Psych 215L:
Language Acquisition

Poverty of the Stimulus: Introduction

About language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

Hoggle is definitely an ornery dwarf. [grammatical]
* Hoggle an dwarf definitely ornery is. [ungrammatical]

About language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

Hoggle is definitely an ornery dwarf. [part of English]
* Hoggle an dwarf definitely ornery is. [not part of English]

About language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

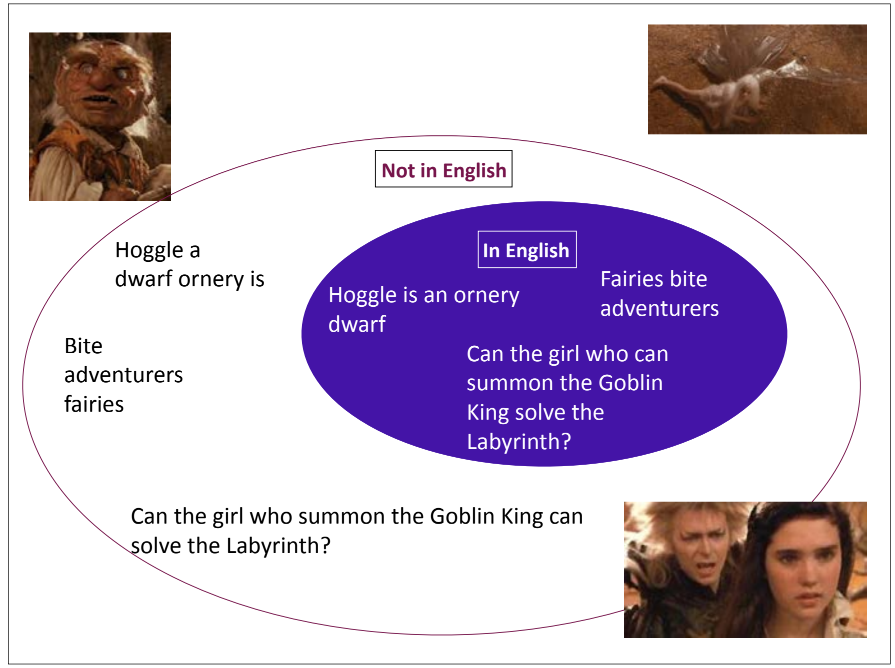
The reason you can do this is because you know the rules & patterns that generate the items that are part of the language.
(mental grammar)

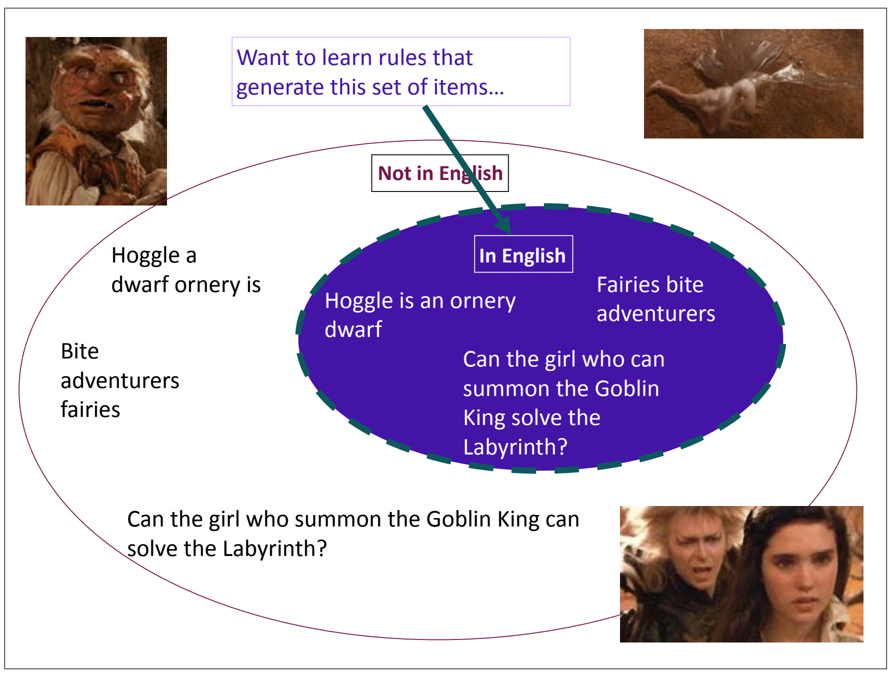
About children learning language

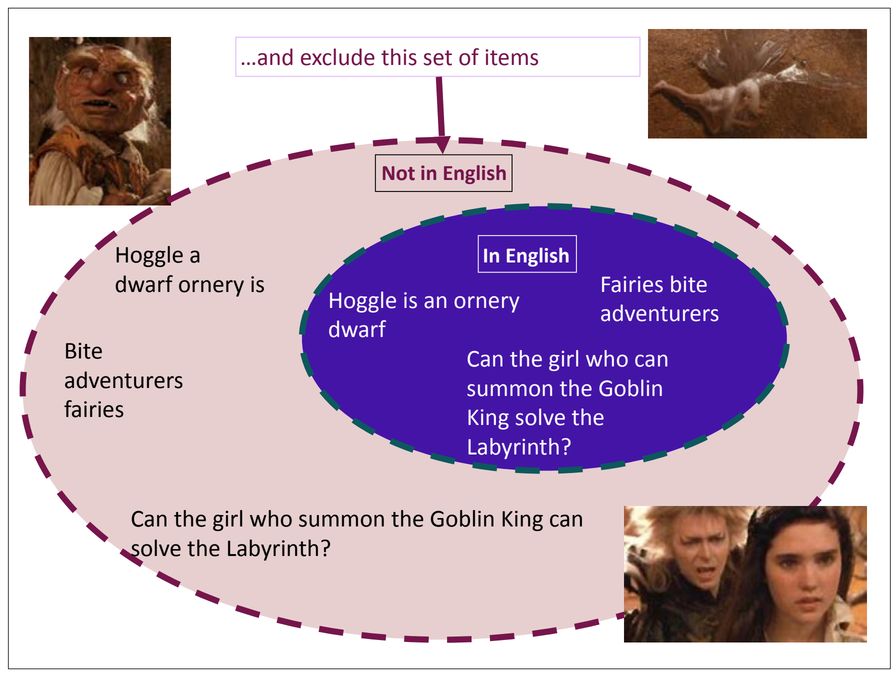
Adult knowledge: rules & patterns that generate the items that are part of the language. (mental grammar)

The child's job: figure out the rules that generate the items that belong in the language and that don't generate items that don't belong in the language.

For example, the child wants rules to generate
"Hoggle is definitely an ornery dwarf" but not
* "Hoggle an dwarf definitely ornery is".



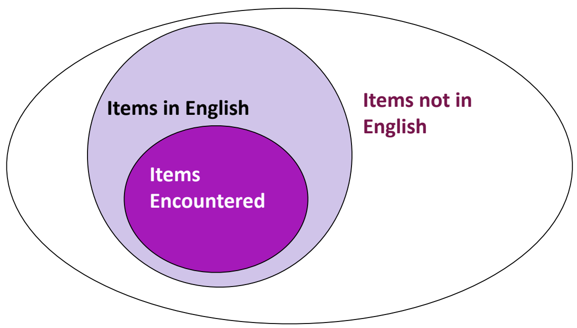




So what's the problem?

