Ling 151/Psych 156A: Acquisition of Language II

Lecture 21

Structure II

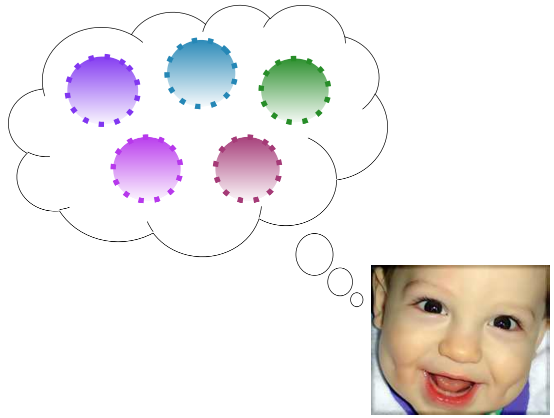
Announcements

Review questions are available for structure

HW8 due 3/16/18

Online course evaluations are available for this class - please fill them out if you haven't already! :)

About linguistic parameters

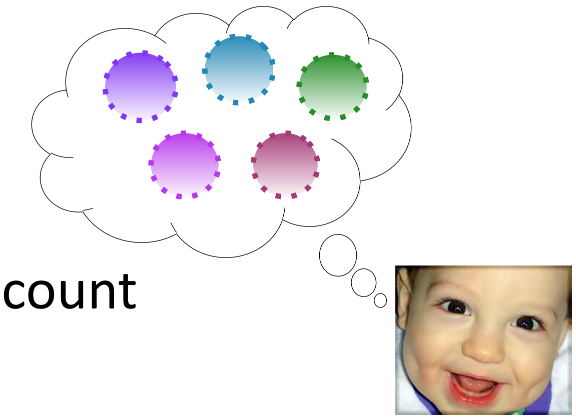


What are linguistic parameters?

How do they work?

What exactly are they supposed to do?

About linguistic parameters

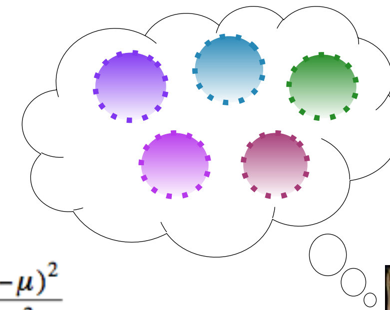


A parameter is meant to be something that can account for multiple observations in some domain.

Parameter for a statistical model: determines what the model predicts will be observed in the world in a variety of situations

Parameter for our mental (and linguistic) model: determines what *we* predict will be observed in the world in a variety of situations

About linguistic parameters

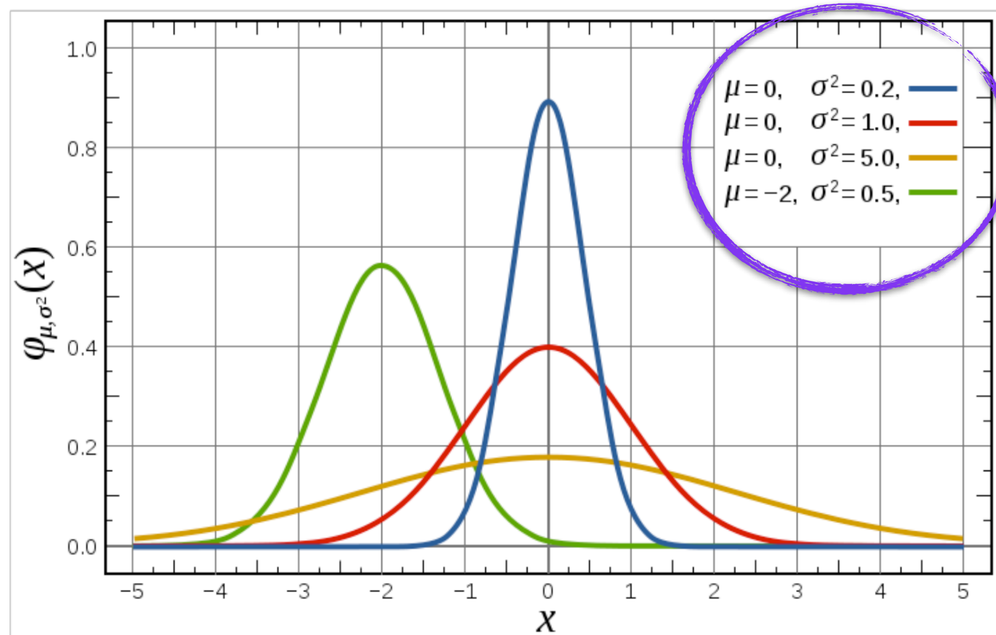


Statistical parameter

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

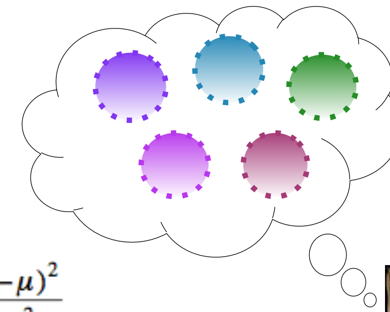
The normal distribution is a statistical model that uses **two parameters**:

- μ for the mean
- σ for the standard deviation



If we know the **values of these parameters**, we can make predictions about the probability of data we rarely or never see.

About linguistic parameters



$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

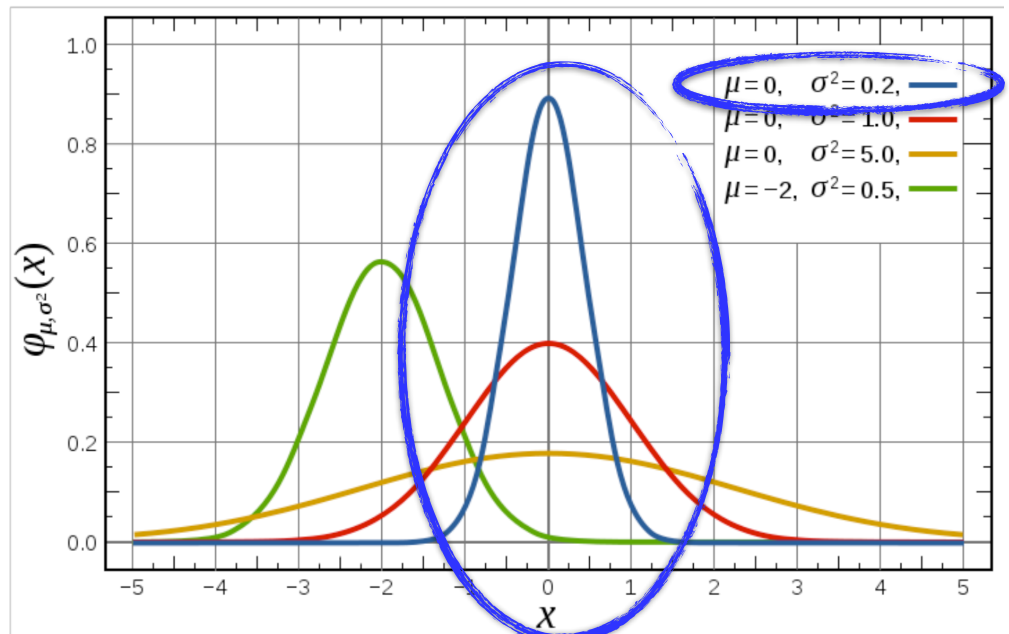
Statistical parameter

μ for the mean

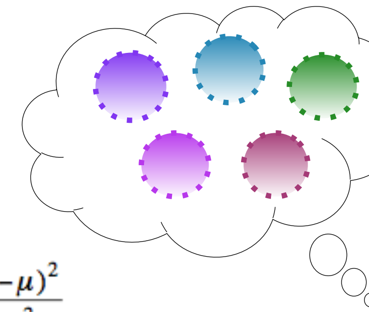
σ for the standard deviation

Suppose this is a model of **how many minutes late** I'll be to class.

Let's use the model with **μ = 0 and σ^2 = 0.2**.



About linguistic parameters



$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

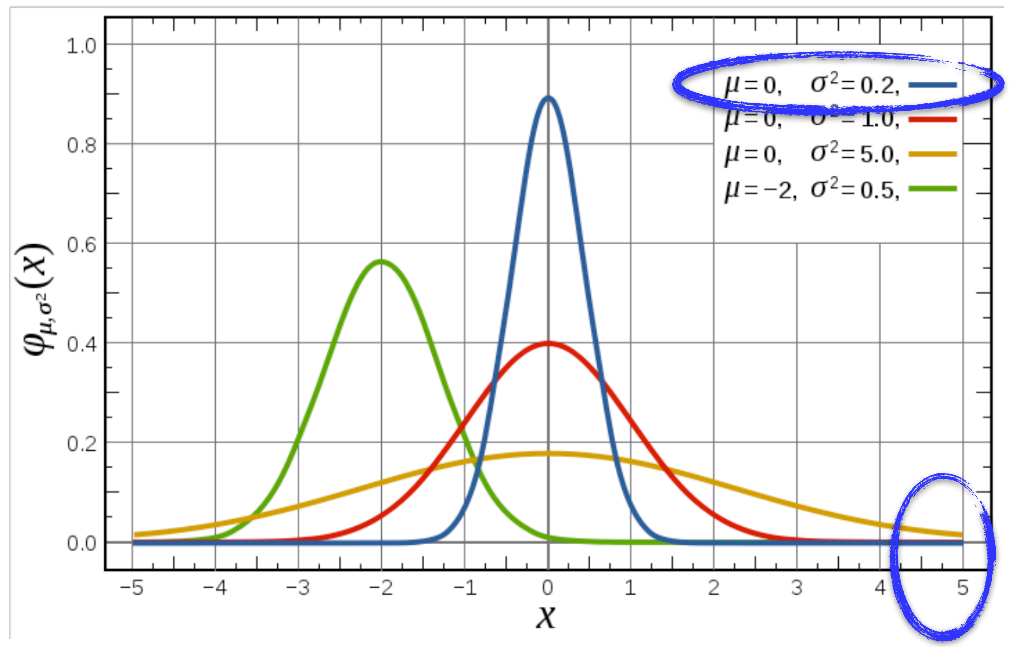
Statistical parameter

μ for the mean

σ for the standard deviation

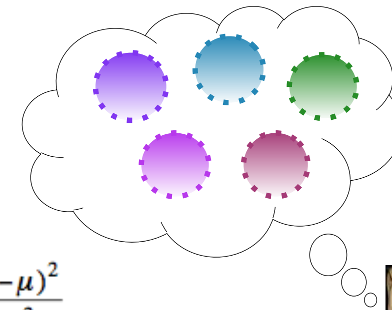
Let's use the model with **μ = 0 and σ^2 = 0.2**.

How probable is it that I'll be 5 minutes late, given these parameter values?



Not very probable!

About linguistic parameters



Statistical parameter

μ for the mean

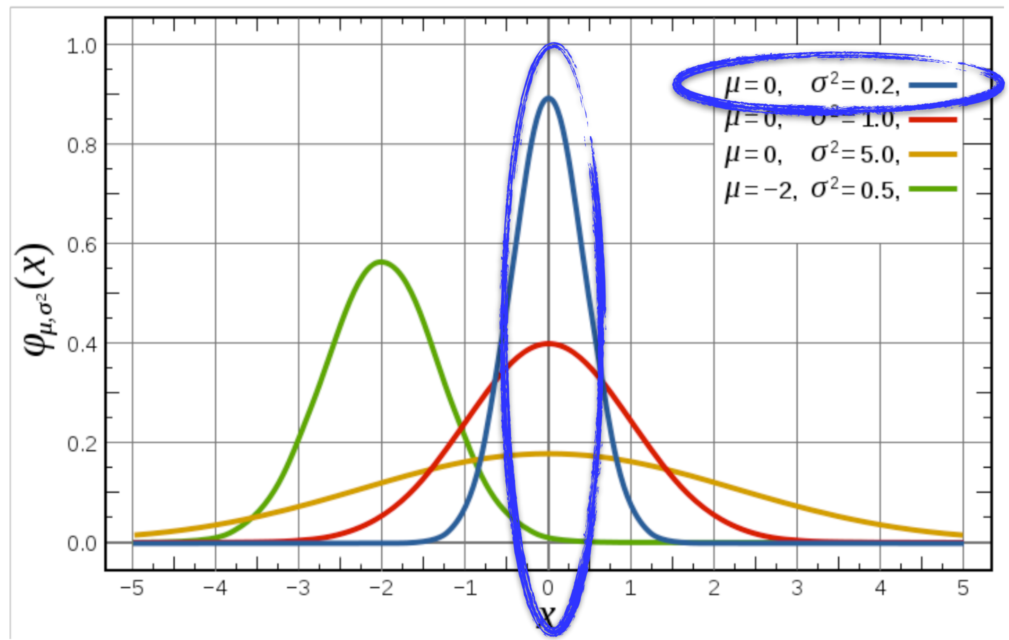
σ for the standard deviation

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Let's use the model with **μ = 0 and σ^2 = 0.2**.

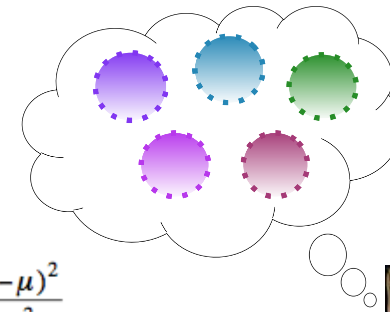
5 minutes late?

What about right on time?



Much more probable!

About linguistic parameters

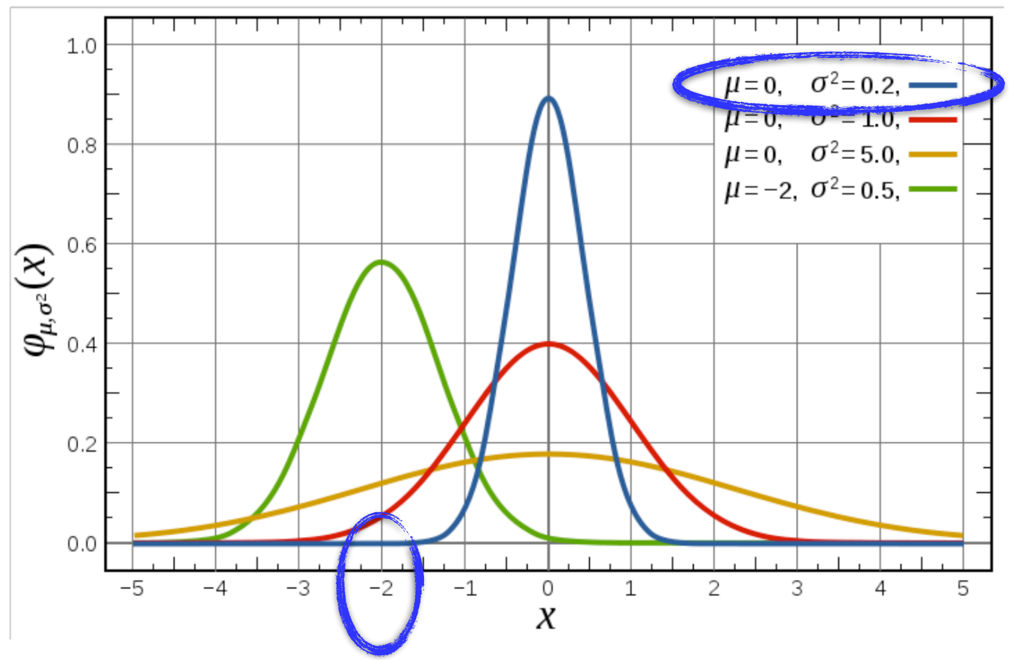


$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Statistical parameter

μ for the mean

σ for the standard deviation



Let's use the model with **μ = 0 and σ^2 = 0.2**.

5 minutes late?

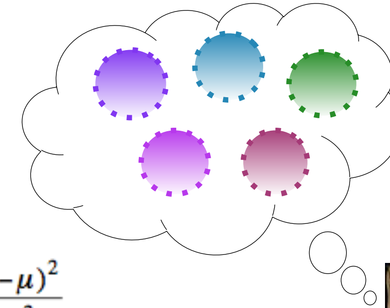
On time?

What about 2 minutes early?

We can tell this just by knowing the values of the two statistical parameters. These parameter values allow us to infer the probability of the observable behavior.

Not very probable!

