Computational models of syntactic acquisition

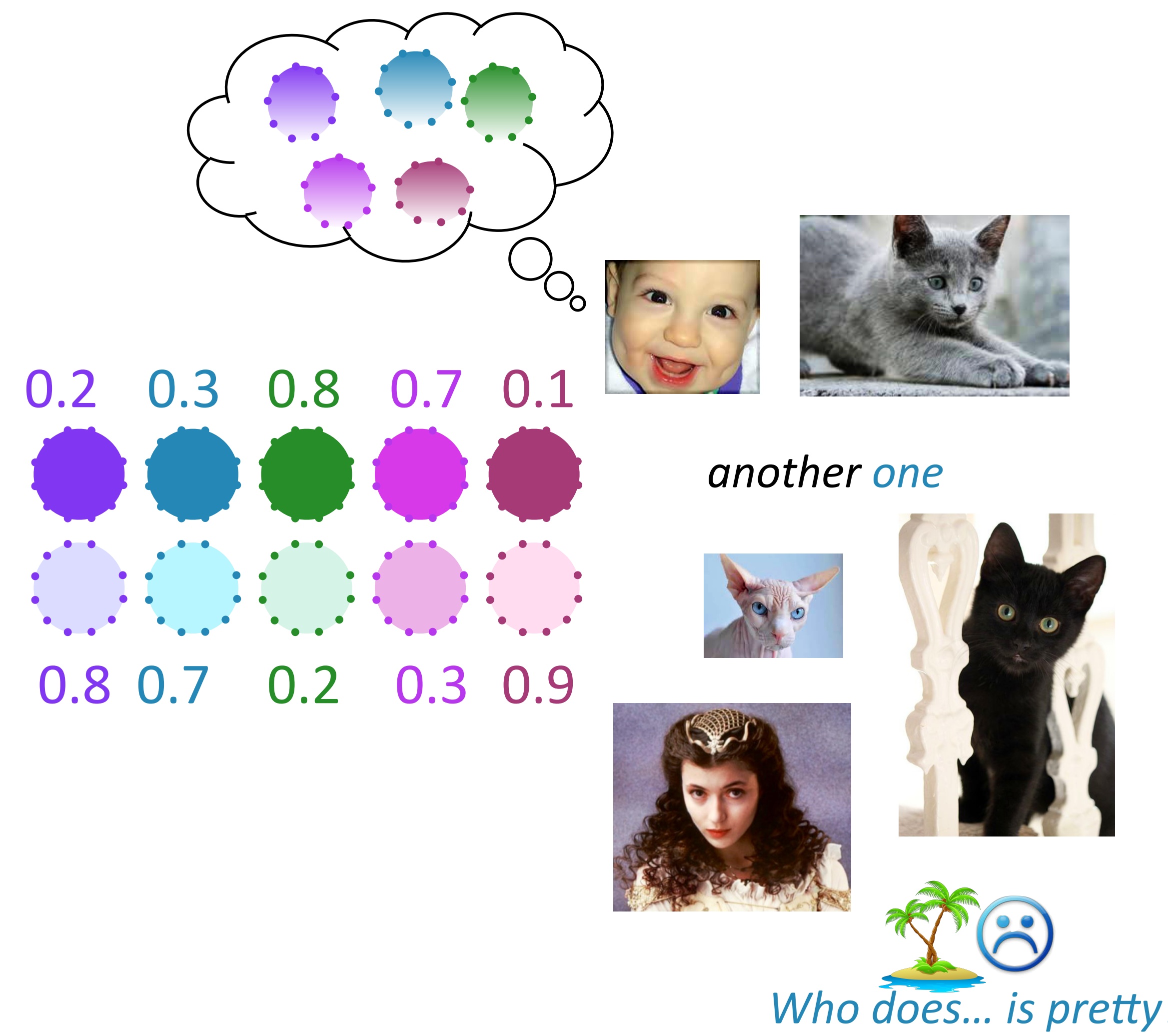
Lisa Pearl

University of California, Irvine

Computation of Language Laboratory

UC Irvine

Lisa S. Pearl
Associate Professor
Department of Linguistics
Department of Cognitive Sciences
SSPB 2219, SBSG 2314
University of California, Irvine
lpearl@uci.edu



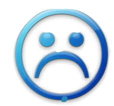
August 4, 2017:

Norwegian Summer Institute on Language & Mind

University of Oslo

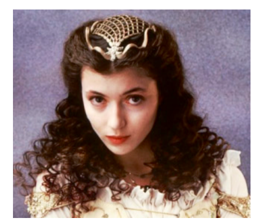
Today's Plan:
Computational models of syntactic acquisition

I. Some non-parametric examples

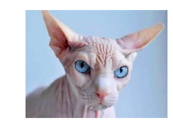
Who does … is pretty?

syntax



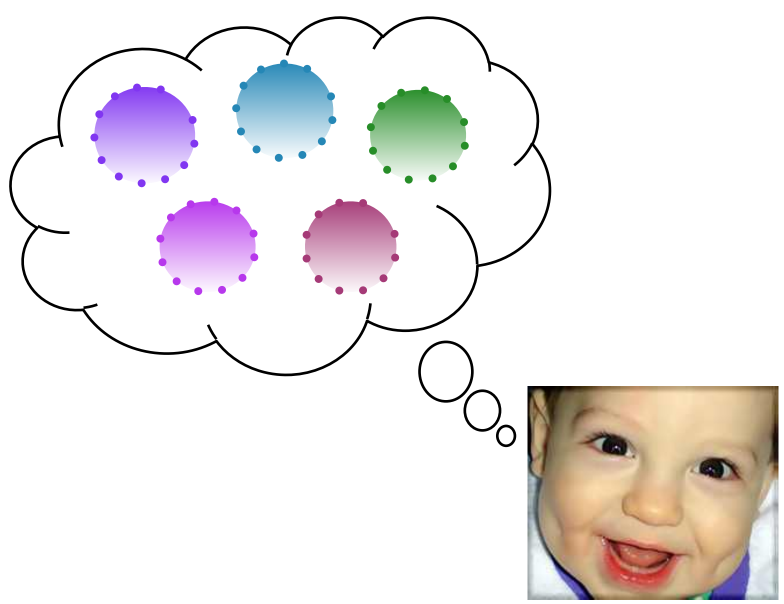
another one



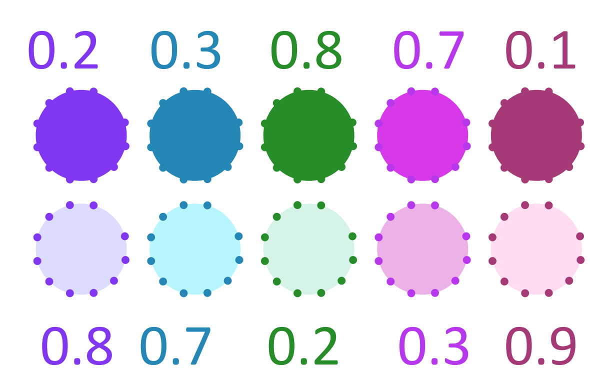


syntax, **semantics**

II. About linguistic parameters



III. Learning with parameters



Today's Plan:
Computational models of syntactic acquisition

I. Some non-parametric examples

Who does … is pretty?

syntax

another one

syntax, semantics

Some non-parametric examples syntax

This kitty was bought as a present for someone.

Lily thinks this kitty is pretty.

What's going on here?

Who does Lily think the kitty for is pretty? 

What does Lily think is pretty, and who does she think it's for? 

Some non-parametric examples $\boxed{\text{syntax}}$

Who does 
Lily think the kitty for is pretty?

What's going on here?

There's a dependency between the *wh*-word *who* and where it's understood (the gap)

Who does Lily think the kitty for ___ is pretty?

This dependency is not allowed in English.

One explanation: The dependency crosses a "syntactic island" (Ross 1967)

Some non-parametric examples syntax

Who does
Lily think the kitty for is pretty?

What's going on here? syntactic island (Ross 1967)

Who does Lily think the kitty for ___ is pretty?

Jack is somewhat tricksy.

He claimed he bought something.

What did Jack make the claim that he bought ___?

Some non-parametric examples syntax

Who does ☹

Lily think the kitty for is pretty?

What's going on here? syntactic island (Ross 1967)

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth wondered if he actually did and what it was.

What did Elizabeth wonder whether Jack bought___ ?

Some non-parametric examples syntax

Who does

Lily think the kitty for is pretty?

What's going on here? syntactic island (Ross 1967)

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth worried it was something dangerous.

What did Elizabeth worry if Jack bought ___ ?

Some non-parametric examples syntax

Who does

Lily think the kitty for is pretty?

What's going on here? syntactic island (Ross 1967)

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

What did Elizabeth worry if Jack bought ___ ?

Jack bought something.

Elizabeth met him afterwards.

What did you meet the pirate who bought ___?

Lily asks Elizabeth about it.

Some non-parametric examples syntax

Who does

Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

What did Elizabeth worry if Jack bought ___ ?

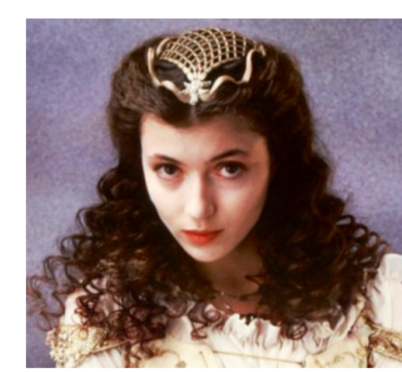
What did you meet the pirate who bought ___?

Jack bought something.

Elizabeth was surprised by it.

What did that Jack bought ___ surprise you ?

Lily asks Elizabeth about it.



Some non-parametric examples syntax

Who does
Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

What did Elizabeth worry if Jack bought ___ ?

What did you meet the pirate who bought ___?

What did that Jack bought ___ surprise you ?

Jack bought two things - a kitty and something else.

Elizabeth wants to know about the other thing.

What did you buy a kitty and ___ ?

Some non-parametric examples syntax

Who does

Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

What did Elizabeth worry if Jack bought ___ ?

What did you meet the pirate who bought ___?

What did that Jack bought ___ surprise you ?

What did you buy a kitty and ___ ?

Jack bought a specific kind of kitty.

Elizabeth wants to know about the kind.

Which did you buy ___ kitty ?

Some non-parametric examples

syntax

Who does
Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___ ?

What did Elizabeth worry if Jack bought ___ ?

What did you meet the pirate who bought ___?

What did that Jack bought ___ surprise you ?

What did you buy a kitty and ___ ?

Which did you buy ___ kitty ?

Important: It's not about the length of the dependency.

(Chomsky 1965, Ross 1967)

Some non-parametric examples syntax

Who does

Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought

What did Elizabeth wonder whether Jack bought

What did Elizabeth worry if Jack bought ___

What did you meet the pirate who bought

What did that Jack bought ___ surprise you

What did you buy a kitty and ___ ?

Which did you buy ___ kitty ?

Elizabeth

What did Elizabeth think ___ ?

✓

It's not about the length of the dependency.

Some non-parametric examples

syntax

Who does
Lily think the kitty for is pretty?

What's going on here? syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought ___ ?

What did Elizabeth wonder whether Jack bought ___

What did Elizabeth worry if Jack bought ___ ?

What did you meet the pirate who bought ___?

What did that Jack bought ___ surprise you?

?

What did you buy a kitty and ___ ?

Which did you buy ___ kitty ?

Jack

Elizabeth

What did Elizabeth think Jack said ___ ?

It's not about the length of the dependency.

Some non-parametric examples $\boxed{\text{syntax}}$

Who does
Lily think the kitty for is pretty?

What's going on here?  syntactic island

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought___ ?

What did Elizabeth wonder whether Jack bought ___

What did Elizabeth worry if Jack bought ___ ?

What did you meet the pirate who bought ___?

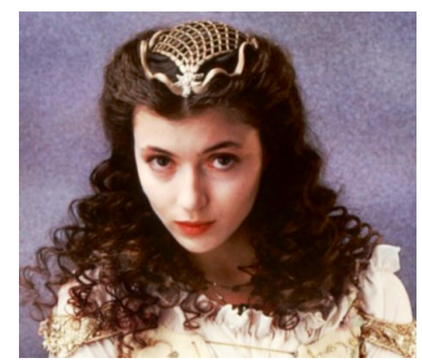
What did that Jack bought ___ surprise you ?

What did you buy a kitty and ___ ?

Which did you buy ___ kitty ?

Jack

Elizabeth



Lily

What did Elizabeth think Jack said Lily saw ___ ?

It's not about the length of the dependency.

Some non-parametric examples

syntax

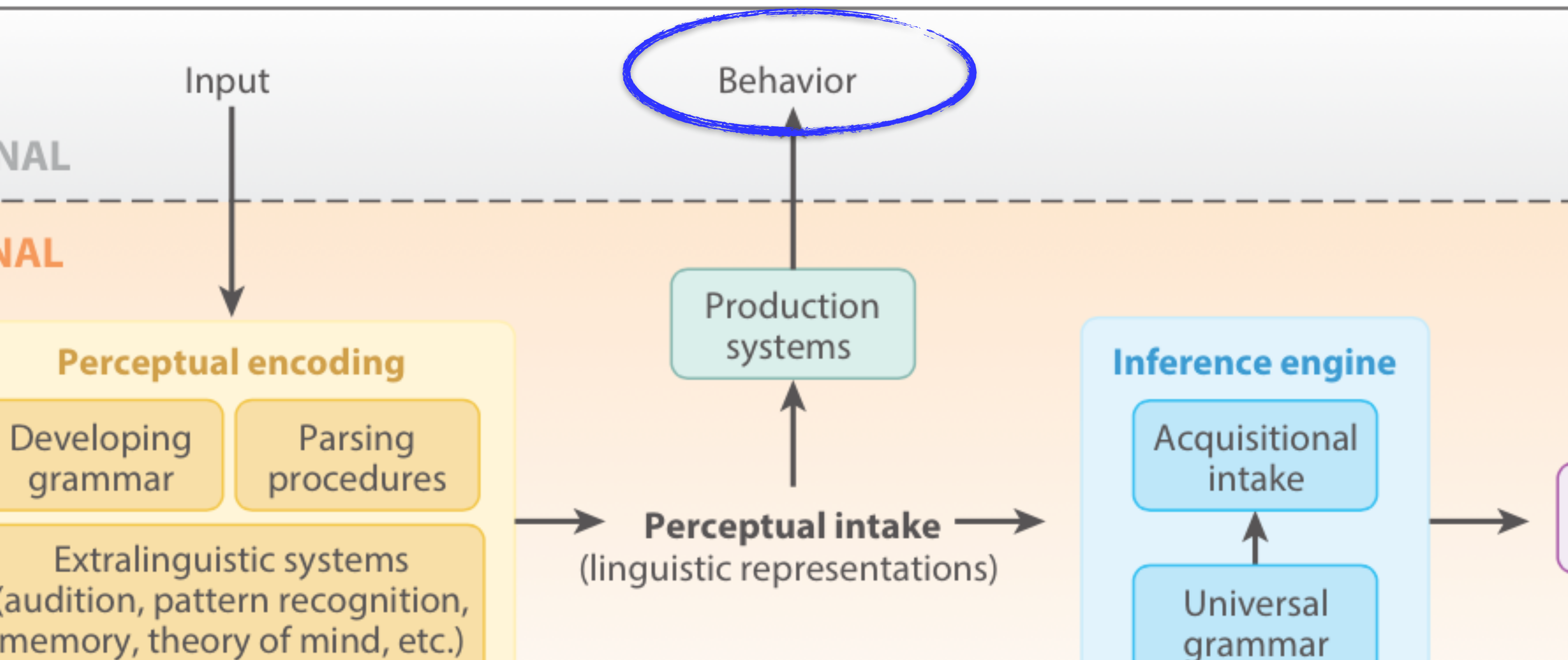
Who does

Lily think the kitty for is pretty?

syntactic island

Who does Lily think the kitty for ___ is pretty?

Adults judge these dependencies to be far worse than many others, including others that are very similar except that they don't cross syntactic islands (Sprouse et al. 2012).



Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

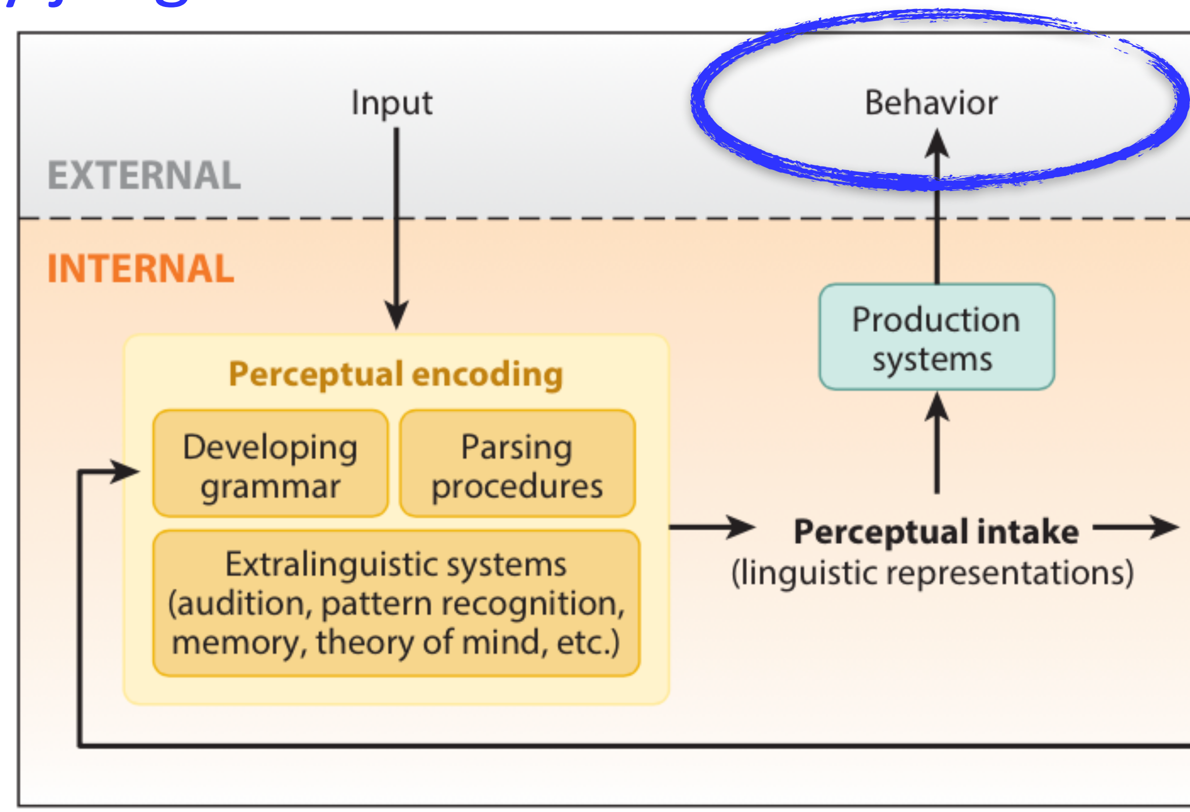
Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency

(**matrix** vs. **embedded**)

- presence of an island structure

(**non-island** vs. **island**)



