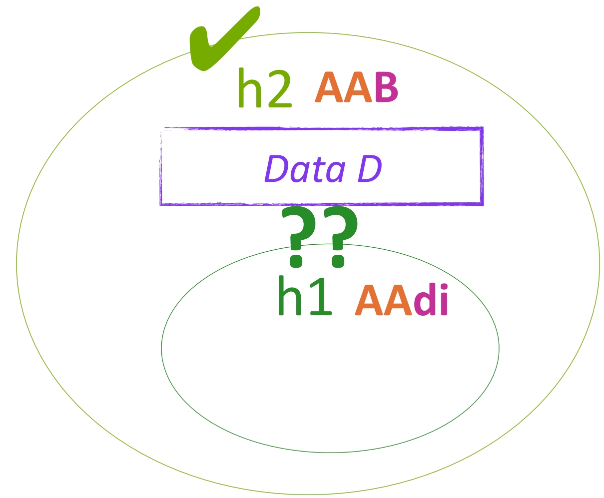
Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

They can notice the AAB pattern and make the generalization from this artificial language data. This task isn't too hard for infants.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

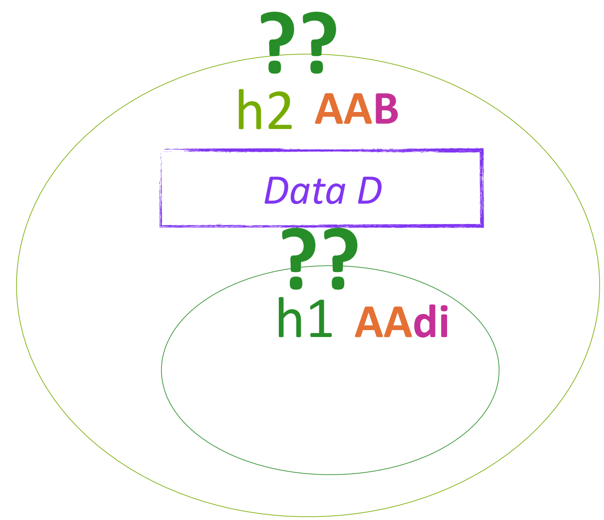
Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

Gerken & Knight 2015, Gerken & Quam 2017: In fact, it might be pretty easy for infants as indicated by their familiarity preference.

Bayesian reasoning

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



Gerken 2006, 2010

artificial language study

Task type: Head Turn Preference Procedure

with 9-month-olds

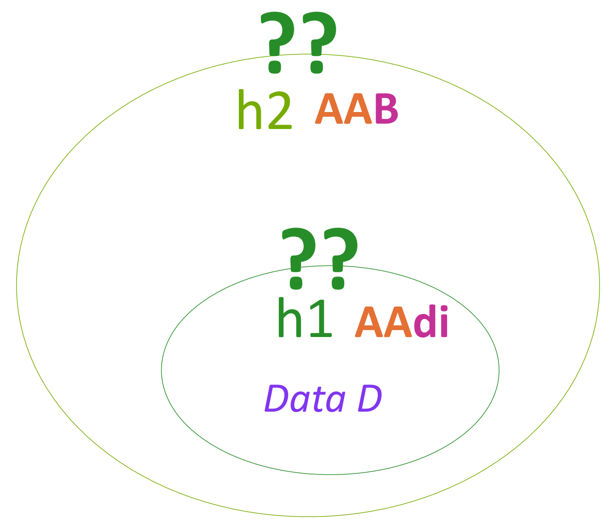
Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

What about the experimental condition?

Bayesian reasoning

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



Gerken 2006, 2010

artificial language study

Task type: Head Turn Preference Procedure

with 9-month-olds

Control condition

Training leledi, wiwije, jijili, dedewe

Test kokoba vs. kobako

Behavior ✔

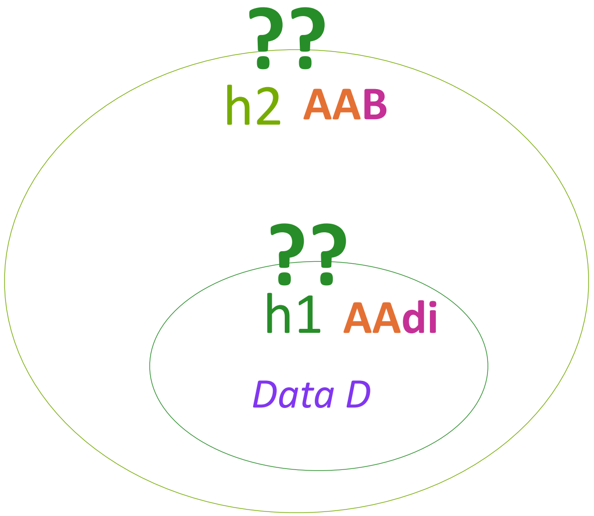
Experimental condition

Training leledi, wiwidi, jijidi, dededi

Consistent with both a less-general hypothesis (h1) and a more-general hypothesis (h2).

Bayesian reasoning

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



Gerken 2006, 2010

artificial language study

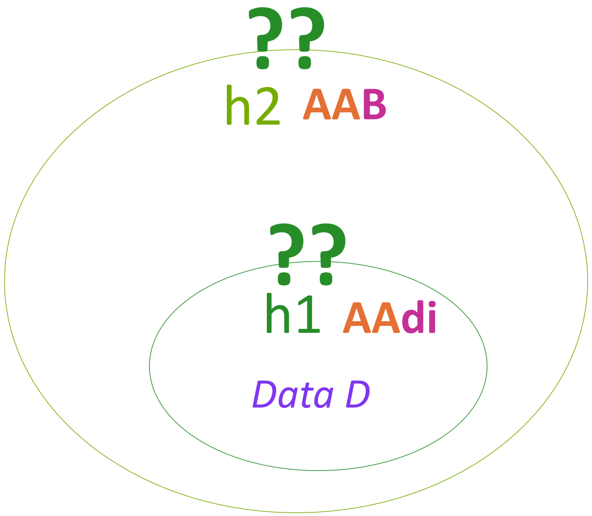
Task type: Head Turn Preference Procedure with 9-month-olds

	Control condition		**Experimental condition**
Training	leledi, wiwije, jijili, dedewe	**Training**	leledi, wiwidi, jijidi, dededi
Test	kokoba vs. kobako	**Test**	kokoba vs. kobako
Behavior	✔		✔

Behavior: If children learn the more-general pattern (AAB), they will prefer to listen to an AAB pattern word like kokoba, over a word that does not follow the AAB pattern, like kobako.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure

with 9-month-olds

Control condition

Training leledi, wiwije, jijili, dedewe

Test kokoba vs. kobako

Behavior ✔

Experimental condition

Training leledi, wiwidi, jijidi, dededi

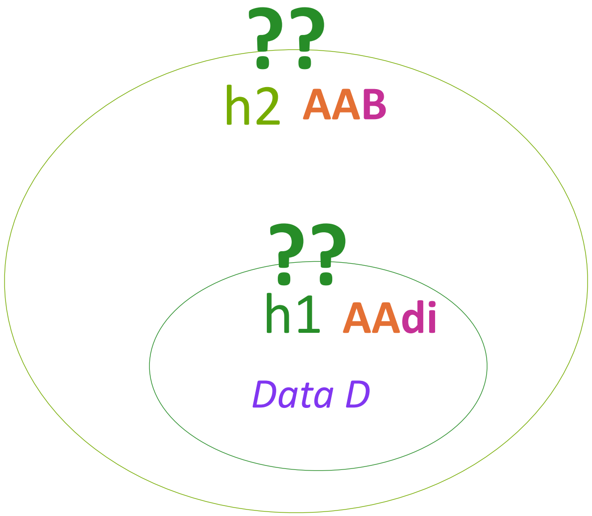
Test kokoba vs. kobako

??

Behavior: If children learn the less-general pattern (AAdi) or no pattern at all, they will not prefer to listen to an AAB pattern word like kokoba, over a word that does not follow the AAB pattern, like kobako.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition

Training leledi, wiwije, jijili, dedewe

Test kokoba vs. kobako

Behavior ✔

Experimental condition

Training leledi, wiwidi, jijidi, dededi

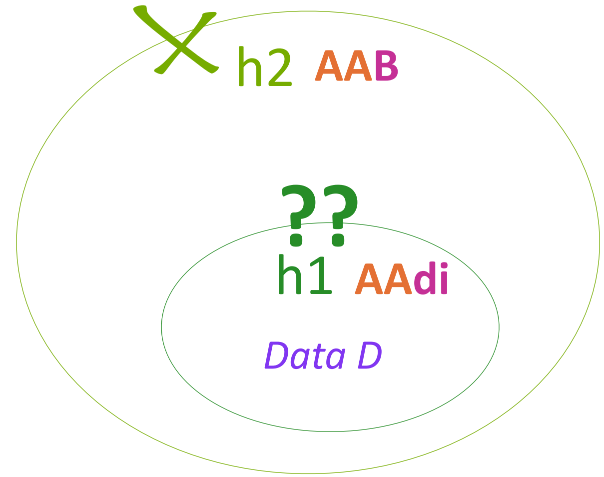
Test kokoba vs. kobako

Behavior ??

Behavior: Children did ***not*** listen longer on average to test items consistent with the AAB pattern [10.74 sec], as opposed to items inconsistent with it [10.18 sec].

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition	
Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

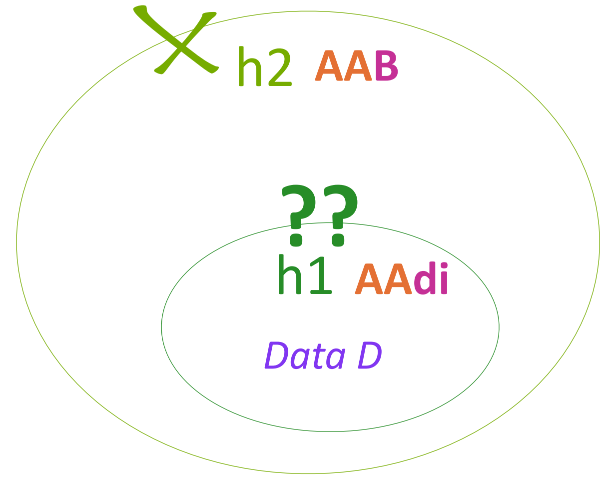
Experimental condition	
Training	leledi, wiwidi, jijidi, profile dededi
Test	kokoba vs. kobako
Behavior	??

They don't learn the more-general pattern. They either learned the less-general pattern or no pattern at all.

Which one is it?