About linguistic parameters



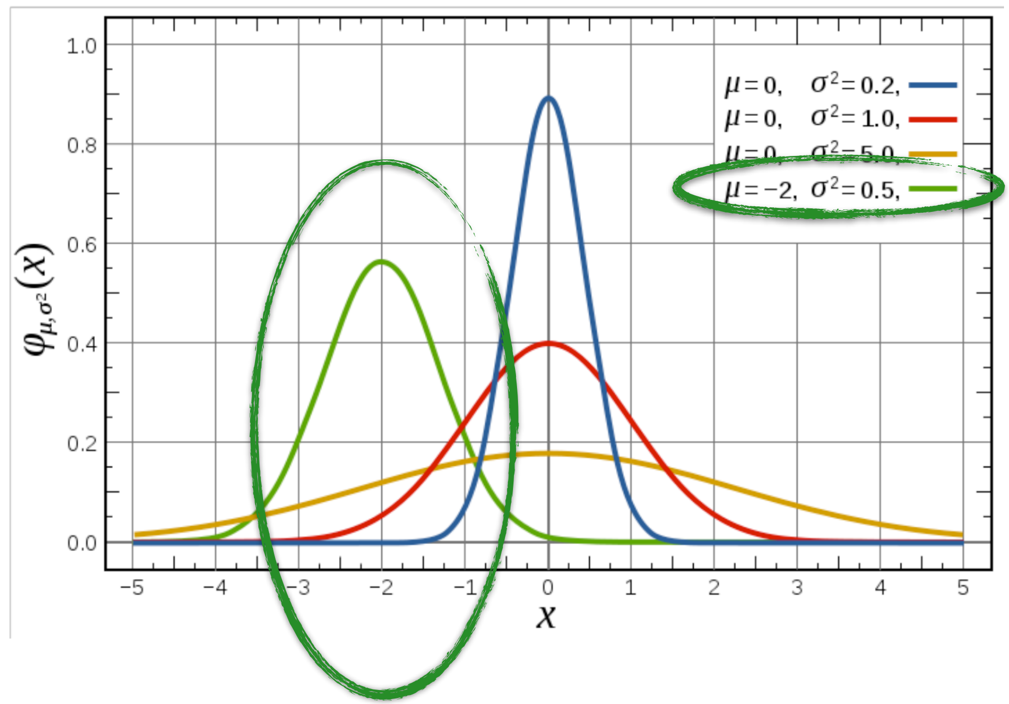
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Statistical parameter

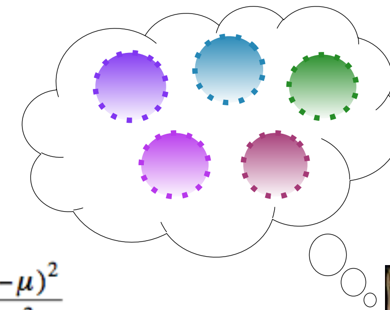
μ for the mean

σ for the standard deviation

Let's shift to the model with
μ = -2 and σ^2 = 0.5.



About linguistic parameters



Statistical parameter

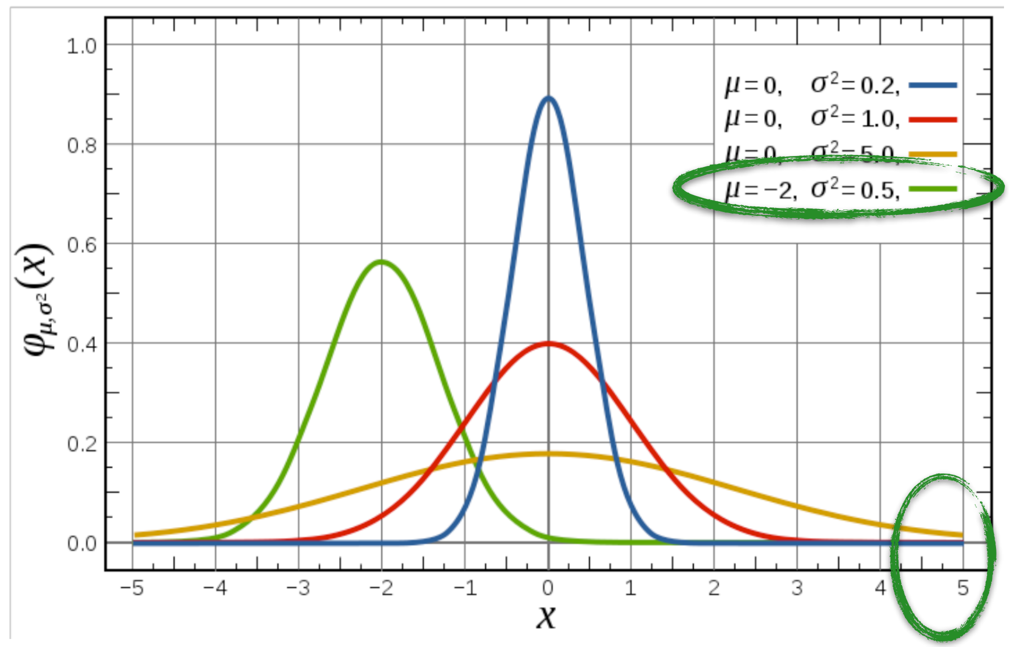
μ for the mean

σ for the standard deviation

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

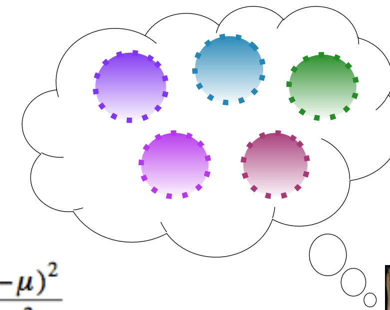
Let's shift to the model with
μ = -2 and σ^2 = 0.5.

How probable is it that I'll be 5 minutes late, given these parameter values?



Not very probable!

About linguistic parameters



Statistical parameter

μ for the mean

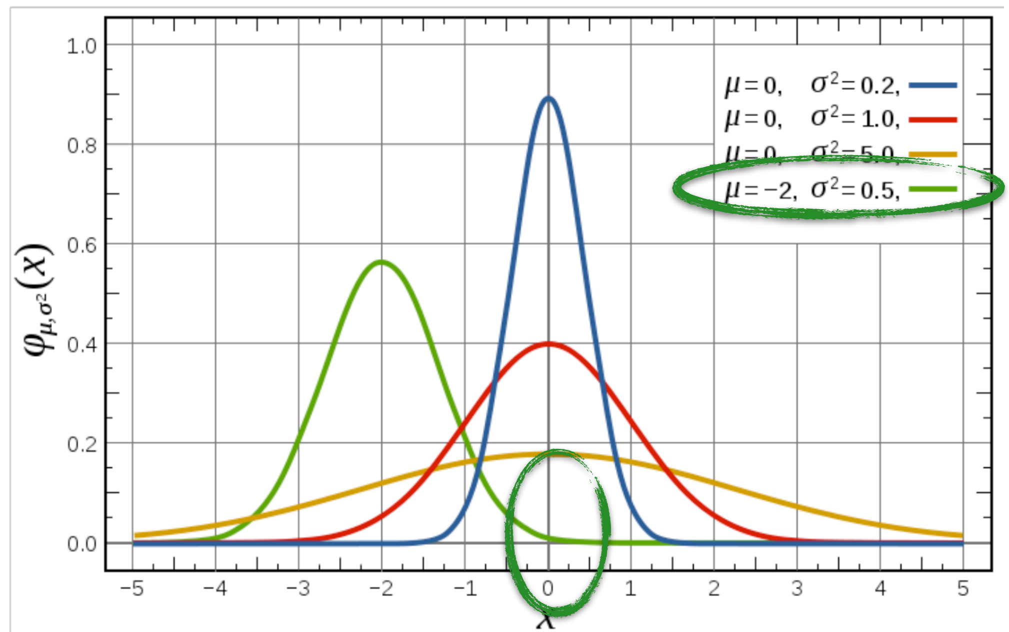
σ for the standard deviation

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Let's shift to the model with
μ = -2 and σ^2 = 0.5.

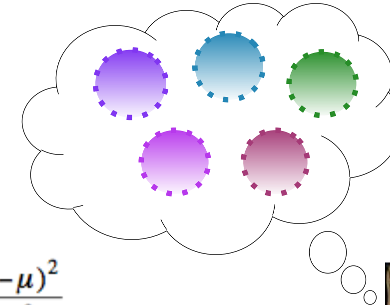
5 minutes late?

What about right on time?



Not very probable!

About linguistic parameters



Statistical parameter

μ for the mean

σ for the standard deviation

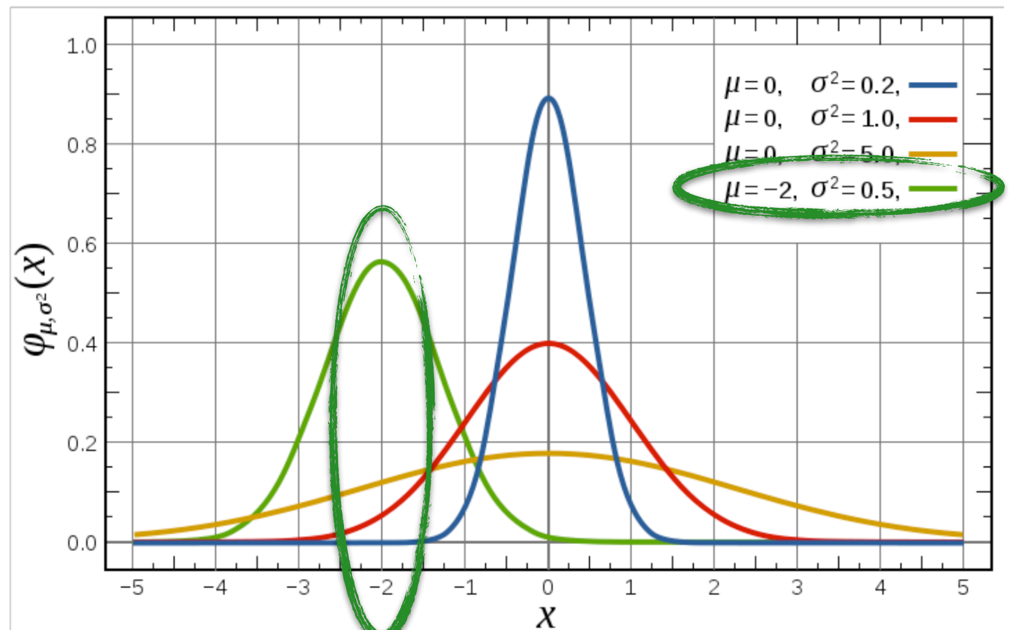
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Let's shift to the model with
μ = -2 and σ^2 = 0.5.

5 minutes late?

On time?

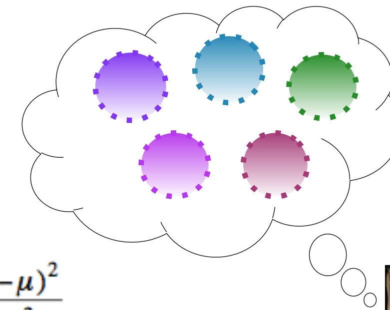
What about 2 minutes early?



Much more probable!

Changing the parameter values changes the behavior we predict we'll observe.

About linguistic parameters



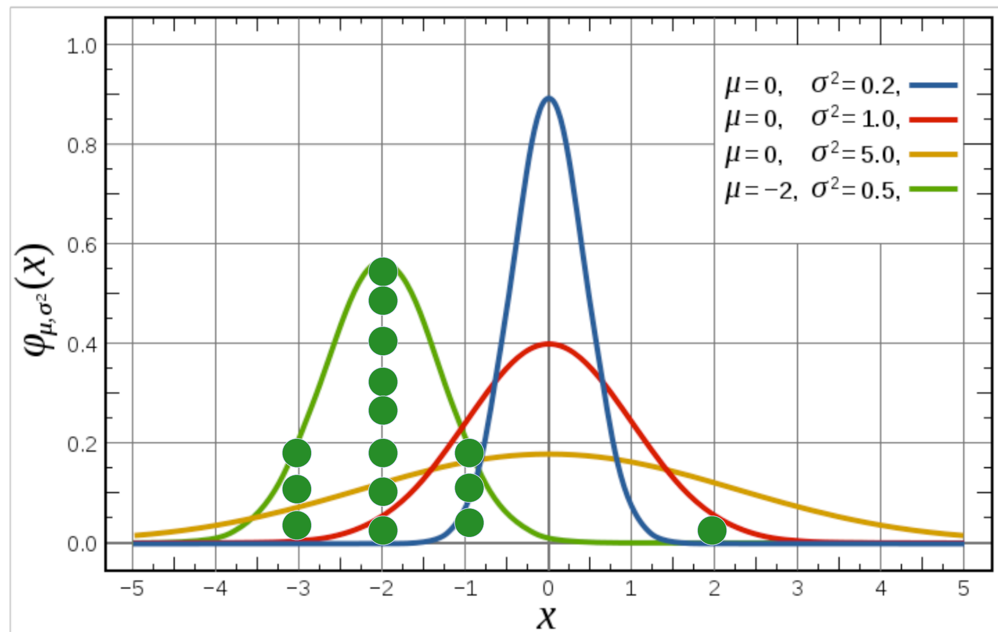
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Statistical parameter

μ for the mean

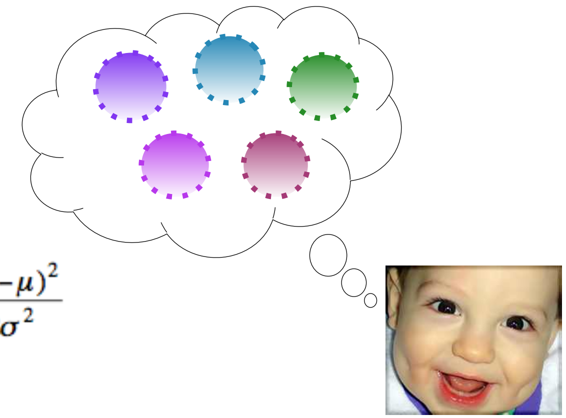
σ for the standard deviation

Observing different quantities of data with particular values can tell us which values of μ and σ^2 are most likely, if we know we're trying to determine the values of μ and σ^2 in function $\phi(X)$



Observing data points distributed like the green curve tells us that μ is likely to be around -2 and σ^2 is likely to be around 0.5.

About linguistic parameters

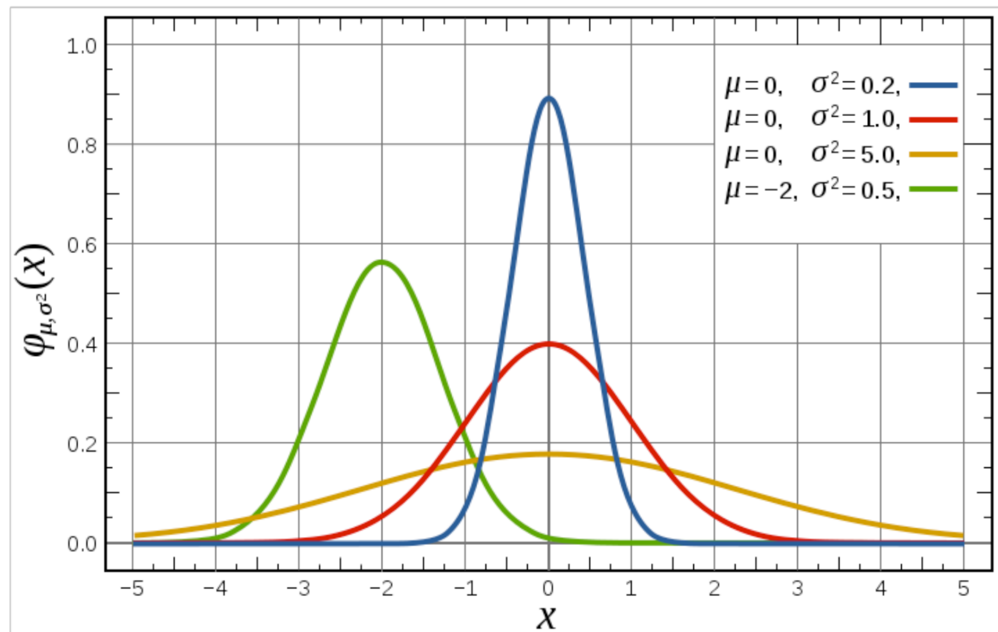


Statistical parameter

μ for the mean

σ for the standard deviation

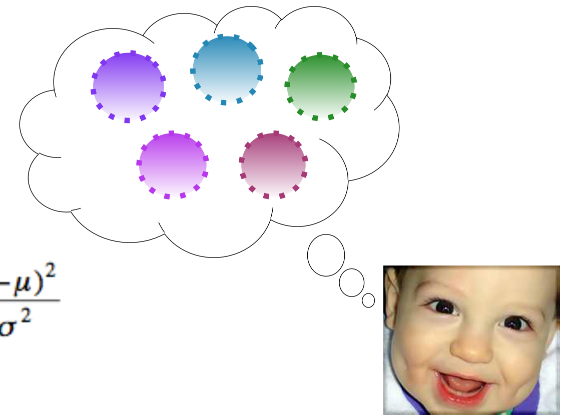
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$



Important similarity to linguistic parameters:

We don't see the process that generates the data, but only the data themselves. This means that in order to form our expectations about X, we are, in effect, reverse engineering the observable data.