Lidz & Gagliardi 2015

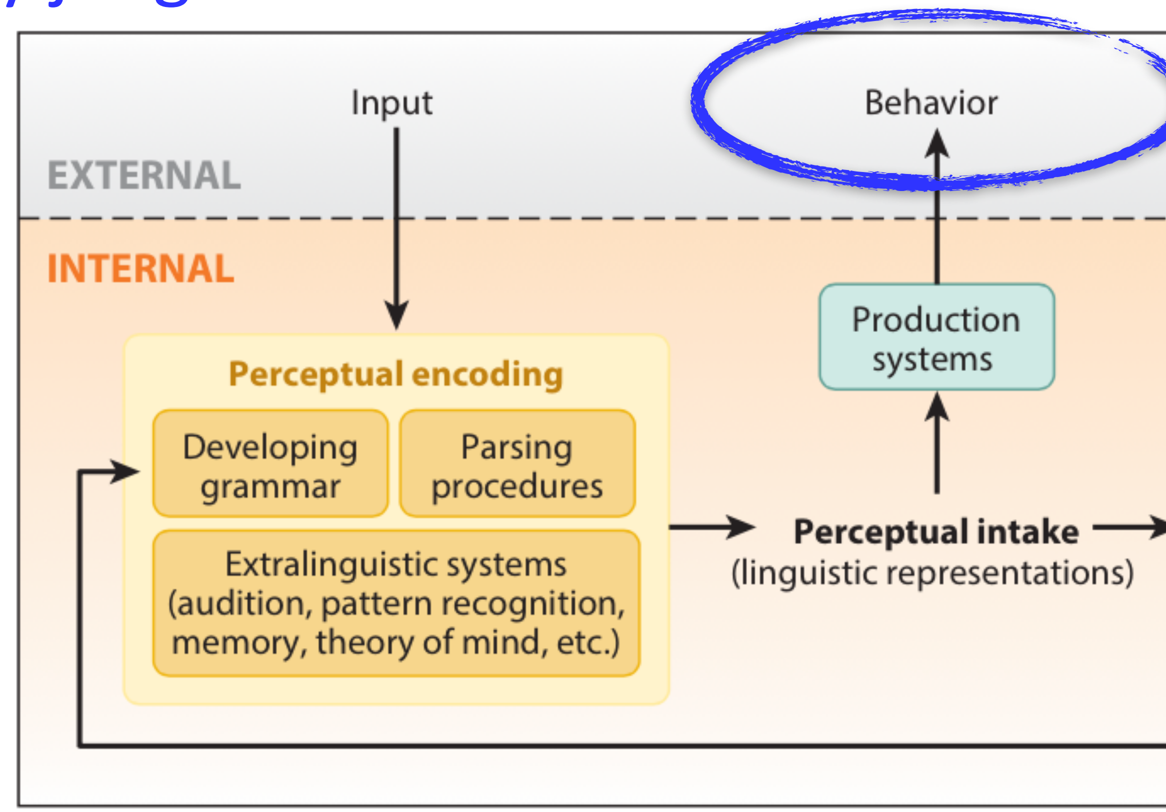
Pearl & Sprouse 2013a, 2013b, 2015

Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012)
length of dependency
(**matrix** vs. **embedded**)
presence of an **island** structure
(**non-island** vs. **island**)



Lidz & Gagliardi 2015

Complex NP island stimuli

Who __ claimed that Lily forgot the necklace?	matrix | non-island
What did the teacher claim that Lily forgot __?	embedded | non-island
Who __ made <u>the claim that Lily forgot the necklace</u>?	matrix | island
*What did the teacher make <u>the claim that Lily forgot __</u>?	embedded | island

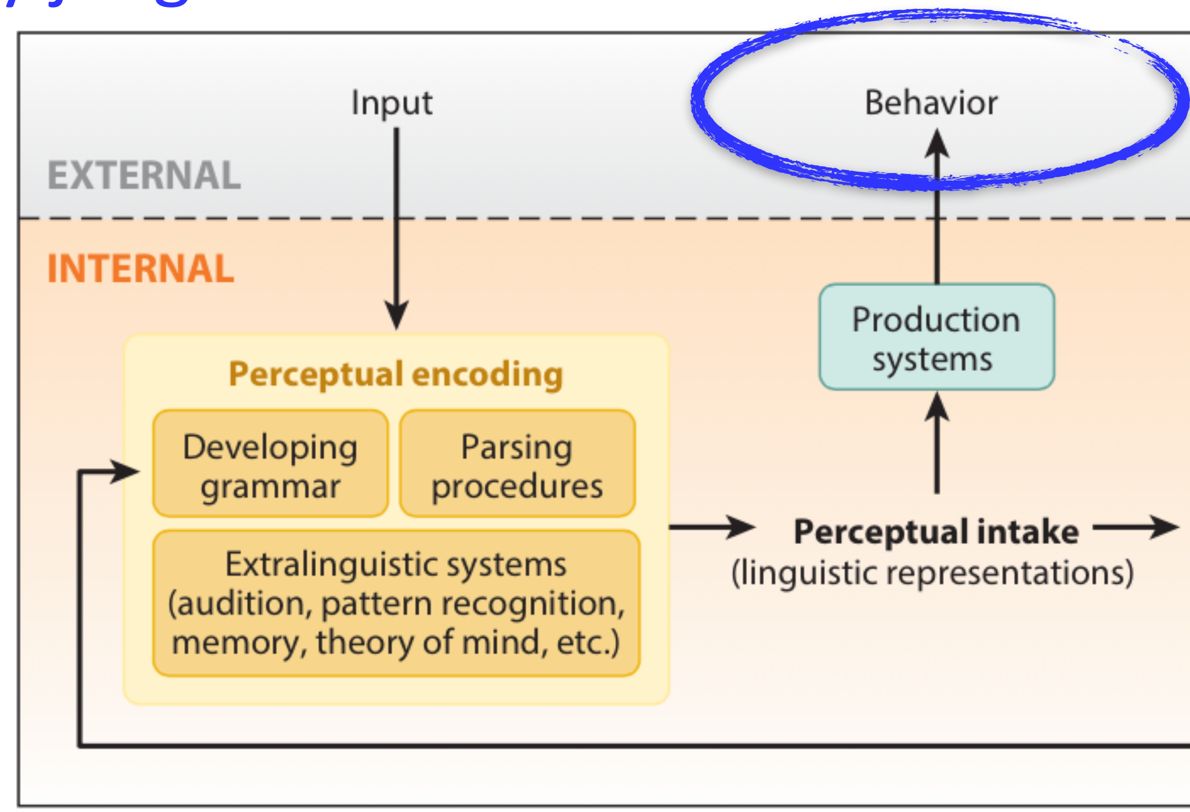
Pearl & Sprouse 2013a, 2013b, 2015

Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012)
length of dependency
(**matrix** vs. **embedded**)
presence of an **island** structure
(**non-island** vs. **island**)



Lidz & Gagliardi 2015

Subject island stimuli

Who __ thinks the necklace is expensive?	matrix | non-island
What does Jack think __ is expensive?	embedded | non-island
Who __ thinks <u>the necklace for Lily</u> is expensive?	matrix | island
*Who does Jack think <u>the necklace for __</u> is expensive?	embedded | island

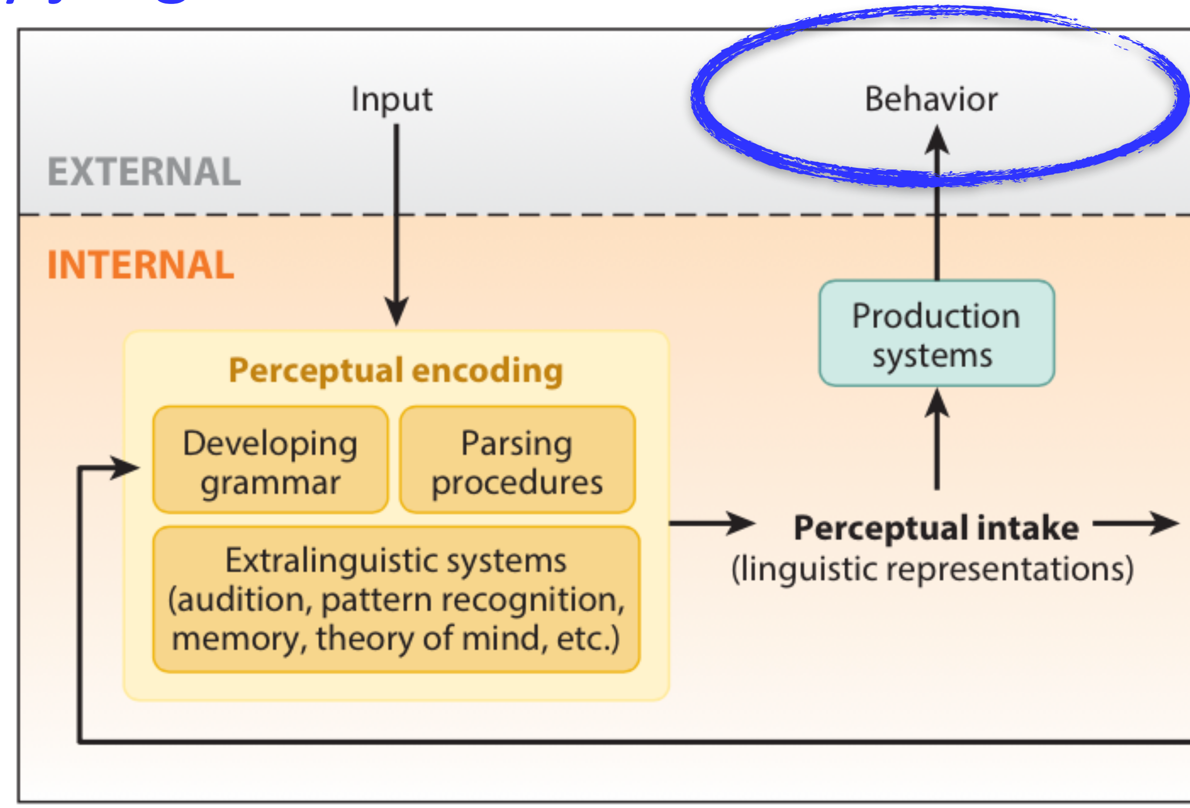
Pearl & Sprouse 2013a, 2013b, 2015

Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012)
length of dependency
(**matrix** vs. **embedded**)
presence of an **island** structure
(**non-island** vs. **island**)



Lidz & Gagliardi 2015

Whether island stimuli

Who __ thinks that Jack stole the necklace?	matrix | non-island
What does the teacher think that Jack stole __ ?	embedded | non-island
Who __ wonders <u>whether Jack stole the necklace</u>?	matrix | island
*What does the teacher wonder <u>whether Jack stole __</u> ?	embedded | island

Pearl & Sprouse 2013a, 2013b, 2015

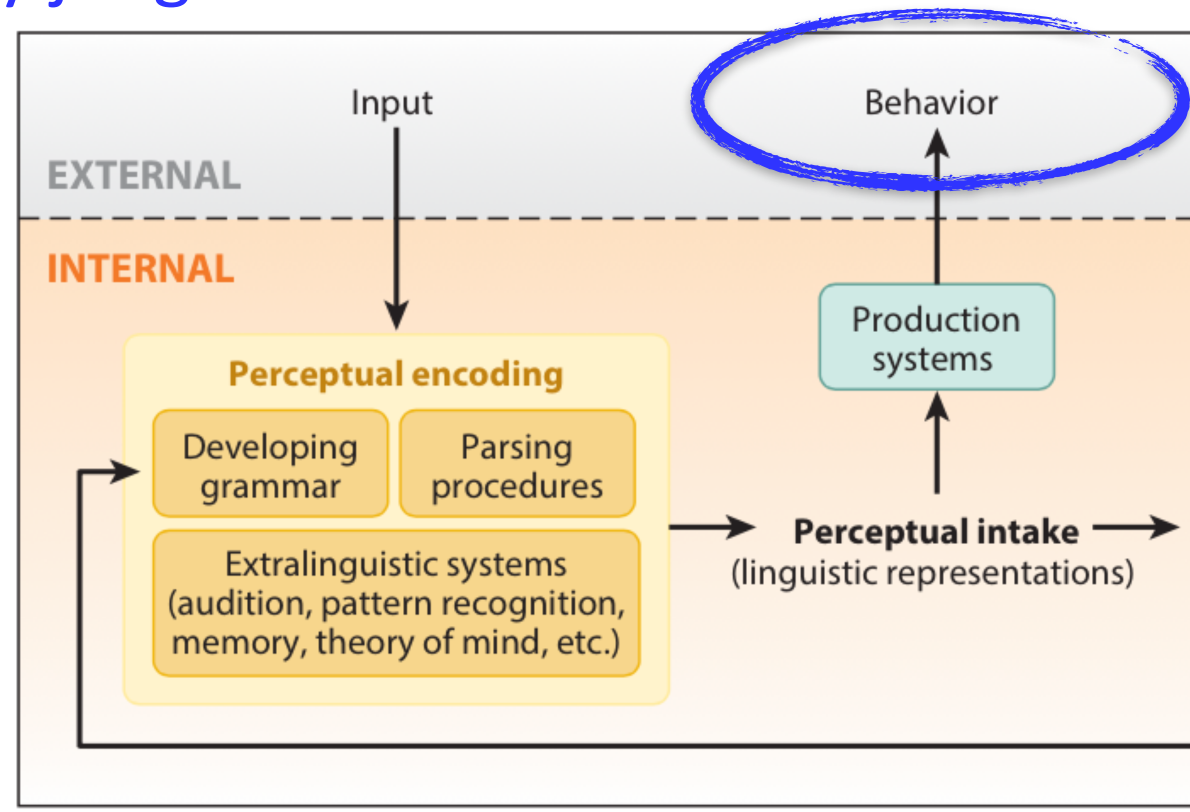
Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012)
length of dependency
(**matrix** vs. **embedded**)
presence of an **island** structure
(**non-island** vs. **island**)

Adjunct island stimuli



Lidz & Gagliardi 2015

Who __ thinks that Lily forgot the necklace?	matrix | non-island
What does the teacher think that Lily forgot __ ?	embedded | non-island
Who __ worries if Lily forgot the necklace?	matrix | island
*What does the teacher worry if Lily forgot __ ?	embedded | island

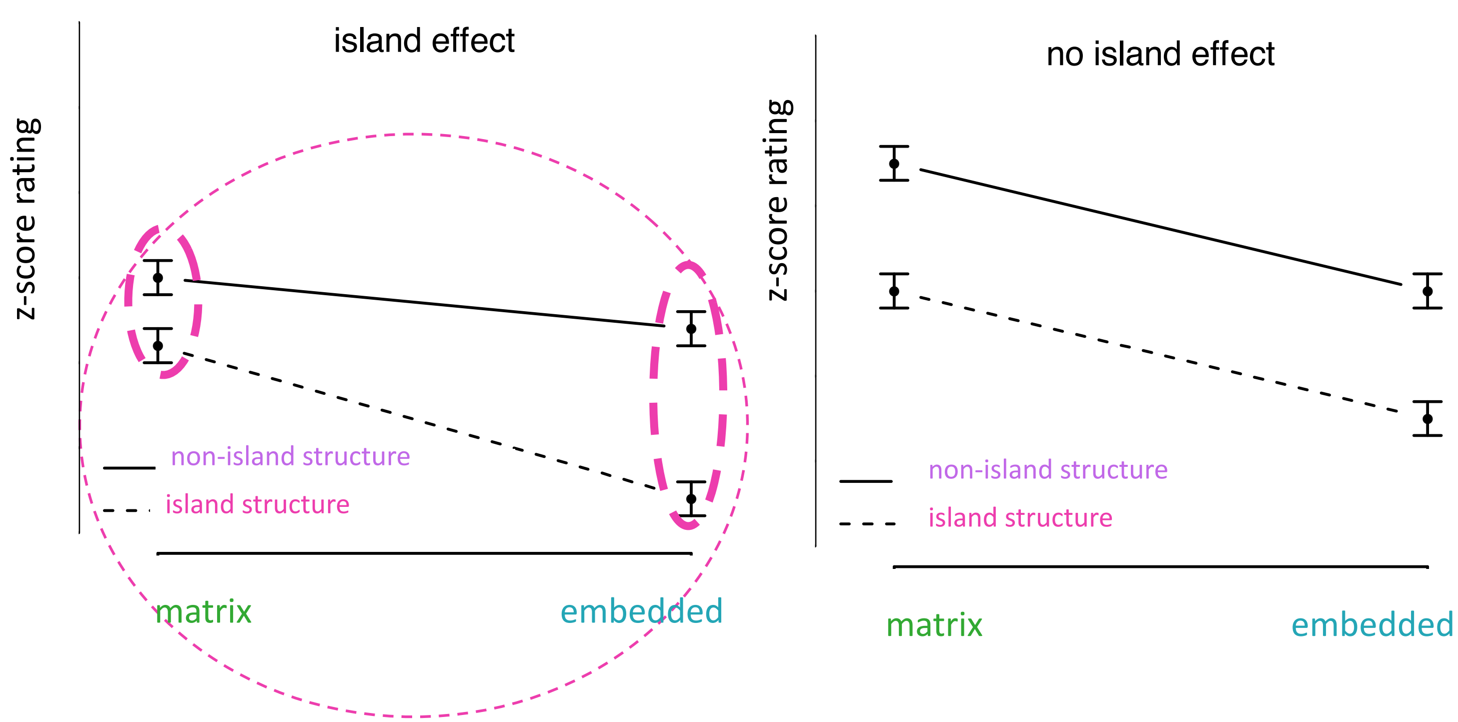
Pearl & Sprouse 2013a, 2013b, 2015

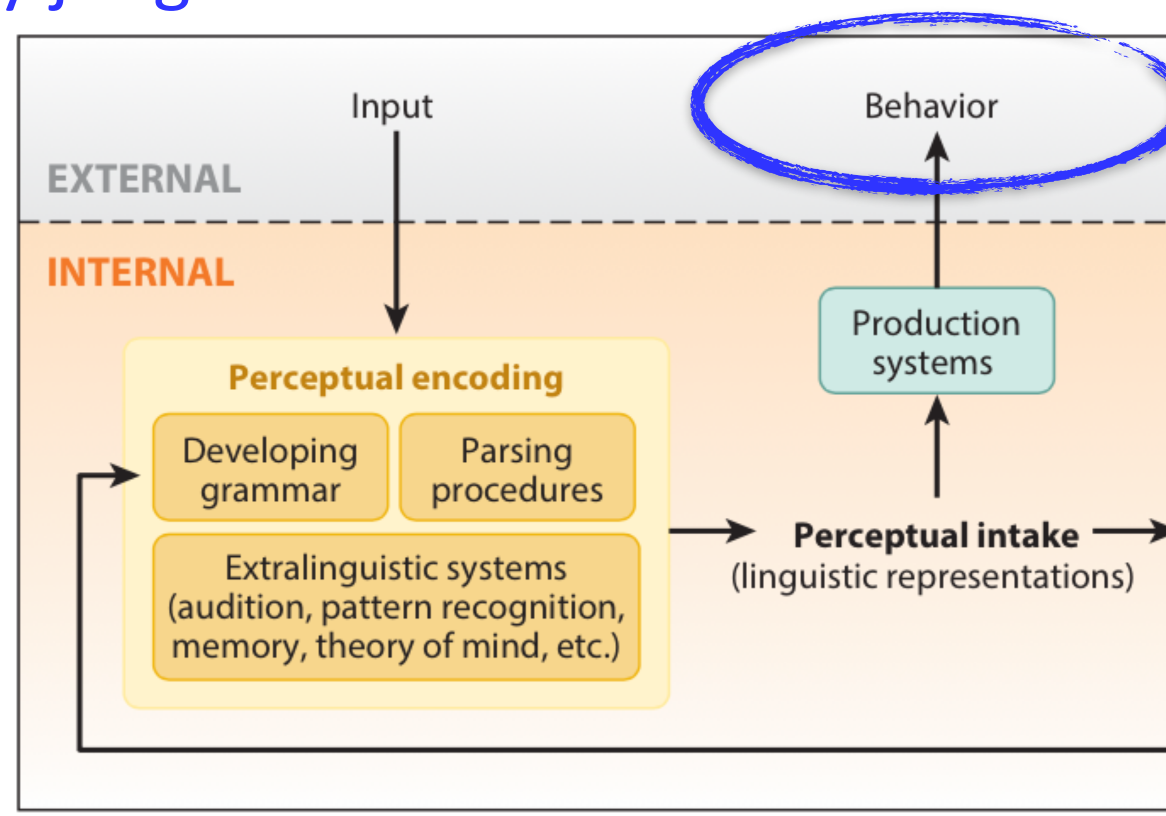
Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Syntactic island = **superadditive** interaction of the two factors (additional unacceptability that arises when the two factors — **length** & presence of an **island** structure — are combined, above and beyond the independent contribution of each factor).