It's not clear that children encounter all the items that are part of the language (they have finite time to learn).

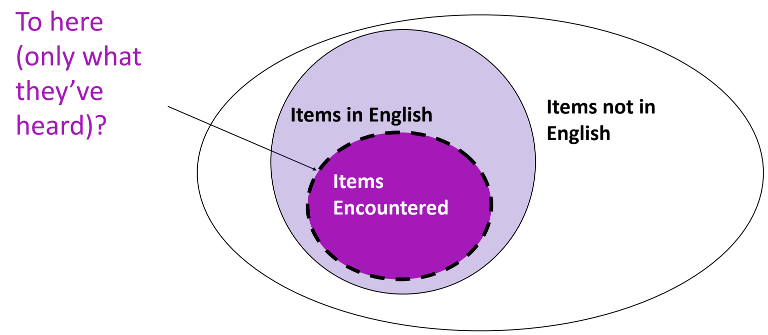
If they only encounter a subset of the language's items, how do they know everything that belongs in the language?



So what's the problem?

One solution: children generalize

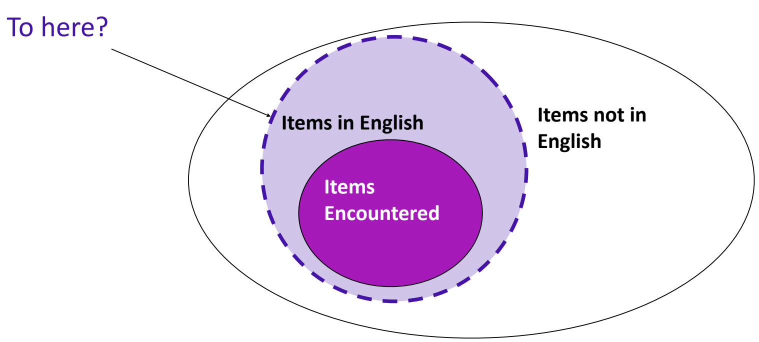
But how do they generalize?



So what's the problem?

One solution: children generalize

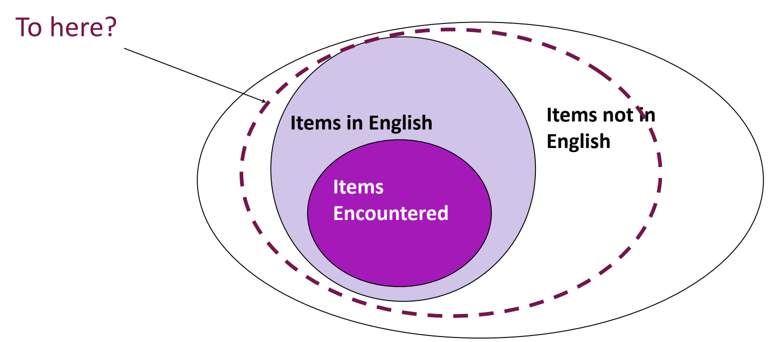
But how do they generalize?



So what's the problem?

One solution: children generalize

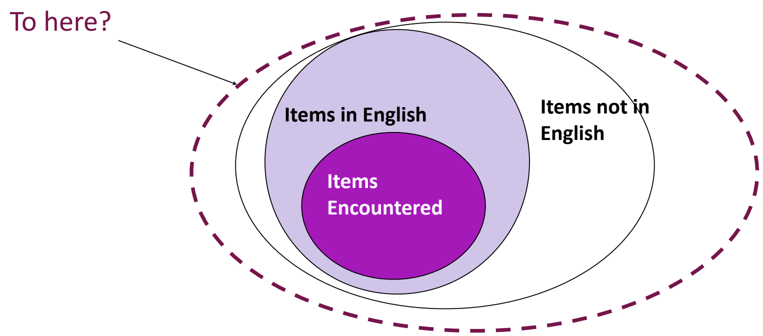
But how do they generalize?



So what's the problem?

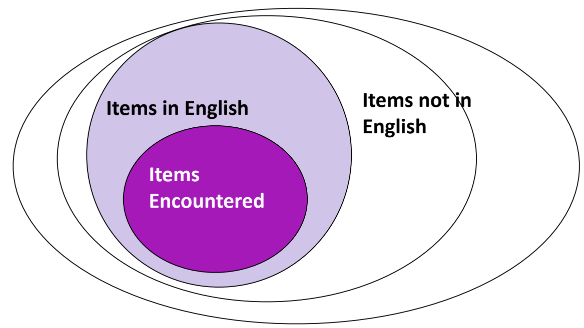
One solution: children generalize

But how do they generalize?



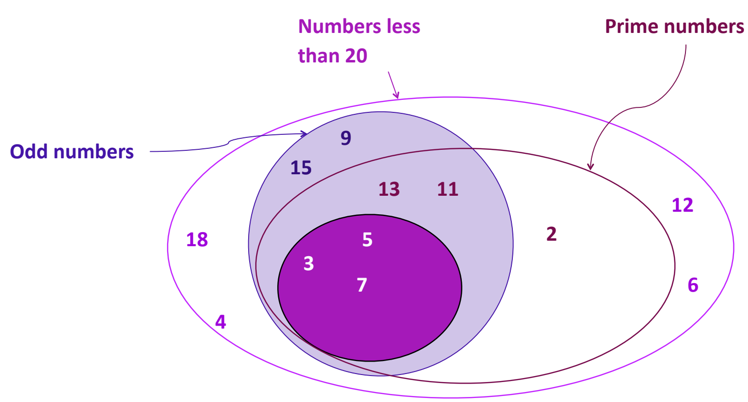
So what's the problem?

The problem is that children must make the right generalization from data that are compatible with multiple generalizations. In this sense, the data (stimulus) encountered are impoverished. They do not single out the correct generalization by themselves.



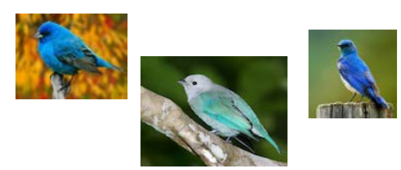
A numerical analogy

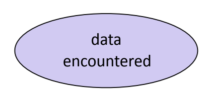
Suppose you encounter the numbers 3, 5, and 7.
What set are these numbers drawn from? That is, what is the right "number rule" for this language that will allow you to predict what numbers will appear in the future?