About linguistic parameters

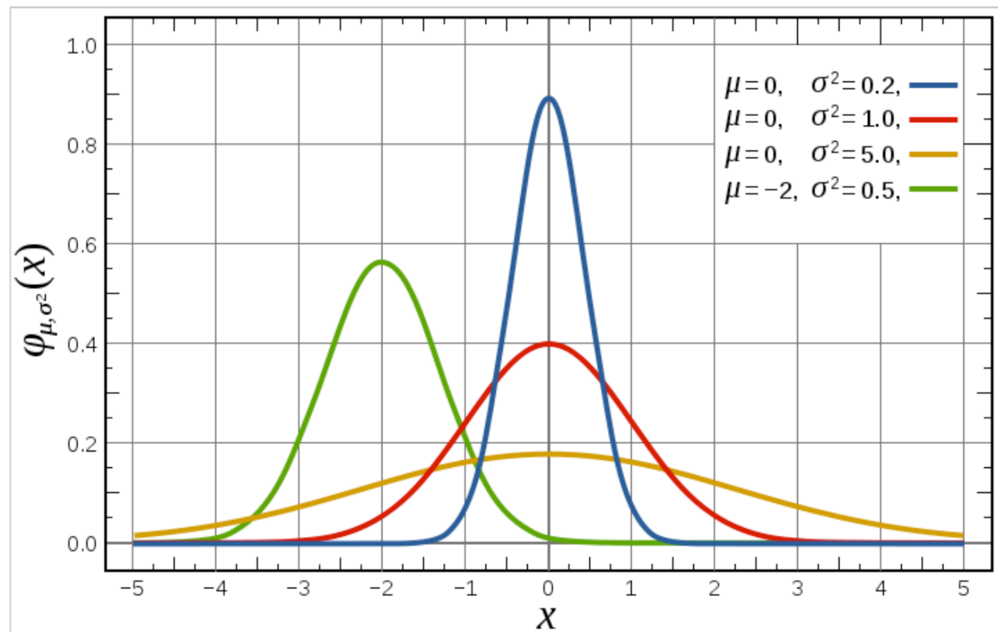


Statistical parameter

μ for the mean

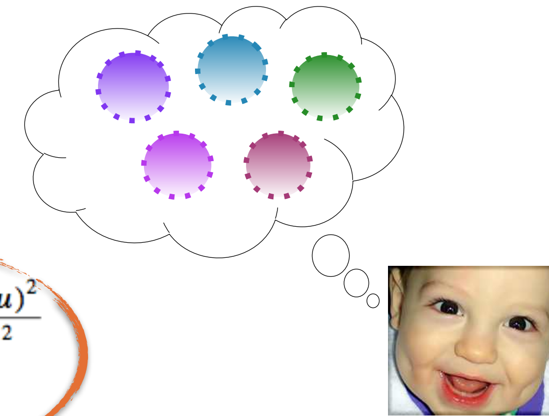
σ for the standard deviation

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$



Our knowledge of the underlying function/principle that generates these data - $\phi(X)$ - as well as the associated parameters - μ, and σ^2 - allows us to represent an infinite number of expectations about the behavior of variable X.

About linguistic parameters



Comparison: **the equation's form** – it's the statistical "principle" that explains the observed data.

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Both linguistic principles and linguistic parameters are often thought of as innate domain-specific abstractions that connect to many structural properties about language.

Linguistic **principles** correspond to the properties that are invariant across all human languages.

About linguistic parameters

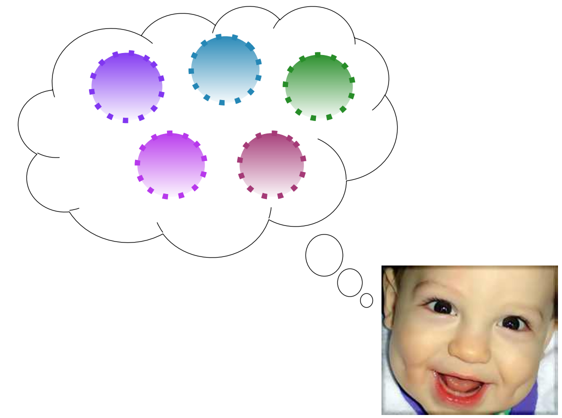
Comparison: μ and σ^2 determine the exact form of the curve that represents the probability of observing certain data. While different values for these parameters can produce many different curves, these curves share their underlying form due to the common invariant function.

$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Both linguistic principles and linguistic parameters are often thought of as innate domain-specific abstractions that connect to many structural properties about language.

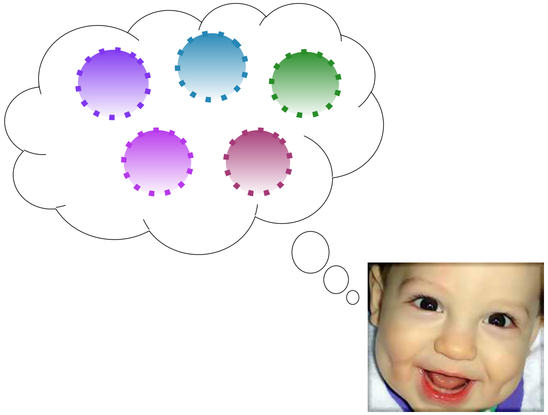
Linguistic parameters correspond to the properties that vary across human languages

About linguistic parameters for language acquisition



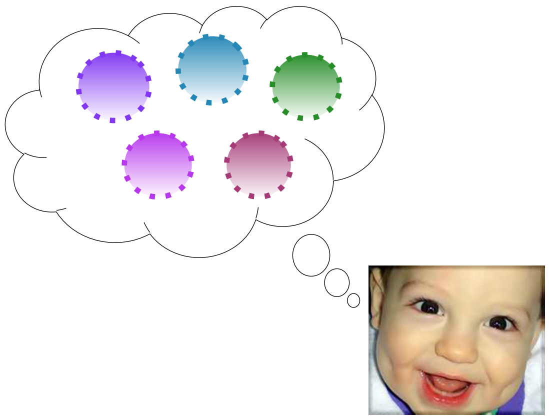
Parameters connecting to multiple structural properties is a very good thing from the perspective of someone trying to acquire language (like a child). This is because a child can learn about a parameter's value by observing many different kinds of examples in the language.

About linguistic parameters for language acquisition



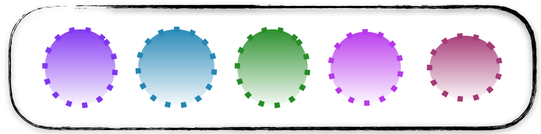
"The richer the deductive structure associated with a particular parameter, the greater the range of potential 'triggering' data which will be available to the child for the 'fixing' of the particular parameter" – Hyams (1987)

About linguistic parameters for language acquisition



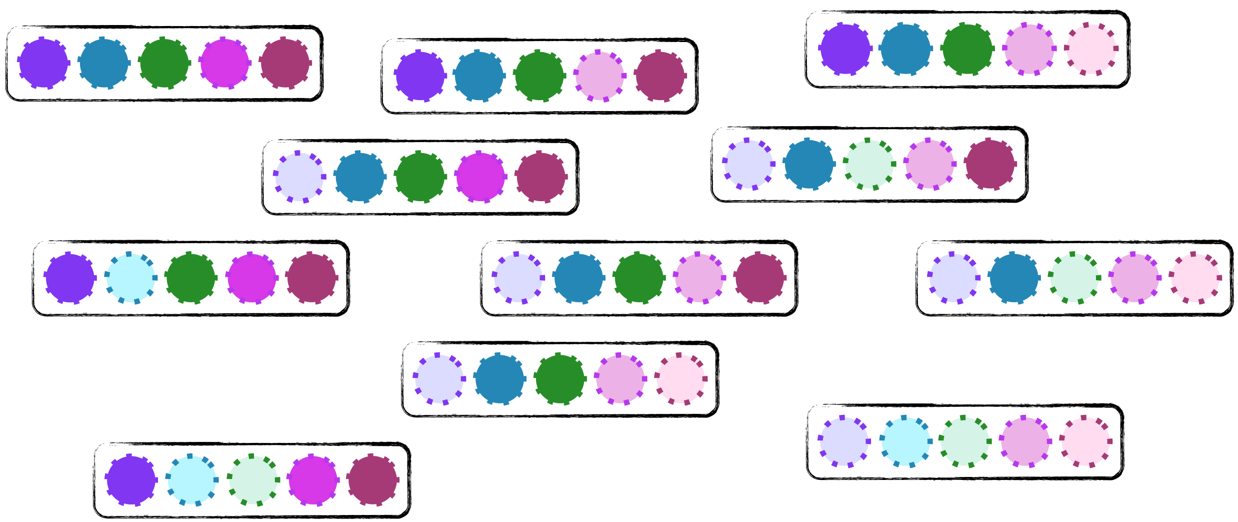
Parameters can be especially useful when a child is trying to learn the things about language structure that are otherwise hard to learn, perhaps because they are very complex properties themselves or because they appear very infrequently in the available data.

Parameters & overhypotheses

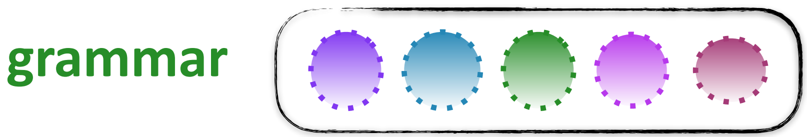


Remember:

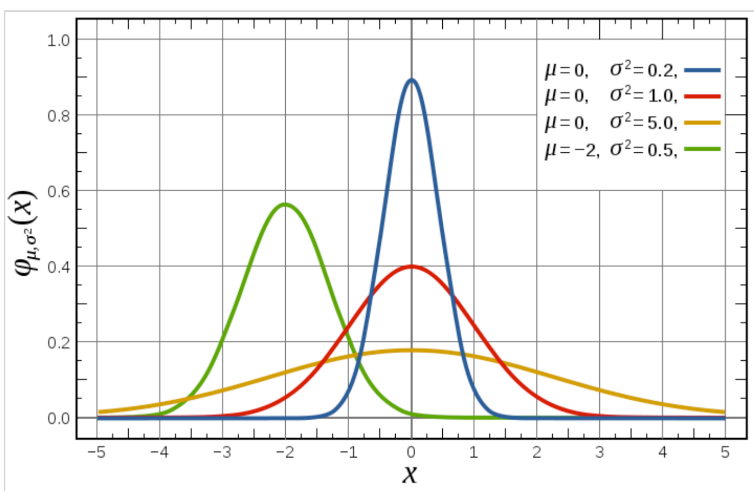
We can think of language systems (grammars) as collections of parameter values.



Parameters & overhypotheses

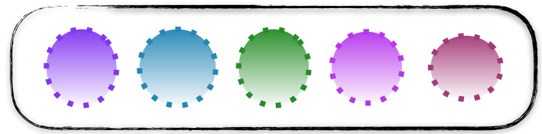


A parameter (and its specific value) determines what we predict will be observed in the world in a variety of situations.



Parameters & overhypotheses

grammar



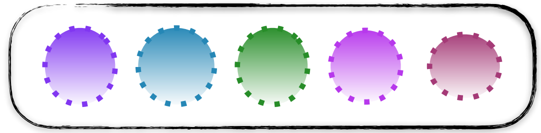
A parameter determines what we predict will be observed.

Example: Head-directionality

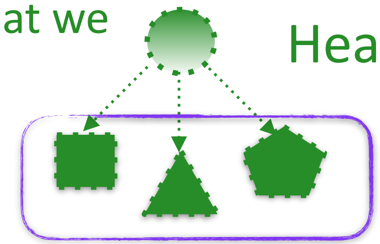
Linguistic parameters correspond to the properties that vary across human languages.

Parameters & overhypotheses

grammar



A parameter determines what we predict will be observed.

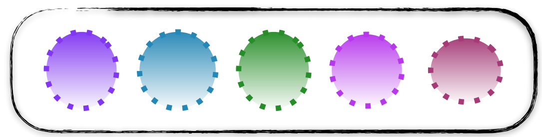


Head-directionality

The fact that parameters connect to multiple structural properties is a very good thing for acquisition. This is because a child can learn about that parameter's value by observing many different kinds of examples in the language.

Parameters & overhypotheses

grammar

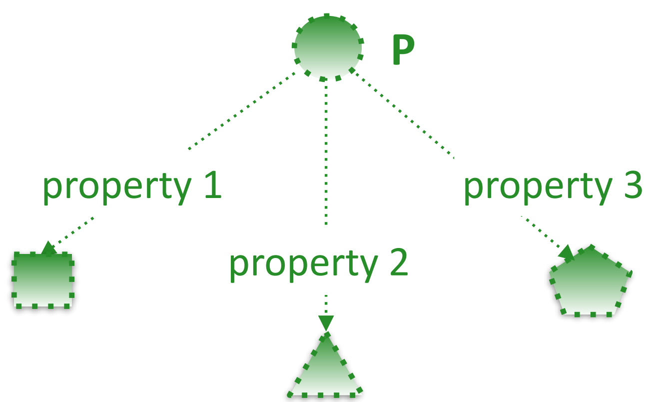


A parameter determines what we predict will be observed.

Head-directionality

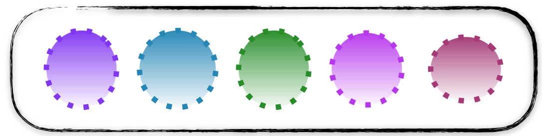
good for acquisition

Let's assume a number of **properties** are all connected to parameter **P**, which can take one of two values: a or b.



Parameters & overhypotheses

grammar

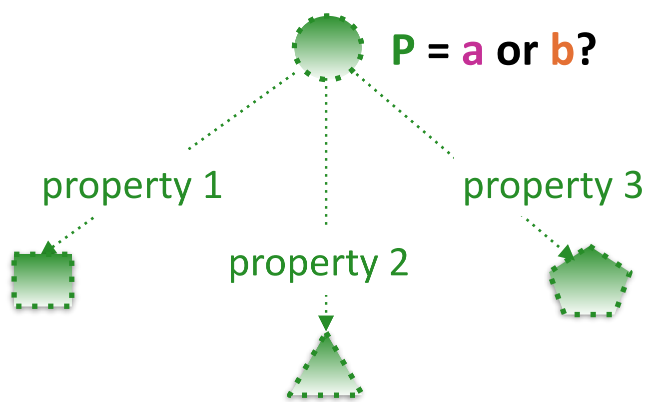


A parameter determines what we predict will be observed.

Head-directionality

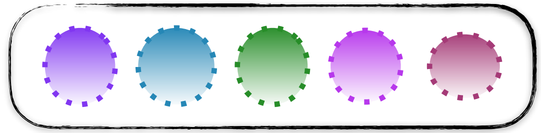
good for acquisition

Let’s assume a number of **properties** are all connected to parameter **P**, which can take one of two values: a or b.



Parameters & overhypotheses

grammar



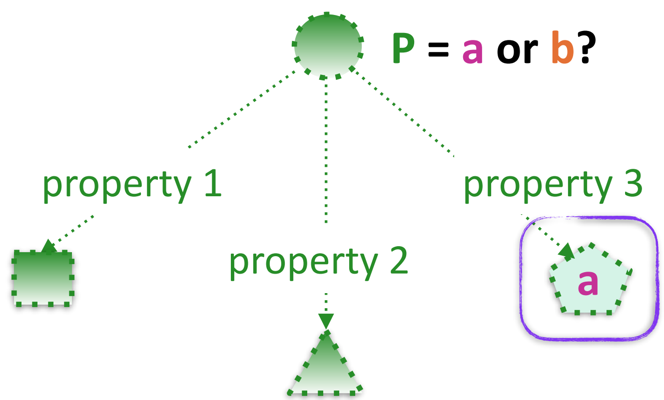
A parameter determines what we predict will be observed.

Head-directionality

good for acquisition

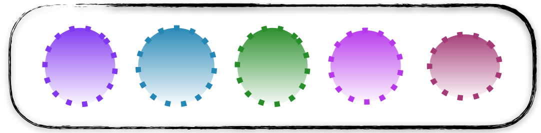
How do we learn whether property 3 shows behavior a or b?

One way is to observe instances of property 3 in the intake.



Parameters & overhypotheses

grammar

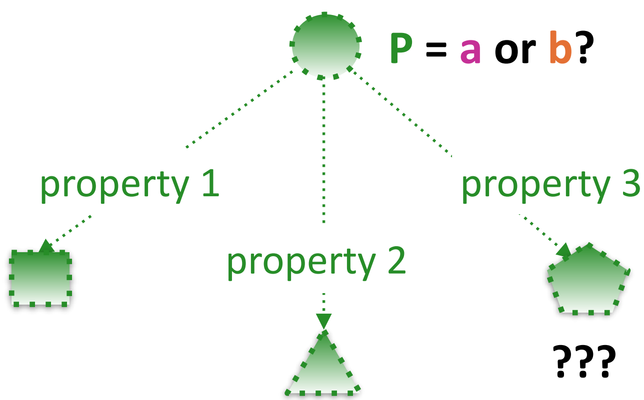


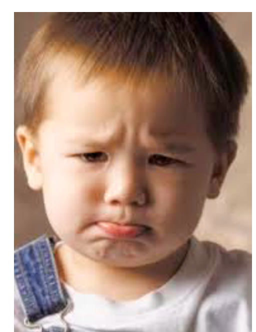
A parameter determines what we predict will be observed.

Head-directionality

good for acquisition

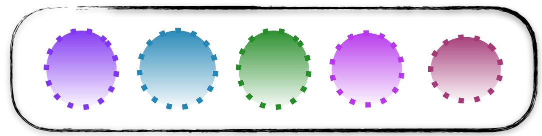
But what if property 3 occurs very rarely? We might never see any examples of property 3.





Parameters & overhypotheses

grammar

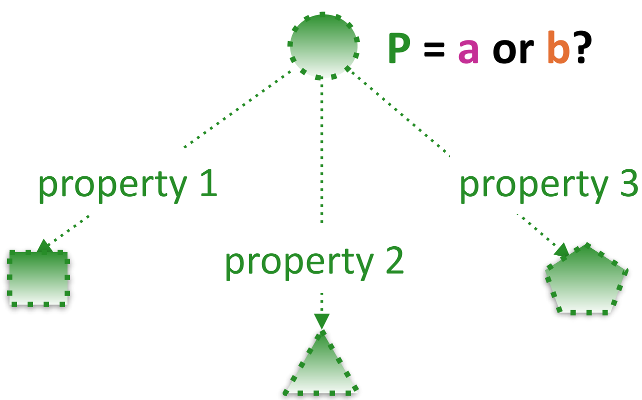


A parameter determines what we predict will be observed.

Head-directionality

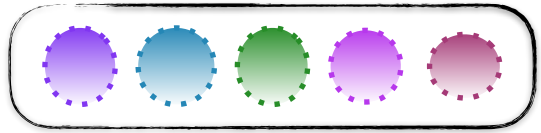
good for acquisition

Fortunately, because property 3 is connected to P, we can learn the value for property 3 by learning the value of P.