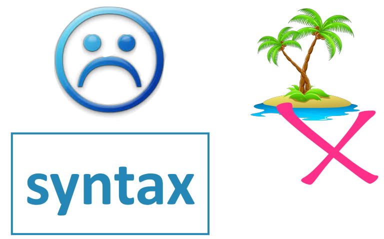




Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

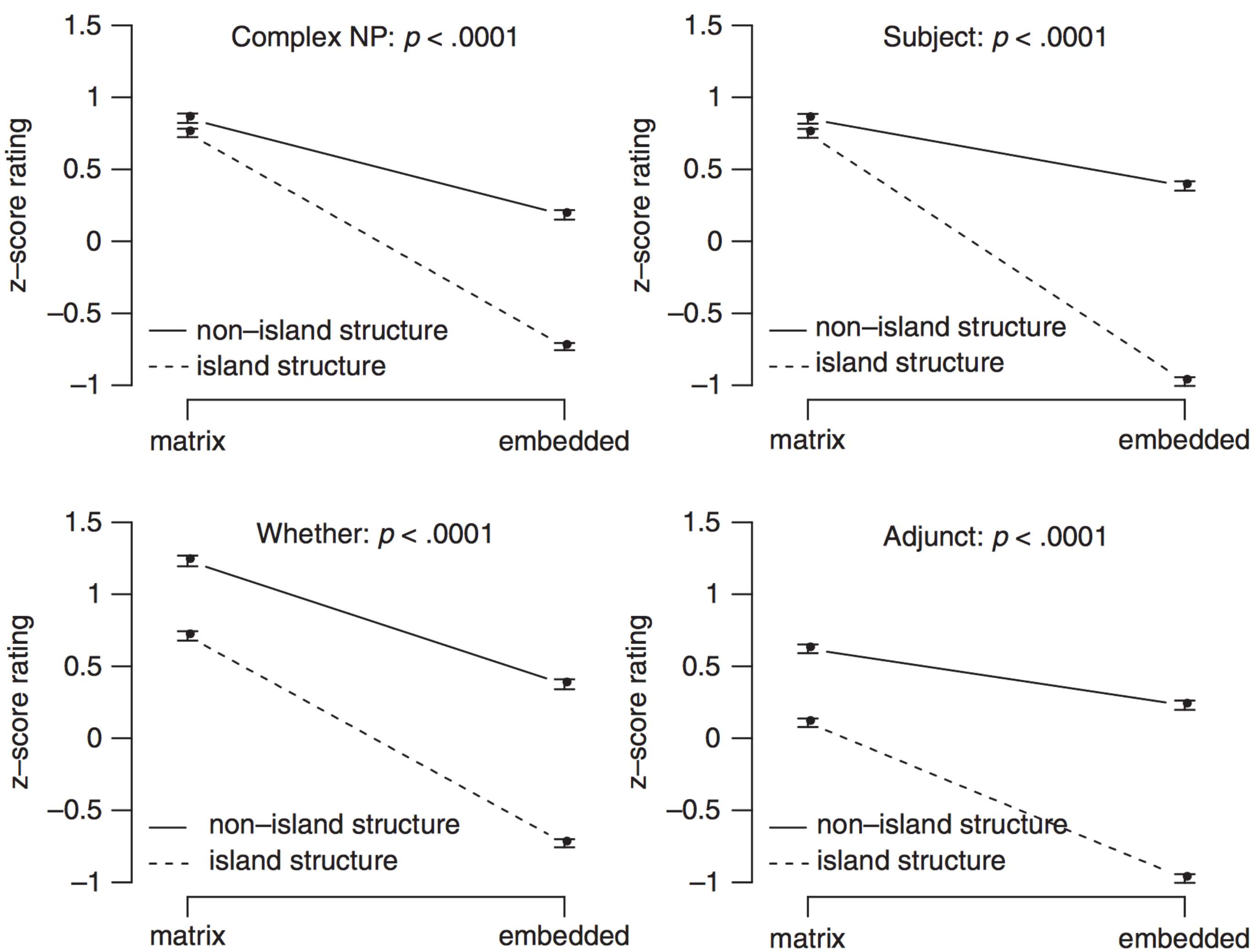
Adult judgments: Target behavior

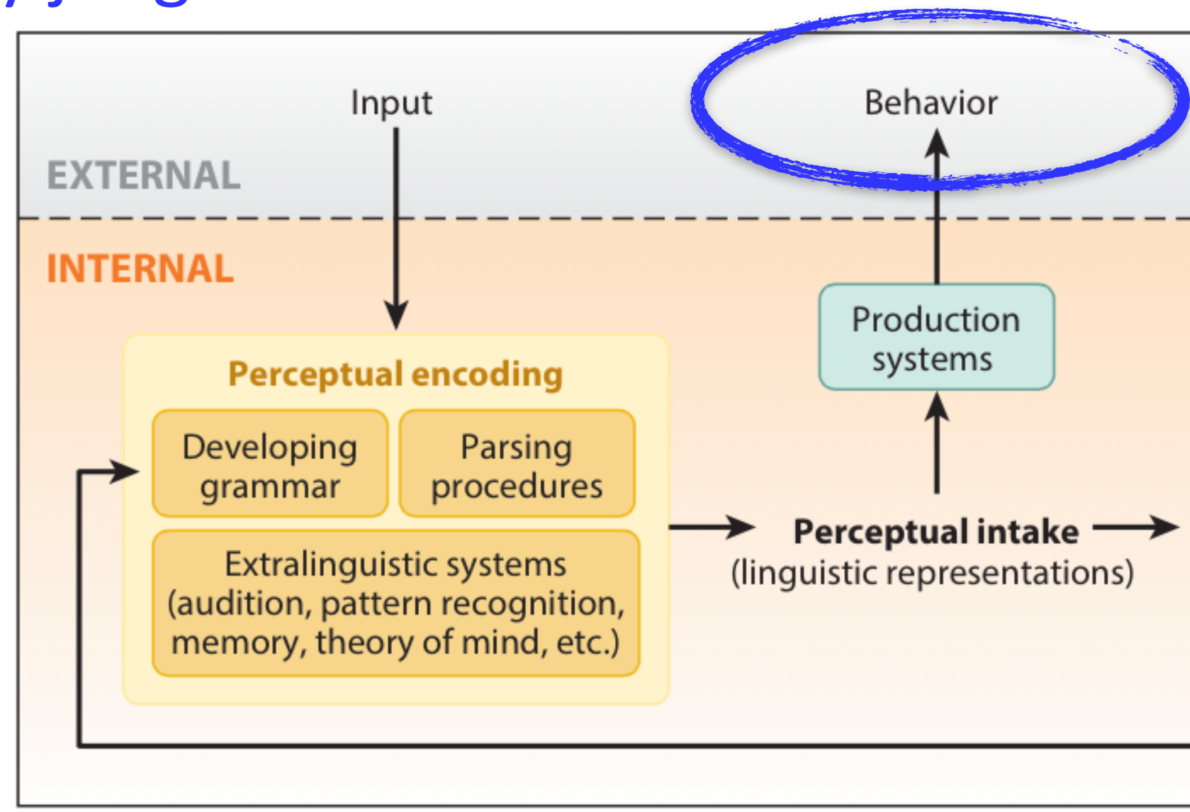


syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012): acceptability judgments from 173 adult subjects





Lidz & Gagliardi 2015

Superadditivity present for all islands tested = Knowledge that dependencies cannot cross these island structures is part of adult knowledge about syntactic islands.

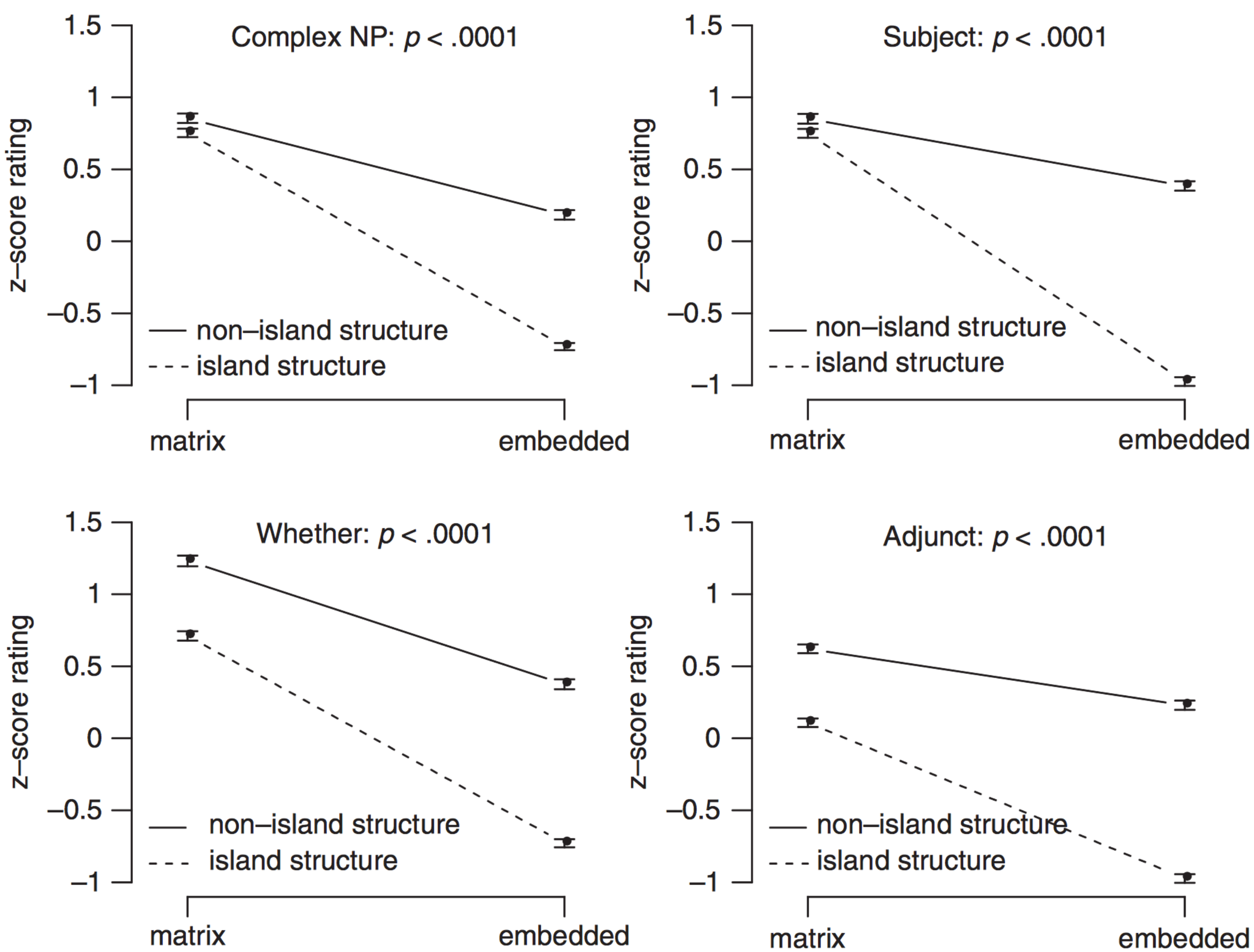
Pearl & Sprouse 2013a, 2013b, 2015

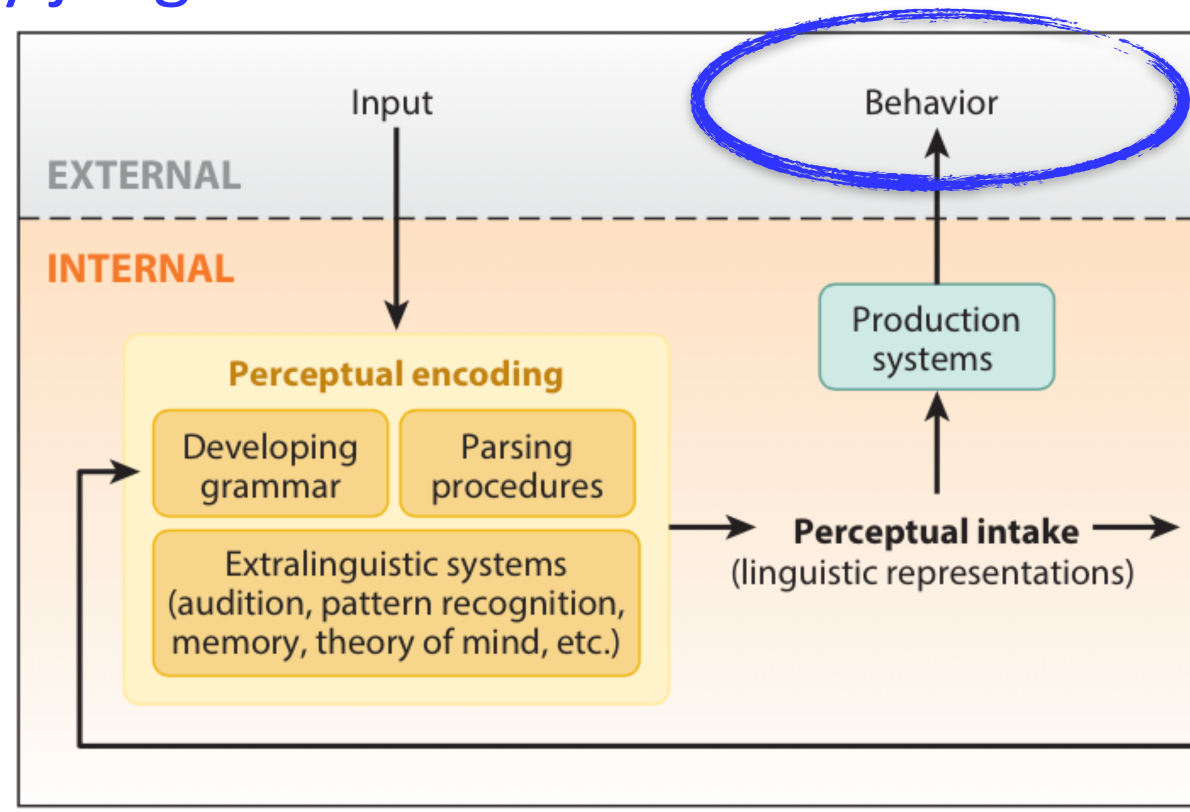
Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012): acceptability judgments from 173 adult subjects





Lidz & Gagliardi 2015

Importance for acquisition: This is one kind of target behavior that we'd like a modeled child to produce.

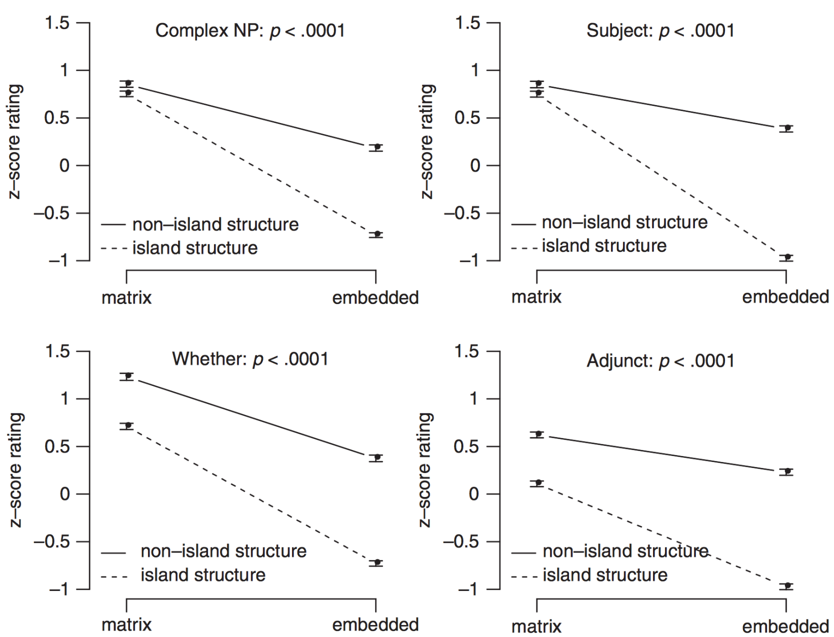
Pearl & Sprouse 2013a, 2013b, 2015

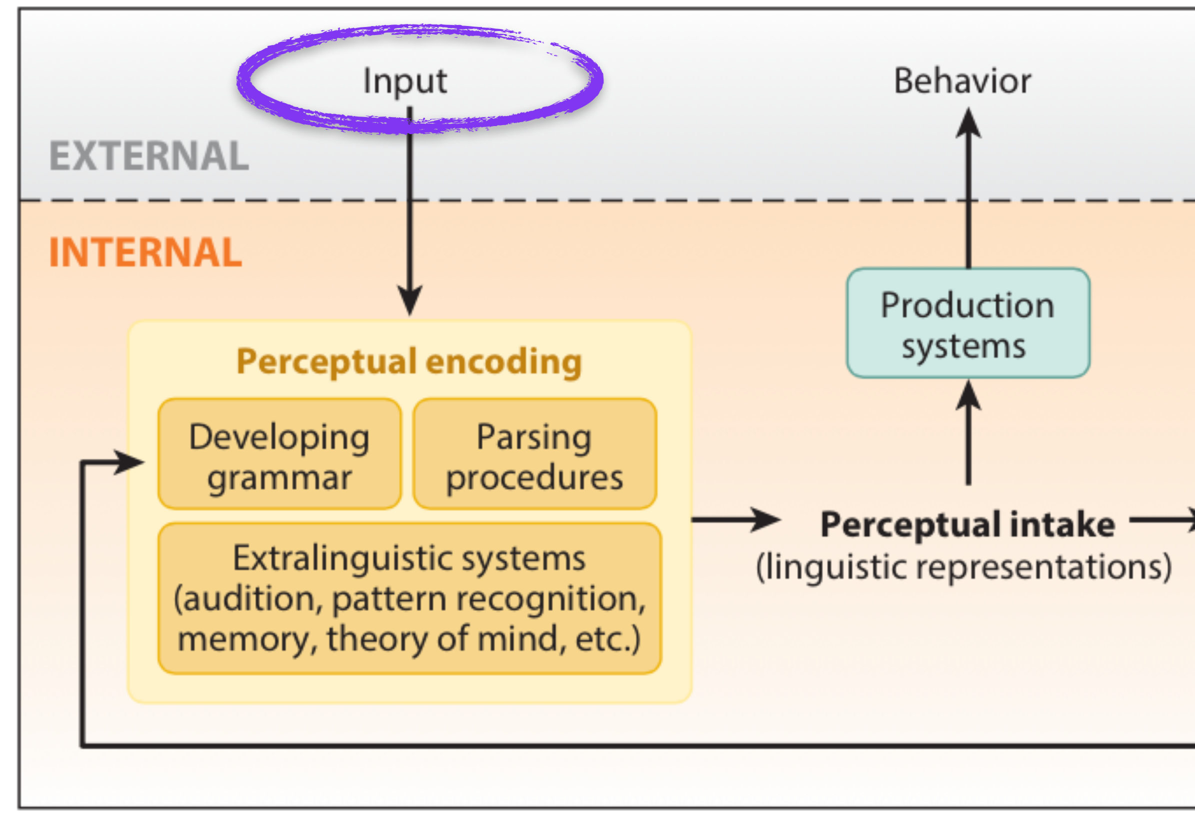
Adult judgments: Target behavior

syntactic island

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012): acceptability judgments from 173 adult subjects





Lidz & Gagliardi 2015

So if we're focusing on these *wh*-dependencies and that specific target state, what does children's input look like?

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

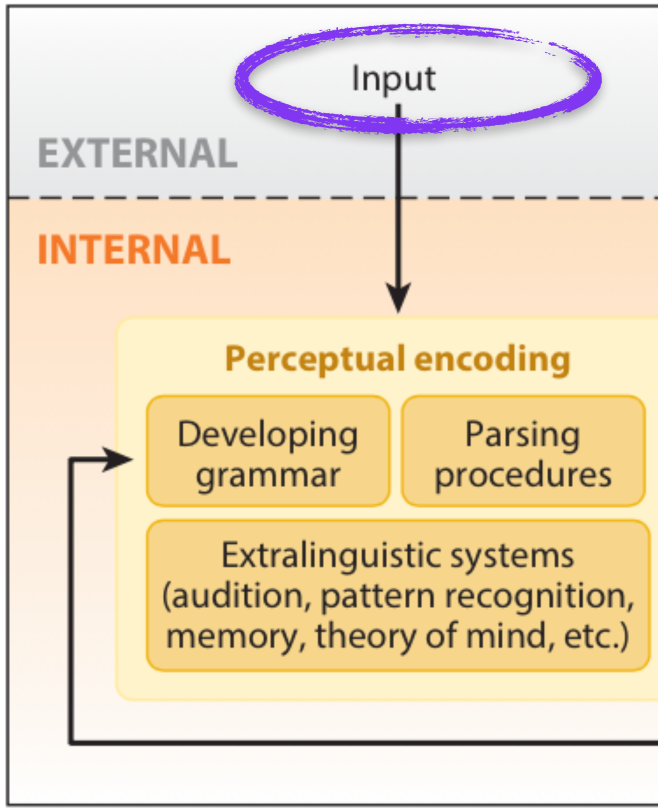
syntactic island

Children's input really doesn't look so helpful

Data from five corpora of child-directed speech (Brown-Adam, Brown-Eve, Brown-Sarah, Suppes, Valian) from CHILDES (MacWhinney 2000): speech to 25 children between the ages of one and five years old.