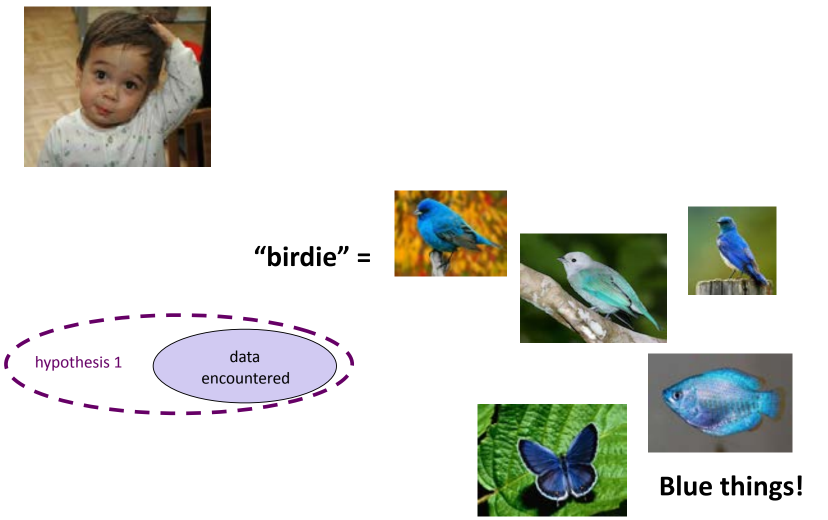
Impoverished data in word learning

"birdie" =

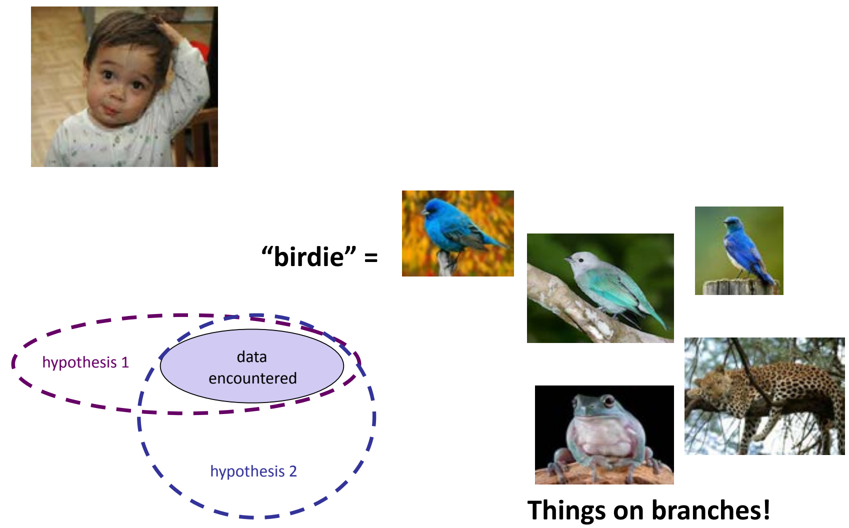




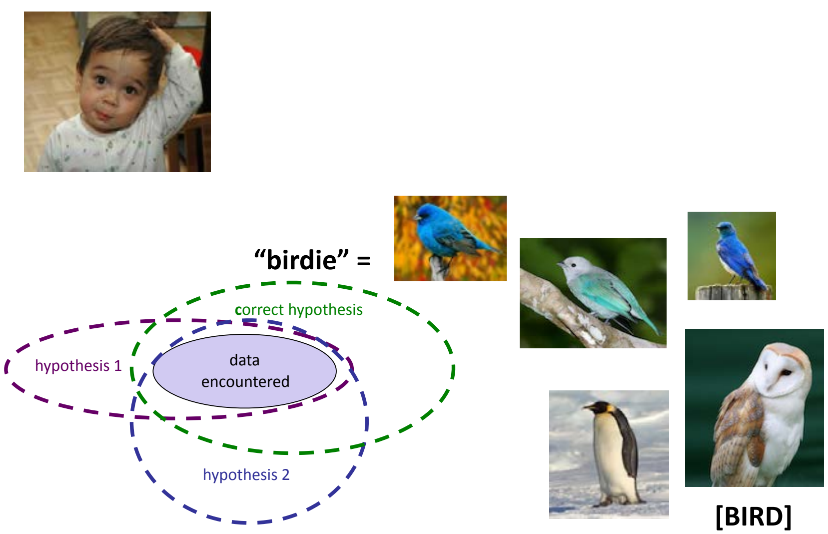
Impoverished data in word learning



Impoverished data in word learning

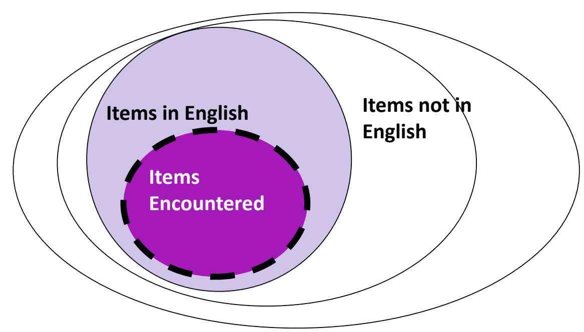


Impoverished data in word learning



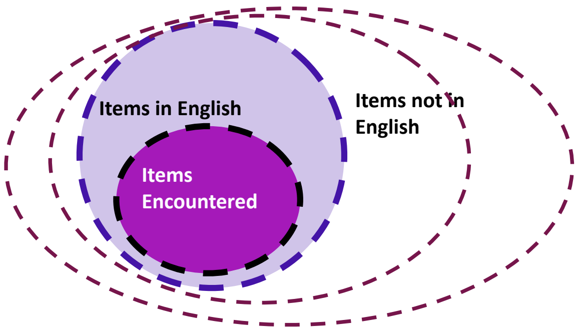
Poverty of the Stimulus: Logic

Children encounter data that are compatible with many hypotheses about the correct rules and patterns of the language.



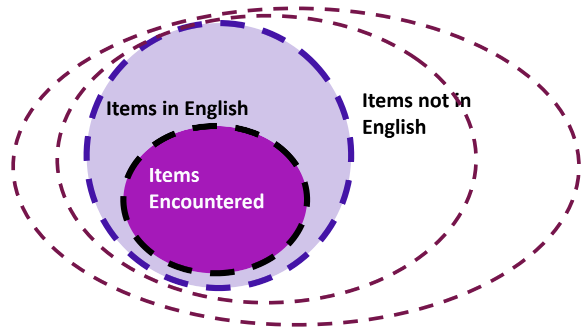
Poverty of the Stimulus: Logic

Specifically, the data encountered are compatible with both the correct hypothesis and other, incorrect hypotheses about the rules and patterns of the language.



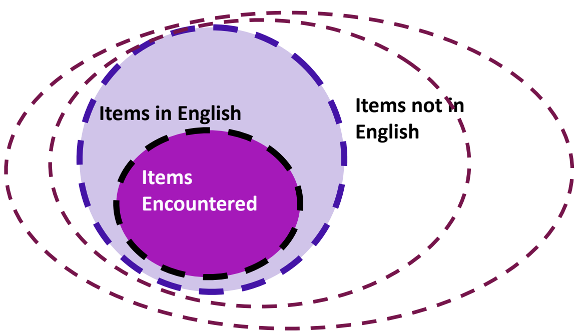
Poverty of the Stimulus: Logic

A rational learner would consider all compatible hypotheses, and perhaps choose the wrong hypothesis in the end, or at least make errors during acquisition.



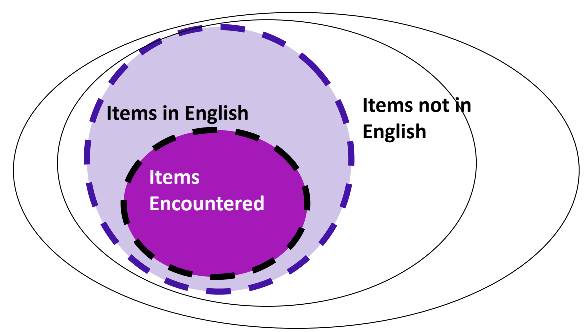
Poverty of the Stimulus: Logic

Expectation for rational learners: errors in performance. Children will behave as if they think ungrammatical items are part of the language at some point in their development.