Parameters & overhypotheses

grammar

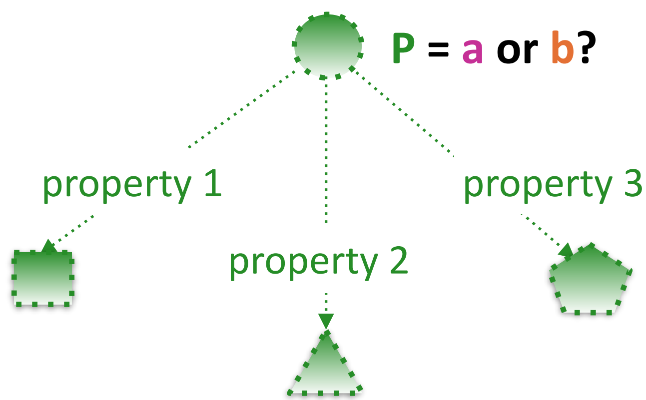


A parameter determines what we predict will be observed.

Head-directionality

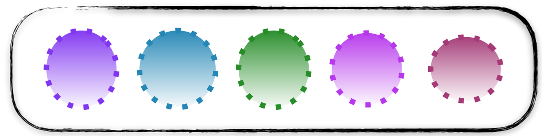
good for acquisition

Also fortunately, P is connected to properties 1 and 2.



Parameters & overhypotheses

grammar

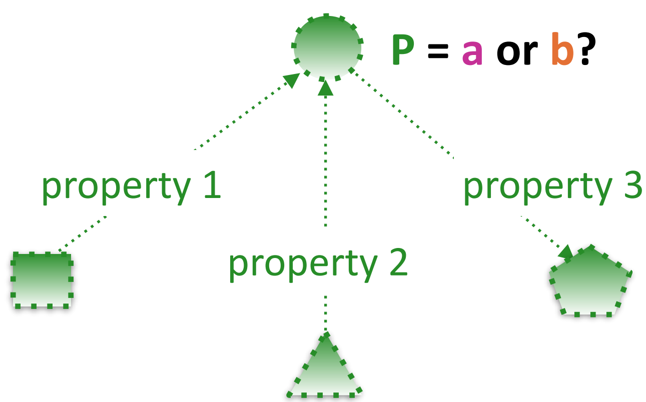


A parameter determines what we predict will be observed.

Head-directionality

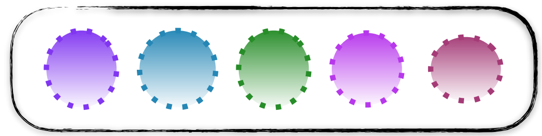
good for acquisition

This means we can learn the value of P from property 1 or property 2.



Parameters & overhypotheses

grammar

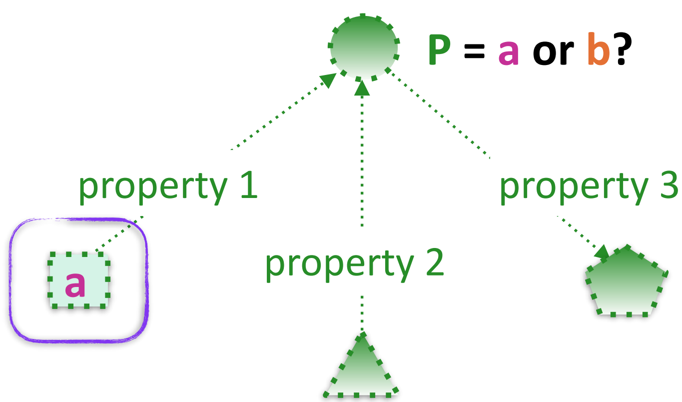


A parameter determines what we predict will be observed.

Head-directionality

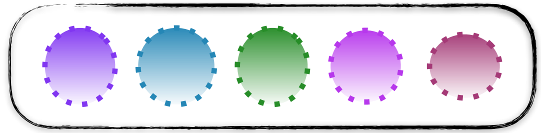
good for acquisition

Suppose we see an example of property 1 with value a.



Parameters & overhypotheses

grammar

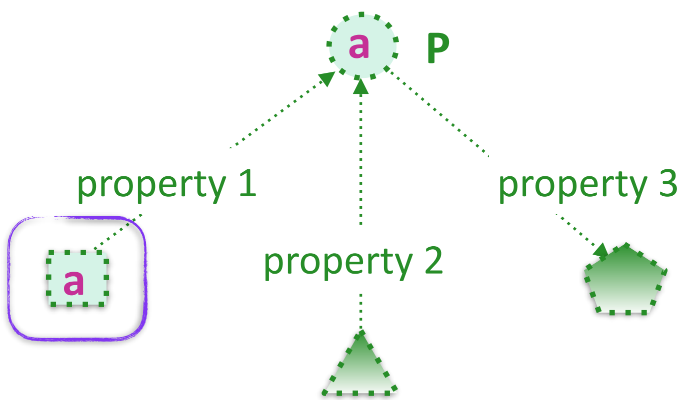


A parameter determines what we predict will be observed.

Head-directionality

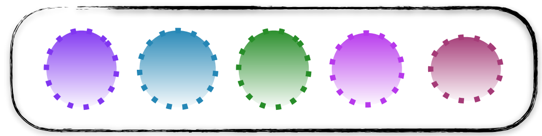
good for acquisition

This means P also should have value a.



Parameters & overhypotheses

grammar



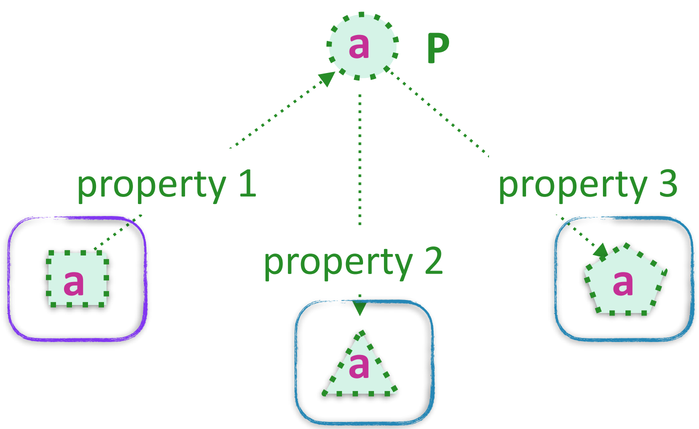
A parameter determines what we predict will be observed.

Head-directionality

good for acquisition

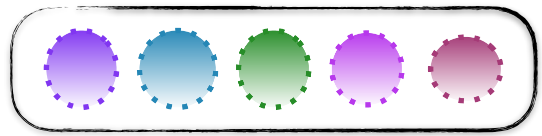
So, we can make predictions for all the other properties connected to P, even if we've never seen examples of them.

This is great!



Parameters & overhypotheses

grammar

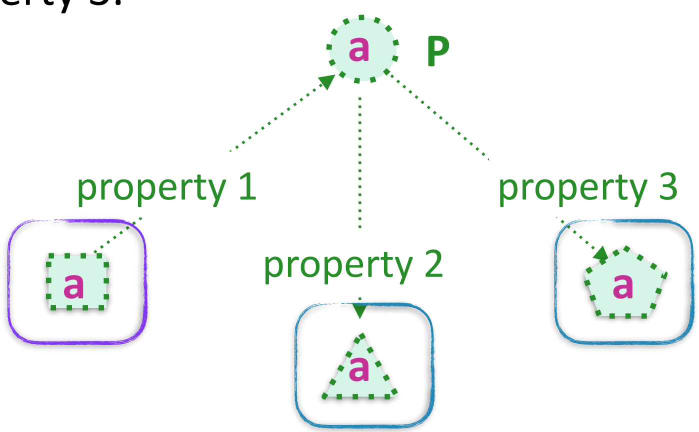


A parameter determines what we predict will be observed.

Head-directionality

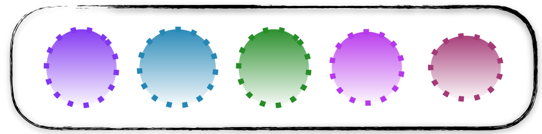
good for acquisition

Note: Property 1 has served as indirect positive evidence for properties 2 and 3. Data about property 1 appearing in the child's input have allowed her to infer things about property 2 and property 3.



Parameters & overhypotheses

grammar

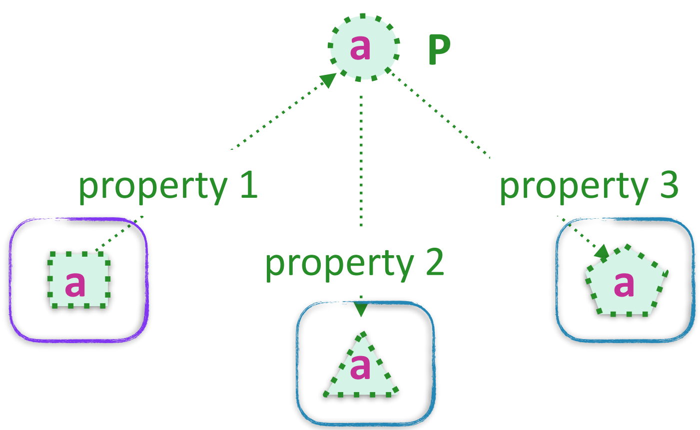


A parameter determines what we predict will be observed.

Head-directionality

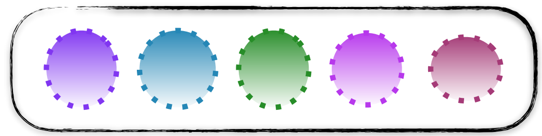
good for acquisition

This highlights another benefit - we don't have to learn the behavior of each structure individually.



Parameters & overhypotheses

grammar

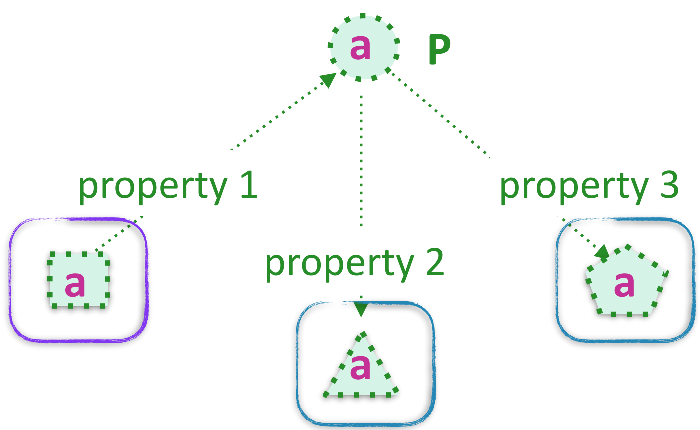


A parameter determines what we predict will be observed.

Head-directionality

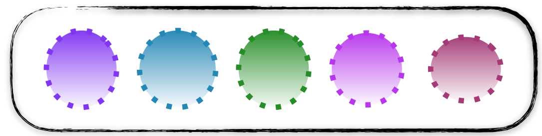
good for acquisition

Instead, we can observe some properties (like property 1) and infer the right behavior for the remaining properties (like property 2 and property 3).



Parameters & overhypotheses

grammar

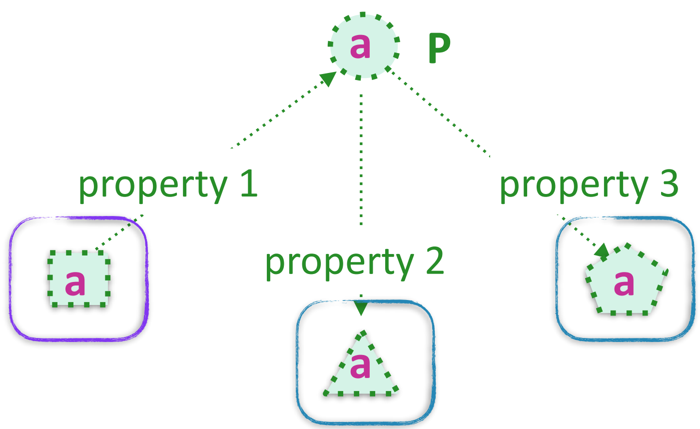


A parameter determines what we predict will be observed.

Head-directionality

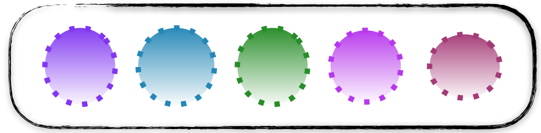
good for acquisition

That is, instead of having to make 3 decisions (one for properties 1, 2, and 3), we actually only need to make one decision - is P a or b?



Parameters & overhypotheses

grammar

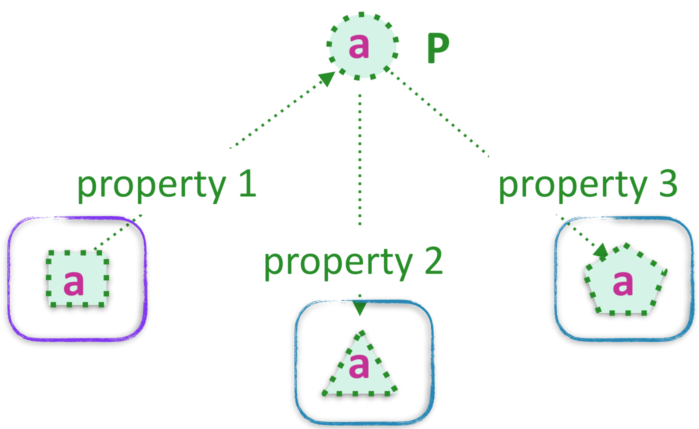


A parameter determines what we predict will be observed.

Head-directionality

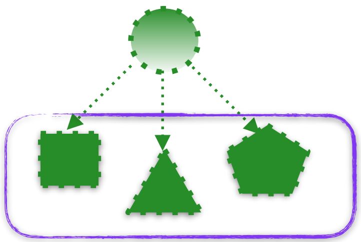
good for acquisition

The intake is used to make this one decision, which generates useful predictions for other properties of the language.



Parameters & overhypotheses

linguistic parameter

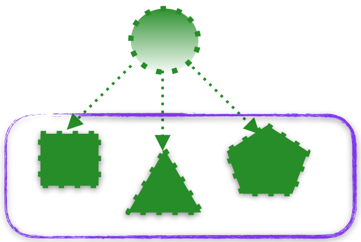


Overhypotheses in hierarchical Bayesian learning are generalizations made at a more abstract level, which cover many different data types.

In this way, they're similar in spirit to linguistic parameters.

Parameters & overhypotheses

linguistic parameter



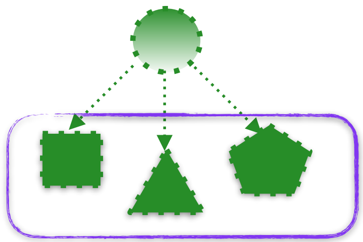
Overhypotheses

Non-linguistic example

Suppose you're observing the contents of marble bags.

Parameters & overhypotheses

linguistic parameter



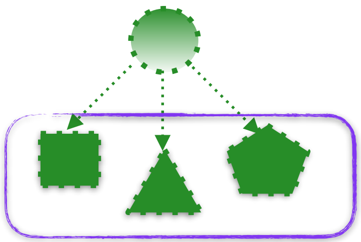
Overhypotheses

Non-linguistic example

The first bag you look at has 20 black marbles.

Parameters & overhypotheses

linguistic parameter



Overhypotheses

Non-linguistic example

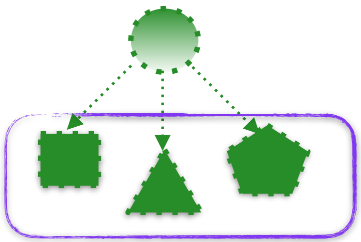
The second bag you look at has 20 white marbles.

20

20

Parameters & overhypotheses

linguistic parameter



Overhypotheses

Non-linguistic example

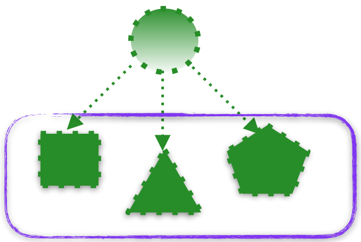
20  20 

Parameters & overhypotheses

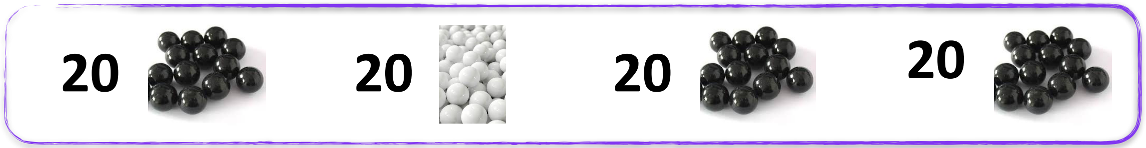
Overhypotheses

Non-linguistic example

linguistic parameter

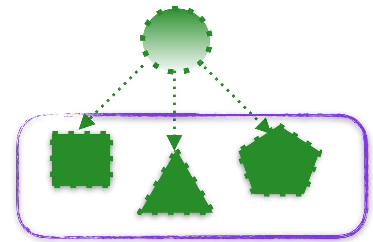


The third and fourth bags you look at have 20 black marbles.