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition

Training leledi, wiwije, jijili, dedewe

Test kokoba vs. kobako

Behavior

Experimental condition

Training leledi, wiwidi, jijidi, dededi

Test kokoba vs. kobako

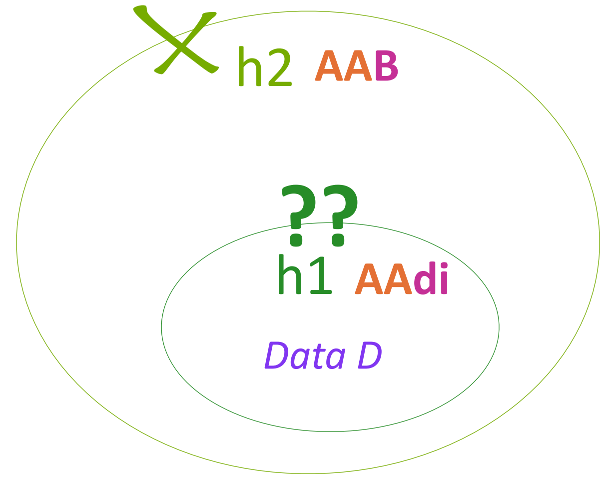
Behavior ??

Test kokodi vs. kodiko

Behavior: If they learn the less-general pattern, they'll prefer to listen to AAdi words like kokodi.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

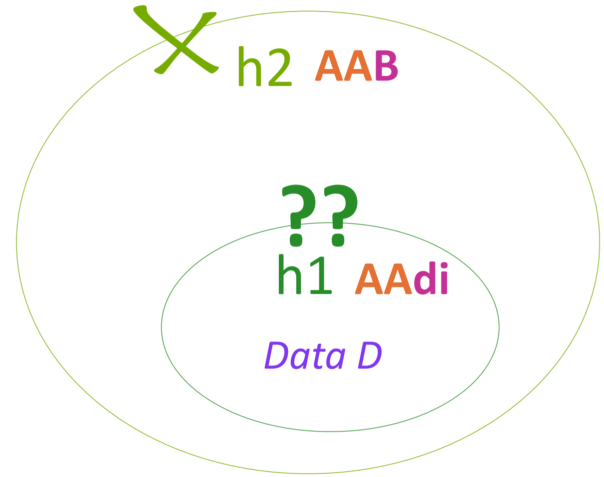
Experimental condition

Training	leledi, wiwidi, jijidi, dededi
Test	kokoba vs. kobako
Behavior	??
Test	kokodi vs. kodiko
	??

Behavior: If they learn no pattern at all, they'll (again) have no preference.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

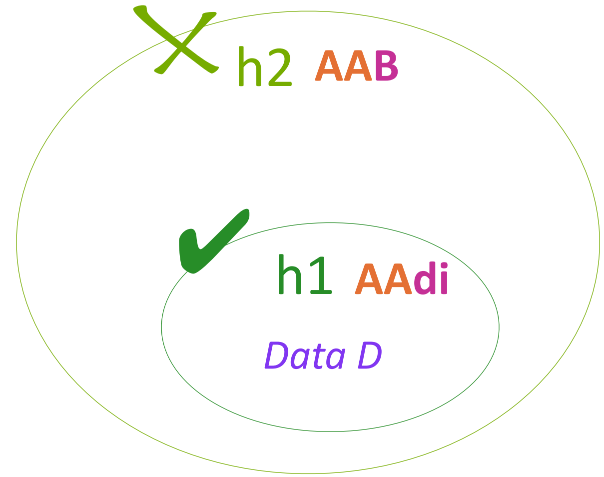
Experimental condition

Training	leledi, wiwidi, jijidi, dededi
Test	kokoba vs. kobako
Behavior	??
Test	kokodi vs. kodiko
Behavior	✔

Children prefer to listen to novel words that follow the less-general AAdi pattern [9.33 sec] over novel words that don't [6.25 sec].

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Task type: Head Turn Preference Procedure with 9-month-olds

Control condition

Training	leledi, wiwije, jijili, dedewe
Test	kokoba vs. kobako
Behavior	✔

Experimental condition

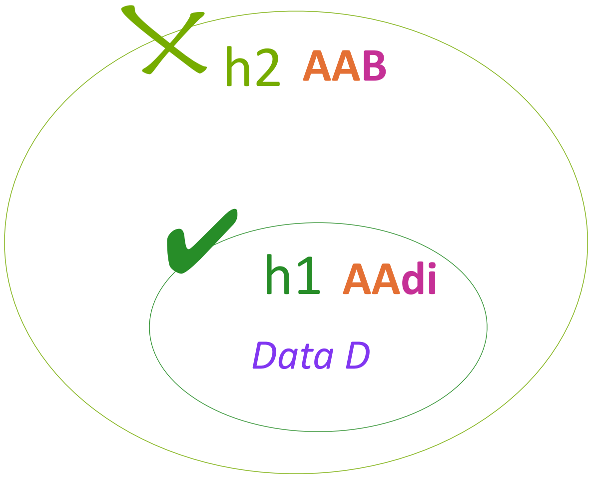
Training	leledi, wiwidi, jijidi, dededi
Test	kokoba vs. kobako
Behavior	??
Test	kokodi vs. kodiko
Behavior	✔

This means that given ambiguous data, they make the less-general generalization (h1) — just like a Bayesian learner would!

Bayesian reasoning

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

Gerken 2006, 2010
artificial language study



Training	leledi, wiwidi, jijidi, dededi
Test	kokodi vs. kodiko
Behavior	✔

Let's remind ourselves why this is

Bayesian reasoning

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

Gerken 2006, 2010

artificial language study

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Bayesian reasoning

Gerken 2006, 2010
artificial language study

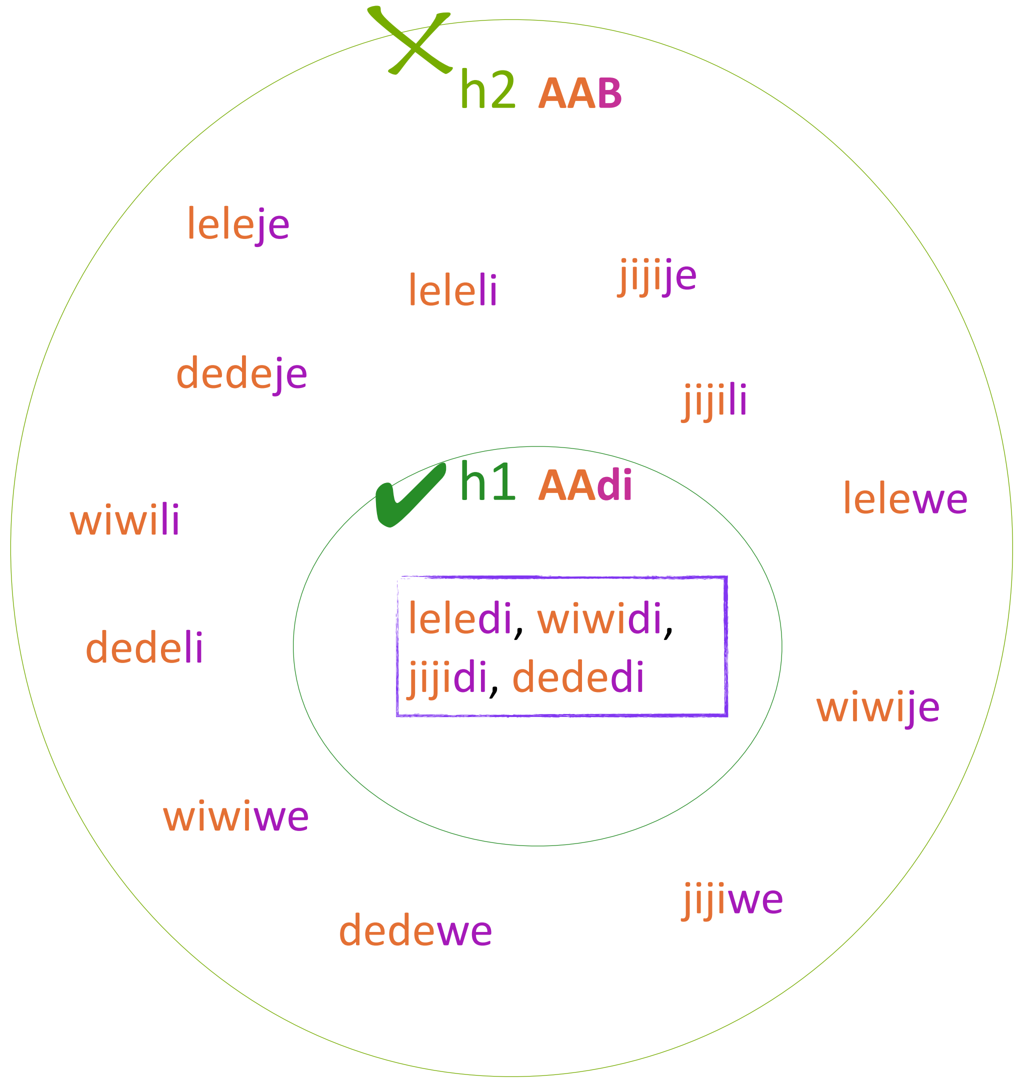
$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)}$$

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

likelihoods

P(D | h1) = 1/4*1/4*1/4*1/4 = 1/256

These are the only 4 data that can be generated, and so the probability of generating each one is 1/4. Let's focus on the types in the data intake, so we just have these four.