Argument about prior knowledge

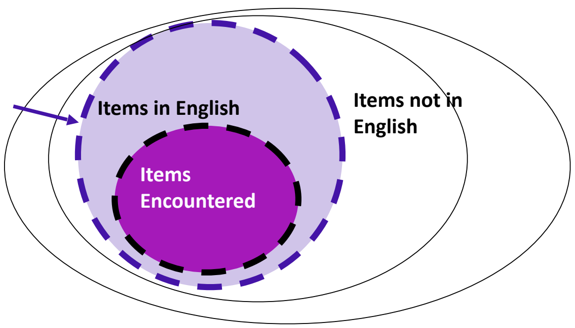
But what if children never behave as if they consider the incorrect hypotheses? That is, they never produce errors compatible with the incorrect hypotheses. They only seem to produce items that are compatible with the correct hypothesis.



Argument about prior knowledge

Conclusion: children have some prior knowledge that causes them never to consider the incorrect hypotheses. Instead, they only consider the correct hypothesis for what the rules and patterns of the language might be.

Prior knowledge restricts children's hypothesis to this



Argument about prior knowledge

The point: Children generalize only in a very specific way. In particular, they don't just generalize everything that they can. Their generalizations appear to be constrained.

Nativist idea for how their generalizations/hypotheses are constrained: innate knowledge.

Linguistic nativist idea for how their generalizations/hypotheses are constrained: innate knowledge about language.

Poverty of the Stimulus leads to prior knowledge about language: Summary of Logic

1) Suppose there are some data.
2) Suppose there is at least one incorrect hypothesis compatible with the data.
3) Suppose children behave as if they never entertain incorrect hypotheses.

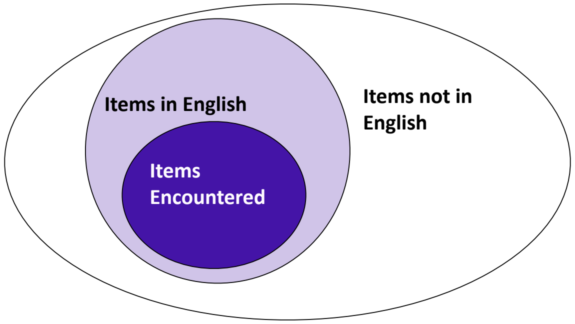
Conclusion: Children possess prior (innate) knowledge ruling out the incorrect hypotheses from consideration.

Hypothesis = Generalization

1) Suppose there are some data.
2) Suppose there are multiple generalizations compatible with the data.
3) Suppose children behave as if they only make one generalization.

Conclusion: Children possess prior (innate) knowledge ruling out the incorrect generalizations from consideration.

Making generalizations that are underdetermined by the data

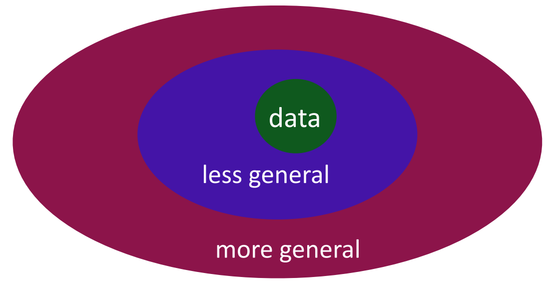


Children encounter a subset of the language's data, and have to decide how to generalize from that data

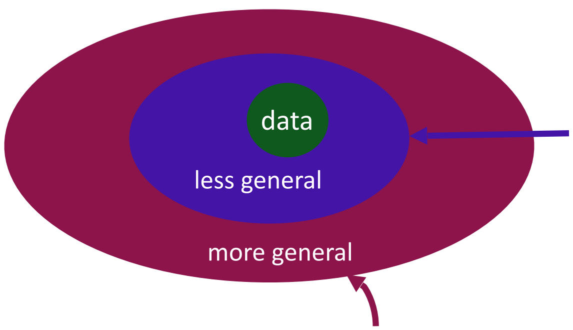
Making generalizations that are underdetermined by the data

Here's a question (Gerken 2006): is there any way to check what kinds of generalizations children prefer to make?

Example: Suppose they're given a data set that is compatible with two generalizations: a less-general one and a more-general one.



Choosing generalizations

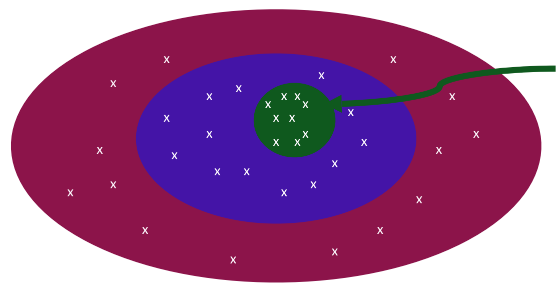


Do children think this generalization is the right one?

Or do children think this generalization is the right one?

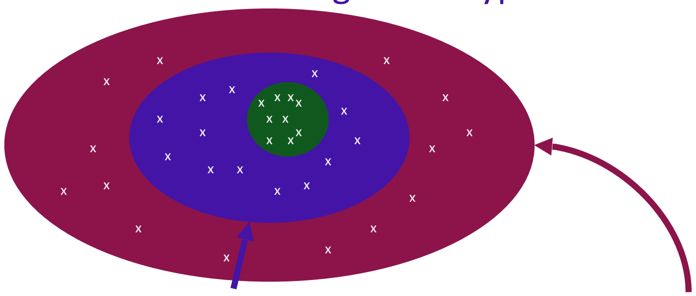
How can we tell?

Generalization = predictions about what data are in the language



Data children encounter

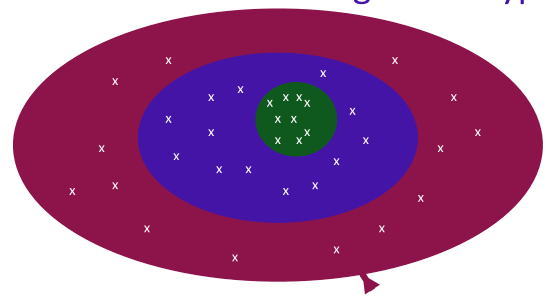
Choosing generalizations: the less general hypothesis



If children think the less-general hypothesis is correct, they will think data covered by that hypothesis are in the language - in addition to the data they encountered.

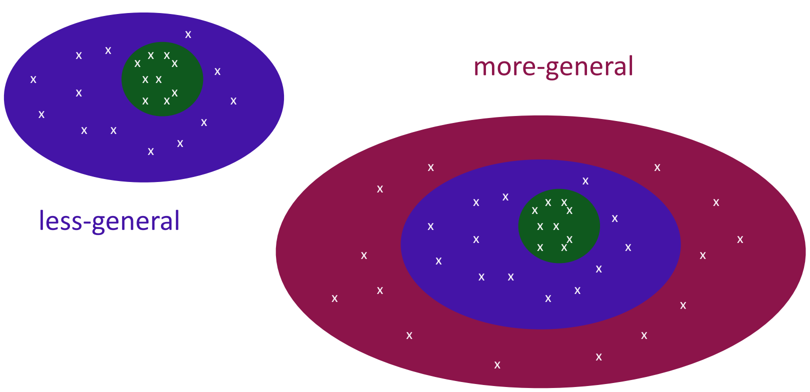
They will not think that data that are in the more-general hypothesis are in the language.