Parameters & overhypotheses

linguistic parameter



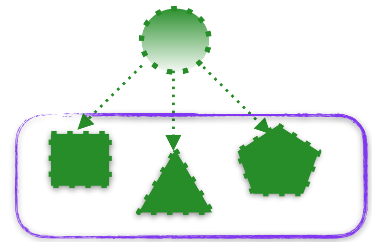
Overhypotheses

Non-linguistic example

You get a fifth bag and pull out a single marble. It's white.

Parameters & overhypotheses

linguistic parameter



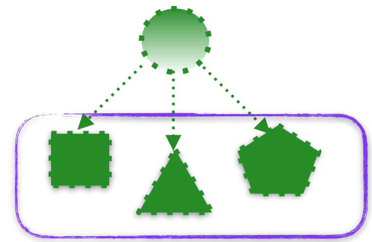
Overhypotheses

Non-linguistic example

What do you predict about the color distribution of the rest of the marbles in the bag?

Parameters & overhypotheses

linguistic parameter



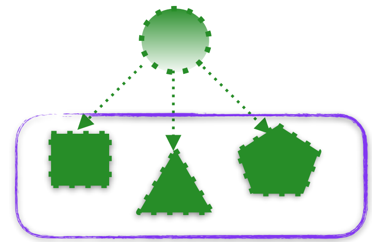
Overhypotheses

Non-linguistic example

Probably that they're all white!

Parameters & overhypotheses

linguistic parameter



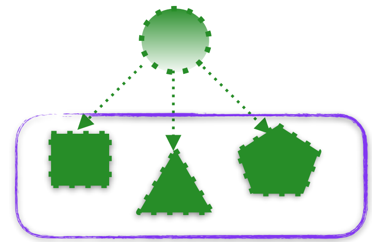
Overhypotheses

Non-linguistic example

What if you then get another bag and pull out a single purple marble from it? What would you predict?

Parameters & overhypotheses

linguistic parameter



Overhypotheses

Non-linguistic example

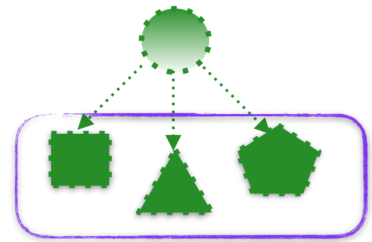
Probably that all the rest of the marbles in the bag are purple, too!

1 

20  **20**  **20**  **20** 

Parameters & overhypotheses

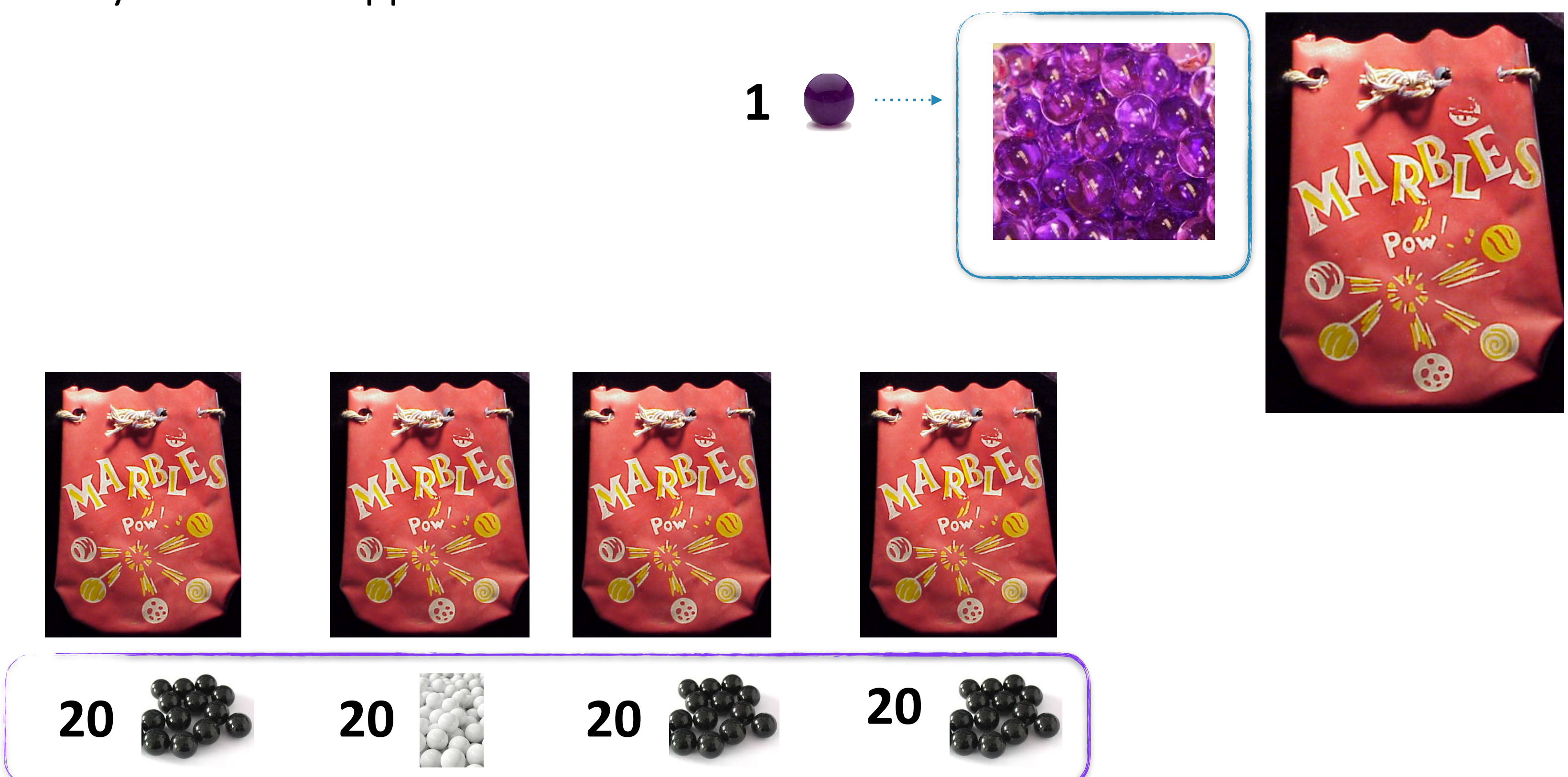
linguistic parameter



Overhypotheses

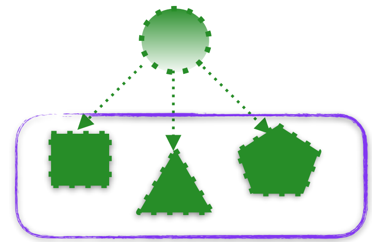
Non-linguistic example

Why does this happen?



Parameters & overhypotheses

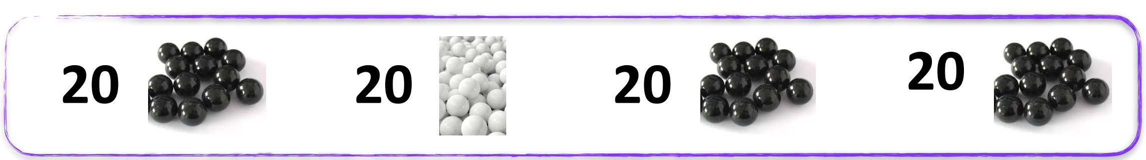
linguistic parameter



Overhypotheses

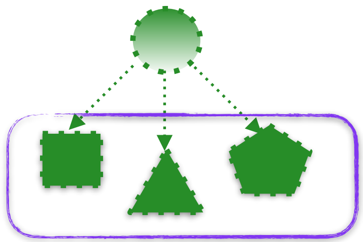
Non-linguistic example

It seems like you're learning something about the color distribution *in general* (not just for a particular bag): all marbles in a bag have the same color.



Parameters & overhypotheses

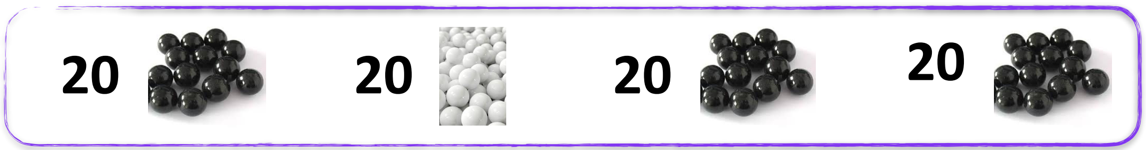
linguistic parameter



Overhypotheses

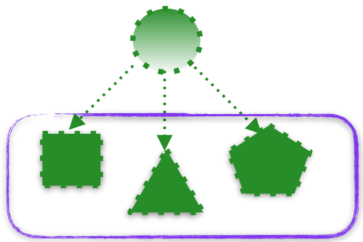
Non-linguistic example

This allows you to make predictions when you've only seen a single marble of whatever color from a bag.