= 813,036 words

= 31,247 utterances containing a *wh*-dependency



Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

syntax

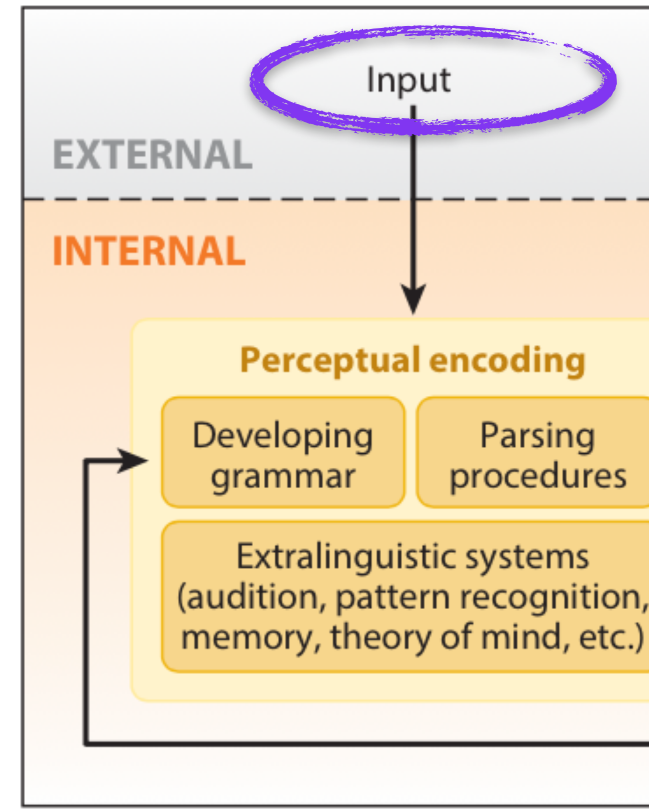
syntactic island

Children's input really doesn't look so helpful

Data from five corpora of child-directed speech = **31,247** utterances containing a *wh*-dependency

	grammatical stimuli		*syntactic island*	
	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

These kinds of utterances are fairly rare in general - the most frequent appears about 0.9% of the time (295 of 31,247.)



Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

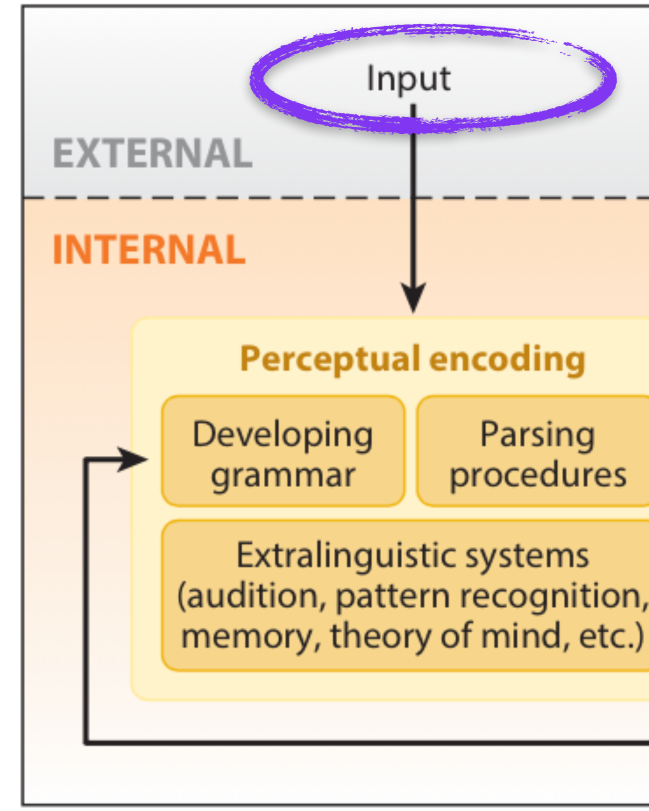
syntactic island

Children's input really doesn't look so helpful

Data from five corpora of child-directed speech = **31,247** utterances containing a *wh*-dependency

	grammatical stimuli		*syntactic island*	
	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

Being grammatical doesn't necessarily mean an utterance will appear in the input at all.



Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

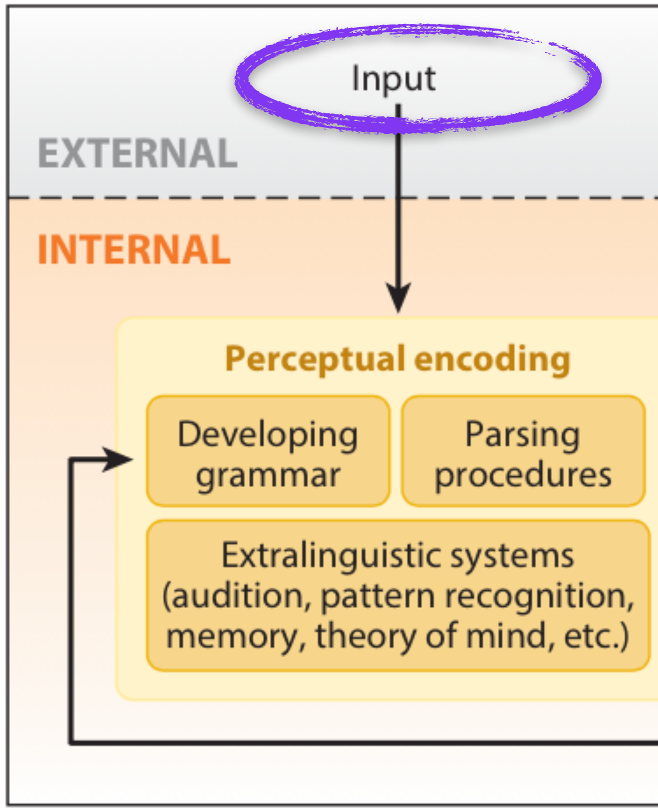
Children's input

syntactic island

Children's input really doesn't look so helpful

Data from five corpora of child-directed speech = **31,247** utterances containing a *wh*-dependency

	grammatical stimuli		*syntactic island*	
	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0



Lidz & Gagliardi 2015

Unless the child is sensitive to very small frequencies, it's difficult to tell the difference between grammatical and ungrammatical dependencies sometimes…

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

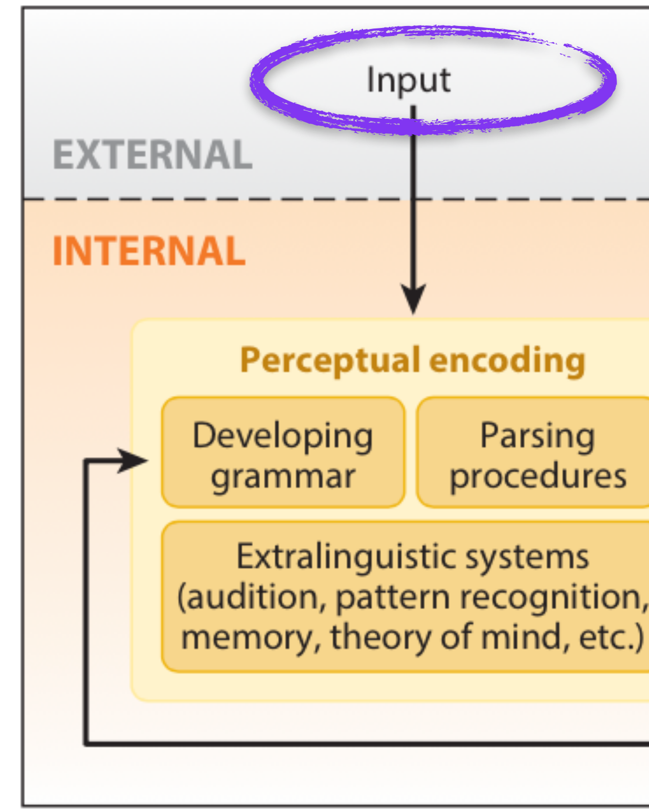

syntax

syntactic island

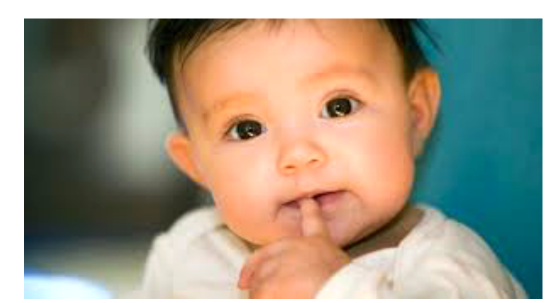
Children's input really doesn't look so helpful

Data from five corpora of child-directed speech =
31,247 utterances containing a *wh*-dependency

	grammatical stimuli		*syntactic island*	
	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0



Lidz & Gagliardi 2015



...and impossible to tell no matter what the rest of the time.
This looks like an **induction problem** for the language learner if we're looking for direct evidence in the input.

Pearl & Sprouse 2013a, 2013b, 2015

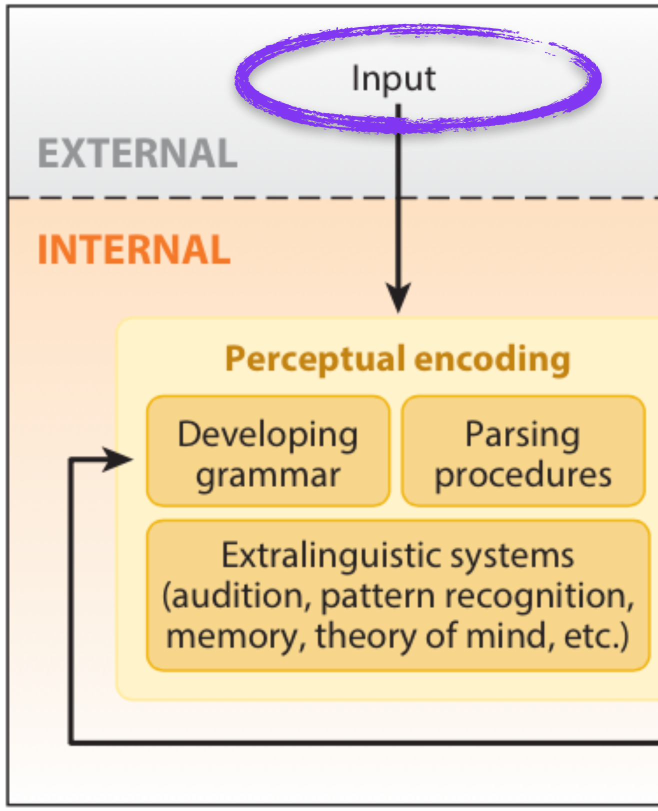
Children's input

syntax

syntactic island

Children's input really doesn't look so helpful

Data from five corpora of child-directed speech =
31,247 utterances containing a *wh*-dependency



Lidz & Gagliardi 2015

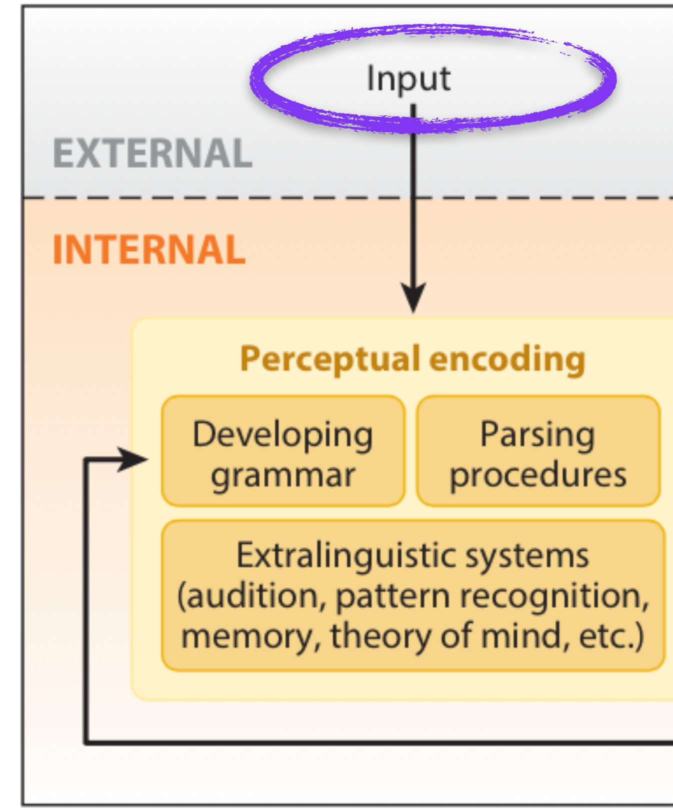
Important: Some grammatical utterances never appeared at all. This means that **only a subset of grammatical utterances appeared**, and the child has to **generalize appropriately from this subset**.

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

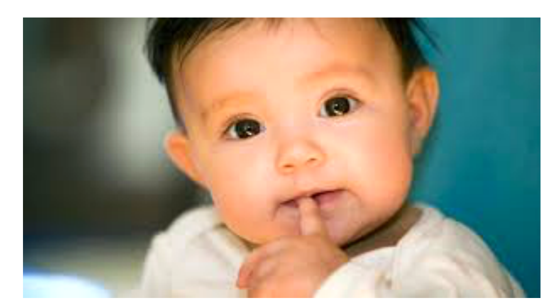
syntactic island

Data from five corpora of child-directed speech =
31,247 utterances containing a *wh*-dependency



Lidz & Gagliardi 2015

So what kinds of dependencies *are* in the input?



Pearl & Sprouse 2013a, 2013b, 2015

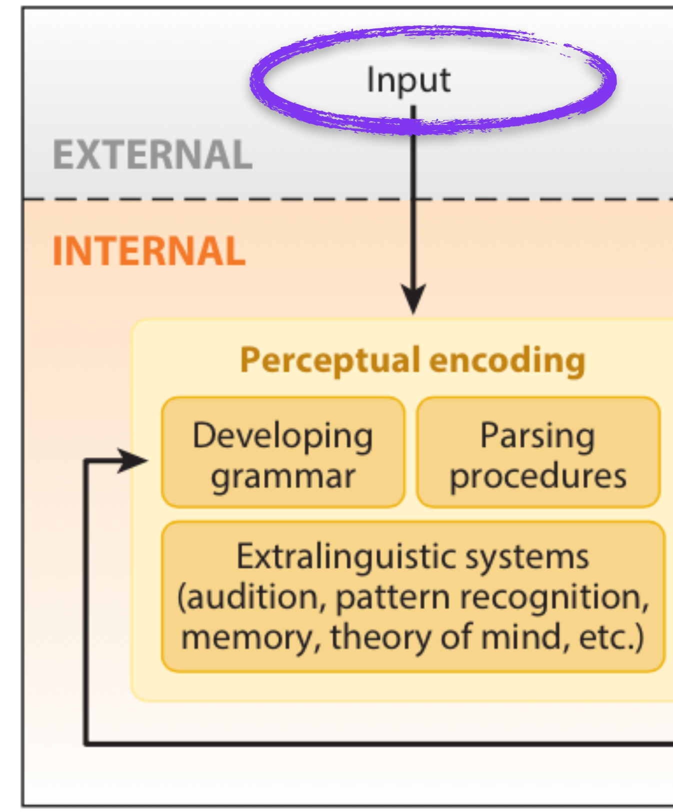
Children's input

syntax

syntactic island

So what kinds of dependencies *are* in the input?

Data from five corpora of child-directed speech =
31,247 utterances containing a *wh*-dependency

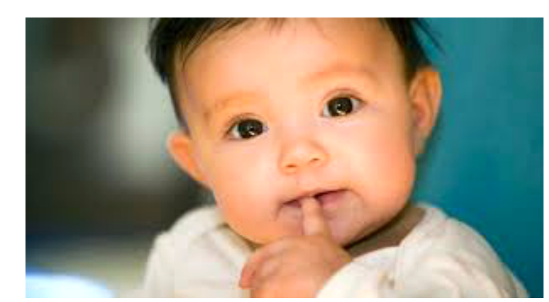


Lidz & Gagliardi 2015

A lot of simpler ones!

76.7%	*What did you see __?*
12.8%	*What __ happened?*
5.6%	*What did she want to do __?*
2.5%	*What did she read from __?*
1.1%	*What did she think he said __?*

…

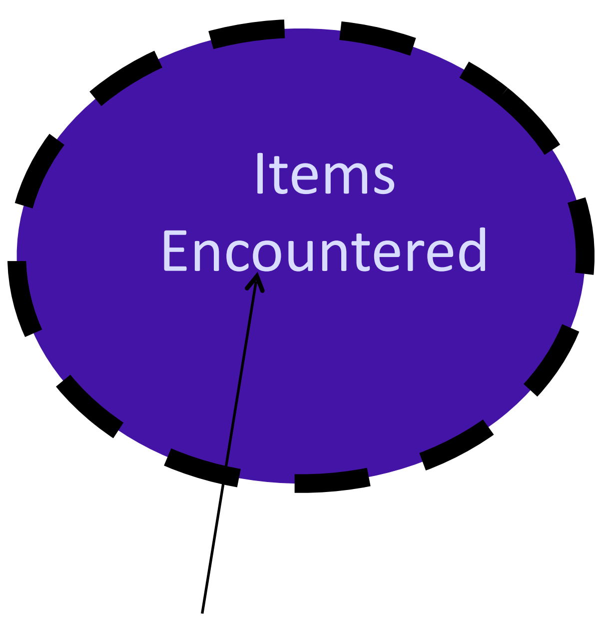


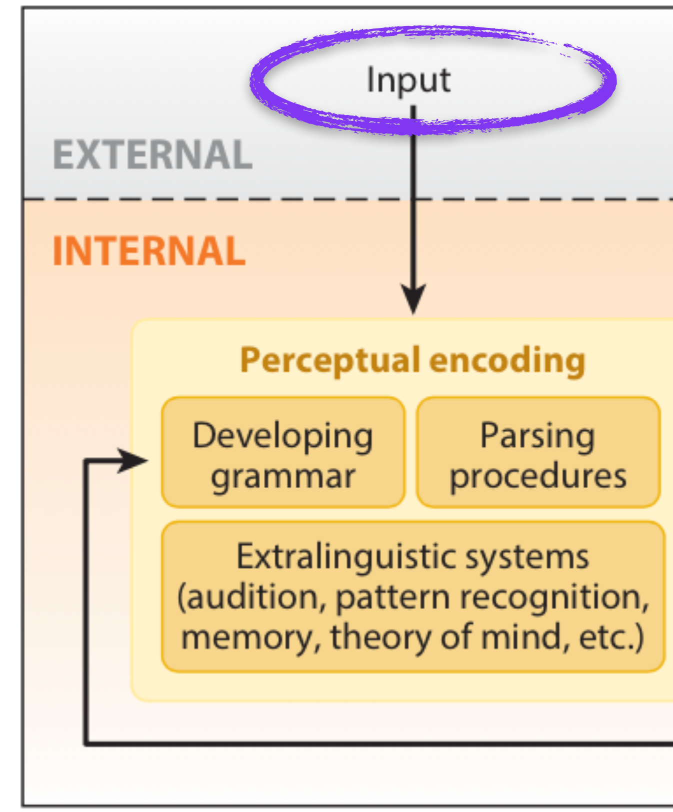
Pearl & Sprouse 2013a, 2013b, 2015

Children’s input

syntactic island

The induction problem





Lidz & Gagliardi 2015

wh-questions in input (usually fairly simple)

What did you see __?

What __ happened?