Choosing generalizations: the more general hypothesis



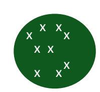
If children think the more-general hypothesis is correct, they will think data covered by that hypothesis are in the language - in addition to the data they encountered and the data in the less-general hypothesis.

Potential child responses when multiple generalizations are possible



Reality check

What do these correspond to in a real language learning scenario?



Data: Simple yes/no questions in English

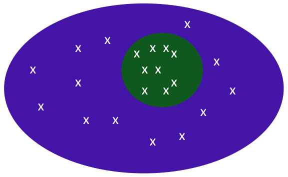
“Is the dwarf laughing?”

“Can the goblin king sing?”

“Will Sarah solve the Labyrinth?”

Reality check

What do these correspond to in a real language learning scenario?



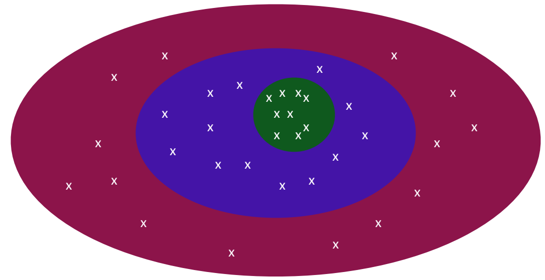
less-general hypothesis:
Some complex grammatical yes-no questions

"Is the dwarf laughing about the fairies he sprayed?"

"Can the goblin king sing whenever he wants?"

Reality check

What do these correspond to in a real language learning scenario?



more-general hypothesis:
Full range of complex grammatical yes-no questions

"Can the girl who ate the peach and forgot everything save her brother?"

"Will the dwarf who deserted Sarah help her reach the castle that's beyond the goblin city?"

Experimental study: Gerken 2006

How can we tell what generalizations children actually make? Gerken (2006) describes an artificial language study that can help us figure this out.

Children will be trained on data from an artificial language. This language will consist of words that follow a certain pattern.

The child's job: determine what the pattern is that allows a word to be part of the artificial language.

Artificial language: AAB/ABA pattern

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

AAB language words: leledi, leleje, wiwidi, wiwije

ABA language words: ledile, lejele, widiwi, wijewi

Artificial language: AAB/ABA pattern

Gerken (2006) decided to test what kind of generalization children would make if they were given particular kinds of data from this same artificial language.

Words in the AAB pattern artificial language.

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

What if children were only trained on a certain subset of the words in the language?

Words in the AAB pattern artificial language.

	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

These data are consistent with a less-general pattern (AAdi) as well as the more-general pattern of the language (AAB)

Question: If children are given this subset of the data that is compatible with both generalizations, which generalization will they make (AAdi or AAB)?

(Experimental Condition) Training on four word types: leledi, wiwidi, jijidi, dededi

These data are consistent with a less-general pattern (AAdi) as well as the more-general pattern of the language (AAB)

Words in the AAB pattern artificial language.

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

(Control Condition) Training on four word types: leledi, wiwije, jijili, dedewe

These data are only consistent with the more-general pattern of the language (AAB)

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

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.)

(Control Condition) Training on four word types: leledi, wiwije, jijili, dedewe

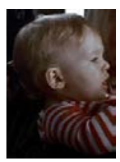
These data are only consistent with the more-general pattern of the language (AAB)

Experiment 1

Task type: Head Turn Preference Procedure

Experimental: leledi…wiwidi…jijidi…dededi

Control: leledi…wiwije…jijili…dedewe



Children: 9-month-olds

Stimuli: 2 minutes of artificial language words.

Test condition words: AAB pattern words using syllables the children had never encountered before in the language. Ex: kokoba (novel syllables: ko, ba)

Experiment 1 predictions

Control: leledi…wiwije…jijili…dedewe

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.

Experiment 1 results

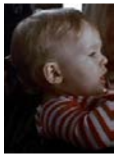
Control: leledi…wiwije…jijili…dedewe

Children listened longer on average to test items consistent with the AAB pattern (like kokoba) [13.51 sec], as opposed to items inconsistent with it (like kobako) [10.14].

Implication: They can notice the AAB pattern and make the generalization from this artificial language data. This task is not too hard for infants.

Experiment 1 predictions

Experimental: leledi…wiwidi…jijidi…dededi



If children learn the less-general pattern (AAdi), they will not prefer to listen to an AAB pattern word that does not end in di, like kokoba, over a word that does not follow the AAB pattern, like kobako.

If children learn the more-general pattern (AAB), they will prefer to listen to an AAB pattern word - even if it doesn't end in di - like kokoba, over a word that does not follow the AAB pattern, like kobako.

Experiment 1 results

Control: leledi…wiwije…jijili…dedewe

They can notice the AAB pattern and make the generalization from this artificial language data.

Experimental: leledi…wiwidi…jijidi…dededi

Children did not listen longer on average to test items consistent with the AAB pattern (like kokoba) [10.74 sec], as opposed to items inconsistent with it (like kobako) [10.18].

Implication: They do not make the more-general generalization (AAB).

Experiment 1 results

Control: leledi…wiwije…jijili…dedewe

They can notice the AAB pattern and make the generalization from this artificial language data.

Experimental: leledi…wiwidi…jijidi…dededi

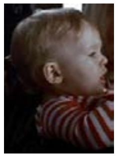
Implication: They do not make the more-general generalization (AAB) from this data

Question: Do they make the less-general generalization (AAdi), or do they just fail completely to make a generalization?

Experiment 2

Task type: Head Turn Preference Procedure

Experimental: leledi…wiwidi…jijidi…dededi



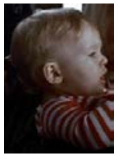
Children:
9-month-olds

Stimuli: 2 minutes of artificial language words.

Test condition words: novel AAdi pattern words using syllables the children had never encountered before in the language. Ex: kokodi (novel syllable: ko)

Experiment 2 predictions

Experimental: leledi…wiwidi…jijidi…dededi



If children learn the less-general pattern (AAdi), they will prefer to listen to an AAdi pattern word, like kokodi, over a word that does not follow the AAdi pattern, like kodiko.

If children don't learn any pattern, they will not prefer to listen to an AAdi pattern word, like kokodi, over a word that does not follow the AAdi pattern, like kodiko.

Experiment 2 results

Experimental: leledi…wiwidi…jijidi…dededi

Children prefer to listen to novel words that follow the less-general AAdi pattern, like kokodi [9.33 sec] over novel words that do not follow the AAdi pattern, like kodiko [6.25 sec].

Implication: They make the less-general generalization (AAdi) from this data. It is not the case that they fail to make any generalization at all.

Gerken (2006) results summary

Expt 1: Control (leledi…wiwije…jijili…dedewe)

Children notice the AAB pattern and make the generalization from artificial language data.

Expt 1: Experimental (leledi…wiwidi…jijidi…dededi)

Children do not make the more-general generalization (AAB) from this data.

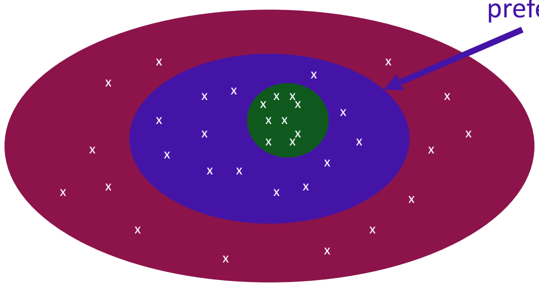
Expt 2: Experimental (leledi…wiwidi…jijidi…dededi)

Children make the less-general generalization (AAdi) from this data. It is not the case that they fail to make any generalization at all.