Parameters & overhypotheses

linguistic parameter



Overhypotheses

Non-linguistic example

overhypothesis

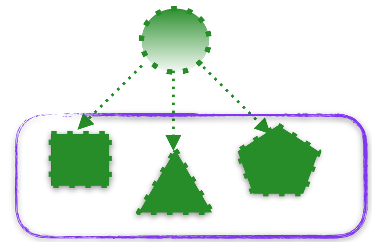
all the same color

all black | all white | all black | all black

20  20  20  20 

Parameters & overhypotheses

linguistic parameter

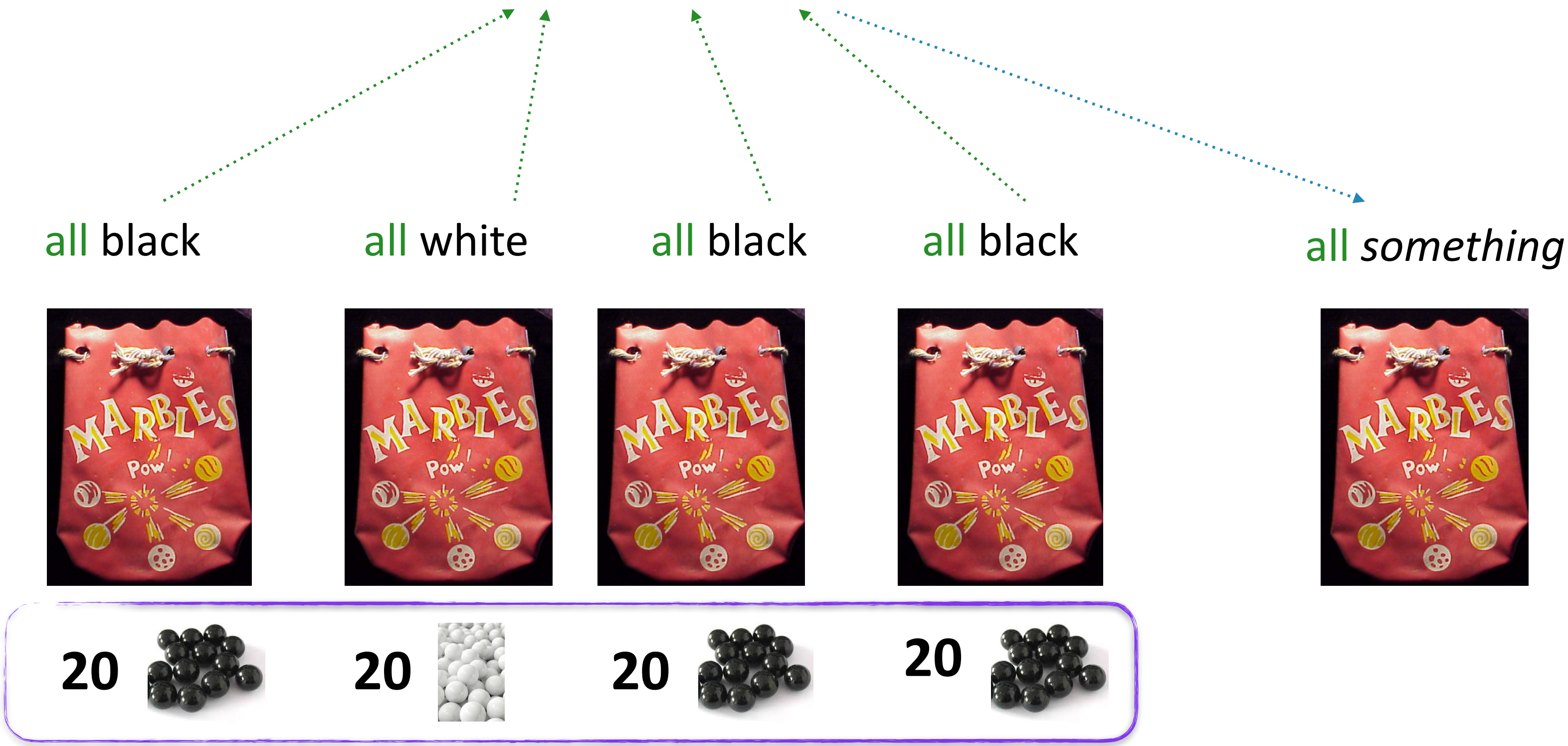


Overhypotheses

Non-linguistic example

overhypothesis

all the same color

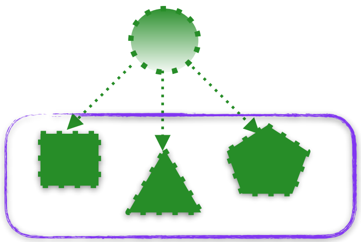


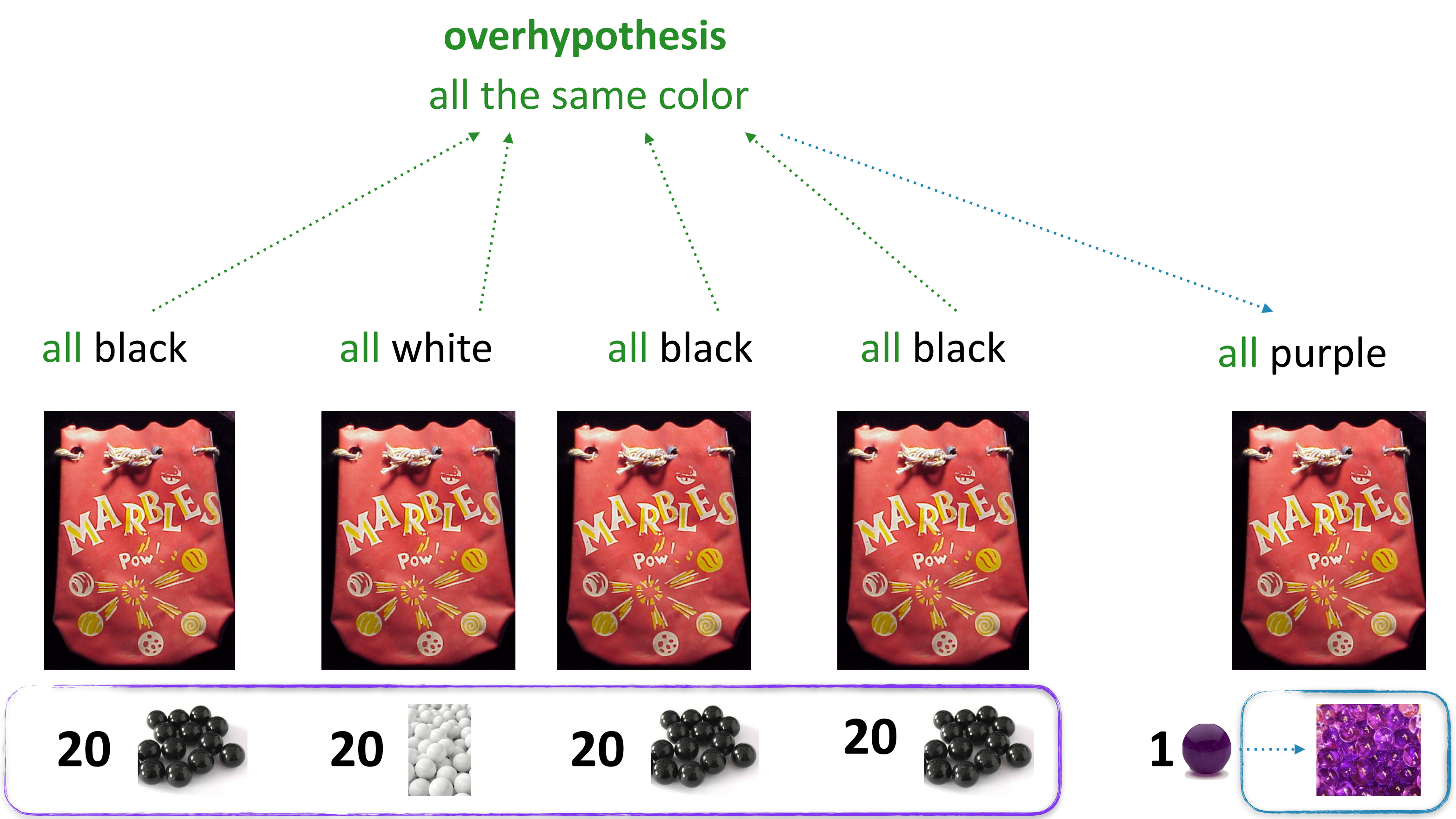
Parameters & overhypotheses

Overhypotheses

Non-linguistic example

linguistic parameter



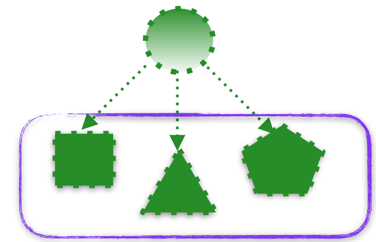


Parameters & overhypotheses

Overhypotheses

Non-linguistic example

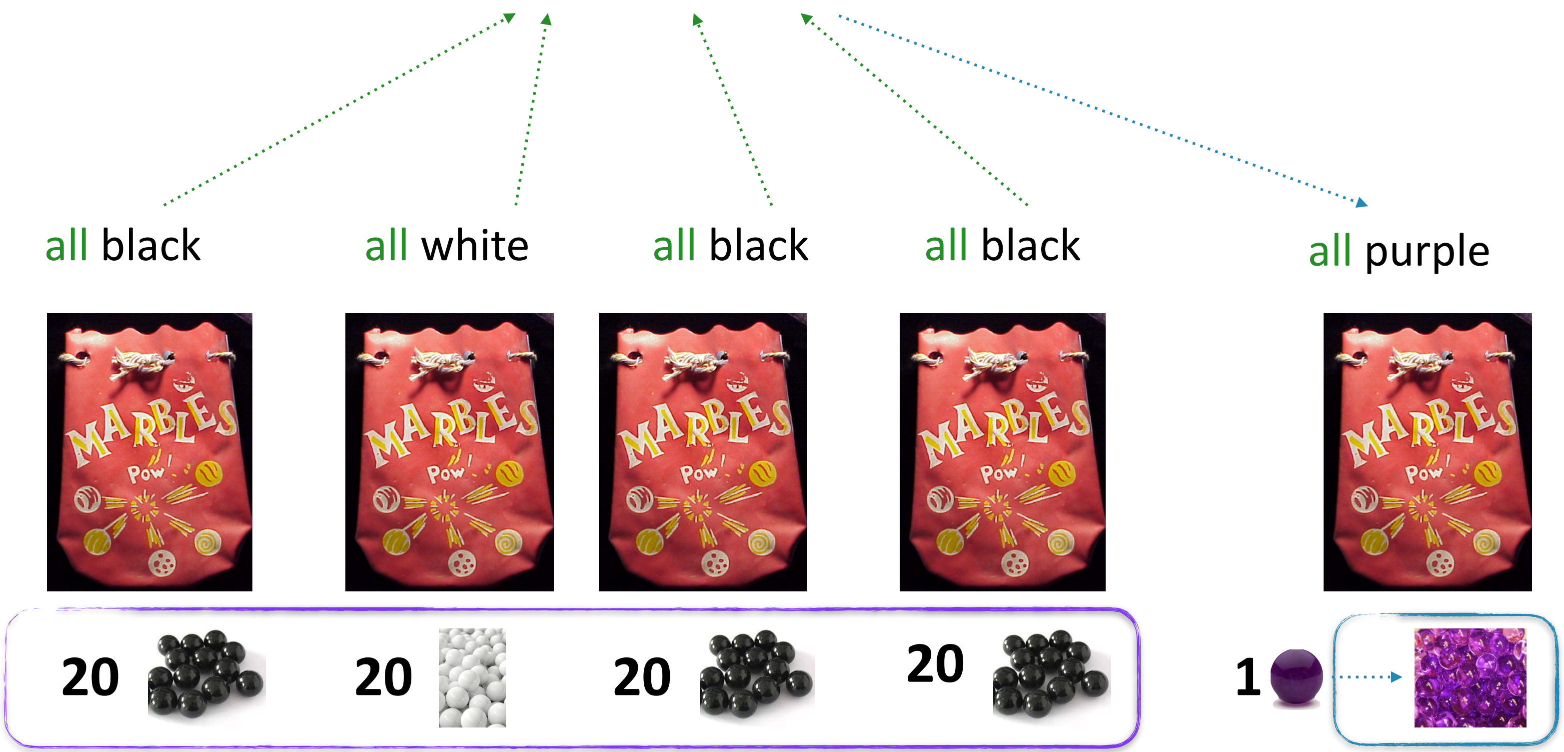
linguistic parameter



Seem familiar?

overhypothesis

all the same color



Parameters & overhypotheses

linguistic parameter

Overhypotheses

Non-linguistic example

Seem familiar?

overhypothesis

all the same color

all black

all white

all black

all black

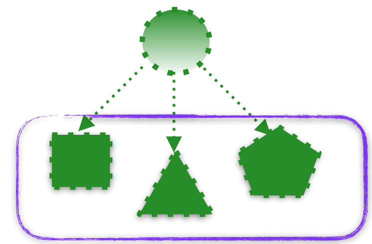
all purple

20 20 20 20

1

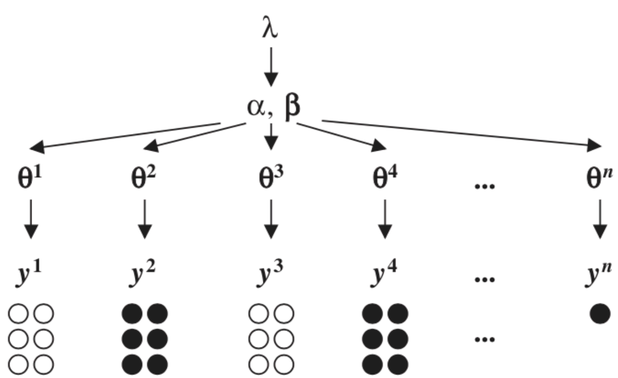
Parameters & overhypotheses

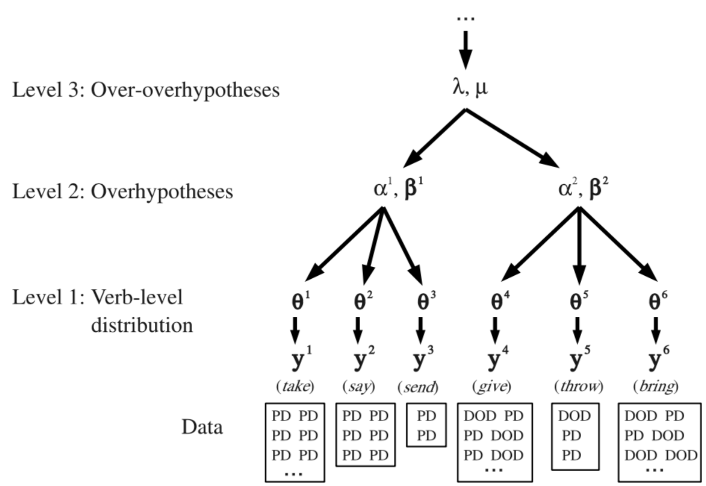
linguistic parameter
overhypothesis



Bayesian learning models are able to learn overhypotheses, provided they know **what the parameters are** and **the range of values those parameters can take**.

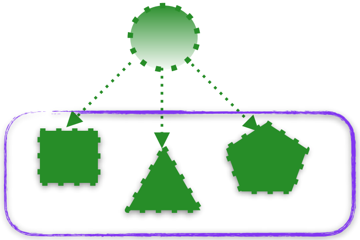
(ex: Kemp, Perfors, & Tenenbaum 2007, Perfors, Tenebaum, & Wonnacott 2010).





Parameters & overhypotheses

linguistic parameter
overhypothesis

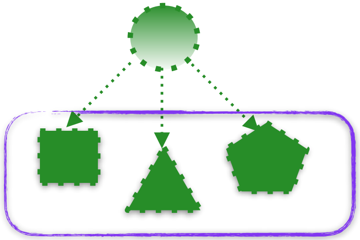


Bayesian learning models are able to learn overhypotheses, provided they know **what the parameters are** and **the range of values those parameters can take**.

What about real learners (children)?

Parameters & overhypotheses

linguistic parameter
overhypothesis



Dewar & Xu 2010
9-month-olds

When provided with partial evidence about a few objects in a few categories, can infants form a more abstract generalization (an overhypothesis) that then applies to a new category?

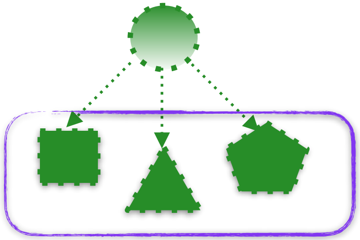
Parameters & overhypotheses

Dewar & Xu 2010

9-month-olds

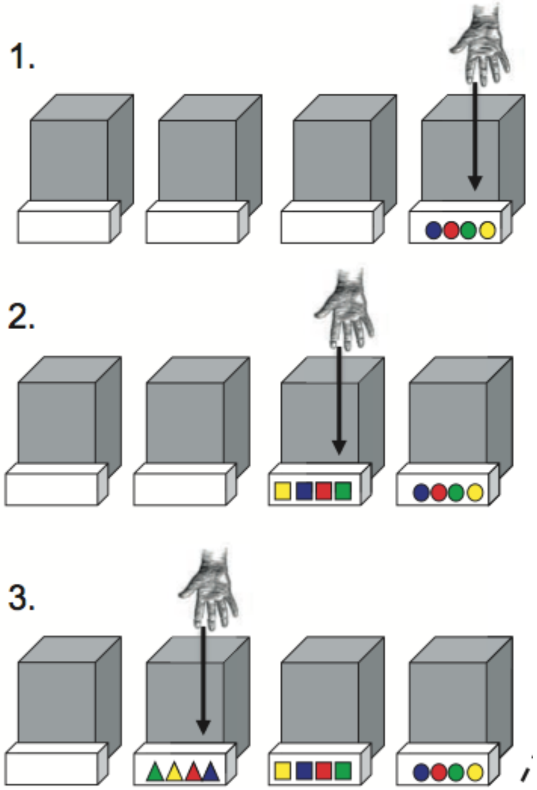
linguistic parameter

overhypothesis



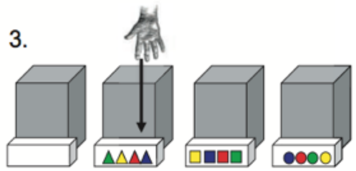
Training trials:

Observe four different objects pulled out by experimenter who had her eyes closed - the objects are **different colors but always have the same shape**.



Parameters & overhypotheses

Dewar & Xu 2010
9-month-olds

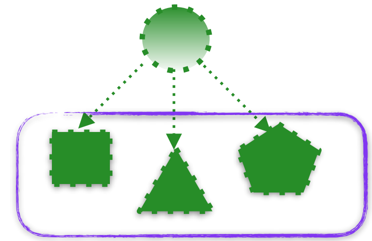


Training: **different colors but same shape**

Experimental condition

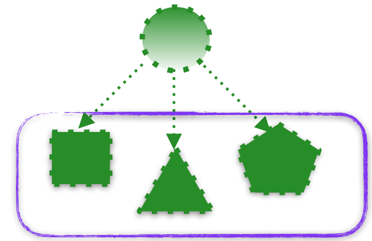
If infants create an overhypothesis that all objects in a box have the same shape…

linguistic parameter
overhypothesis

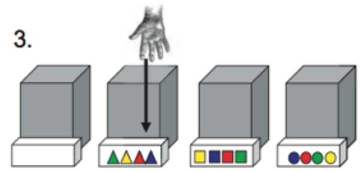


Parameters & overhypotheses

linguistic parameter
overhypothesis



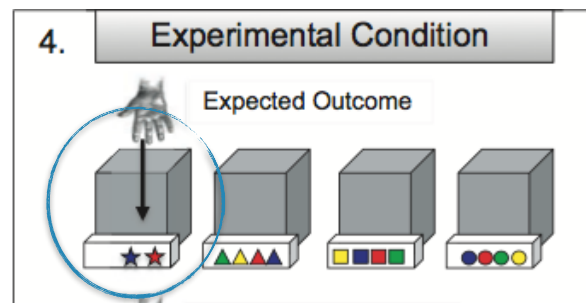
Dewar & Xu 2010
9-month-olds



Training: different colors but same shape

Experimental condition

If infants create an overhypothesis that all objects in a box have the same shape…

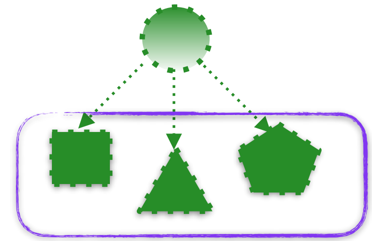


they should expect the experimenter to pull out all the same shape from a new box.

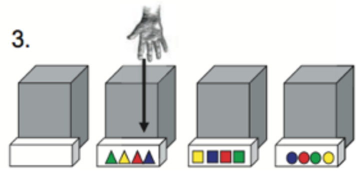
This shouldn't be surprising, and so infants shouldn't look as long at it.

Parameters & overhypotheses

linguistic parameter
overhypothesis



Dewar & Xu 2010
9-month-olds

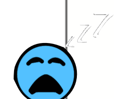


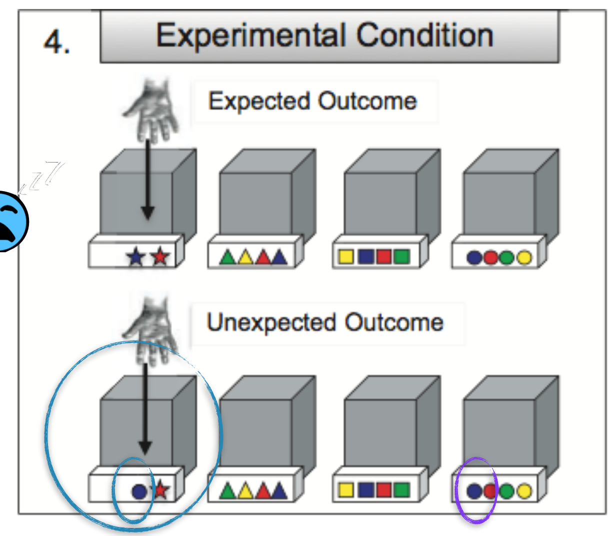
Training: **different colors but same shape**

Experimental condition

If infants create an overhypothesis that all objects in a box have the same shape…

they shouldn't expect the experimenter to pull out different shapes from a new box, even if one is a shape they've seen before.

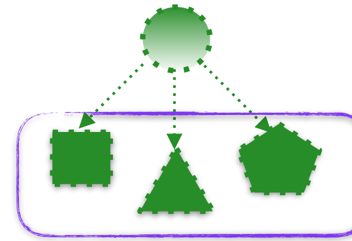




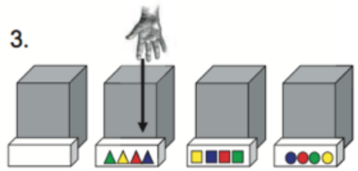
This should be surprising, and so infants should look longer at it.

Parameters & overhypotheses

linguistic parameter
overhypothesis

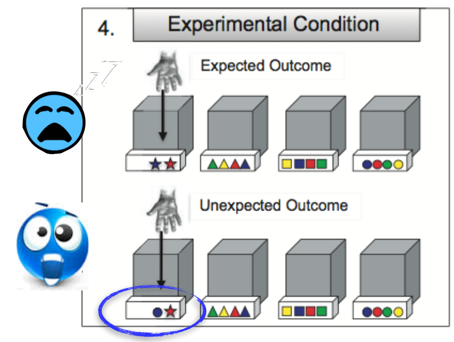


Dewar & Xu 2010
9-month-olds



Training: different colors but same shape

Experimental condition

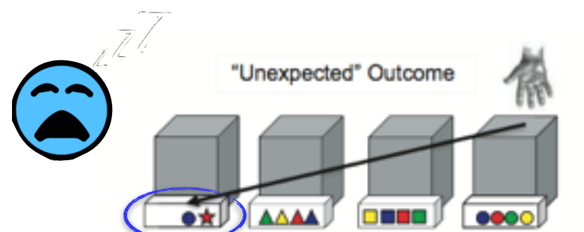


Control condition

If infants create an overhypothesis that all objects in a box have the same shape…

they should expect the experimenter to pull out different shapes from different boxes.

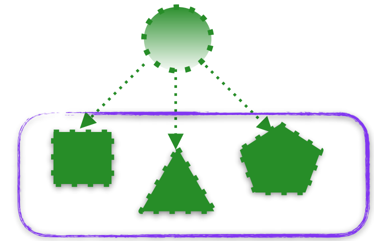
This shouldn't be surprising, and so infants shouldn't look as long at it.



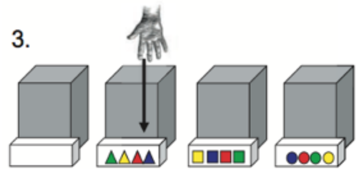
Note how this outcome looks identical to the experimental condition outcome.

Parameters & overhypotheses

linguistic parameter
overhypothesis

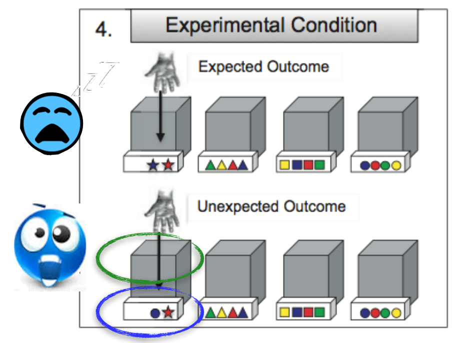


Dewar & Xu 2010
9-month-olds



Training: **different colors but same shape**

Experimental condition

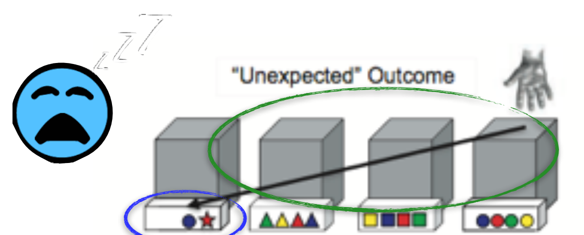


Control condition

If infants create an overhypothesis that all objects in a box have the same shape…

they should expect the experimenter to pull out different shapes from different boxes.

This shouldn't be surprising, and so infants shouldn't look as long at it.