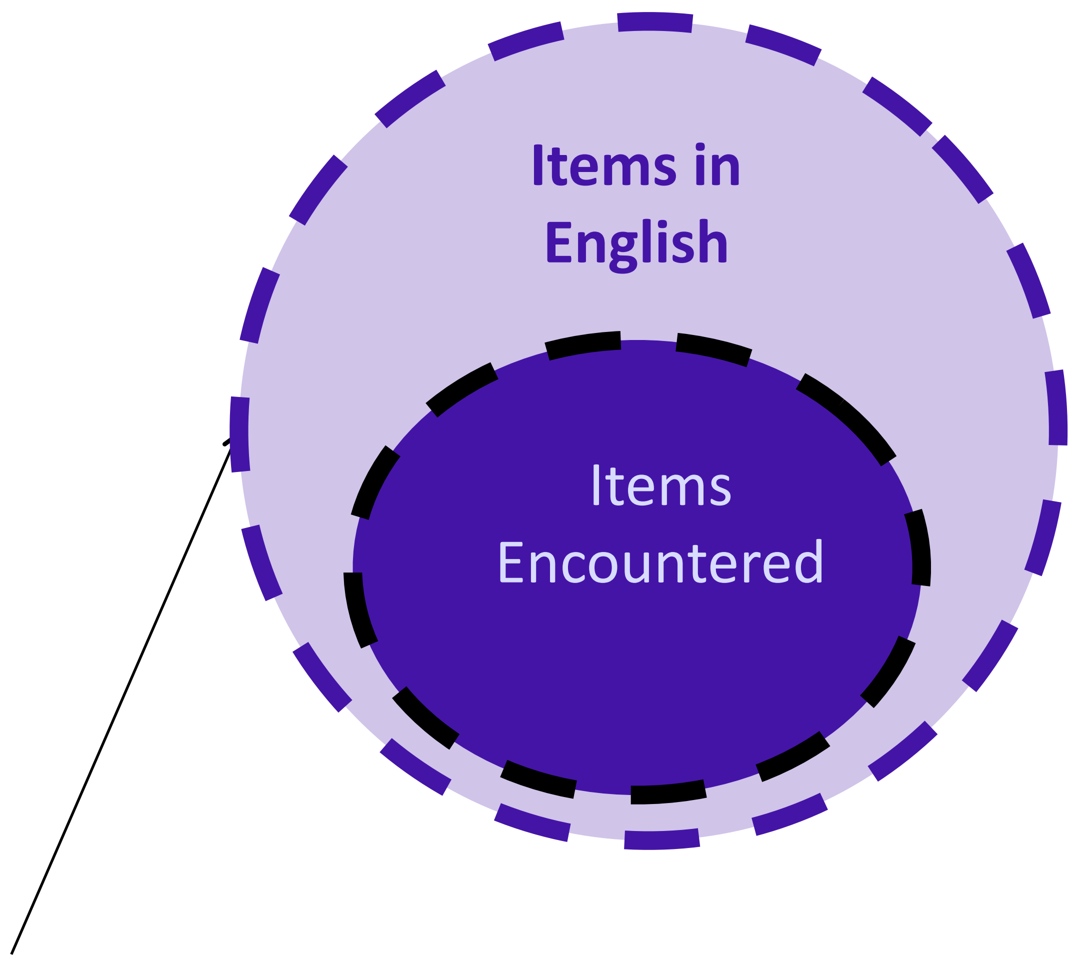
…

Pearl & Sprouse 2013a, 2013b, 2015

Children's input

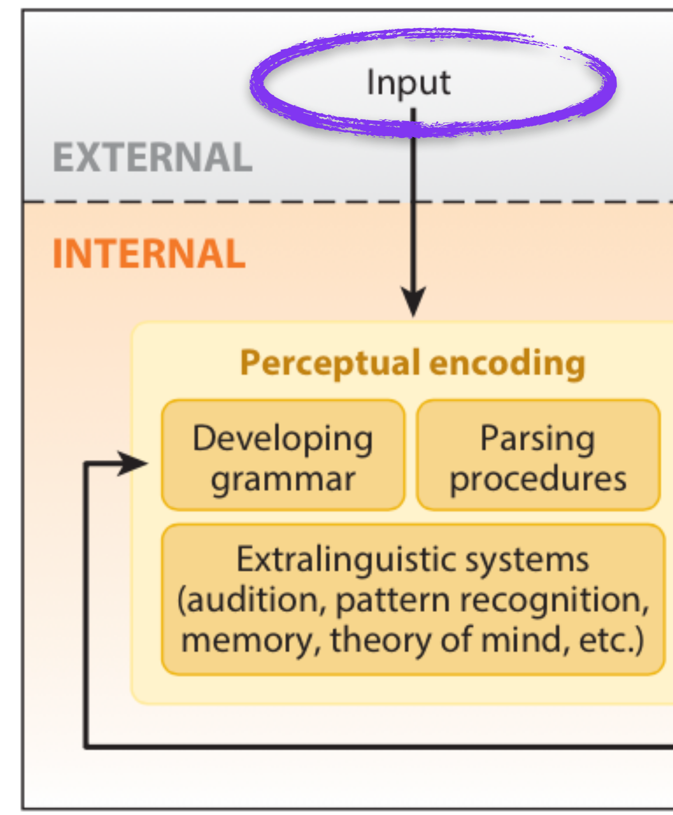
syntactic island

The induction problem



Grammatical *wh*-questions

What did you see __?
What __ happened?
Who did Jack think that Lily saw __?
What did Jack think __ happened?



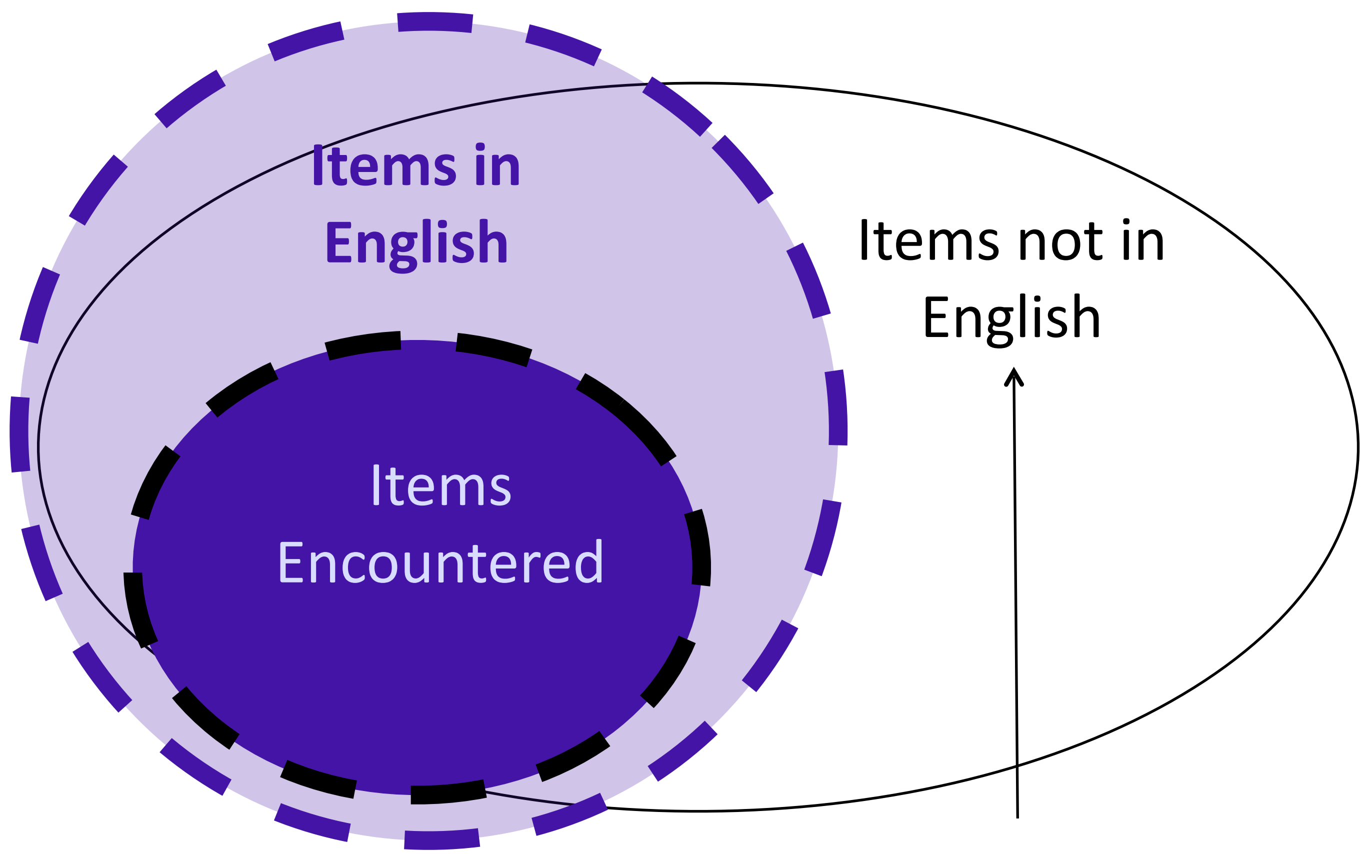
Lidz & Gagliardi 2015

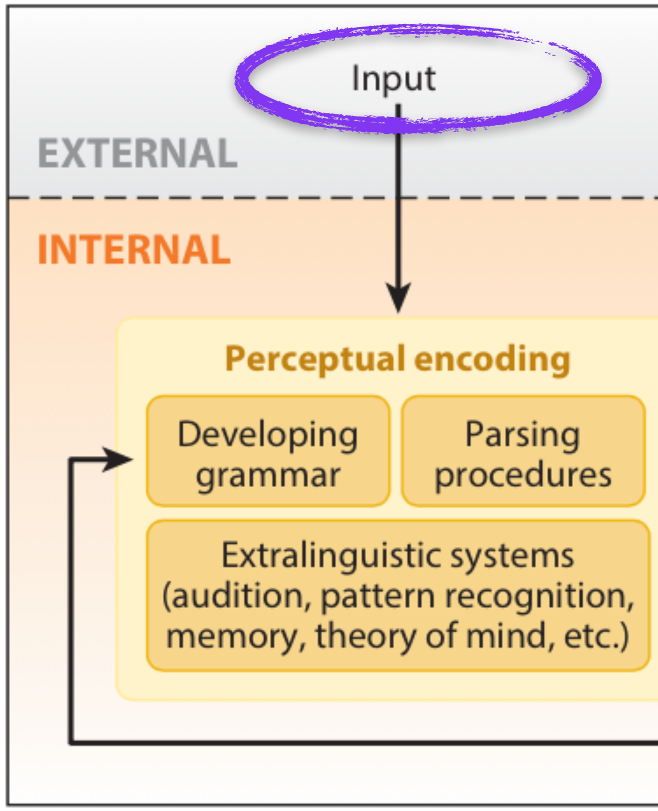
Pearl & Sprouse 2013a, 2013b, 2015

Children's input

syntactic island

The induction problem





Lidz & Gagliardi 2015

Ungrammatical *wh*-questions: Syntactic islands

Who does Lily think the kitty for ___ is pretty?

What did Jack make the claim that he bought ___ ?

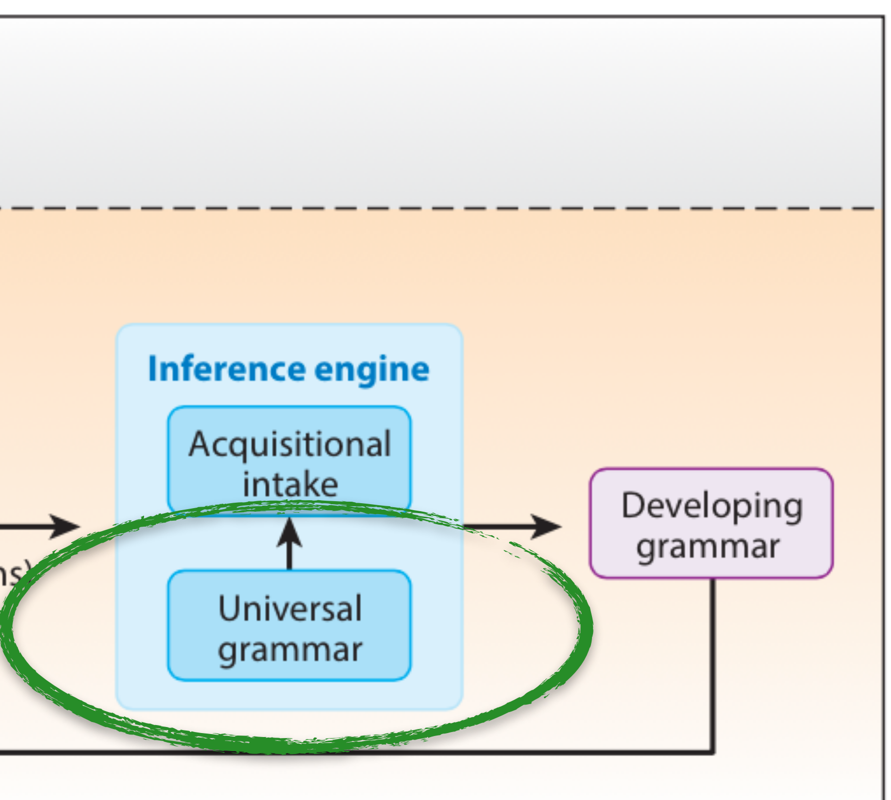
What did Elizabeth wonder whether Jack bought ___

What did Elizabeth worry if Jack bought ___ ?

Pearl & Sprouse 2013a, 2013b, 2015

Learning strategies

Previous learning theories suggested children need syntactic-island-specific innate knowledge.



Learning strategies

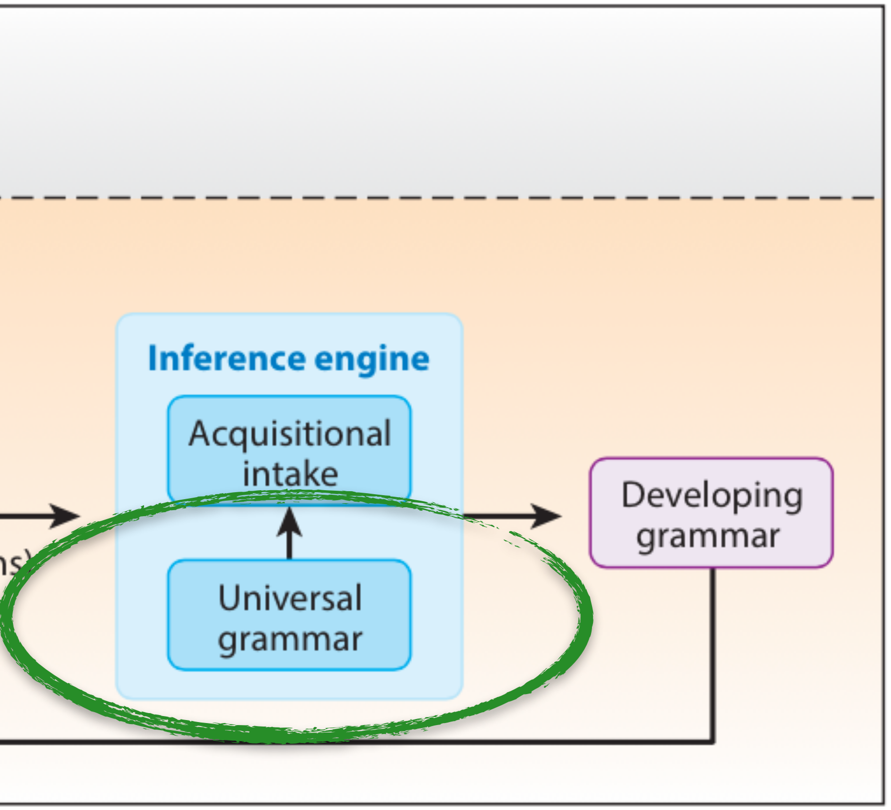
syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

A dependency cannot cross two or more bounding nodes.

Wh … $[_{BN1}$ … $[_{BN2}$ … ___]]



Learning strategies


syntax

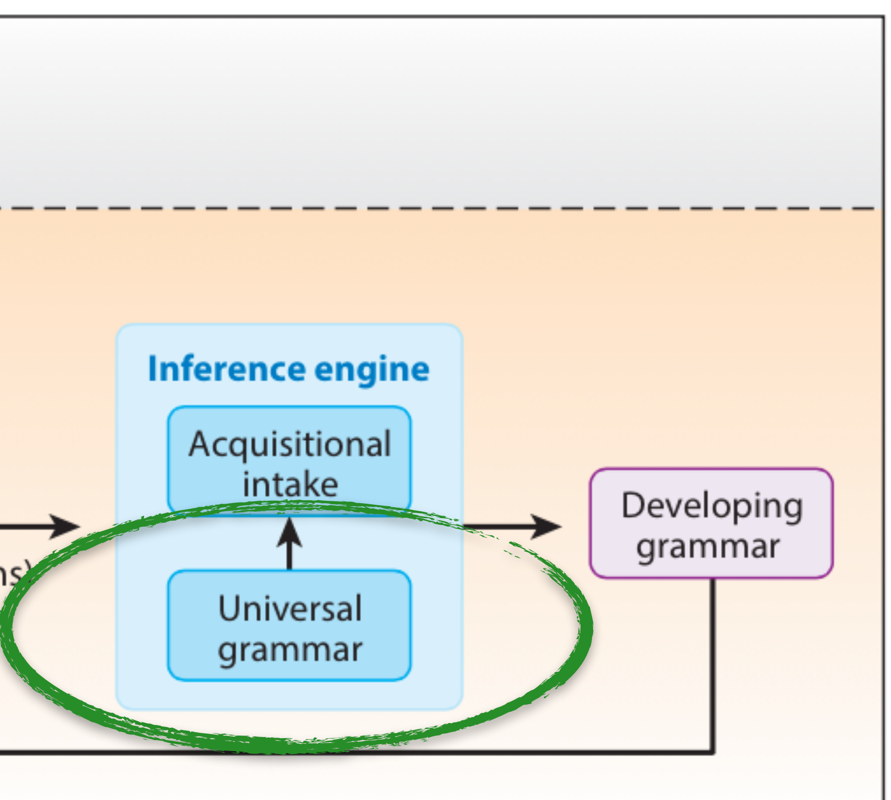
syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

Bounding nodes come from a fixed set (CP, IP, and/or NP). The ones that act as a bounding nodes for a given language must be learned.

Wh … $[_{BN1}$ … $[_{BN2}$ … ___]]

{CP, IP, NP}?