Gerken (2006) Results

When children are given data that is compatible with a less-general and a more-general generalization, they prefer to be conservative and make the less-general generalization.

prefer this one



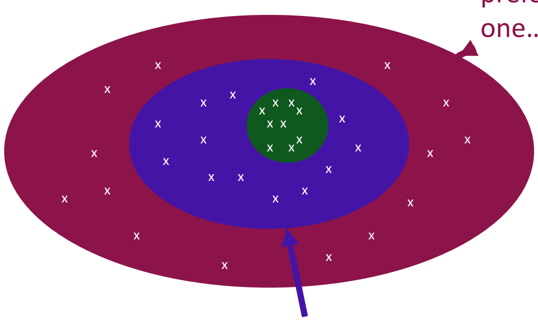
Gerken (2006) Results

When children are given data that is compatible with a less-general and a more-general generalization, they prefer to be conservative and make the less-general generalization.

Specifically for the artificial language study conducted, children prefer not to make unnecessary abstractions about the data. They prefer the AAdi pattern over a more abstract AAB pattern when the AAdi pattern fits the data they have encountered.

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

What if children preferred this one…

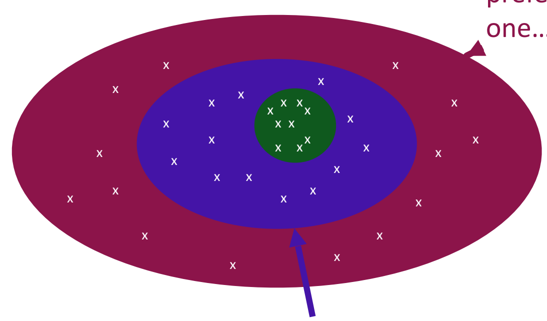


…but the language really was this one?

Problem: There are no data the child could receive that would clue them in that the less-general generalization is right. All data compatible with the less-general one are compatible with the more-general one.

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

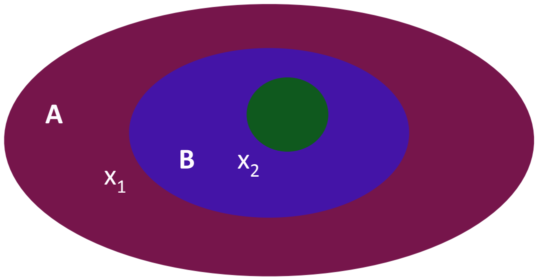
What if children preferred this one…



…but the language really was this one?

This is known as the Subset Problem for language learning.

Let's take a closer look at the Subset Problem



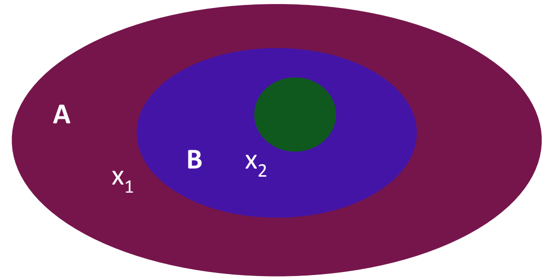
A is the superset
B is the subset

x_1 and x_2 are examples of data points

What data are compatible with A? x_1, x_2

What data are compatible with B? x_2

Let's take a closer look at the Subset Problem



A is the superset
B is the subset

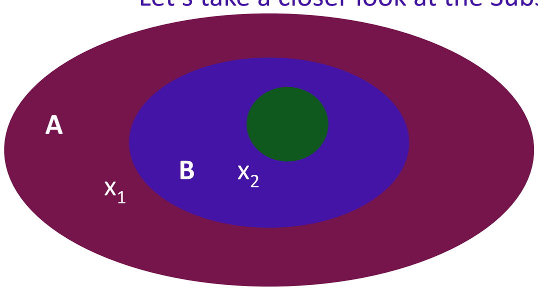
x_1 and x_2 are examples of data points

Suppose A is the correct generalization, and the child's hypothesis is that A is correct. (No fixing necessary.)

What data will the child see? x_1, x_2

What data will the child expect to see? x_1, x_2

Let's take a closer look at the Subset Problem



A is the superset
B is the subset

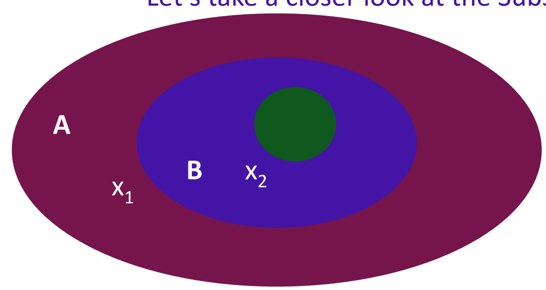
x_1 and x_2 are examples of data points

Suppose B is the correct generalization, and the child's hypothesis is that B is correct. (No fixing necessary.)

What data will the child see? x_2

What data will the child expect to see? x_2

Let's take a closer look at the Subset Problem



A is the superset
B is the subset

x_1 and x_2 are examples of data points

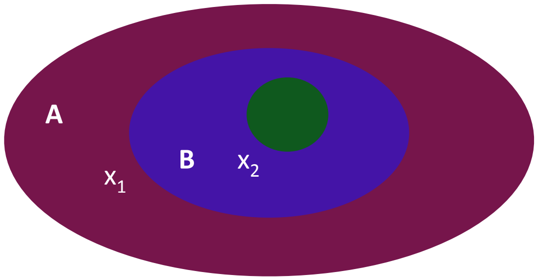
Suppose A is the correct generalization, and the child's hypothesis is that B is correct. (Fixing required.)

What data will the child see? x_1, x_2

What data will the child expect to see? x_2

Data like x_1 let the child realize that B is incorrect.

Let's take a closer look at the Subset Problem



A is the superset
B is the subset

x_1 and x_2 are examples of data points

Suppose B is the correct generalization, and the child's hypothesis is that A is correct. (Fixing required.)

What data will the child see? x_2

What data will the child expect to see? x_1, x_2

There are no data the child will see that indicate A is incorrect. This is the Subset Problem - when the subset is correct but the superset is chosen.

Solutions to the Subset Problem

Subset Principle (Wexler & Manzini 1987): In order to learn correctly in this scenario where one generalization covers a subset of the data another generalization covers, children should prefer the less-general generalization.

This is a learning strategy that can result very naturally from a Bayesian learner which uses the Size Principle (Tenenbaum & Griffiths 2001).