The only difference is how the outcome was **generated** (from the same box or from different boxes — which is what the overhypothesis is about).

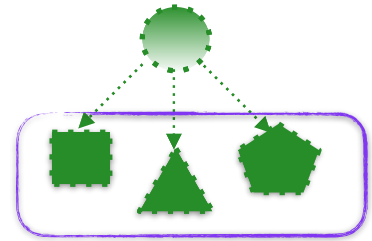
Parameters & overhypotheses

Dewar & Xu 2010

9-month-olds

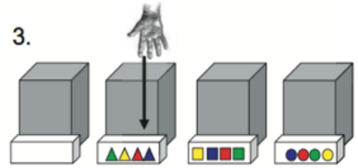
linguistic parameter

overhypothesis



Training:

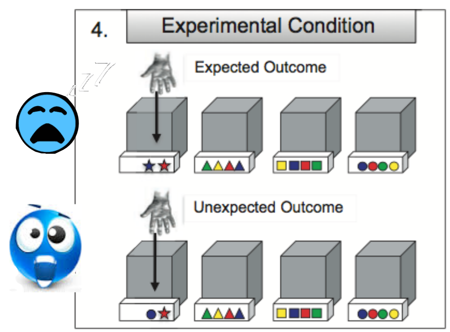
different colors but same shape



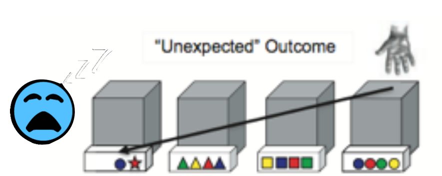
If infants create an overhypothesis that all objects in a box have the same shape

Experimental condition

This is what we expect.

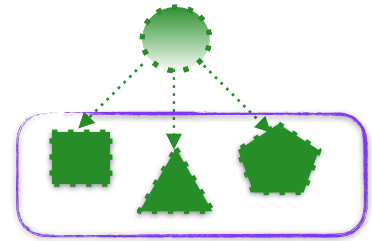


Control condition



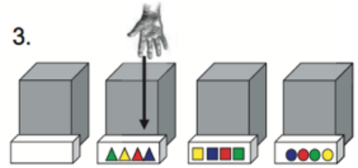
Parameters & overhypotheses

linguistic parameter
overhypothesis



Dewar & Xu 2010
9-month-olds

Training:
different colors but same shape

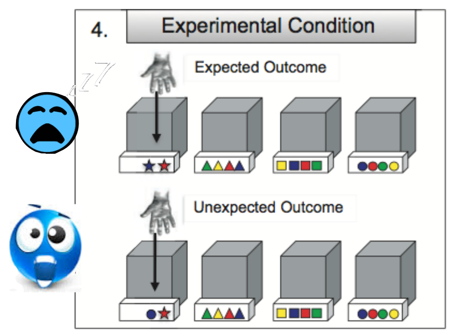


If infants create an overhypothesis that all objects in a box have the same shape

Experimental condition

~11.32 sec

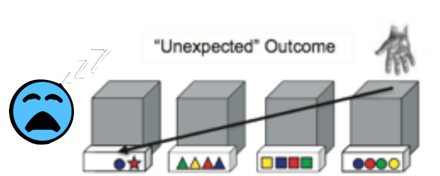
~14.28 sec



And this is exactly what happened!

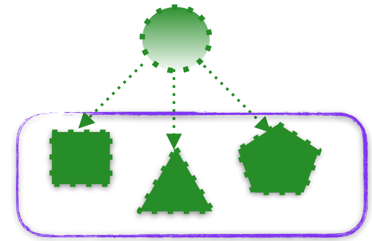
Control condition

~10.3-11.0 sec



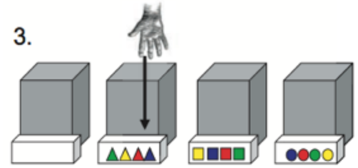
Parameters & overhypotheses

linguistic parameter
overhypothesis



Dewar & Xu 2010
9-month-olds

Training:
different colors but same shape

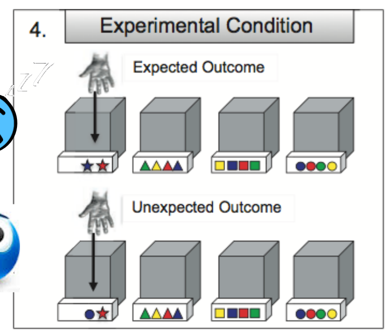


If infants create an overhypothesis that all objects in a box have the same shape

Experimental condition

~11.32 sec 

~14.28 sec 



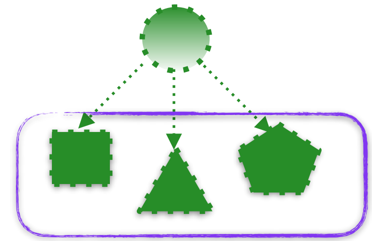
9-month-olds appear able to form overhypotheses from very limited data sets.

Control condition

~10.3-11.0 sec 

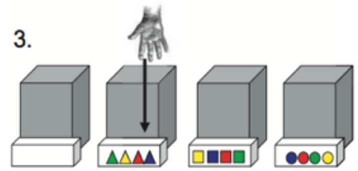
Parameters & overhypotheses

linguistic parameter
overhypothesis



Dewar & Xu 2010
9-month-olds

Training:
different colors but same shape

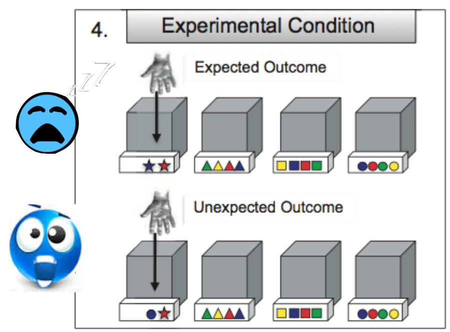


If infants create an overhypothesis that all objects in a box have the same shape

Experimental condition

~11.32 sec

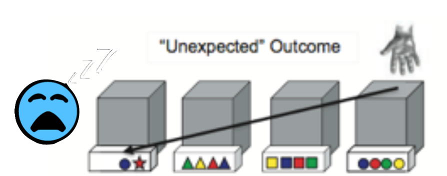
~14.28 sec



Hopefully, this means they can also use linguistic parameters to learn, since parameters are similar to overhypotheses about language!

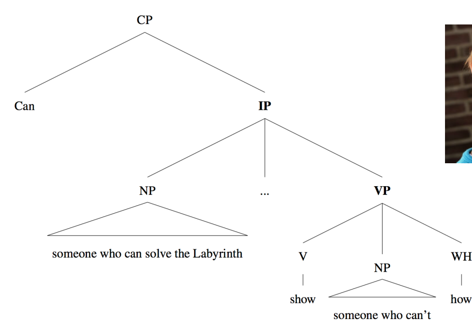
Control condition

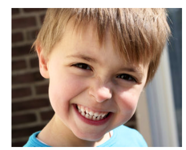
~10.3-11.0 sec

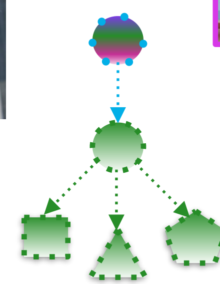


Parameters & overhypotheses

Structure dependence

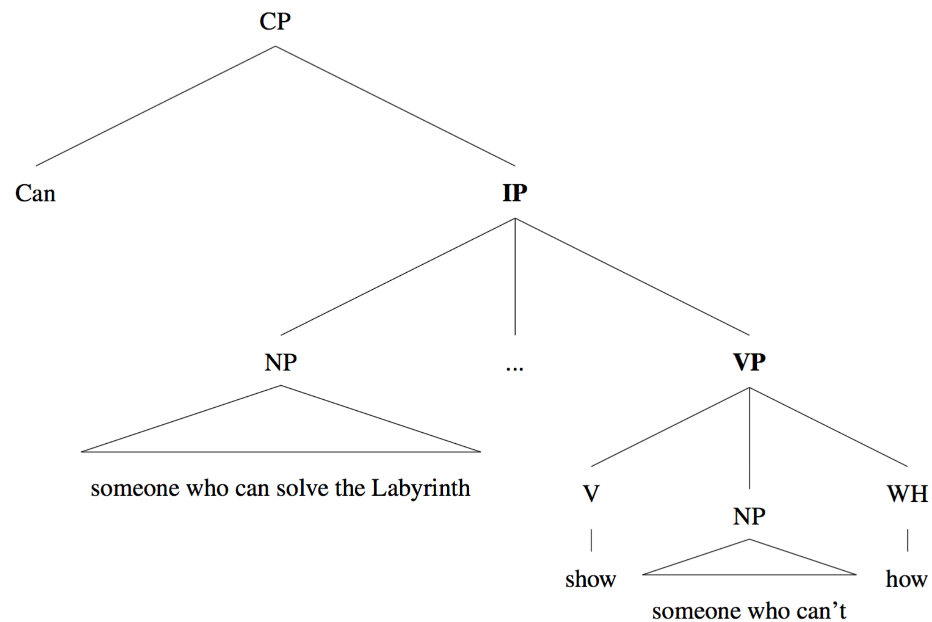






Structure dependence

Idea: Rules for word order **depend on linguistic structure**



Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

How could they learn this?

Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue

Most of the yes/no question data children encounter (particularly before the age of 3) consists of simple yes/no questions compatible with many different rules.

Jareth can alter time.

Can Jareth alter time?

Rule? Swap the order of the first two words

Rule? Swap the order of the subject and the auxiliary

Rule? Move the first noun to the second position

Rule? Move the auxiliary to the first position

Rule? Move the main clause auxiliary to the first position

Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue

Most of the yes/no question data children encounter (particularly before the age of 3) consists of simple yes/no questions compatible with many different rules.

Jareth can alter time.

Can Jareth alter time?

But structure-dependence is a very ***general* property** about language…

Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue

Most of the yes/no question data children encounter (particularly before the age of 3) consists of simple yes/no questions compatible with many different rules.

Jareth can alter time.

Can Jareth alter time?

It could be an **overhypothesis** about language.

Structure dependence

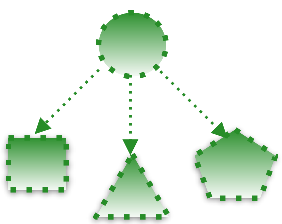
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue

Most of the yes/no question data children encounter (particularly before the age of 3) consists of simple yes/no questions compatible with many different rules.



Jareth can alter time.

Can Jareth alter time?

The girl in the Labyrinth thought something.

What did the girl in the Labyrinth think?

And this overhypothesis would connect to many other structures besides yes/no questions.

Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

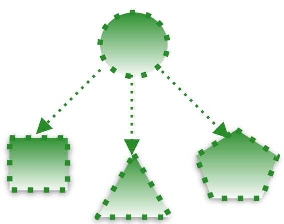
By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue

Most of the yes/no question data children encounter (particularly before the age of 3) consists of simple yes/no questions compatible with many different rules.

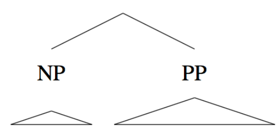
Jareth can alter time.

Can Jareth alter time?



The girl in the Labyrinth thought something.

What did the girl in the Labyrinth think?



The girl in the Labyrinth

And this overhypothesis would connect to many other structures besides yes/no questions.

Structure dependence

Rules for word order **depend on linguistic structure**

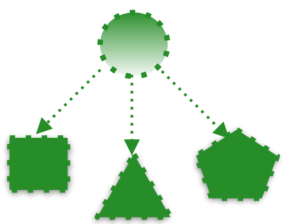
Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

A potential input issue - may not be as bad

Children could encounter a lot of data that might favor structured representations over unstructured ones (e.g., linear structures)

overhypothesis

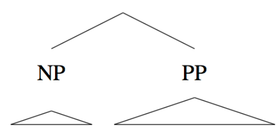


Jareth can alter time.

Can Jareth alter time?

The girl in the Labyrinth thought something.

What did the girl in the Labyrinth think?



The girl in the Labyrinth

Structure dependence

Rules for word order **depend on linguistic structure**

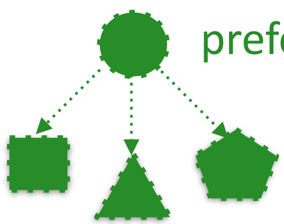
Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

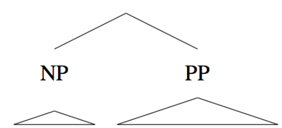
A potential input issue - may not be as bad

Children could encounter a lot of data that might favor structured representations over unstructured ones (e.g., linear structures)

overhypothesis



prefer structured representations



Jareth can alter time.

Can Jareth alter time?

The girl in the Labyrinth thought something.

What did the girl in the Labyrinth think?

The girl in the Labyrinth

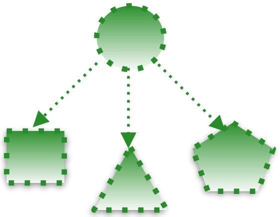
Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011



computational-level modeled learner

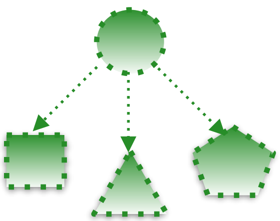
Structure dependence

Rules for word order **depend on linguistic structure**

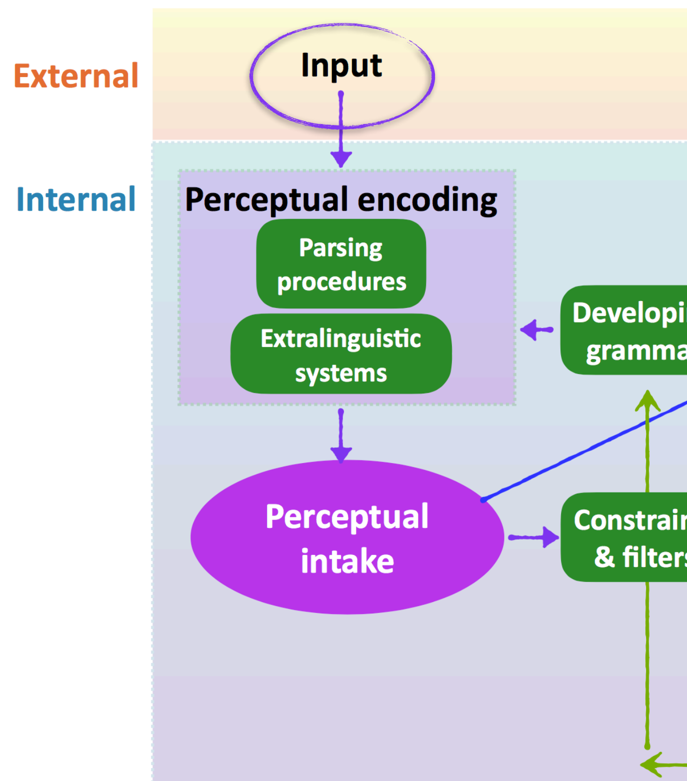
Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011



Learned from realistic samples of child-directed English speech



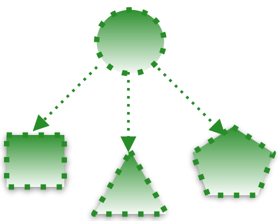
Structure dependence

Rules for word order **depend on linguistic structure**

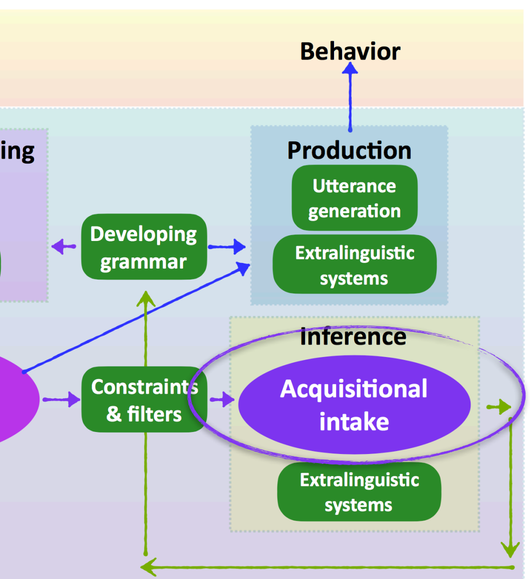
Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011



Learned from realistic samples of child-directed English speech abstracted into syntactic category sequences



Structure dependence

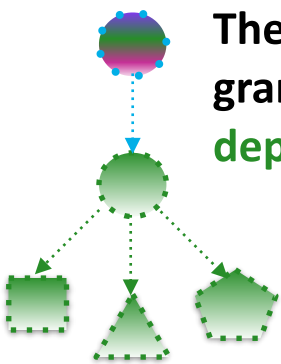
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

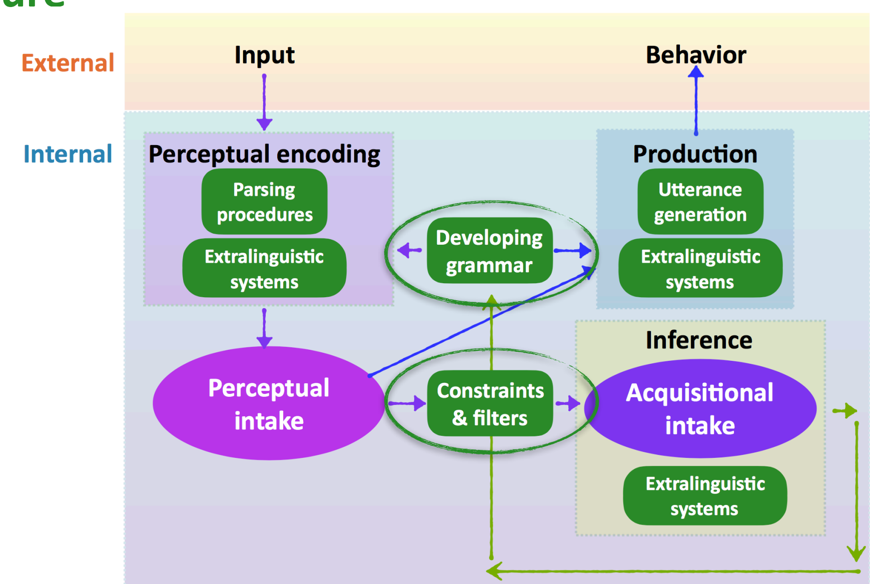
By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011