Learning strategies

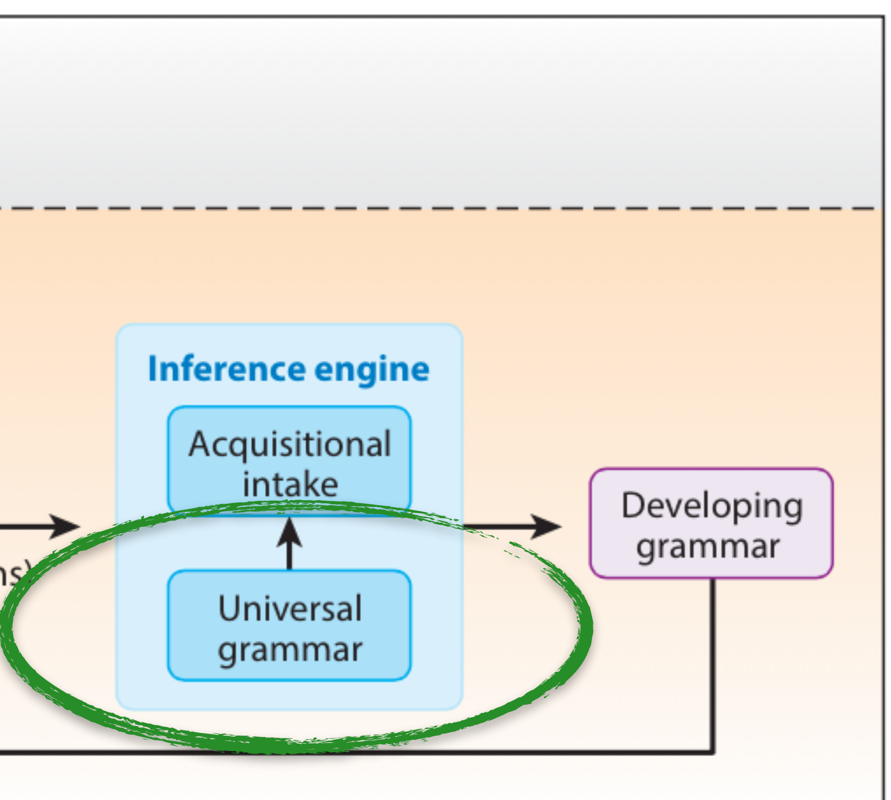
syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … ___]]



Learning strategies

syntax

syntactic island

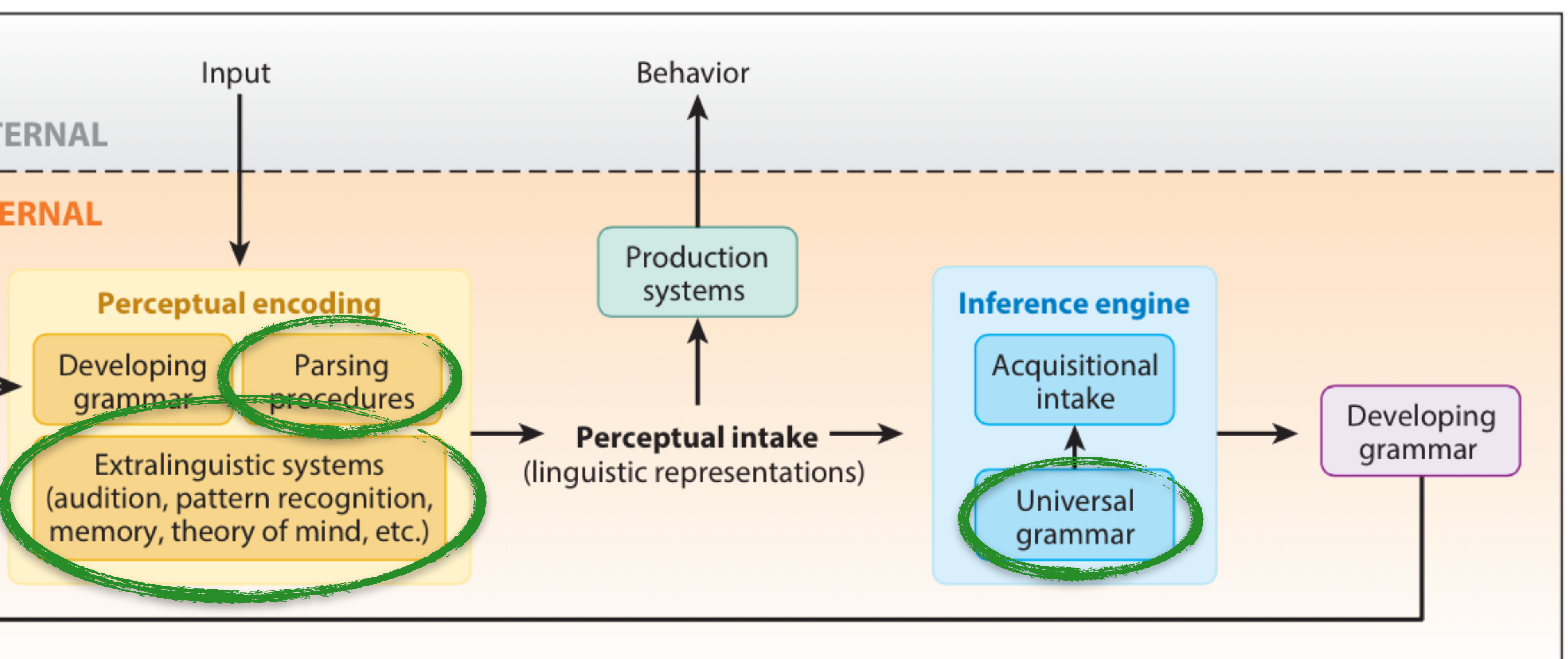
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

An alternative learning strategy proposes children need less-specific linguistic prior knowledge along with probabilistic learning.

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)



Learning strategies

syntax

syntactic island

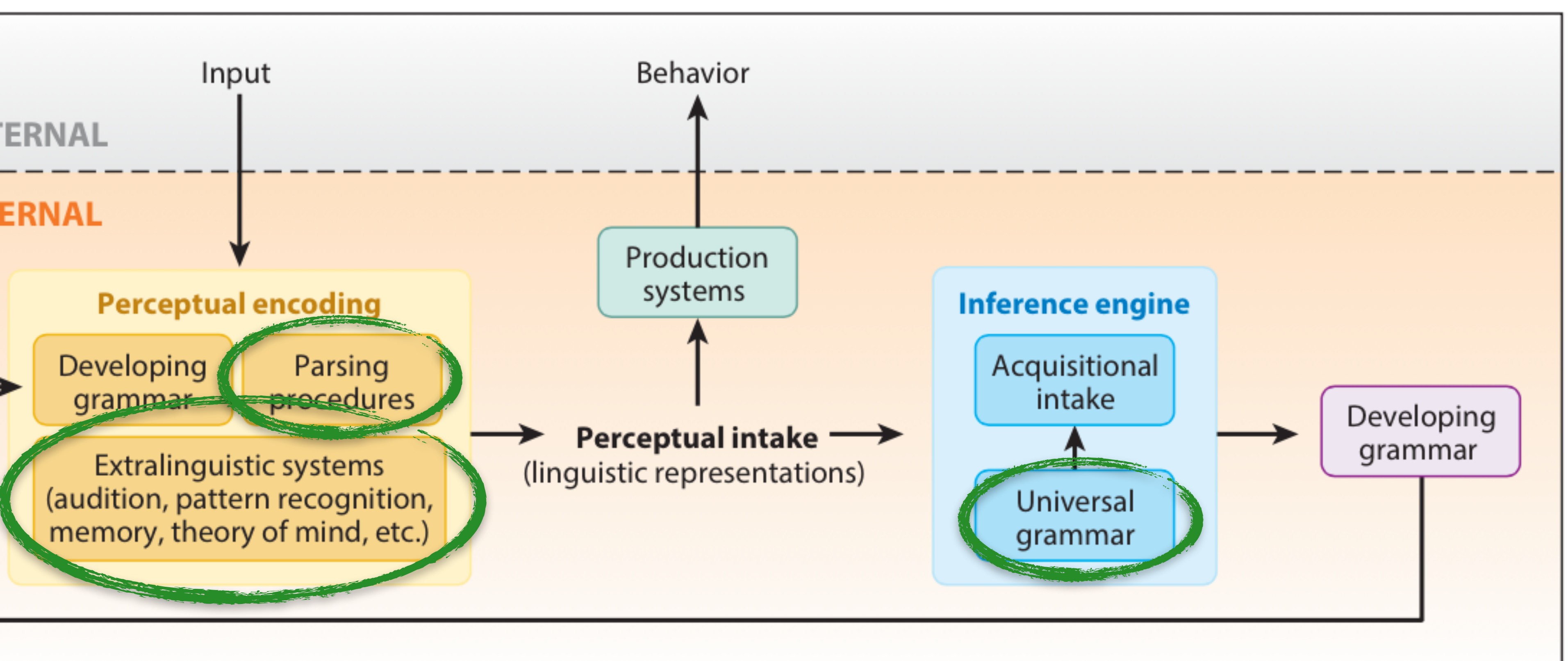
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … ___]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure



Learning strategies

syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

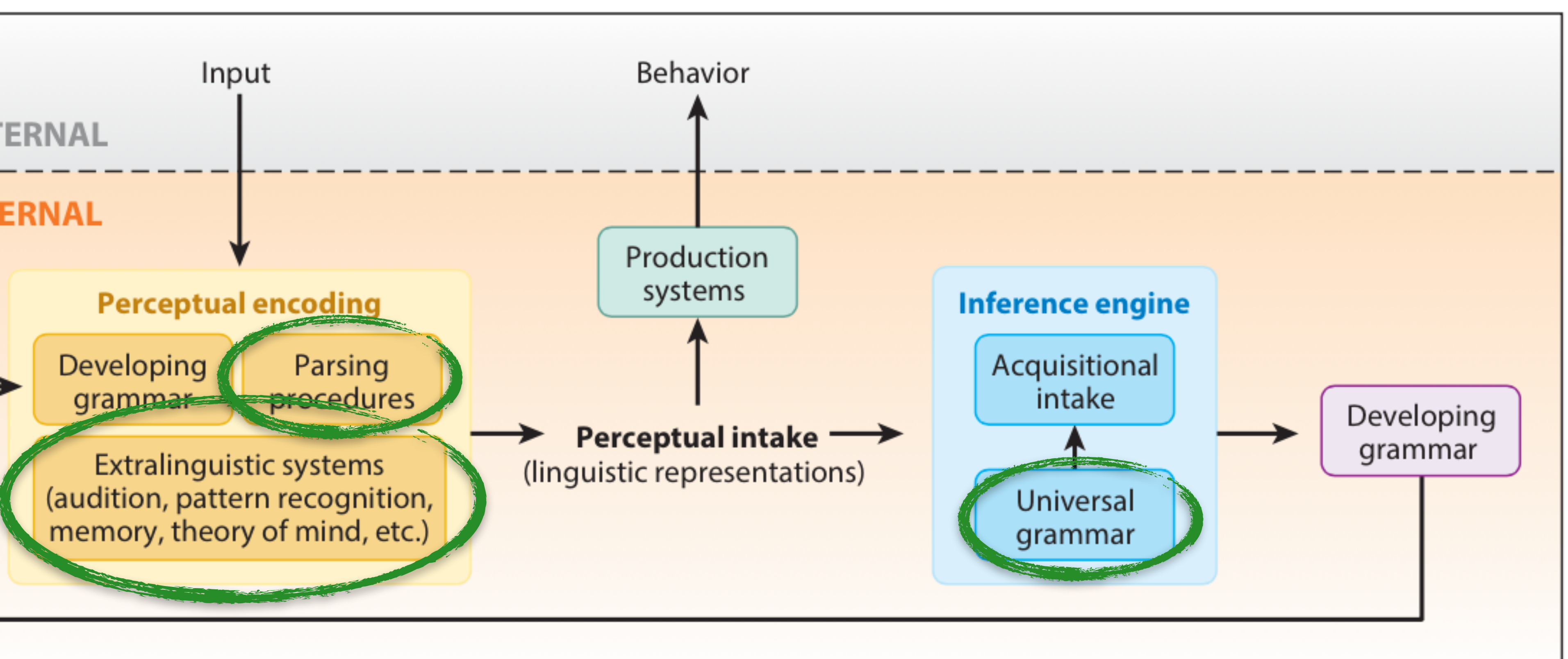
can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … [$_{BN1}$ … [$_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Container nodes

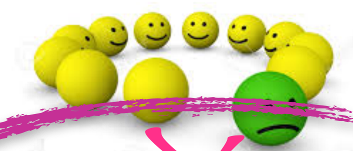
syntax

syntactic island

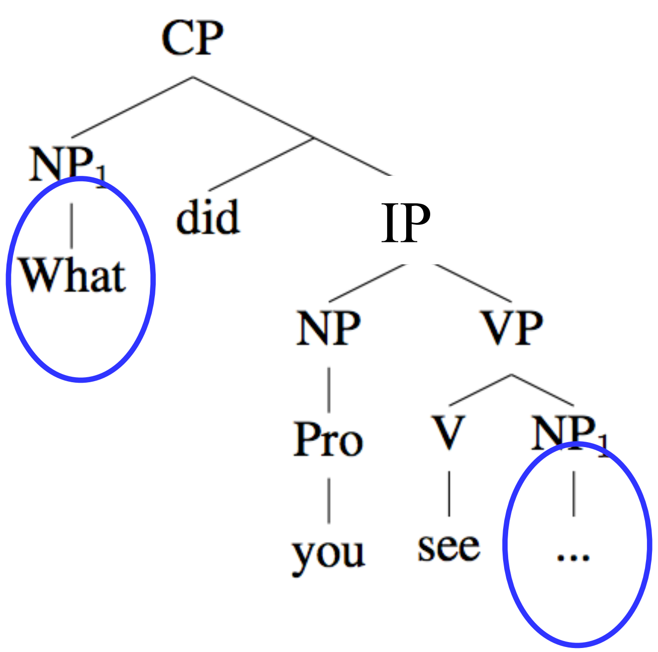
Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]



How to describe this dependency:
What phrases is the gap inside but the *wh*-word isn't inside?

Container nodes

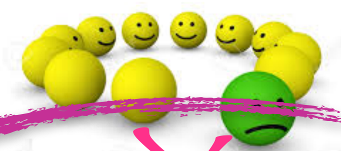

syntax

syntactic island

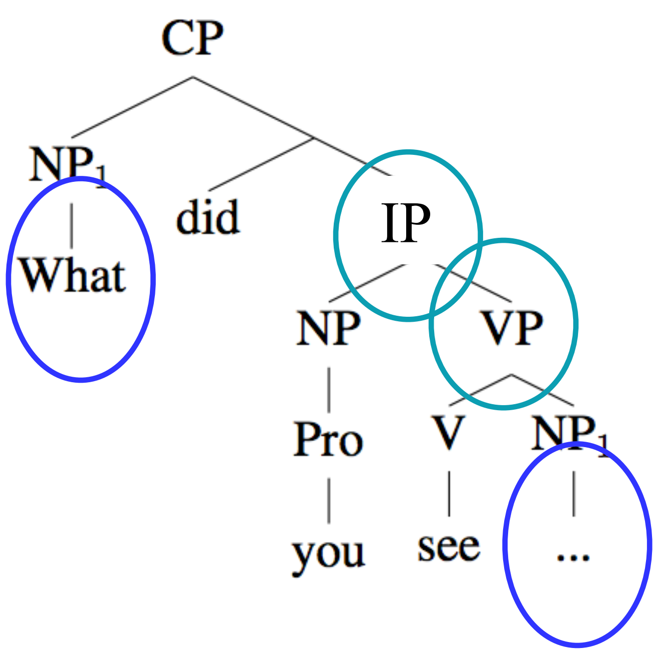
Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]



How to describe this dependency:
What phrases is the gap inside but the *wh*-word isn't inside?

What did you see __?
= What did [$_{IP}$ you [$_{VP}$ see __]]?
= IP-VP

Container nodes

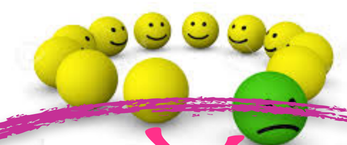
syntax

syntactic island

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

What did you see __?

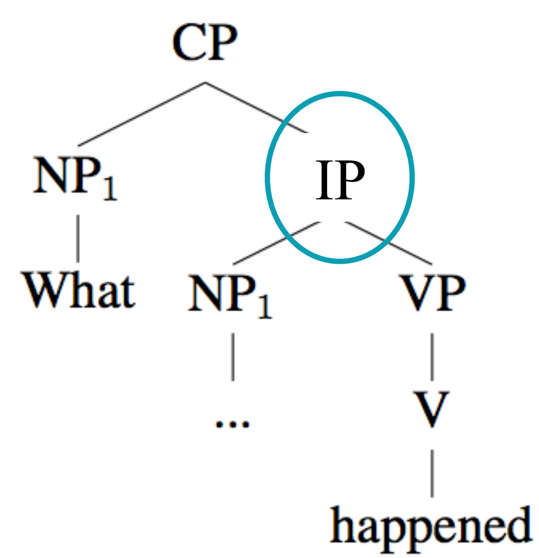
= What did [$_{IP}$ you [$_{VP}$ see __]]?

= IP-VP

What __ happened?

= What [$_{IP}$ __ happened]?

= IP



Container nodes

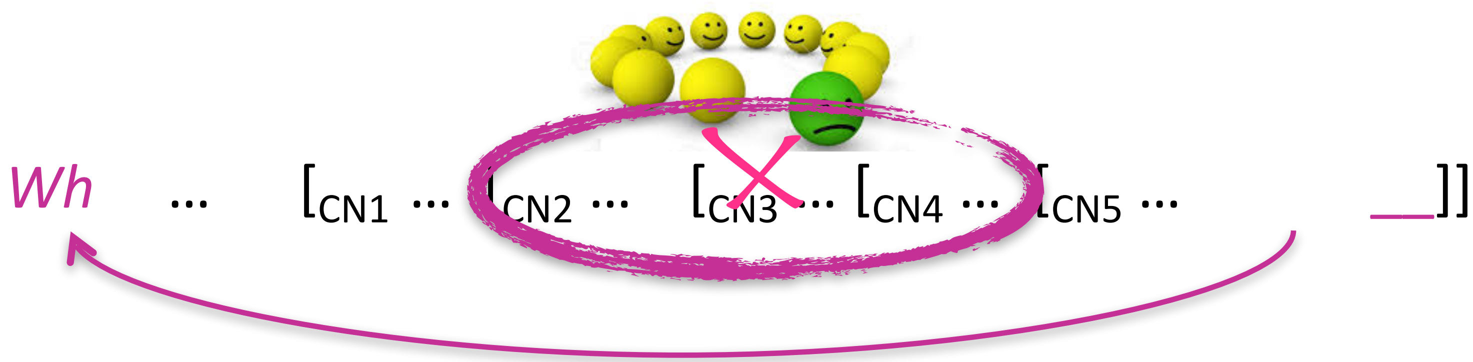
syntax

syntactic island

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



What did you see __?

= What did [$_{IP}$ you [$_{VP}$ see __]]?

= IP-VP

What __ happened?

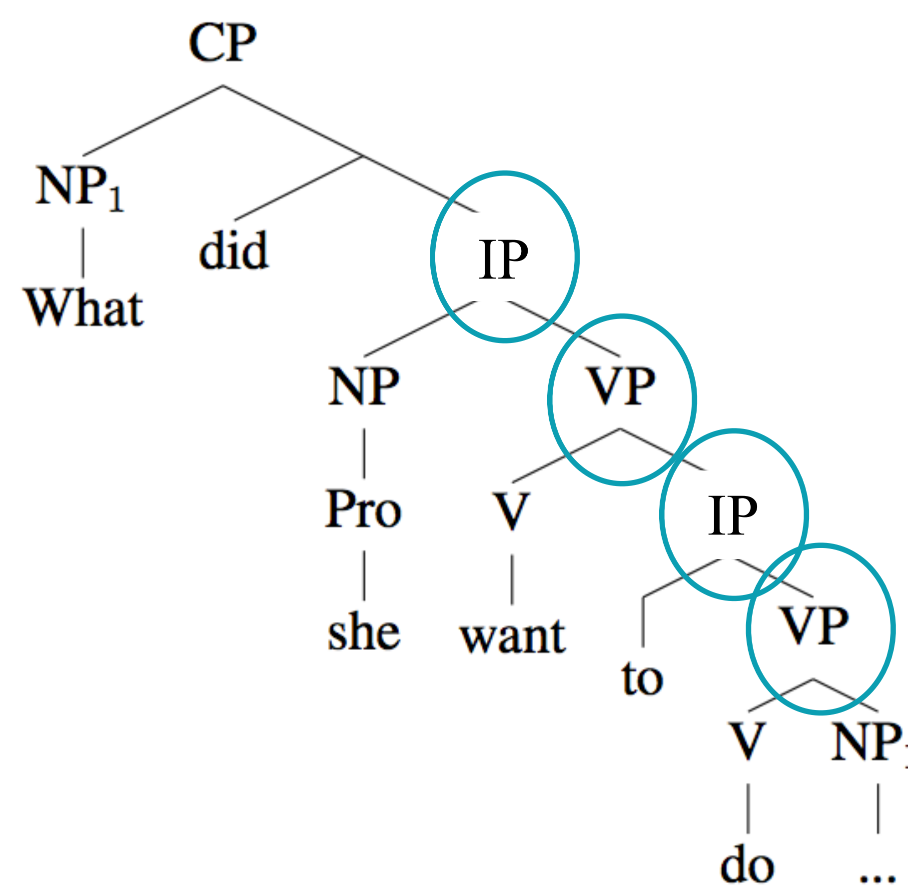
= What [$_{IP}$ __ happened]?

= IP

What did she want to do __ ?

= What did [$_{IP}$ she [$_{VP}$ want [$_{IP}$ to [$_{VP}$ do __]]]]?