Bayesian reasoning

Gerken 2006, 2010
artificial language study

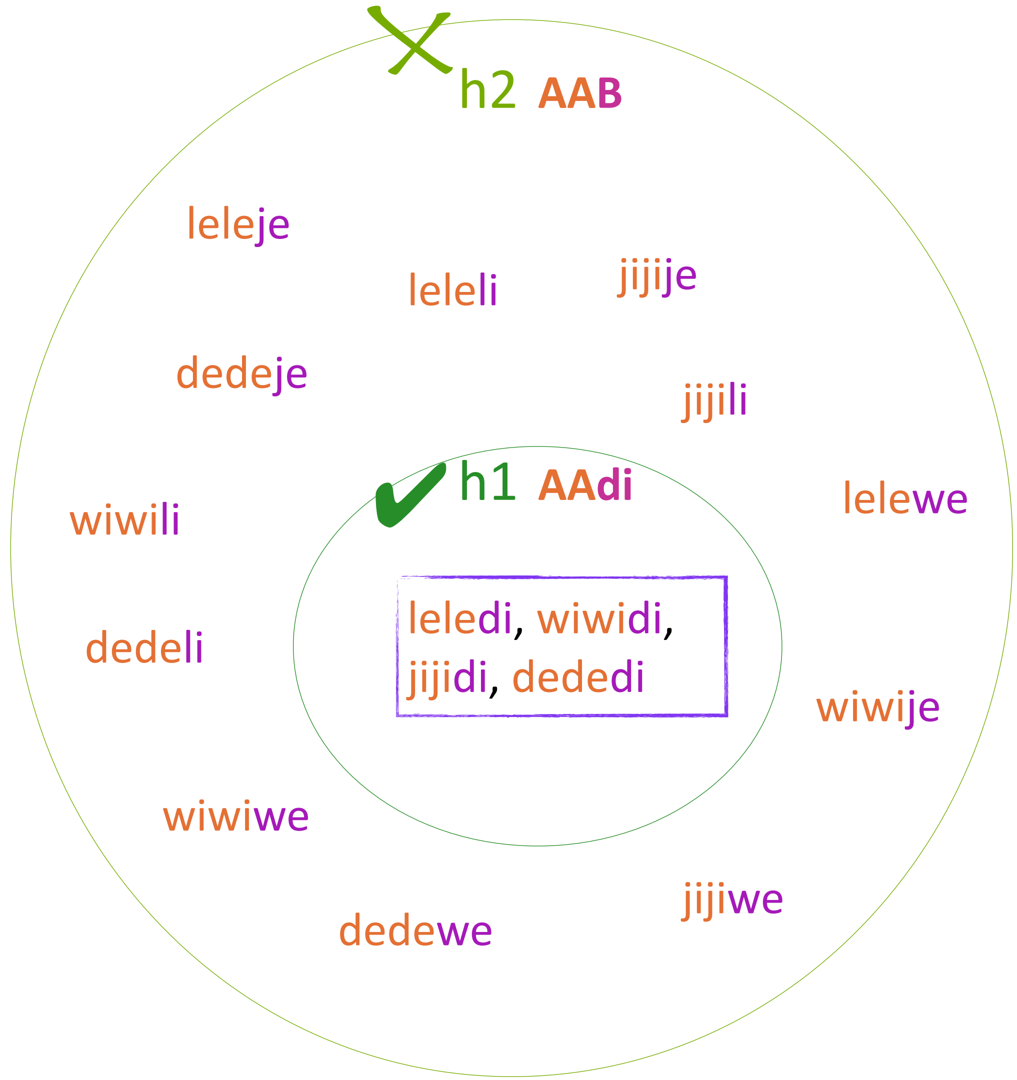
$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)}$$

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

likelihoods

P(D | h1) = 1/256

P(D | h2) = 1/16*1/16*1/16*1/16
= 1/65536



These are 16 data that can be generated, and so the probability of generating each one is 1/16.

Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

likelihoods

P(D | h1) = 1/256

P(D | h2) = 1/65536

priors

Let's assume the hypotheses are equally complex a priori, so they have uniform prior probability.

Bayesian reasoning

Gerken 2006, 2010

artificial language study

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

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

likelihoods

P(D | h1) = 1/256

P(D | h2) = 1/65536

priors

P(h1) = 1/2

P(h2) = 1/2

Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

likelihoods

P(D | h1) = 1/256

P(D | h2) = 1/65536

priors

P(h1) = 1/2

P(h2) = 1/2

posteriors

P(h1 | D) ∝ 1/256 * 1/2

P(h2 | D) ∝ 1/65536 * 1/2

Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

likelihoods

P(D | h1) = 1/256

P(D | h2) = 1/65536

priors

P(h1) = 1/2

P(h2) = 1/2

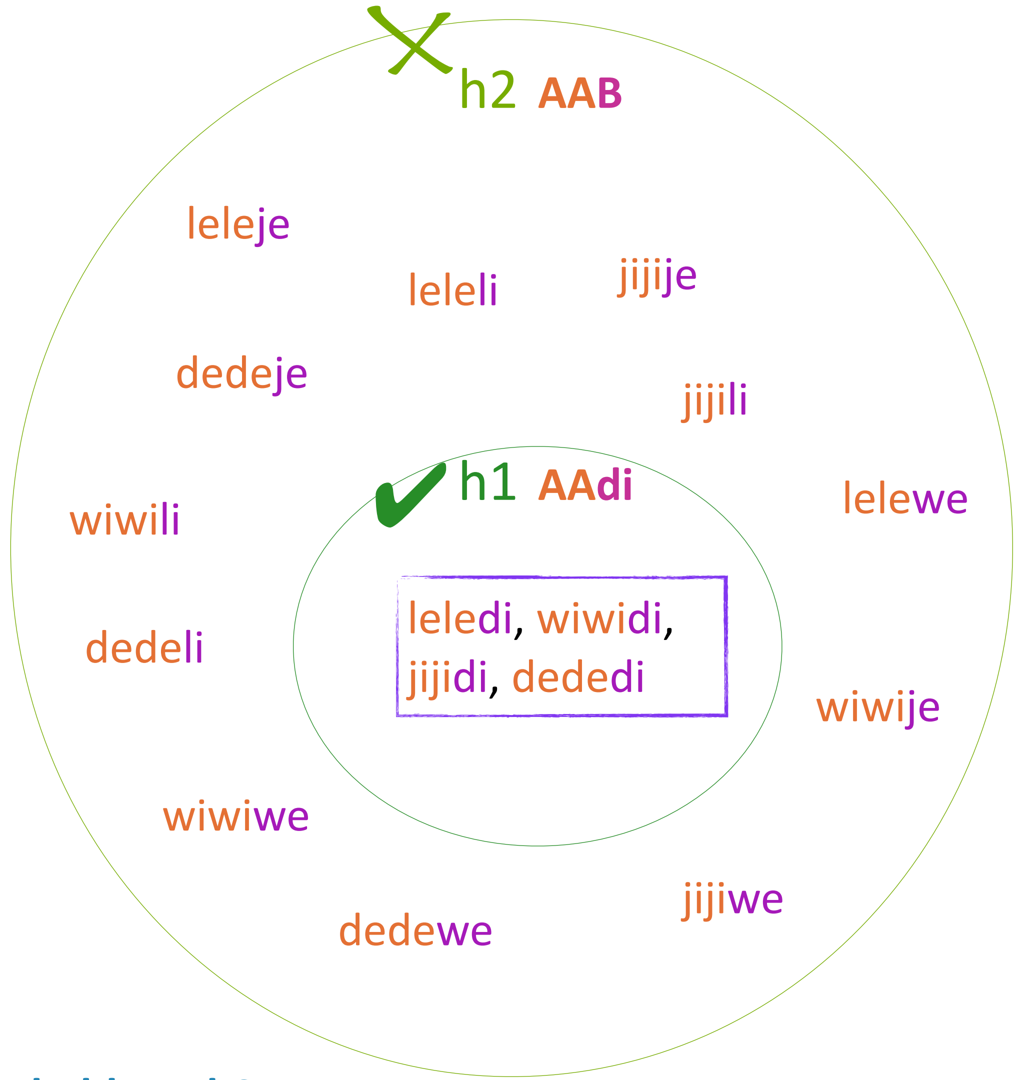
posteriors

P(h1 | D) ∝ **1/256** * 1/2

P(h2 | D) ∝ **1/65536** * 1/2

h1 is 256 times (1/256 vs. 1/65536) as probable as h2

Therefore, prefer h1.



Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

likelihoods

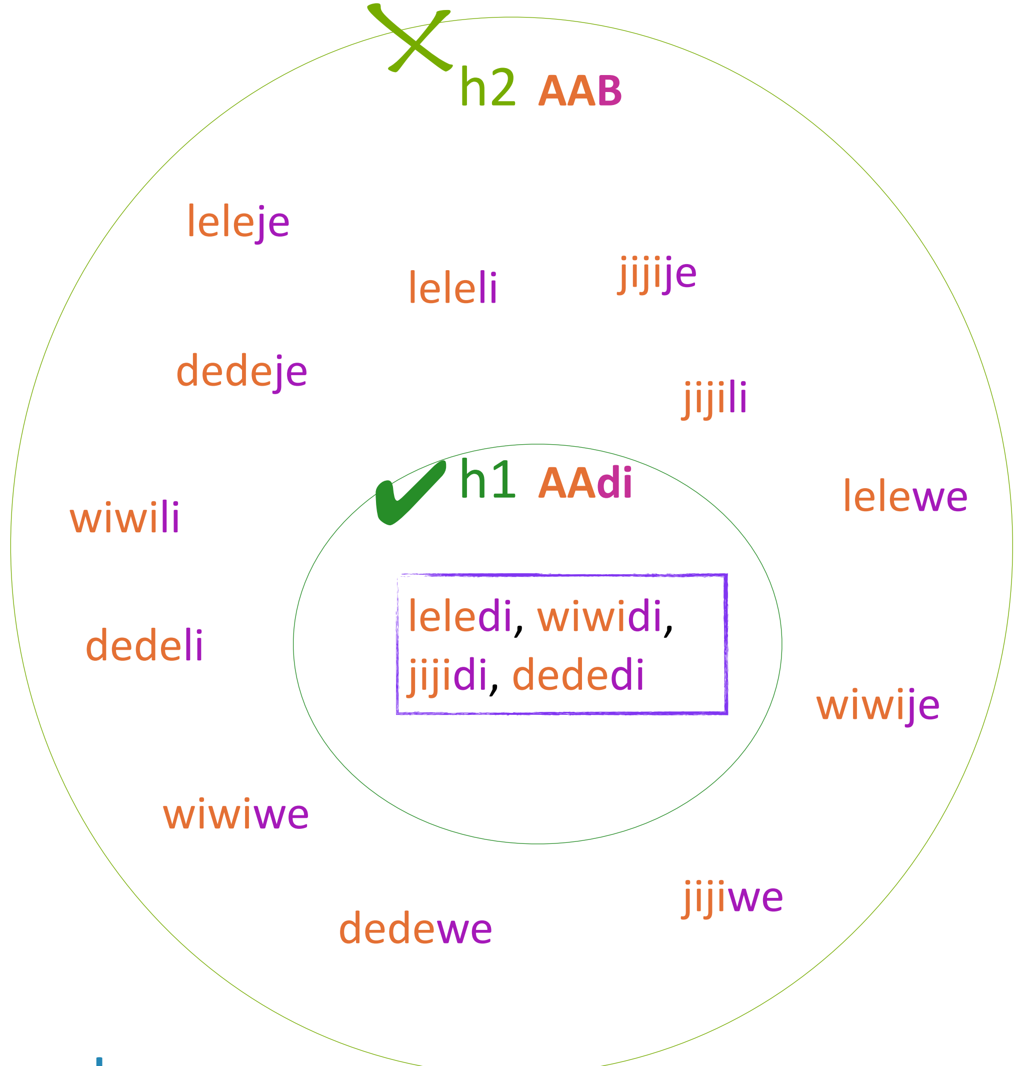
P(D | h1) = 1/256

P(D | h2) = 1/65536

priors

P(h1) = 1/2

P(h2) = 1/2



posteriors

P(h1 | D) ∝ **1/256** * 1/2

P(h2 | D) ∝ **1/65536** * 1/2

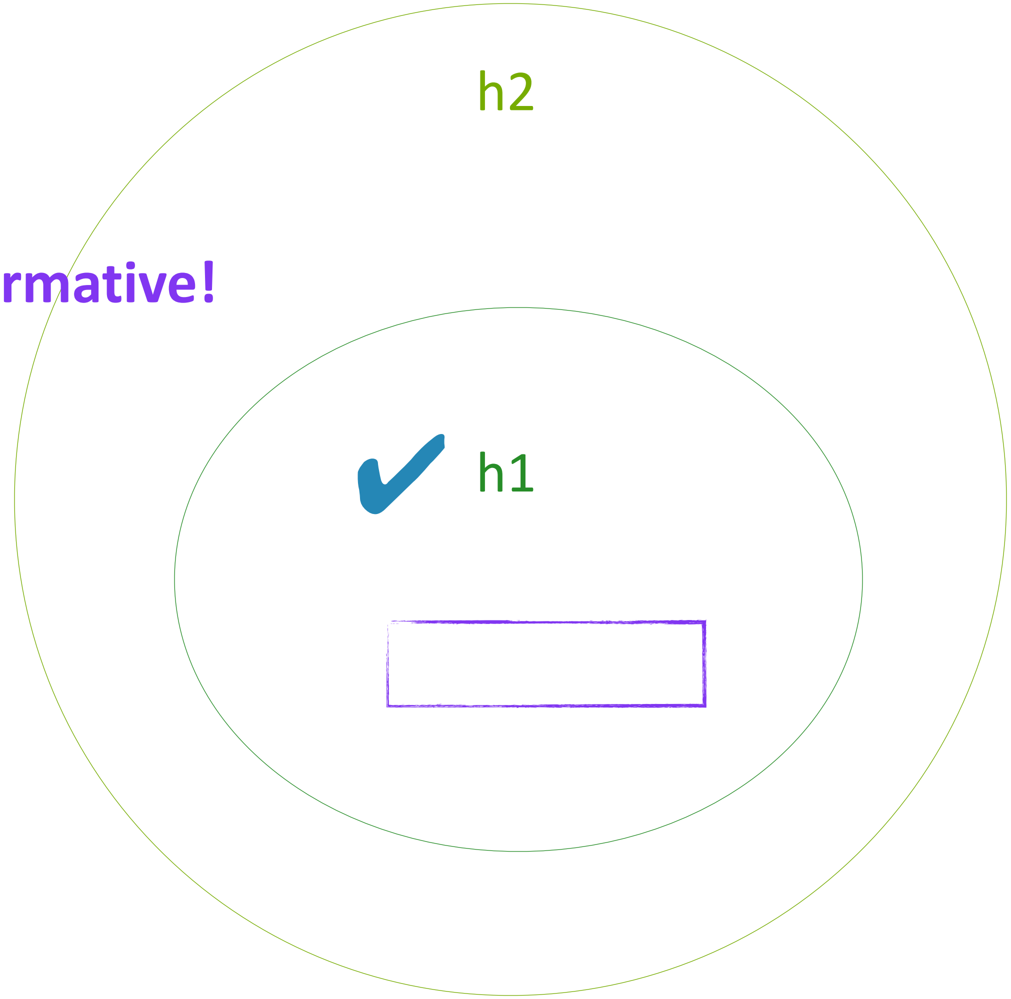
Note how it's the likelihood doing all the work.

Therefore, prefer h1.

Why would a preference for the less-general generalization be a sensible preference to have?

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

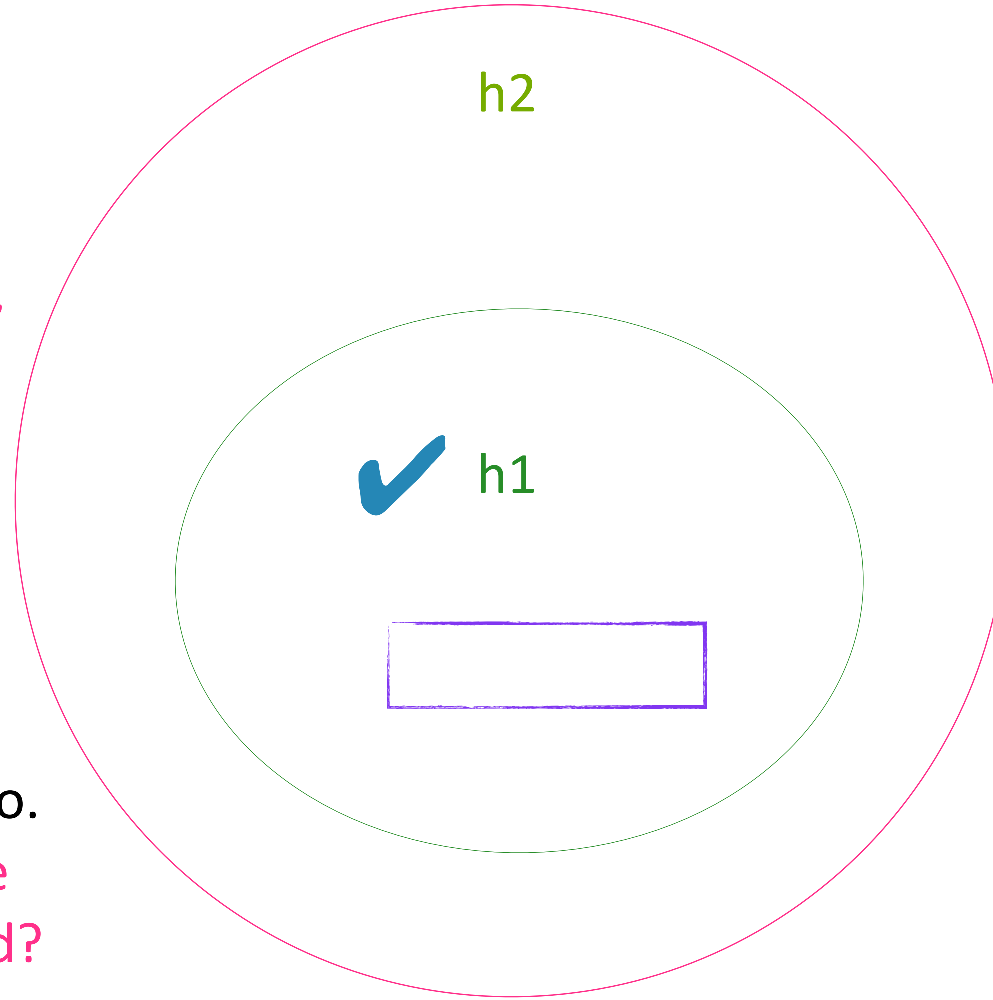
It makes ambiguous data informative!



This solves something called the Subset Problem.

The Subset Problem

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



What if children preferred h2, but the right generalization was h1?

Ambiguous data: All data compatible with the subset are compatible with the superset too. How would the child ever realize her generalization was too broad? There are no unambiguous data to indicate this.

The Subset Problem

What if children preferred h2, but the right generalization was h1?

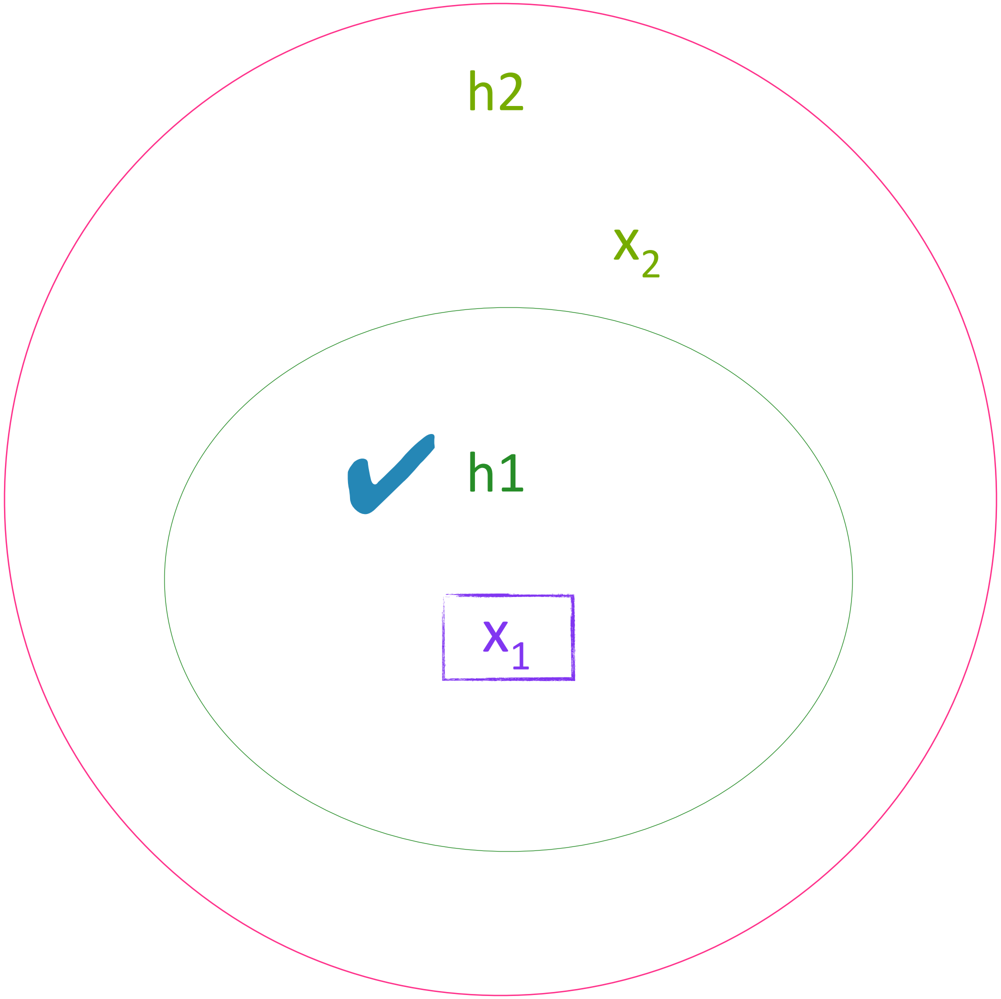
$$P(h|D) \propto P(D|h) * P(h)$$

Ambiguous data: There are no unambiguous data to indicate h1.

x_1 is the only data point that will appear if h1 is true.

(x_2 won't show up)

But x_1 is also compatible with h2.