The Size Principle & suspicious coincidences

A Bayesian learner can assign a probability to any hypothesis under consideration by balancing two things:
The prior probability of that hypothesis being correct
The likelihood of that hypothesis producing the observed data

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{hypothesis}) * P(\text{data} \mid \text{hypothesis})}{P(\text{data})}$$

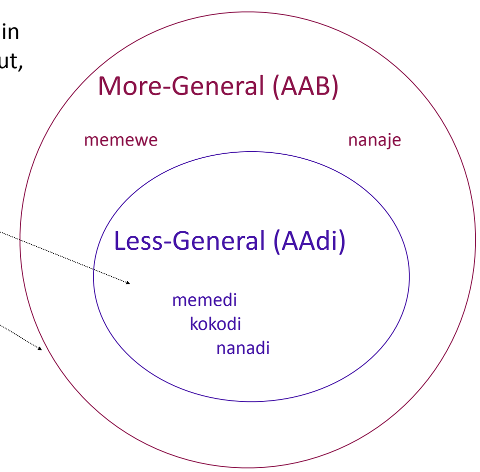
The likelihood calculation allows a Bayesian learner to follow the Size Principle (Tenenbaum & Griffiths 2001), and automatically prefer less-general hypotheses (which correspond to sets of smaller size) to more-general hypotheses (which correspond to sets of larger size). This is sometimes referred to as a sensitivity to "suspicious coincidences" (Xu & Tenenbaum 2007).

Formal instantiation of "suspicious coincidence"

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

Hypothesis 1 (H1): The less-general hypothesis is true, and *AAdi* is the pattern.

Hypothesis 2 (H2): The more-general hypothesis is true, and *AAB* is the pattern.



Formal instantiation of “suspicious coincidence”

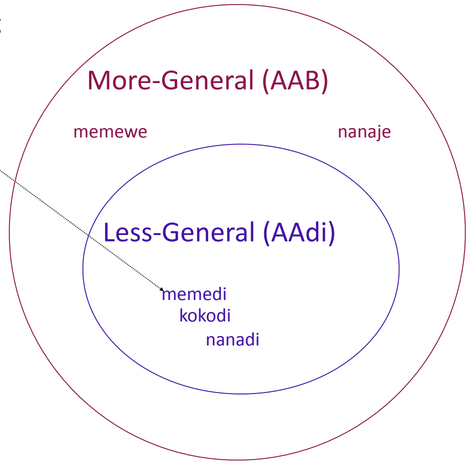
What’s the likelihood of selecting this word for each hypothesis?

$p(\text{memedi} \mid H1) = 1/3$

(since only three words are possible)

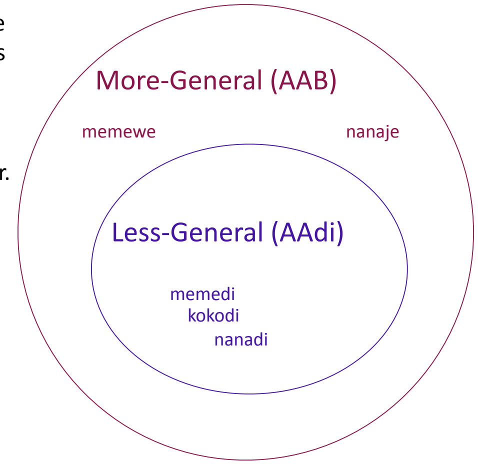
$p(\text{memedi} \mid H2) = 1/5$

(since all five words are possible)



Formal instantiation of “suspicious coincidence”

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



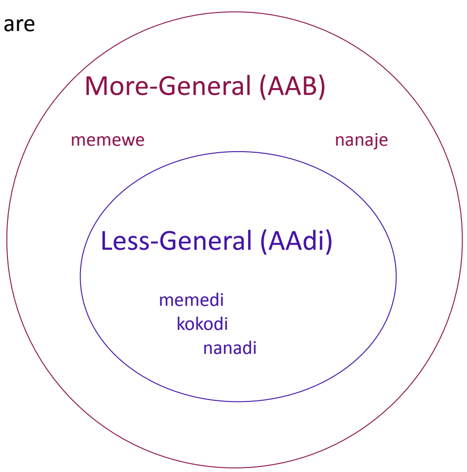
Formal instantiation of “suspicious coincidence”

If the prior is equal

(ex: before any data, both hypotheses are equally likely [p(H1) = p(H2) = *p*])…

$$p(H1 \mid \text{memedi}) = \frac{p(\text{memedi} \mid H1) * p(H1)}{p(\text{memedi})}$$

$$p(H2 \mid \text{memedi}) = \frac{p(\text{memedi} \mid H2) * p(H2)}{p(\text{memedi})}$$



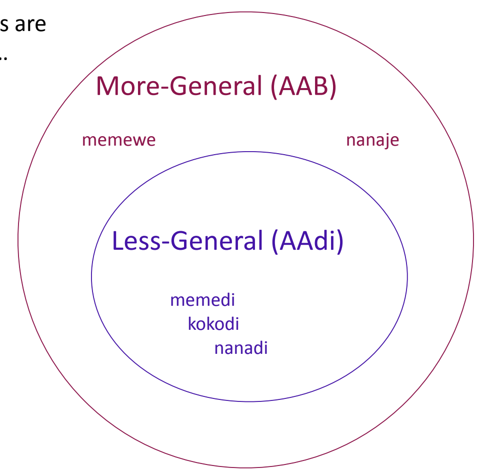
Formal instantiation of “suspicious coincidence”

If the prior is equal

(ex: before any data, both hypotheses are equally likely [p(H1) = p(H2) = *p*])…

$$p(H1 \mid \text{memedi}) = \frac{p(\text{memedi} \mid H1) * p}{p(\text{memedi})}$$

$$p(H2 \mid \text{memedi}) = \frac{p(\text{memedi} \mid H2) * p}{p(\text{memedi})}$$

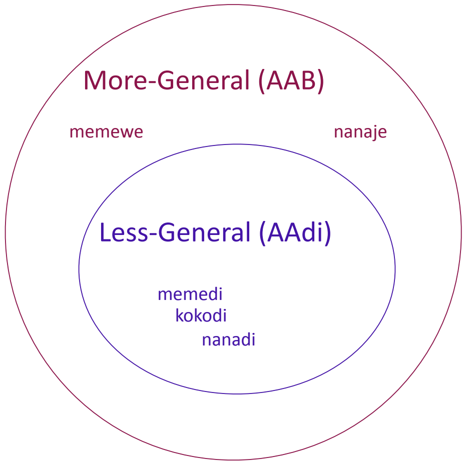


Formal instantiation of “suspicious coincidence”

...and p(data) is the same for both
[= p(memedi | H1) * *p* + p(memedi | H2) * *p*]
[= p(data) = *d*] ...

$$p(H1 \mid memedi) = \frac{p(memedi \mid H1) * p}{p(memedi)}$$

$$p(H2 \mid memedi) = \frac{p(memedi \mid H2) * p}{p(memedi)}$$

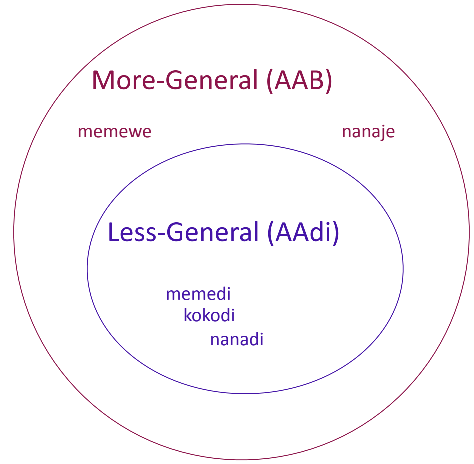


Formal instantiation of “suspicious coincidence”

...and p(data) is the same for both
[= p(memedi | H1) * *p* + p(memedi | H2) * *p*]
[= p(data) = *d*] ...