= IP-VP-IP-VP



Container nodes

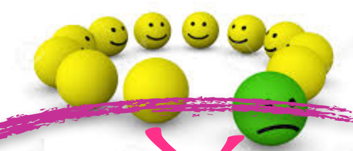

syntax

syntactic island

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



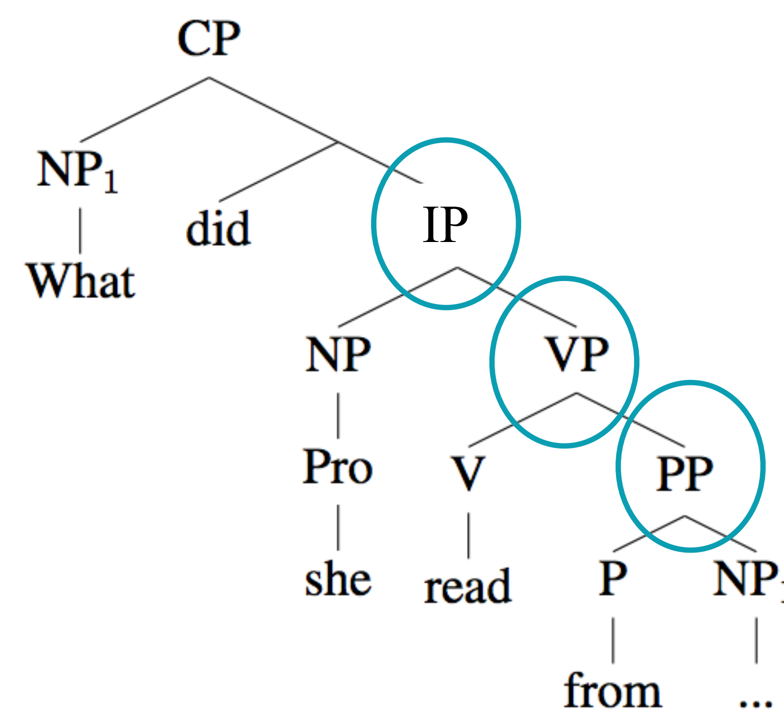
Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

What did you see __?
= What did $[_{IP}$ you $[_{VP}$ see __]]?
= IP-VP

What __ happened?
= What $[_{IP}$ __ happened]?
= IP

What did she want to do __ ?
= What did $[_{IP}$ she $[_{VP}$ want $[_{IP}$ to $[_{VP}$ do __]]]]?
= IP-VP-IP-VP

What did she read from __ ?
= What did $[_{IP}$ she $[_{VP}$ read $[_{PP}$ from __]]]]?
= IP-VP-PP



Learning strategies


syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

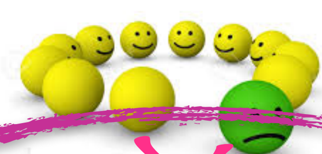
can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Container node: phrase structure node that contains dependency

$[_{CP}$ *What* *do* $[_{IP}$ *you* $[_{VP}$ *like* __ $[_{PP}$ *in this picture?*]]]]

Learning strategies


syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

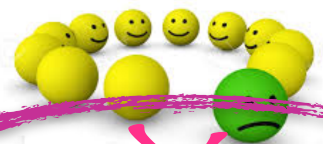
can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Sequence of container nodes characterizes dependencies

$[_{CP}$ *What* *do* $[_{IP}$ *you* $[_{VP}$ *like* __ $[_{PP}$ *in this picture?*]]]]

start-IP-VP-*end*

Learning strategies

syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

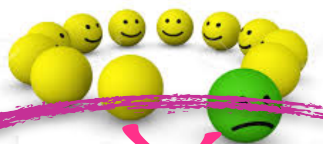
can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Ungrammatical dependencies have low probability segments

$[_{CP}$ *Who* *did* $[_{IP}$ *Lily* $[_{VP}$ *think* $[_{CP}$ $[_{IP}$ $[_{NP}$ *the kitty* $[_{PP}$ *for* __]] *was pretty ?*]]]]

start-IP-VP-CP-IP-NP-PP-*end*

Learning strategies


syntax

syntactic island

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability region of structure

Dependencies represented as a sequence of **container nodes**

Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Low probability container node sequences have to be learned for the language

Learning strategies


syntax

syntactic island

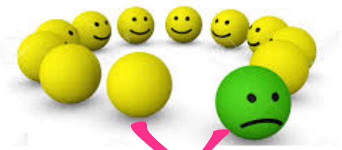
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability sequence of container nodes



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

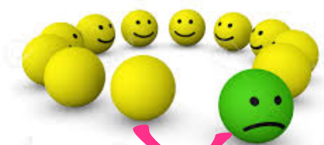
In common: Local structural anomaly is the problem

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

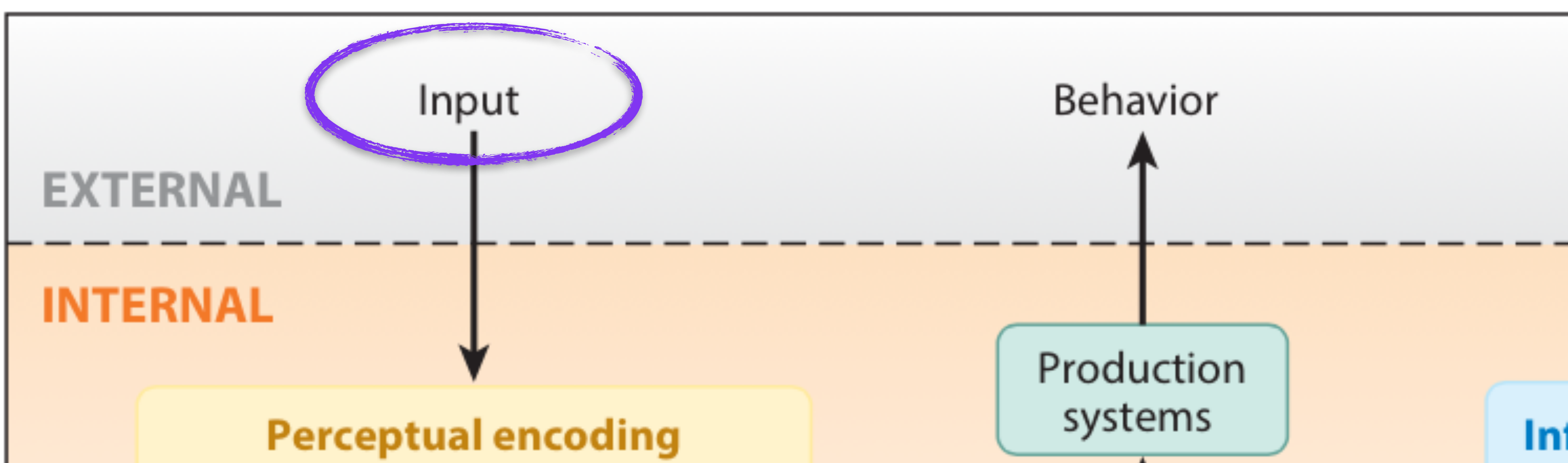
syntactic island

A dependency can't cross a very low probability sequence of container nodes



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Implemented in an algorithmic-level learning model that learned from realistic samples of child-directed speech.

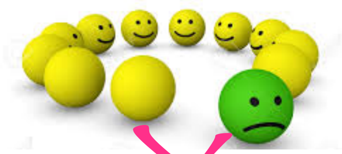


Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island

A dependency can't cross a very low probability sequence of container nodes



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

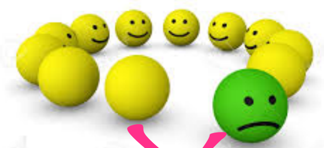
Intuition: Learn what you can from the dependencies you do actually observe in the data and apply it to make a judgment about the dependencies you haven't seen before, like these syntactic islands.

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island

A dependency can't cross a very low probability sequence of container nodes



Wh ... [$_{CN1}$ ... [$_{CN2}$ ... [$_{CN3}$ ... [$_{CN4}$ ... [$_{CN5}$ ... __]]

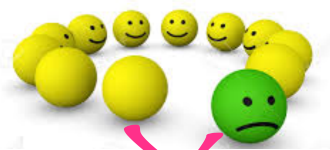
Intuition: Learn what you can from the dependencies you do actually observe in the data and apply it to make a judgment about the dependencies you haven't seen before, like these syntactic islands.

That is, leverage a broader set of data to make syntactic generalizations.

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

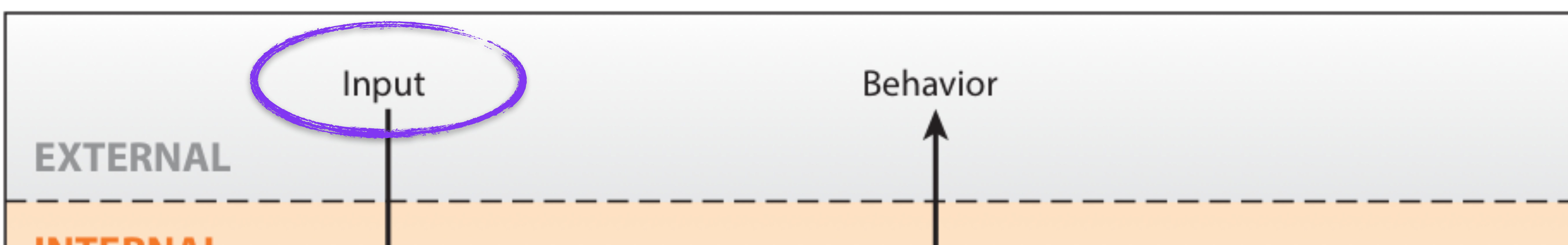
$\boxed{\text{syntax}}$

syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … ___]]

What information is there to leverage exactly?

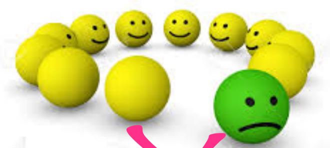


syntax

syntactic island

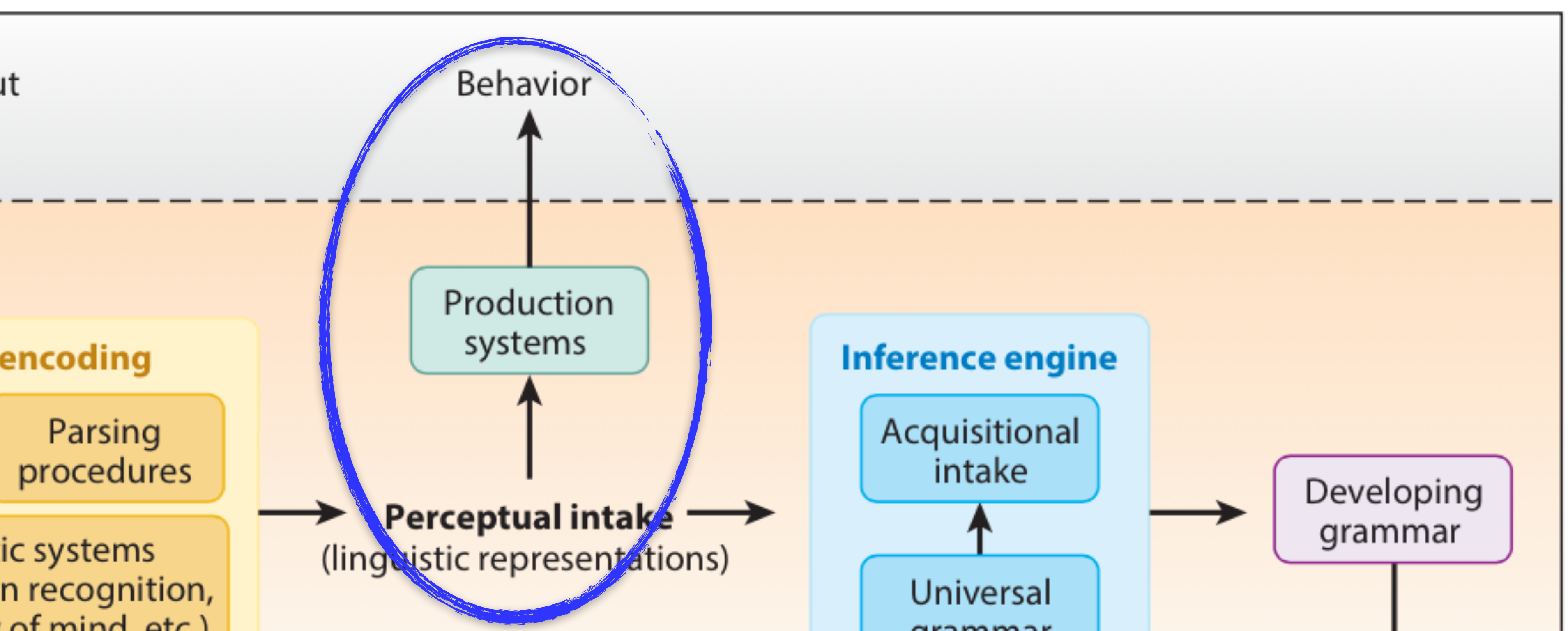
Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

What information is there to leverage exactly?

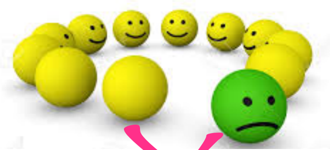
This relates to the strategy children use for learning and then generating predictions about the grammaticality of dependencies.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

What information is there to leverage exactly?

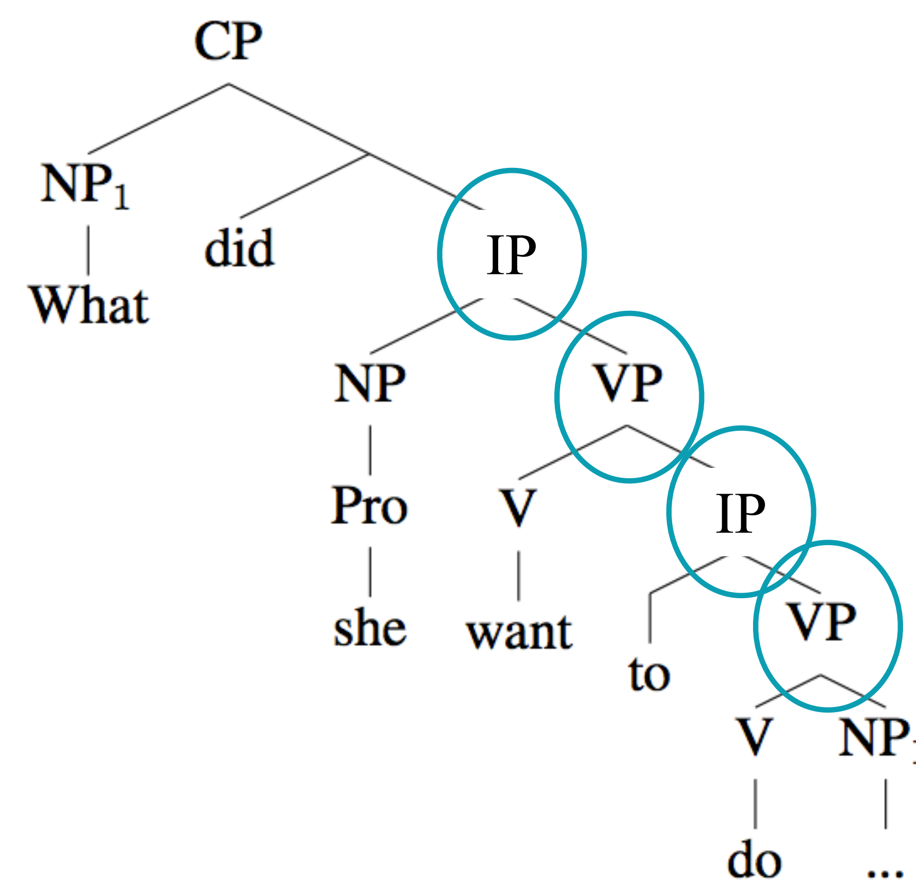
Strategy

(1) Pay attention to the structure of dependencies.

What did she want to do __ ?

= What did [$_{IP}$ she [$_{VP}$ want [$_{IP}$ to [$_{VP}$ do __]]]]?

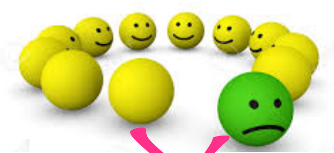
= IP-VP-IP-VP



Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … ___]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break these dependency structures into smaller pieces made up of three units (trigrams) that you can track the frequency of in the input you encounter.

IP-VP =
begin-IP-VP
IP-VP-*end*

IP =
begin-IP-*end*

IP-VP-IP-VP
= *begin*-IP-VP
IP-VP-IP
VP-IP-VP
IP-VP-*end*

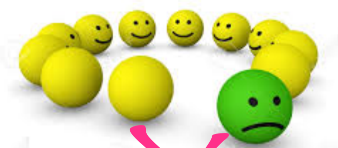
IP-VP-PP
= *begin*-IP-VP
IP-VP-PP
VP-PP-*end*

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break these dependency structures into smaller pieces made up of three units (trigrams) that you can track the frequency of in the input you encounter.

IP-VP =
begin-IP-VP
IP-VP-*end*

IP =
begin-IP-*end*

IP-VP-IP-VP
= *begin*-IP-VP
IP-VP-IP
VP-IP-VP
IP-VP-*end*

IP-VP-PP
= *begin*-IP-VP
IP-VP-PP
VP-PP-*end*

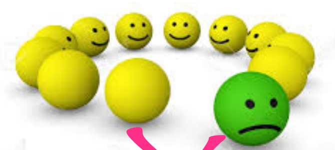
begin-IP-VP = 86/225
IP-VP-*end* = 83/225
begin-IP-*end* = 13/225
IP-VP-IP = 6/225
VP-IP-VP = 6/225
IP-VP-PP = 3/225
VP-PP-*end* = 3/225
…

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break these dependency structures into smaller pieces made up of three units (trigrams) that you can track the frequency of in the input you encounter.

IP-VP =
begin-IP-VP
IP-VP-*end*

IP =
begin-IP-*end*

IP-VP-IP-VP
= *begin*-IP-VP
IP-VP-IP
VP-IP-VP
IP-VP-*end*

IP-VP-PP
= *begin*-IP-VP
IP-VP-PP
VP-PP-*end*

begin-IP-VP = 86/225
IP-VP-*end* = 83/225
begin-IP-*end* = 13/225
IP-VP-IP = 6/225
VP-IP-VP = 6/225
IP-VP-PP = 3/225
VP-PP-*end* = 3/225
…

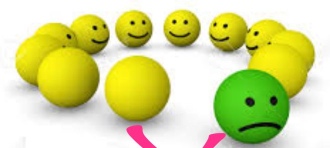
Note that some of these trigrams appear in multiple dependencies that commonly occur in children's input. This will be helpful!

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break dependency structures into trigrams that you can track the frequency of.

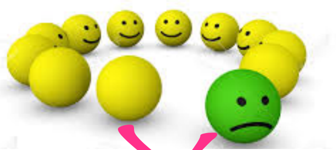
(3) Use trigram frequency to calculate the probability of that trigram occurring in a dependency.

begin-IP-VP = 86/225	p(*begin*-IP-VP) = 0.38
IP-VP-*end* = 83/225	p(IP-VP-*end*) = 0.37
begin-IP-*end* = 13/225	p(*begin*-IP-*end*) = 0.06
IP-VP-IP = 6/225	p(IP-VP-IP) = 0.03
VP-IP-VP = 6/225	p(VP-IP-VP) = 0.03
IP-VP-PP = 3/225	p(IP-VP-PP) = 0.01
VP-PP-*end* = 3/225	p(VP-PP-*end*) = 0.01
…	

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … [$_{\text{CN1}}$ … [$_{\text{CN2}}$ … [$_{\text{CN3}}$ … [$_{\text{CN4}}$ … [$_{\text{CN5}}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break dependency structures into trigrams that you can track the frequency of.

(3) Calculate the trigram probability in a dependency.

(4) When you see a new dependency, break it down into its trigrams and then calculate its probability, based on the trigram probabilities.

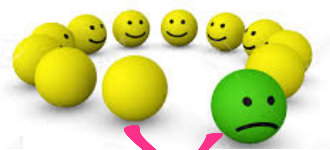
What does Jack want __?
= What does [$_{\text{IP}}$ Jack [$_{\text{VP}}$ want __]]?
= IP-VP
= *begin*-IP-VP
IP-VP-*end*

p(IP-VP) = p(*begin*-IP-VP)*p(IP-VP-*end*)
= 0.38 * 0.37 = 0.14

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break dependency structures into trigrams that you can track the frequency of.