The Subset Problem

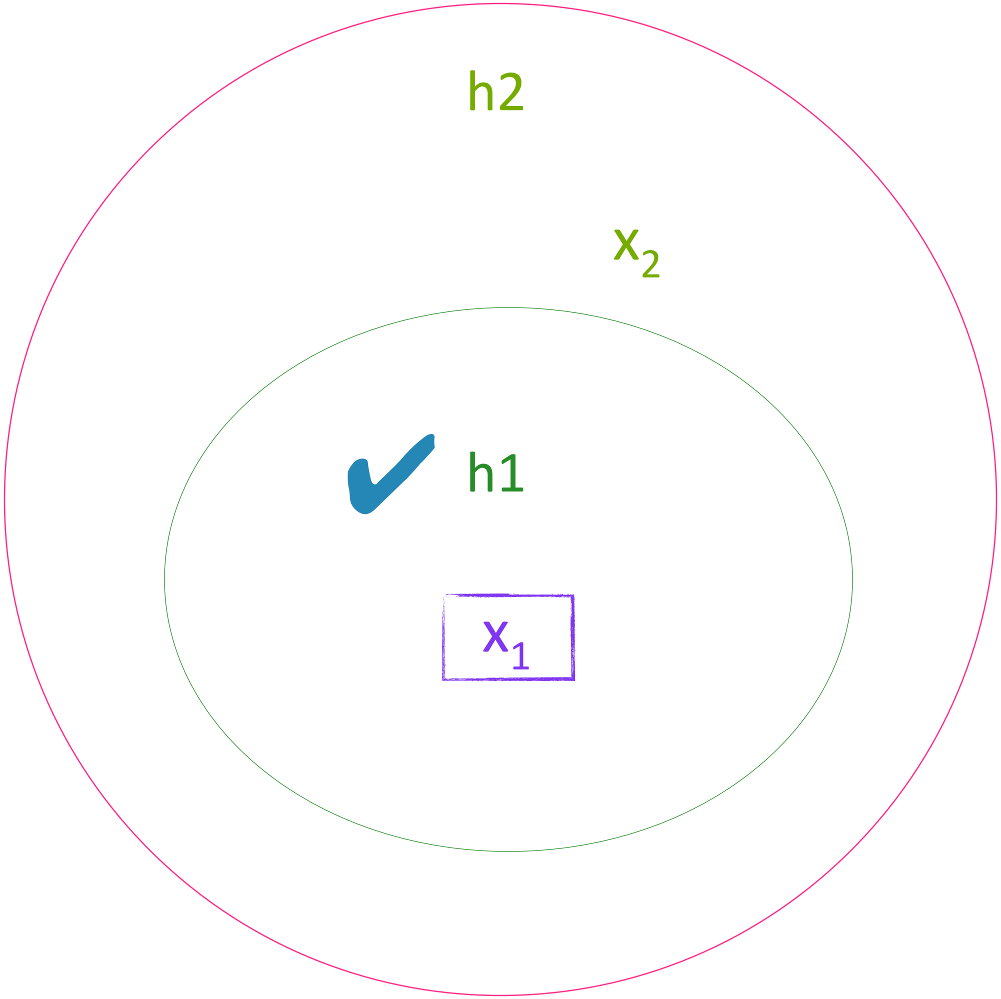
What if children preferred h2, but the right generalization was h1?

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

Ambiguous data: There are no unambiguous data to indicate h1.

Note that no other situation is a problem. If h2 is true, both x_1 and x_2 should appear.

If a child thinks h1 is right, seeing x_2 is an unambiguous signal to revise her hypothesis.



The Subset Problem

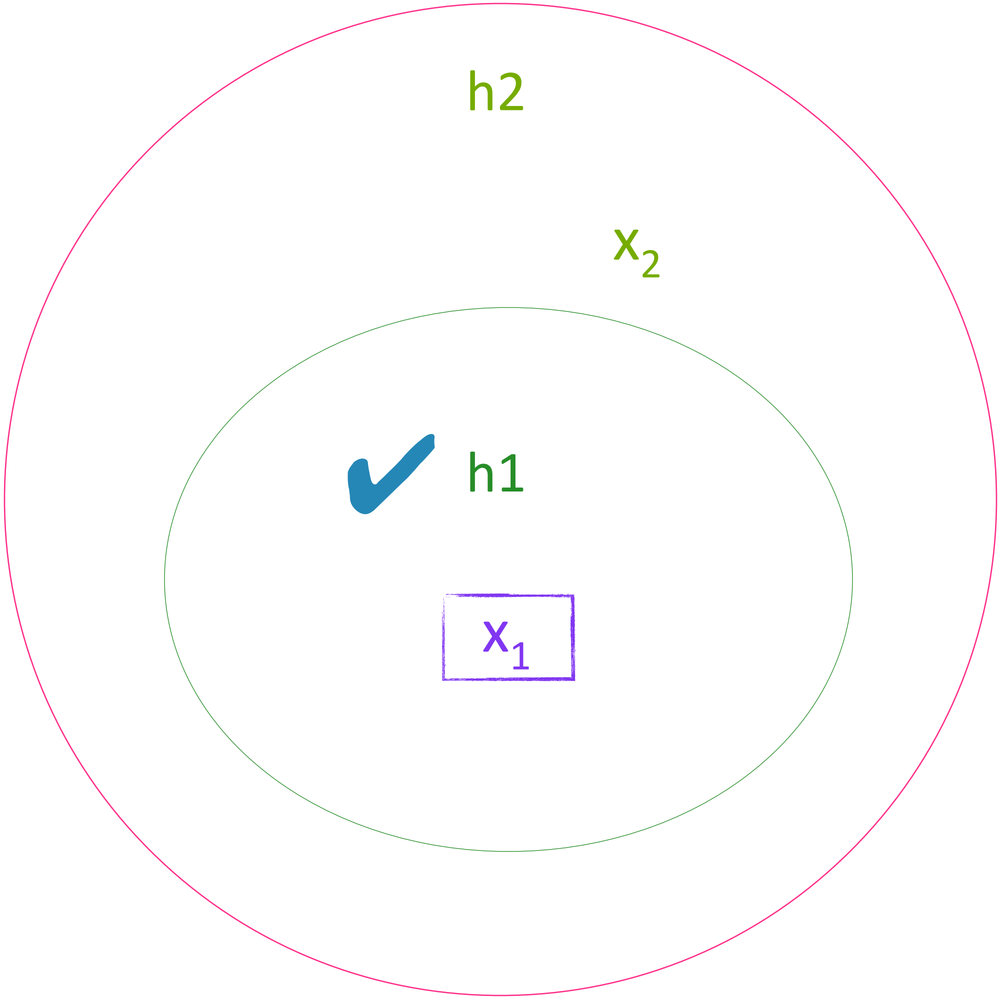
What if children preferred h2, but the right generalization was h1?

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

Ambiguous data: There are no unambiguous data to indicate h1.

So what to do when h1 is right, but the child thinks h2 is?

Have a **bias to prefer the subset hypothesis h1**. This can be implemented by the general-purpose probabilistic reasoning mechanism of **Bayesian inference**.



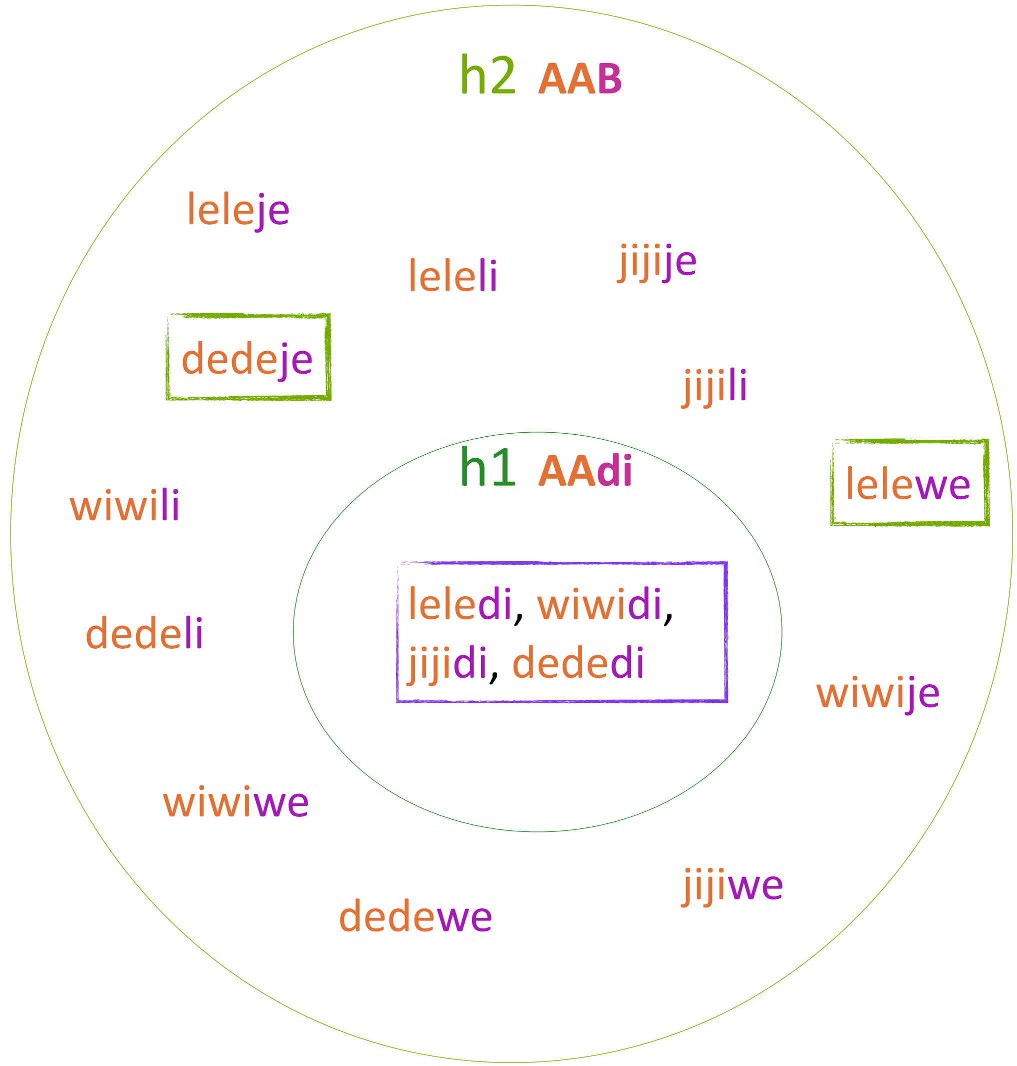
Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

Another way to think about this:
A child who thinks h2 is true will **expect to see data** that correspond to that hypothesis but not the subset hypothesis h1.



Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

If this child keeps *not* seeing those data that are counterexamples to h1 — that is, those that are compatible with h2, like she would expect if h2 were true — this is evidence that h1 is the right hypothesis.

This is an example of indirect negative evidence.