$$p(H1 \mid memedi) = \frac{p(memedi \mid H1) * p}{d}$$

$$p(H2 \mid memedi) = \frac{p(memedi \mid H2) * p}{d}$$

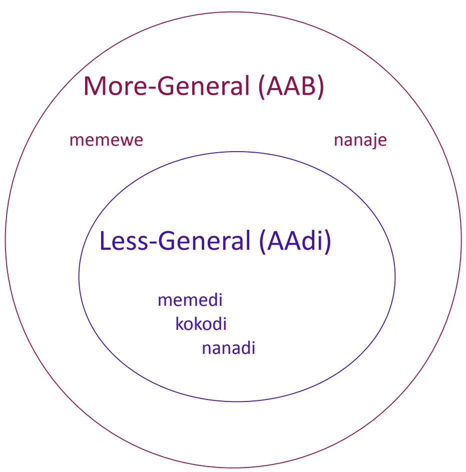


Formal instantiation of “suspicious coincidence”

...then the likelihood is what
determines the posterior.

$$p(H1 \mid memedi) = \frac{p(memedi \mid H1) * p}{d}$$

$$p(H2 \mid memedi) = \frac{p(memedi \mid H2) * p}{d}$$

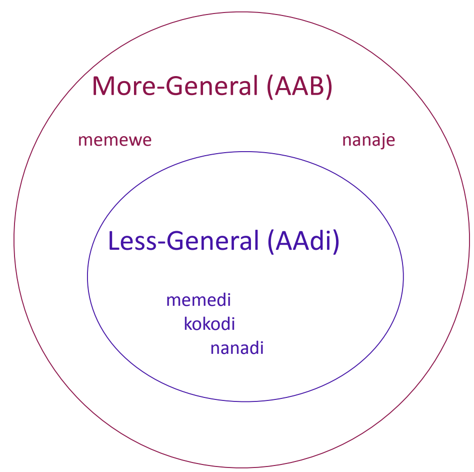


Formal instantiation of “suspicious coincidence”

...then the likelihood is what
determines the posterior.

$$p(H1 \mid memedi) = \frac{1/3 * p}{d}$$

$$p(H2 \mid memedi) = \frac{1/5 * p}{d}$$

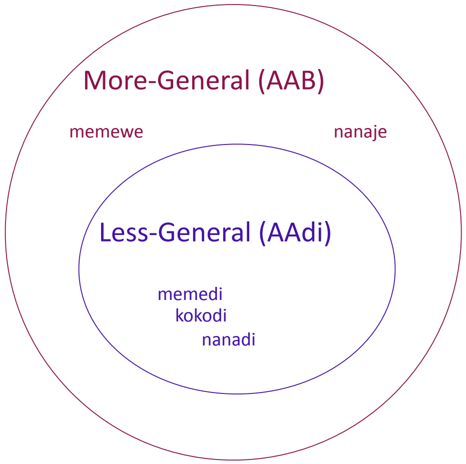


Formal instantiation of “suspicious coincidence”

So, the posterior probability will be greater for the less-general hypothesis.

$$p(H1|_{memedi}) = \frac{1/3 * p}{d}$$

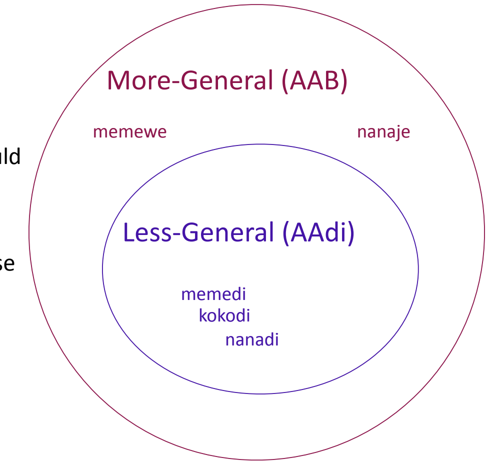
$$p(H2|_{memedi}) = \frac{1/5 * p}{d}$$



Another way to think about it

Has to do with children’s expectation of the data points that they should encounter in the input

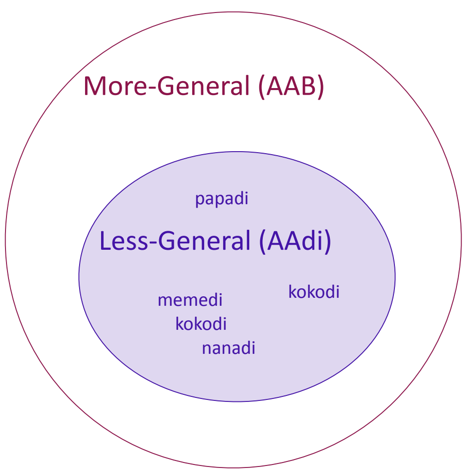
If more-general generalization (AAB) is correct, the child should encounter some data that can only be accounted for by the more-general generalization (like memewe or nanaje). These data would be incompatible with the less-general generalization (AAdi).



Another way to think about it

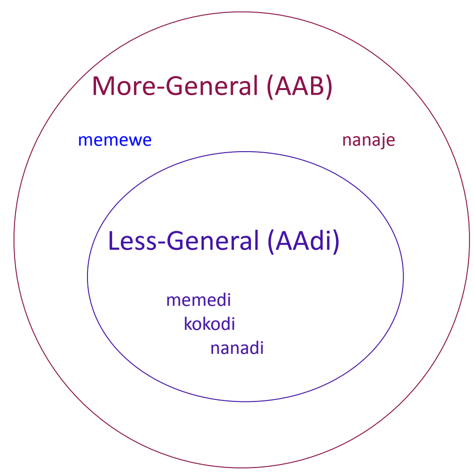
Has to do with children’s expectation of the data points that they should encounter in the input

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



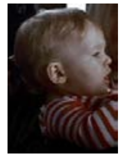
Children as rational learners

Gerken (2006) suggests that children behave like rational (Bayesian) learners. If so, this means that if children do receive counterexamples to the less-general hypothesis, they should update their beliefs about its probability. In particular, they should believe it is less probable than the more-general hypothesis. Is this true?



Gerken (2010)

Experimental: leledi…wiwidi…jijidi…dededi + 3 AAB



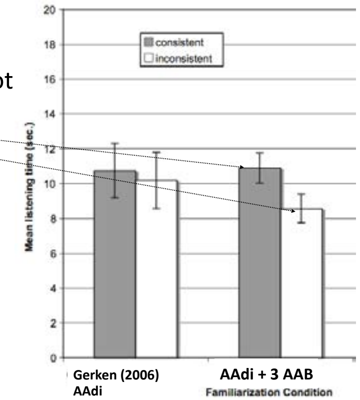
Children:
9-month-olds

Stimuli: 2 minutes of artificial language words following the AAdi pattern, with three of the last stimuli heard being examples of the AAB pattern (like memewe)

Test condition words: novel AAB pattern words using syllables the children had never encountered before in the language. Ex: kokoba (novel syllable: ko)

Gerken (2010)

Children prefer to listen to novel words that follow the more-general AAB pattern, like kokoba [~11 sec] over novel words that do not follow the AAB pattern, like kobako [~8 sec].



Implication: They update their beliefs about which hypothesis is more probable, given a few data that implicate the more-general AAB hypothesis. They are sensitive to counterexamples.