Hypotheses



There are different types of grammars available (e.g., structure-dependent vs. linear)



Structure dependence

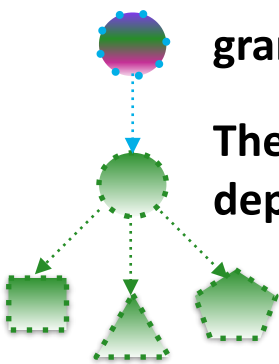
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

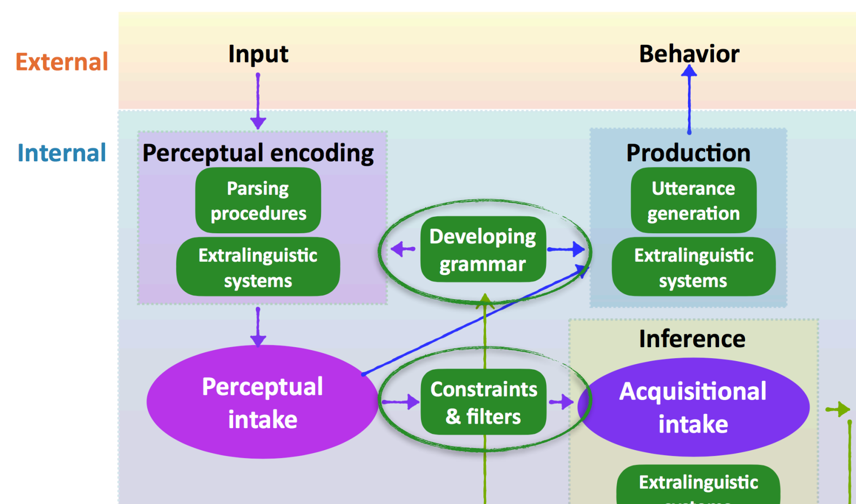
Perfors, Tenenbaum, & Regier 2011

Hypotheses



grammar type

There are specific grammars of each type (e.g., different structure-dependent grammars)



Structure dependence

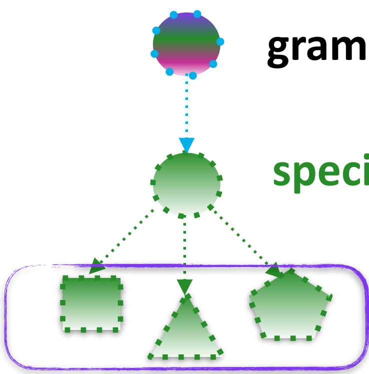
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011

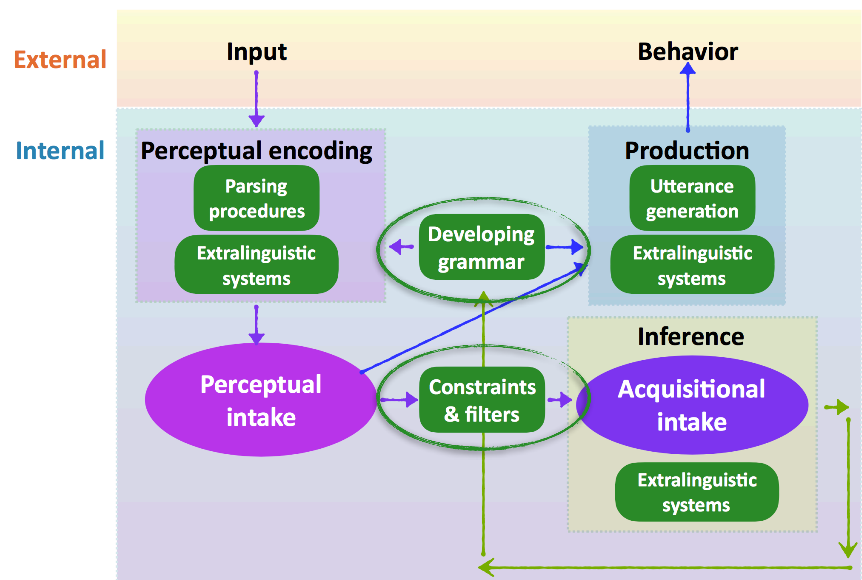
Hypotheses



grammar type

specific grammar

Each grammar connects to specific structures in the observable data



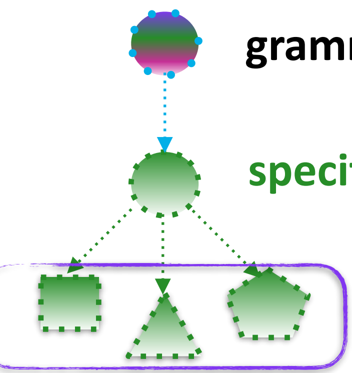
Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

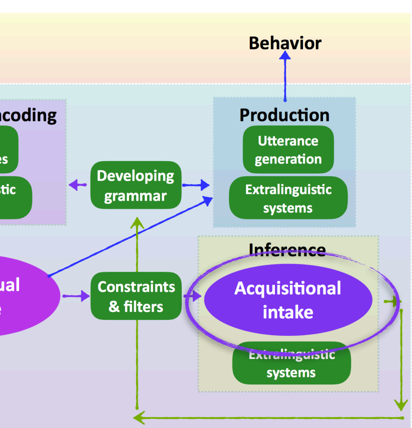
Perfors, Tenenbaum, & Regier 2011



grammar type

specific grammar

structures in observable data



Use Bayesian inference to infer the best grammar type & specific grammar, given the child-directed speech data.

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

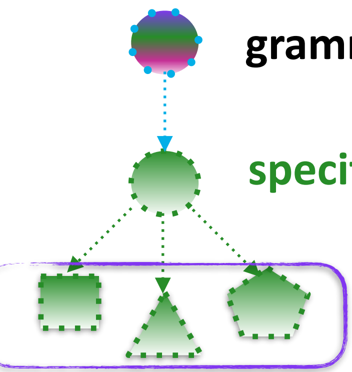
Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011

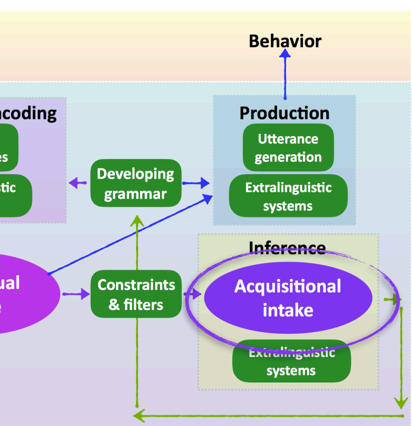


grammar type

specific grammar

structures in observable data

Note: The priors for different grammars aren't equal. Structure-dependent grammars are more complex than other grammar types being considered, and so have lower prior probability.



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

This means structure-dependent grammars are actually *disfavored* a priori!

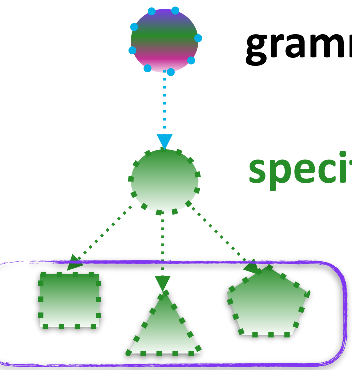
Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

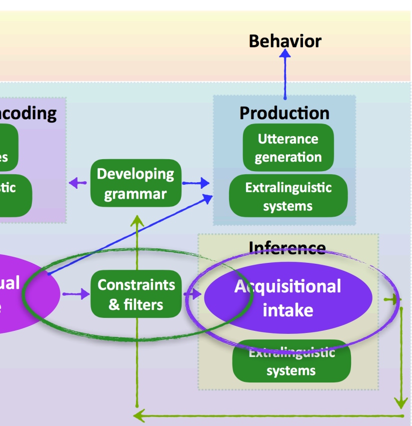
Perfors, Tenenbaum, & Regier 2011



grammar type

specific grammar

structures in observable data



Note: The priors for different grammars aren't equal. Structure-dependent grammars are more complex than other grammar types being considered, and so have lower prior probability.

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

This means they really have to do a better job accounting for the data to be preferred!

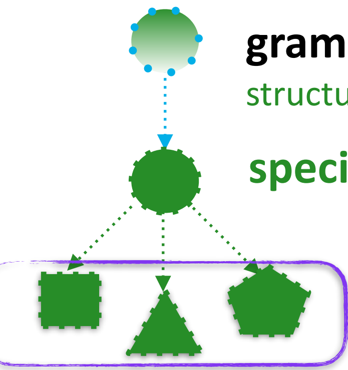
Structure dependence

Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Perfors, Tenenbaum, & Regier 2011

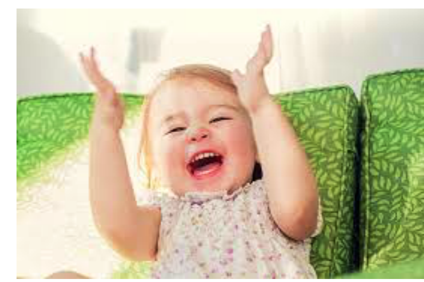


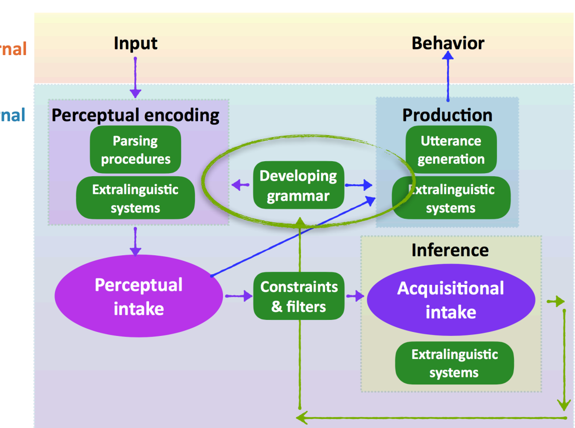
grammar type
structure-dependent

specific grammar

structures in observable data

And this is exactly what happens!





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

Structure dependence

Rules for word order **depend on linguistic structure**

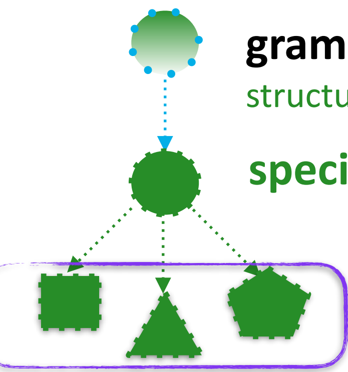
Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

Pronoun Verb Noun

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

Perfors, Tenenbaum, & Regier 2011

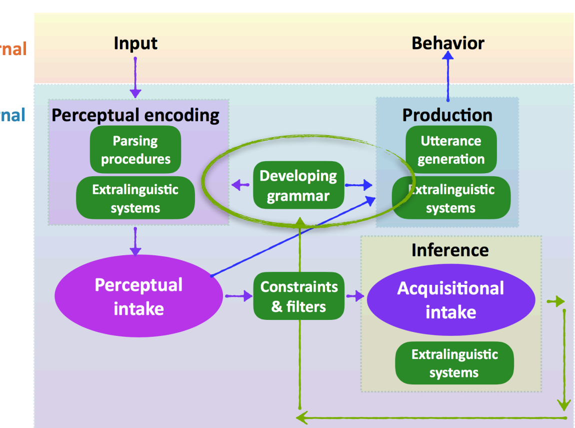


grammar type

structure-dependent

specific grammar

structures in observable data



Even for the earliest child-directed speech samples (directed at children **two years old**), the structure-dependent grammar types are preferred.

Structure dependence

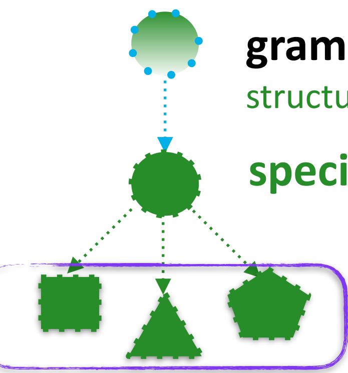
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

By three years old, children have some very specific constraints on hypotheses about word order.

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

Perfors, Tenenbaum, & Regier 2011

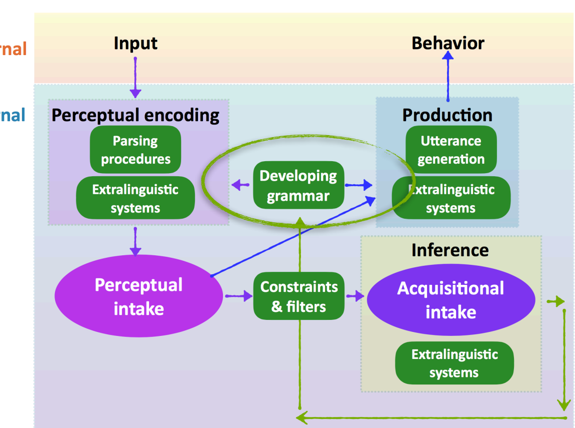


grammar type
structure-dependent

specific grammar

structures in observable data

two years old



Why? Because many different data types favor structure-dependent representations over other simpler representations.

Structure dependence

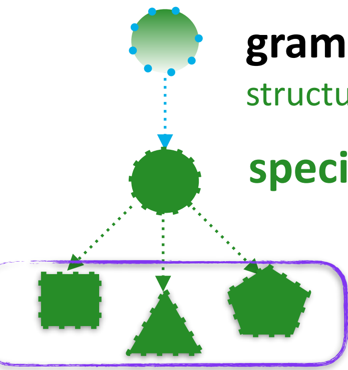
Rules for word order **depend on linguistic structure**

Yes/No question formation in English

Perfors, Tenenbaum, & Regier 2011

Pronoun Verb Noun

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

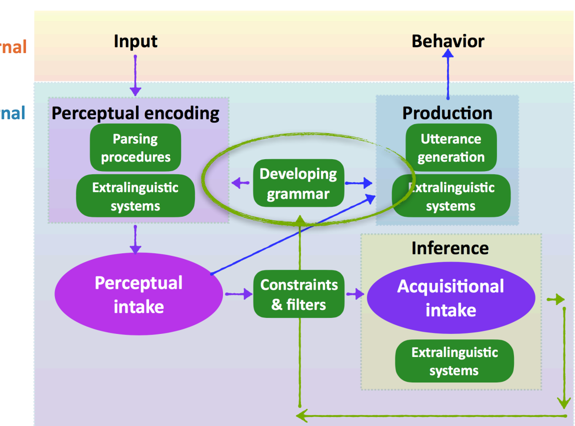


grammar type
structure-dependent

specific grammar

structures in observable data

two years old



By three years old, children have some very specific structure-dependent constraints on hypotheses about word order.

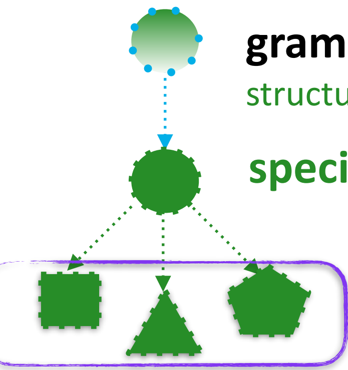
Structure dependence

Rules for word order **depend on linguistic structure**

By three years old, children have some very specific constraints on hypotheses about word order.

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

Perfors, Tenenbaum, & Regier 2011

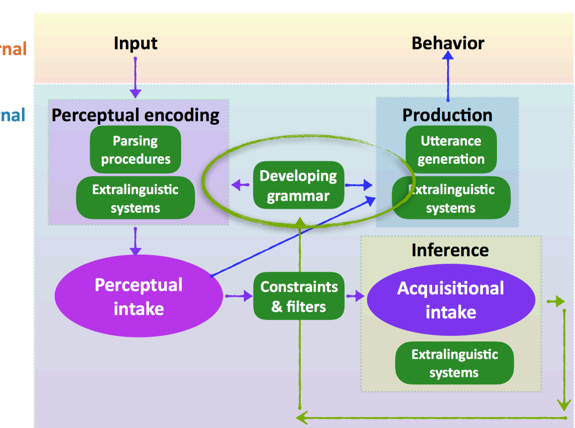


grammar type
structure-dependent

specific grammar

structures in observable data

two years old



Yes/No question formation in English

And so these structure-dependent representations make hypothesizing structure-dependent rules much more probable.

Summary: Linguistic parameters

Parameters make acquisition easier because hard-to-learn structures can be learned by observing easy-to-learn structures that are connected to the same parameters.

Linguistic parameters are similar to statistical parameters in that they are abstractions about the observable data. For linguistic parameters, the observable data are language data.

Parameters may be similar to overhypotheses, which Bayesian learners and 9-month-olds are capable of learning.

An overhypothesis about structure-dependence may not be so hard to learn from the available data for a child using Bayesian inference.

Questions?

You should be able to do up through question 12 on the structure review questions
and up through 3 on HW8.