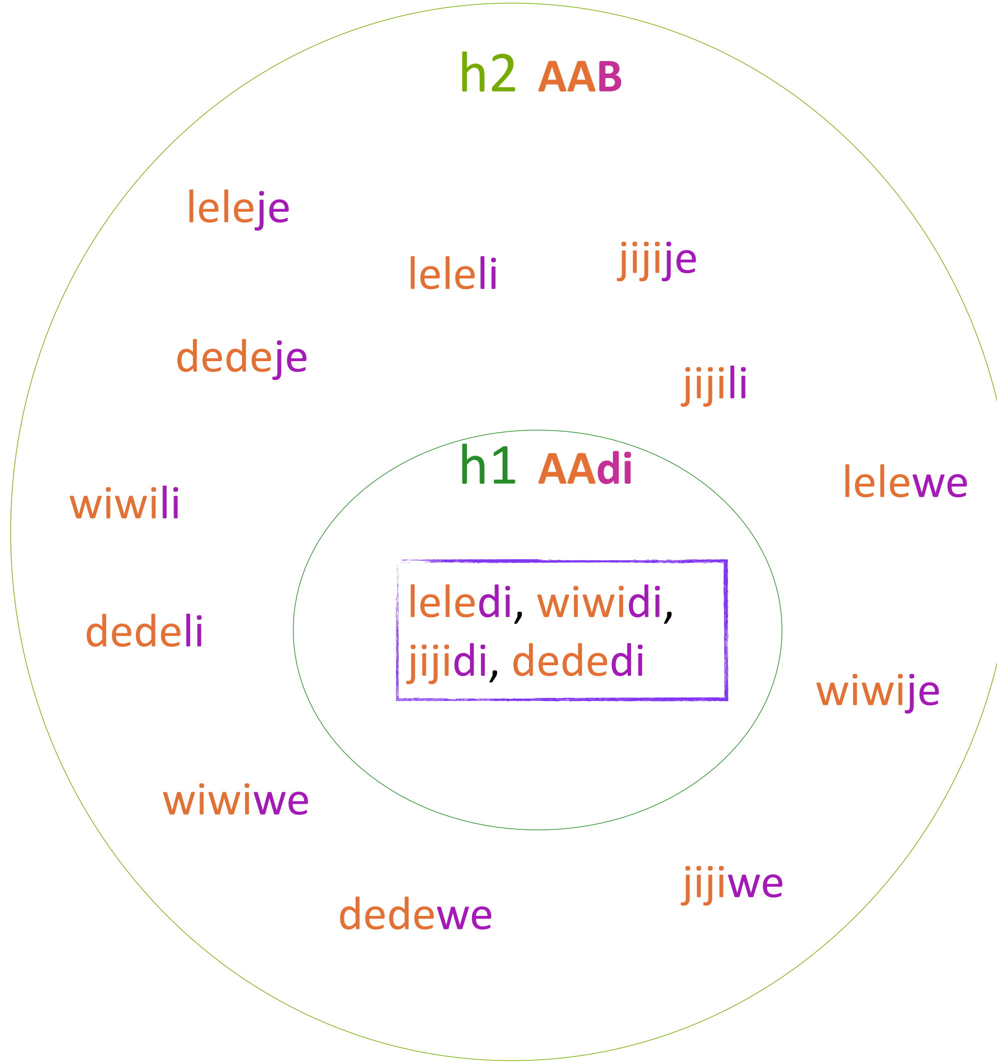
Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

Another important point:
Bayesian learners are sensitive to **counterexamples**.



Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

sensitive to **counterexamples**

If even one word in the intake **wasn't compatible** with the less-general AAdi pattern, a Bayesian learner would notice that and shift beliefs.

Bayesian reasoning

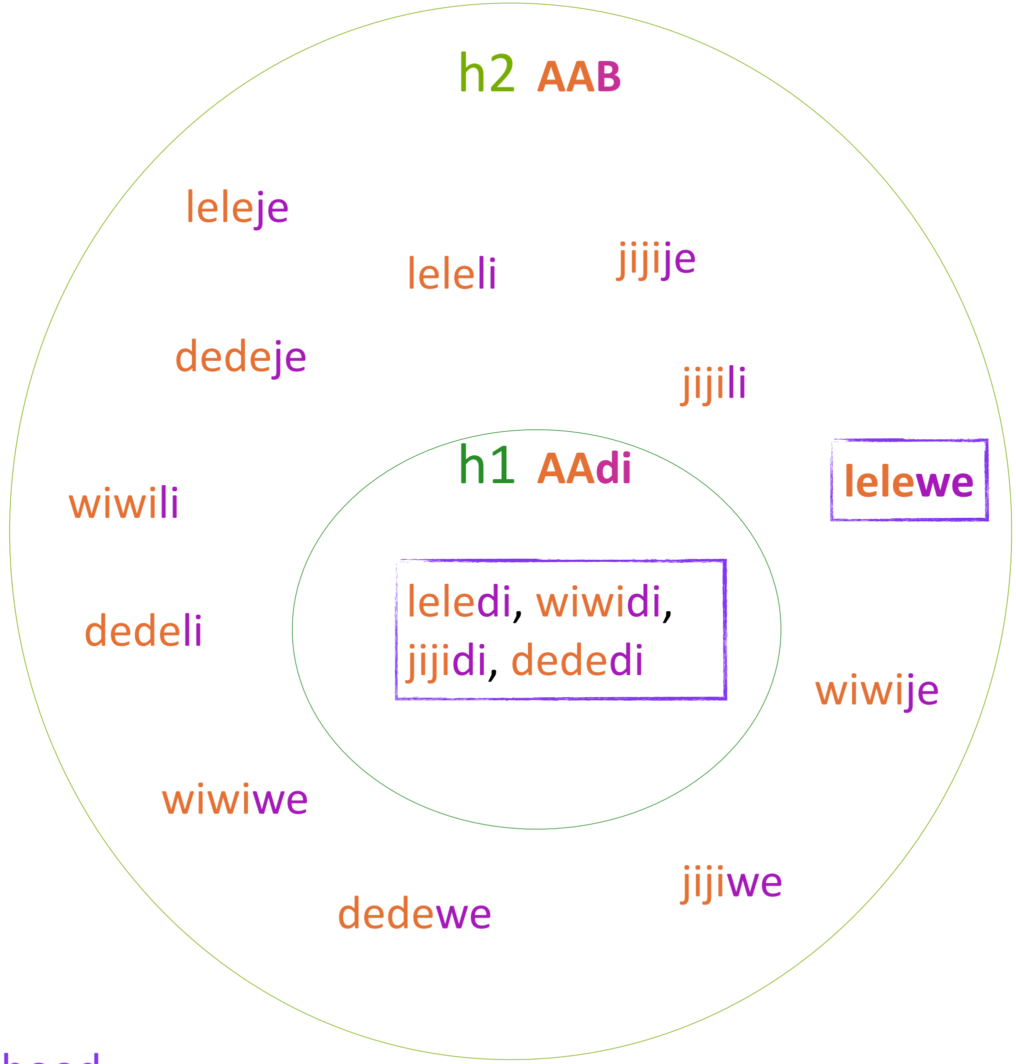
Gerken 2006, 2010
artificial language study

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

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

sensitive to **counterexamples**

If even one word in the intake **wasn't compatible** with the less-general AAdi pattern, a Bayesian learner would notice that and shift beliefs.



Why? This has to do with the likelihood.

Bayesian reasoning

Gerken 2006, 2010
artificial language study

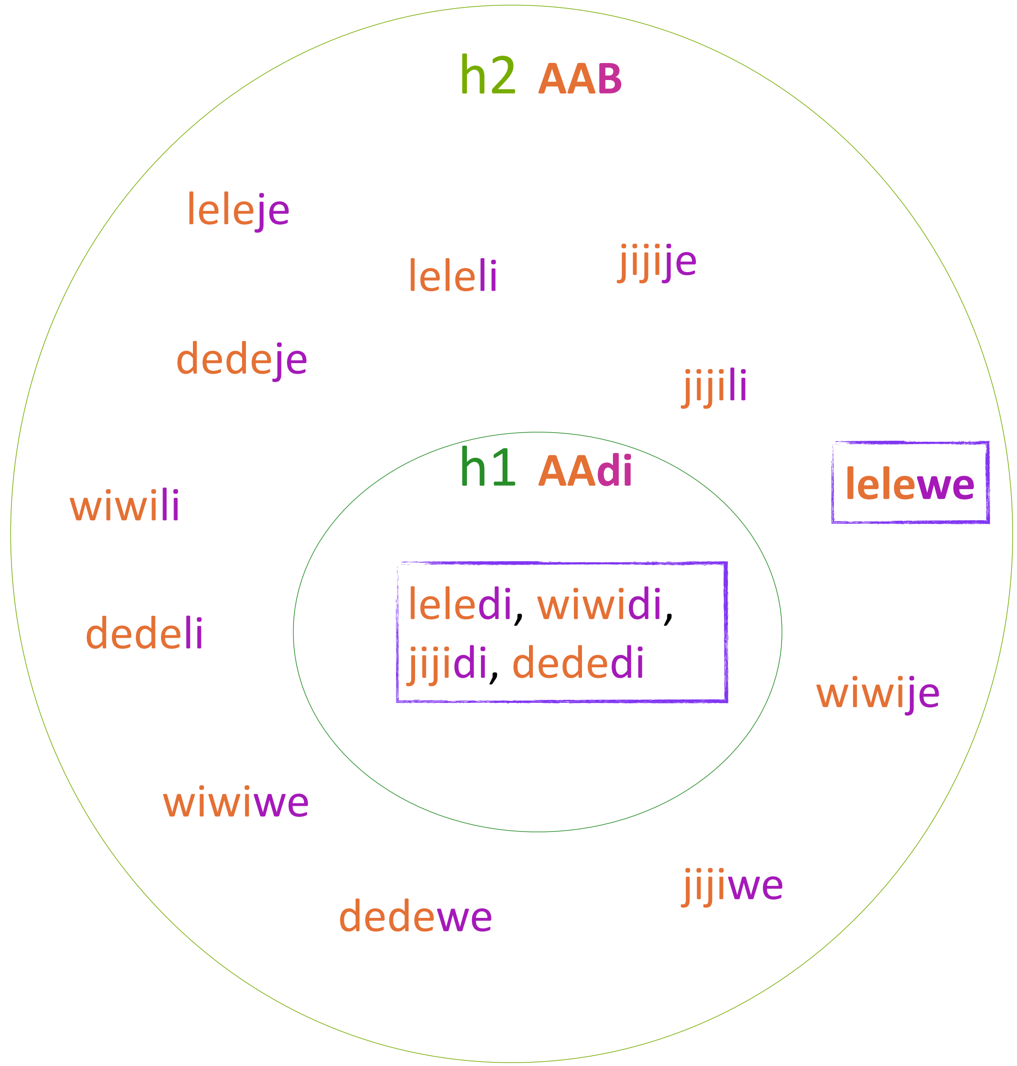
$$P(h|D) = \frac{P(D|h)*P(h)}{P(D)}$$

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

sensitive to **counterexamples**

likelihood
P(D | h1) = 1/4*1/4*1/4*1/4 * **0** = 0

These are the only 4 data that can be generated, and so the probability of generating each one is 1/4 **except the last one, which can't be generated.**



Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

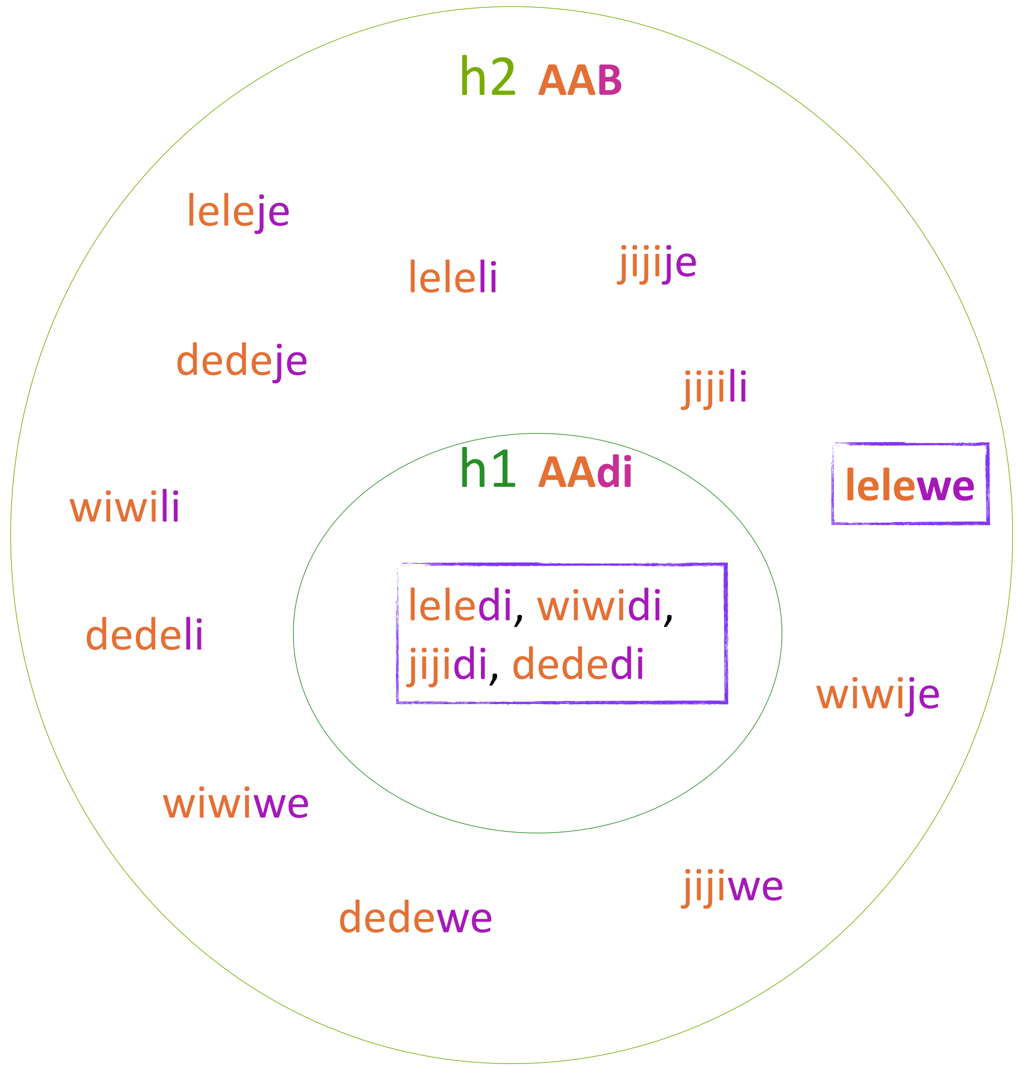
sensitive to **counterexamples**

likelihood

P(D | h1) = 0

P(D | h2) = 1/16*1/16*1/16*1/16*1/16
= 1/1048576

In contrast, even though the other data points have a smaller probability of being generated by h2, the last one *can* be generated, so **the likelihood isn't 0**.



Bayesian reasoning

Gerken 2006, 2010
artificial language study

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

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

sensitive to **counterexamples**

likelihood

P(D | h1) = 0

P(D | h2) = 1/1048576

This means only h2 will have a non-zero posterior, and so the Bayesian learner prefers h2.

Bayesian reasoning

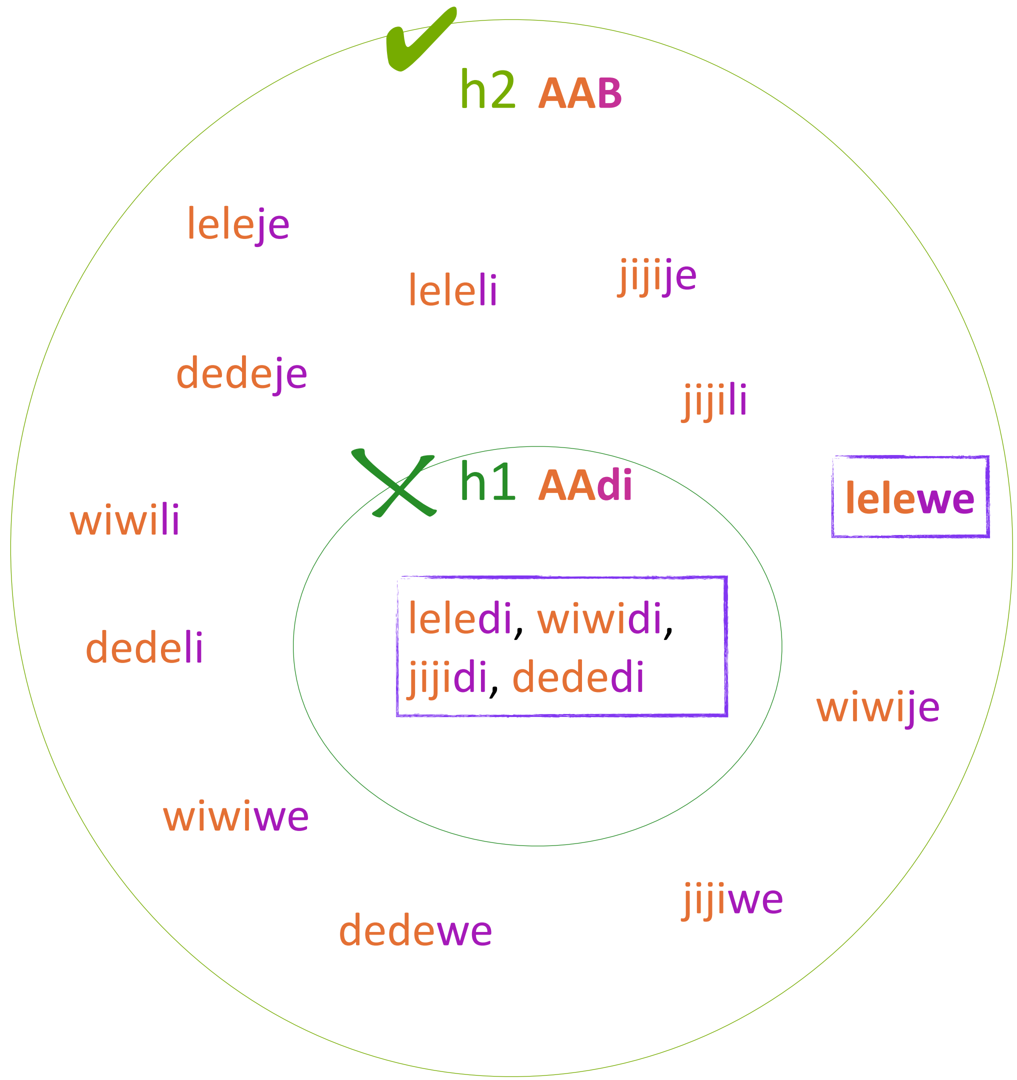
Gerken 2006, 2010
artificial language study

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

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

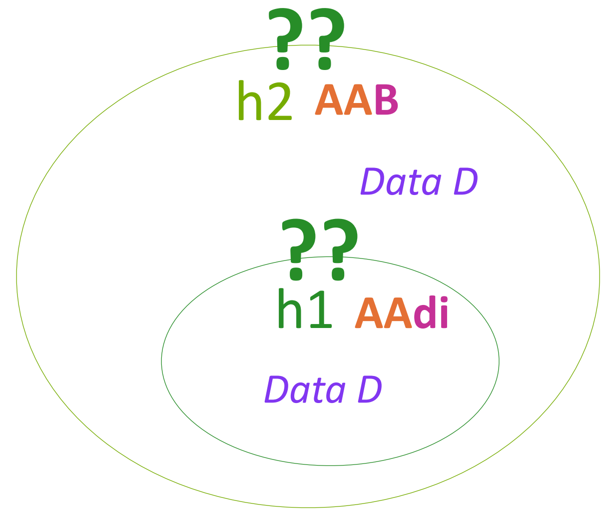
sensitive to **counterexamples**

Do 9-month-olds reason this way too?



Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

sensitive to **counterexamples**

Task type: Head Turn Preference Procedure

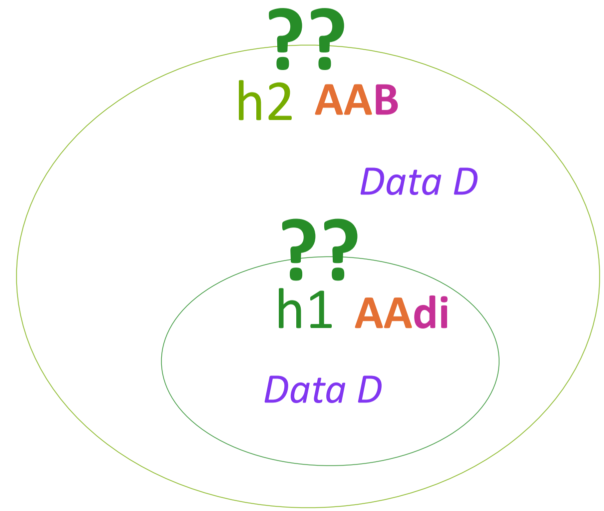
with 9-month-olds

Training leledi, wiwidi, jijidi, dededi + 3 AAB examples (like lelewe)

2 minutes **a few seconds at the end**

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

sensitive to **counterexamples**

Task type: Head Turn Preference Procedure

with 9-month-olds

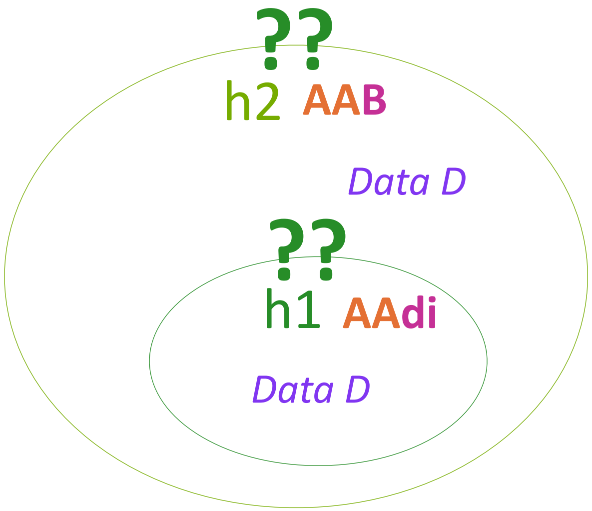
Training leledi, wiwidi, jijidi, dededi + 3 AAB examples (like lelewe)

2 minutes **a few seconds at the end**

Test kokoba vs. kobako

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

sensitive to **counterexamples**

Task type: Head Turn Preference Procedure

with 9-month-olds

Training leledi, wiwidi, jijidi, dededi + 3 AAB examples (like lelewe)

2 minutes **a few seconds at the end**

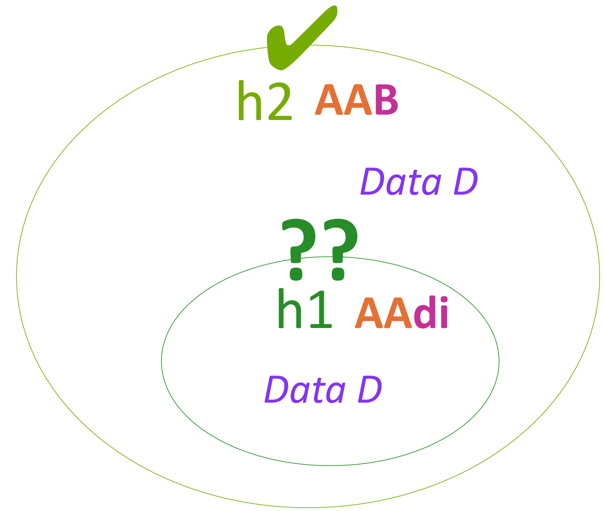
Test kokoba vs. kobako

Behavior ✔

Behavior: If they learn the more-general pattern from these three counterexamples, they'll prefer to listen to AAB words like kokoba.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

sensitive to **counterexamples**

Task type: Head Turn Preference Procedure

with 9-month-olds

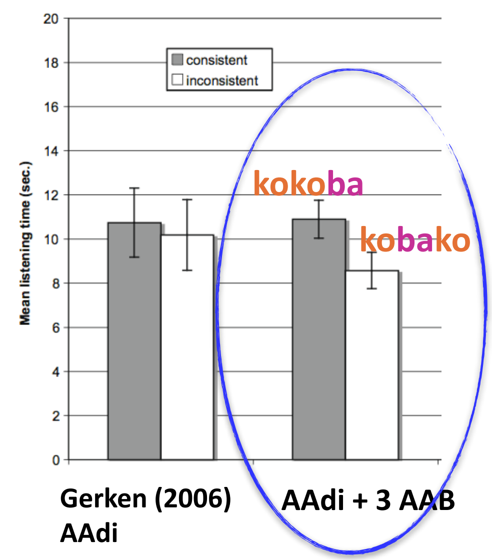
Training leledi, wiwidi, jijidi, dededi + 3 AAB examples (like lelewe)

2 minutes **a few seconds at the end**

Test kokoba vs. kobako

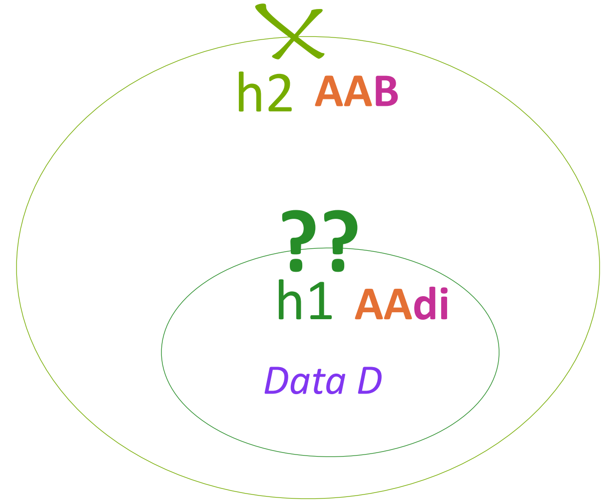
Behavior ✔

Children prefer to listen to novel words that follow the more-general AAB pattern [~11 sec] over novel words that don't [~8 sec]



Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

sensitive to **counterexamples**

Task type: Head Turn Preference Procedure

with 9-month-olds

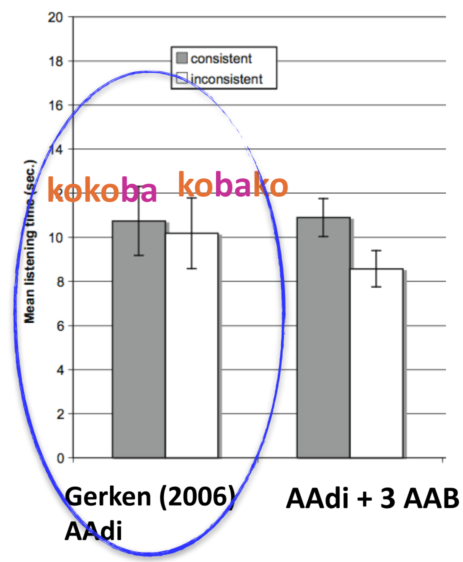
Training leledi, wiwidi, jijidi, dededi + 3 AAB examples (like lelewe)

2 minutes **a few seconds at the end**

Test kokoba vs. kobako

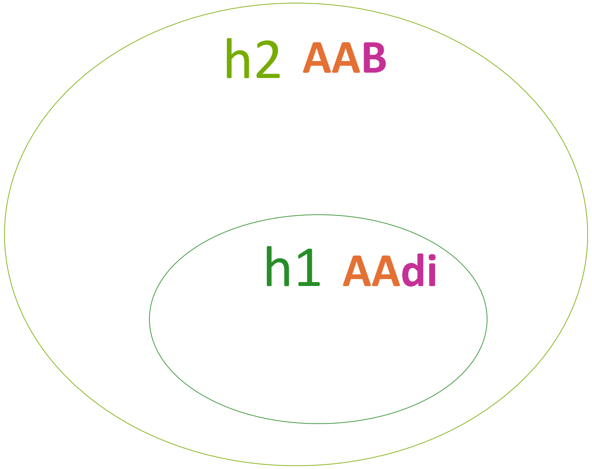
Behavior ✔

This is noticeably different than their behavior when they only hear AAdi examples in their intake.



Bayesian reasoning

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

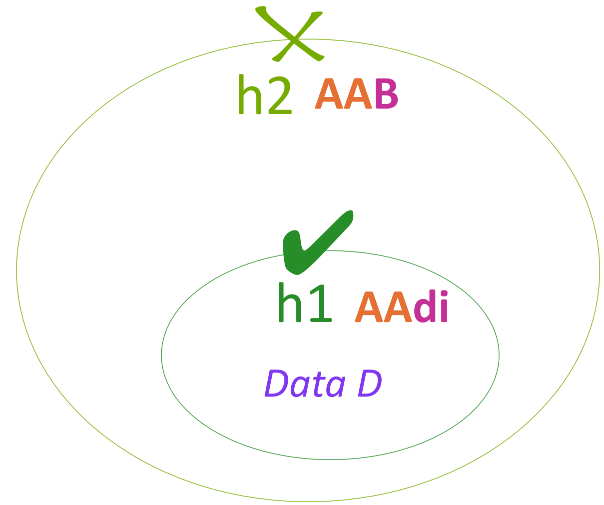


Gerken 2006, 2010
artificial language study

Takeaway: At 9 months, infants show probabilistic reasoning abilities similar to a Bayesian learner.

Bayesian reasoning

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



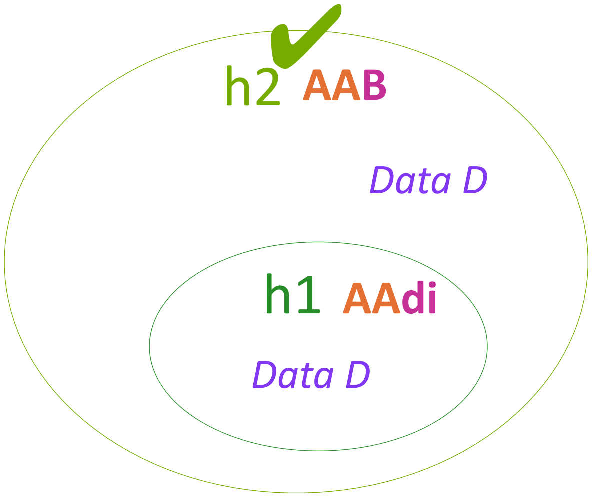
Gerken 2006, 2010
artificial language study

Takeaway: At 9 months, infants show probabilistic reasoning abilities similar to a Bayesian learner.

When given ambiguous data compatible with two hypotheses, a less-general and more-general one, they choose the less-general one (which gives a higher likelihood to the data).

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

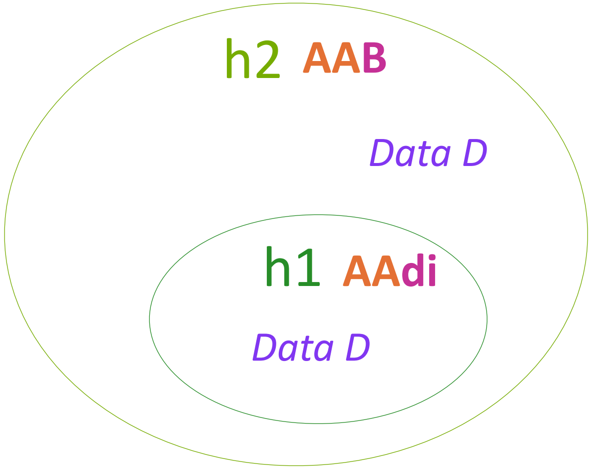
Takeaway: At 9 months, infants show probabilistic reasoning abilities similar to a Bayesian learner.

✔ambiguous data = less-general hypothesis

When given even a very few counterexamples that are only compatible with the more-general hypothesis, they shift their beliefs accordingly.

Bayesian reasoning

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



Gerken 2006, 2010
artificial language study

Takeaway: At 9 months, infants show probabilistic reasoning abilities similar to a Bayesian learner.

✔ambiguous data = less-general hypothesis

✔counterexamples = shift beliefs accordingly to more-general hypothesis

Recap

Children will often be faced with multiple generalizations that are compatible with the language data they encounter. In order to learn their native language, they must choose the correct generalizations.

Experimental research on artificial languages suggests that children prefer the more conservative generalization compatible with the data they encounter, but will update their beliefs based on the data available.

This learning strategy is one that a Bayesian learner may be able to take advantage of quite naturally. So, if children are probabilistic learners of this kind (and experiments by Gerken suggest they may be), they may automatically follow this conservative generalization strategy.

Questions?

You should be able to do all the review questions for poverty of the stimulus, and all of HW7.