(3) Calculate the trigram probability in a dependency.

(4) When you see a new dependency, break it down into its trigrams and then calculate its probability, based on the trigram probabilities.

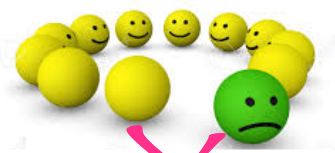
What does Jack want to do that for __?
= What does $[_{IP}$ Jack $[_{VP}$ want $[_{IP}$ to $[_{VP}$ do that $[_{PP}$ for __]]?
= IP-VP-IP-VP-PP
= *begin*-IP-VP
IP-VP-IP
VP-IP-VP
IP-VP-PP
VP-PP-*end*

p(IP-VP-IP-VP-PP) = p(*begin*-IP-VP)*p(IP-VP-IP)*p(VP-IP-VP)*p(IP-VP-PP)*p(VP-PP-*end*)
= 0.38*0.03*0.03*0.01*0.01 = 0.000000034

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break dependency structures into trigrams that you can track the frequency of.

(3) Calculate the trigram probability in a dependency.

(4) When you see a new dependency, break it down into its trigrams and then calculate its probability, based on the trigram probabilities.

Subject island dependency

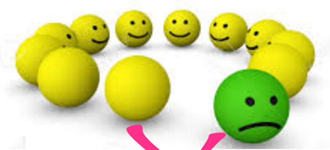
What do you think that the joke about __ offended Jack?
= What do $[_{IP}$ you $[_{VP}$ think $[_{CP}$ that $[_{IP}$ $[_{NP}$ the joke $[_{PP}$ about __]]]]]] offended Jack?
= IP-VP-CP-NP-PP
= *begin*-IP-VP
IP-VP-CP
VP-CP-IP
CP-IP-NP
IP-NP-PP
NP-PP-*end*

p(IP-VP-CP-IP-NP-PP) = p(*begin*-IP-VP)*p(IP-VP-CP)*p(VP-CP-S)*p(CP-IP-NP)*p(IP-NP-PP)*p(NP-PP-*end*)
= 0.86*0.01*0.001*0.00*0.00*0.02 = 0.00

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{\cancel{CN3}}$ … $[_{CN4}$ … $[_{CN5}$ … ___]]

What information is there to leverage exactly?

Strategy

(1) Pay attention to dependency structure.

(2) Break dependency structures into trigrams that you can track the frequency of.

(3) Calculate the trigram probability in a dependency.

(4) Break a new dependency into its trigrams and calculate its probability.

(5) Use calculated dependency probabilities as the basis for grammaticality judgments. Lower probability dependencies are dispreferred, compared to higher probability dependencies.

p(IP-VP) = 0.14

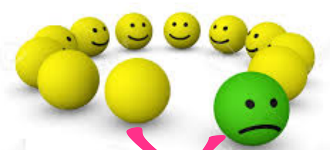
p(IP-VP-IP-VP-PP) = 0.000000034

p(IP-VP-CP-IP-NP-PP) = 0.00

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

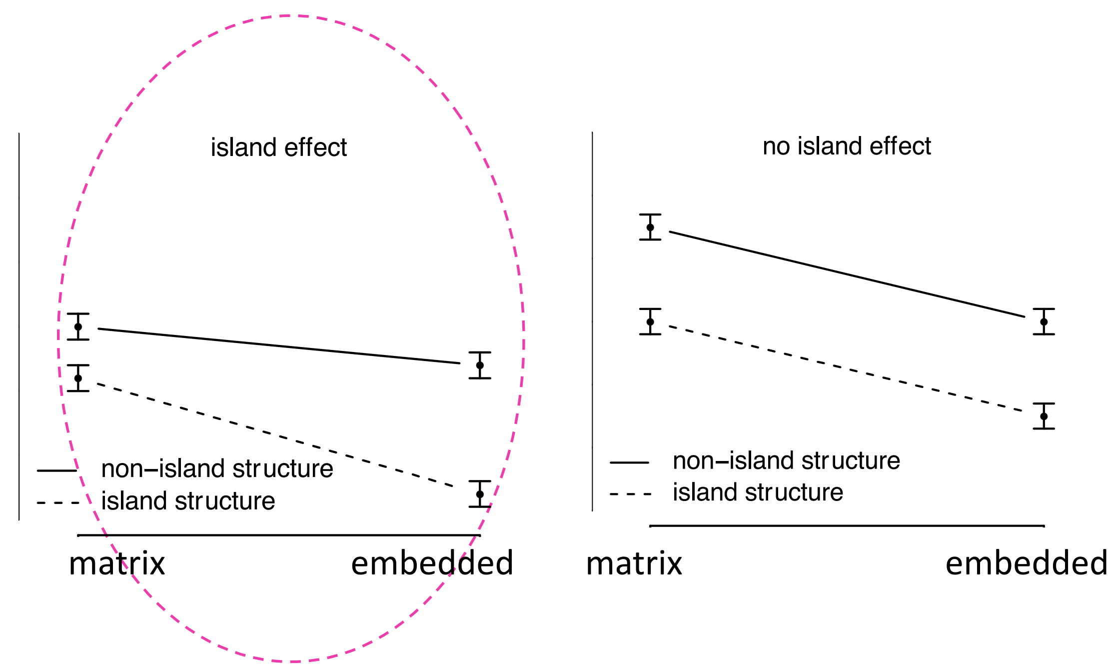
syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … ___]]

Use calculated dependency probabilities as the basis for grammaticality judgments. Lower probability dependencies are dispreferred, compared to higher probability dependencies.

For each set of island stimuli from Sprouse et al. (2012), we generate grammaticality preferences for the modeled learner based on the dependency's perceived probability and use this as a stand-in for acceptability.

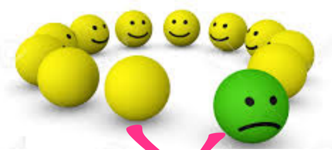


Looking for superadditivity as a sign of syntactic island knowledge

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

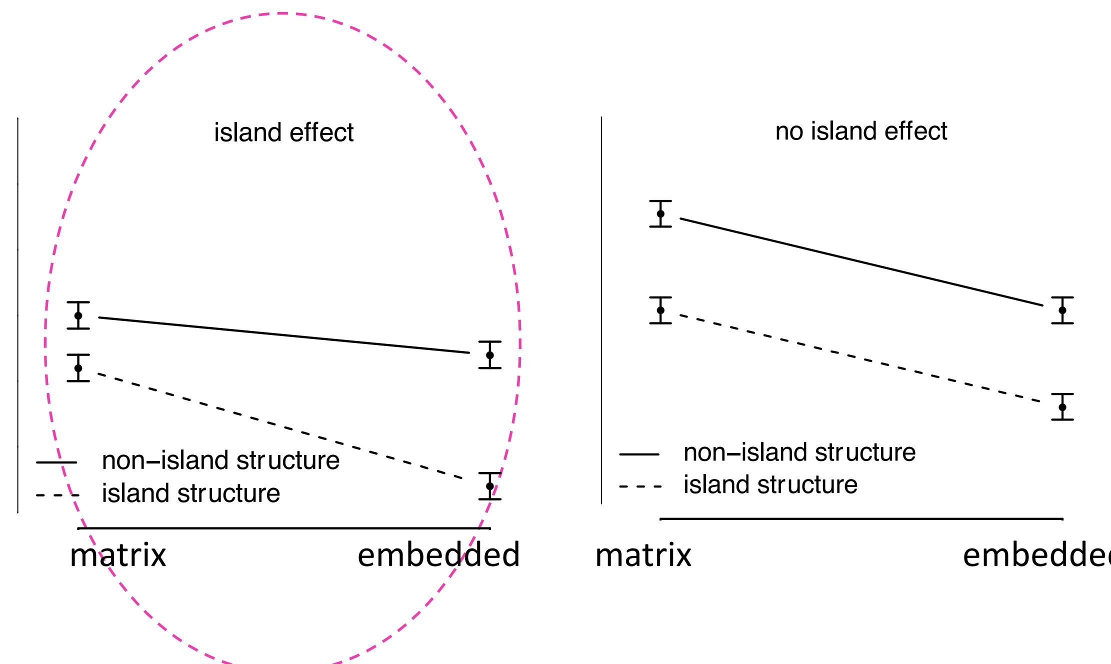
syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

Use calculated dependency probabilities as the basis for grammaticality judgments. Lower probability dependencies are dispreferred, compared to higher probability dependencies.

non-island	*Who __ claimed that Lily forgot the necklace?*	*What did the teacher claim that Lily forgot __?*
island	*Who __ made the claim that Lily forgot the necklace?*	**What did the teacher make the claim that Lily forgot __?*
	matrix	embedded

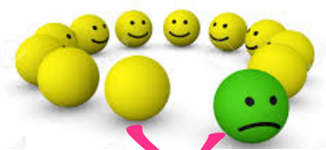


Looking for superadditivity as a sign of syntactic island knowledge

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

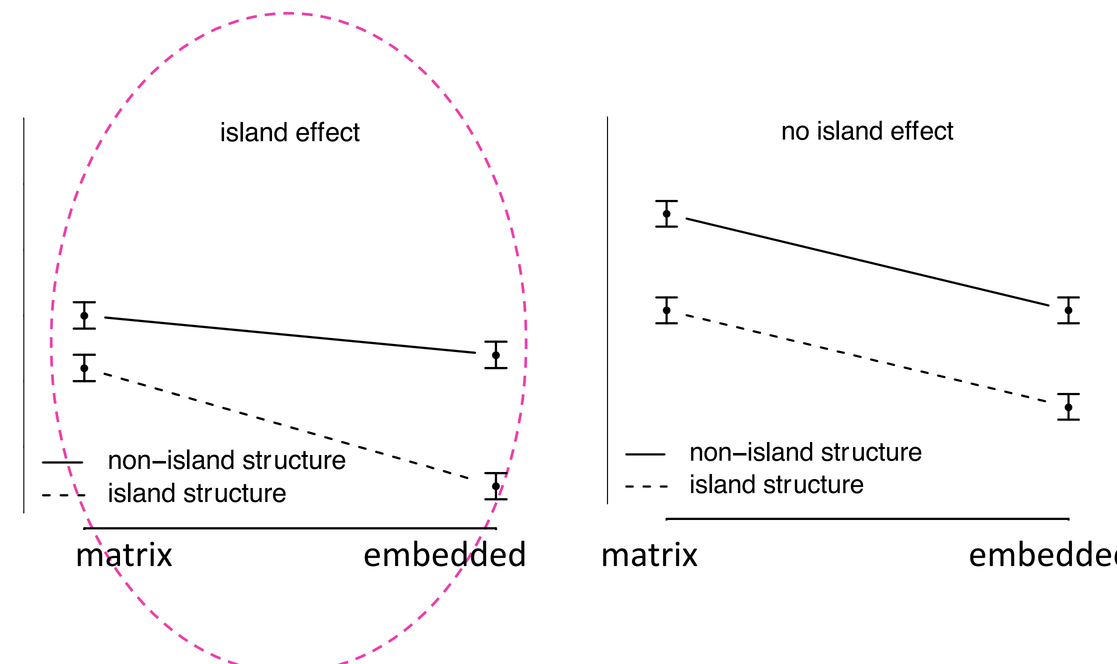
syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … ___]]

Use calculated dependency probabilities as the basis for grammaticality judgments. Lower probability dependencies are dispreferred, compared to higher probability dependencies.

	matrix	embedded
non-island	 *IP*	 *IP-VP-CP$_{that}$-IP-VP*
island	 *IP*	 **IP-VP-NP-CP$_{that}$-IP-VP*



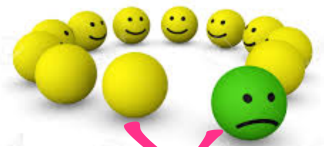
Each dependency is characterized by a container node sequence, whose probability can be calculated and then plotted.

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island

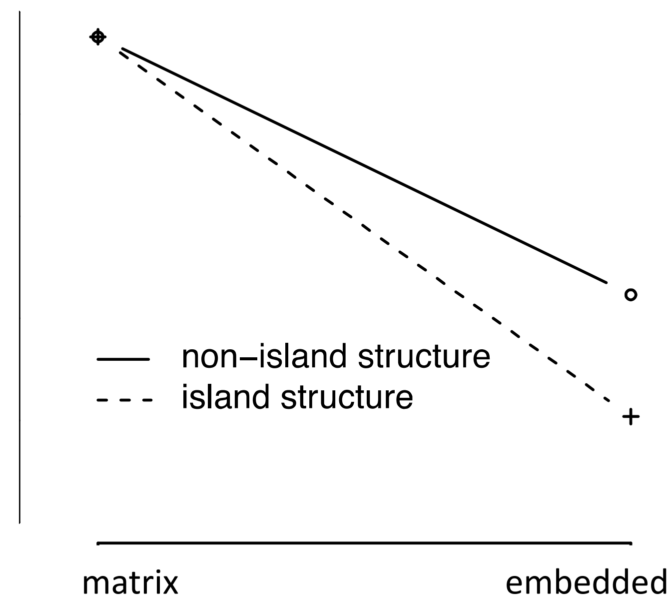


Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … ___]]

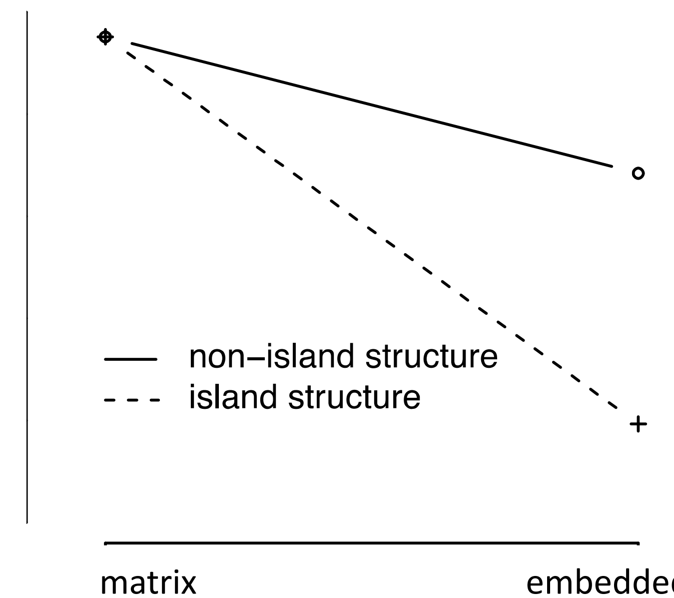
Superadditivity observed for all four islands — the qualitative behavior suggests that this learner has knowledge of these syntactic islands.

The Subjacency-ish representation that relies on container node trigram probabilities can solve this learning problem using this learning strategy.

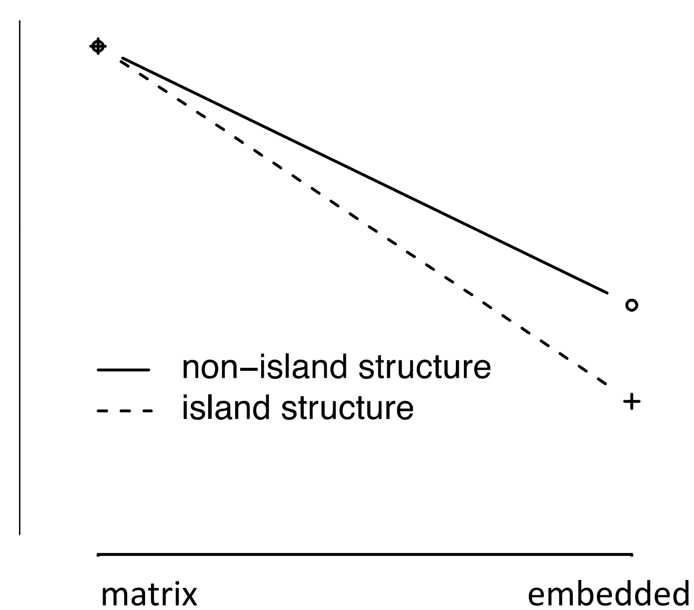
Complex NP



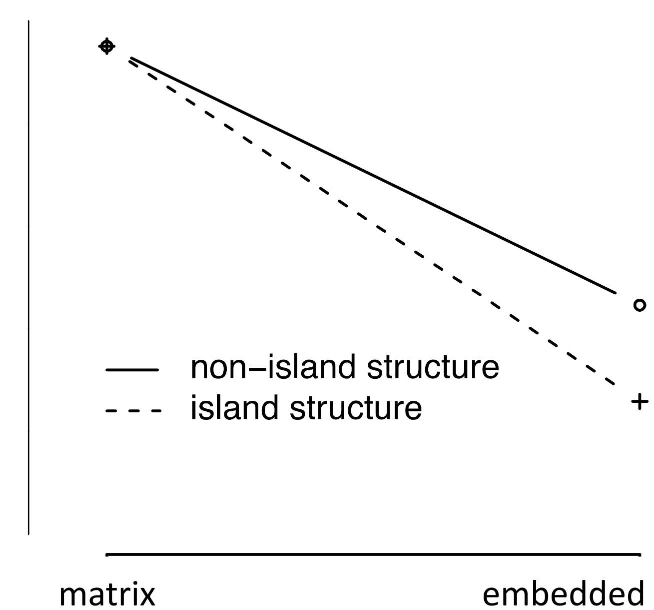
Subject



Whether



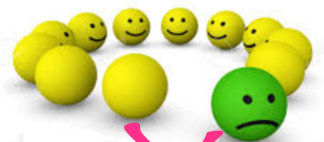
Adjunct



✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

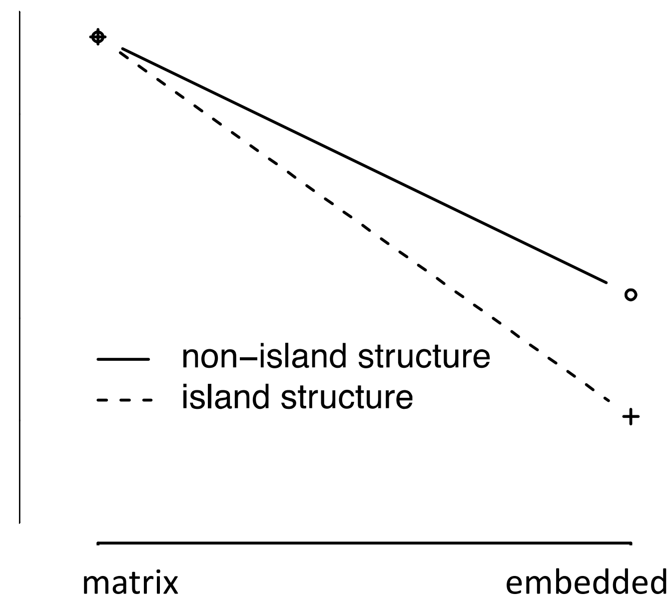
syntactic island



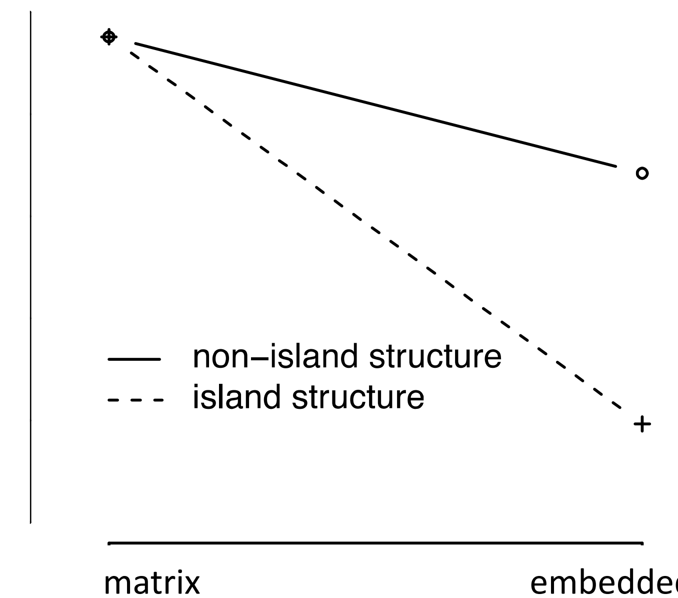
Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

Note: We're careful to say "qualitative" behavior fit because there are lots of other factors that impact acceptability judgment behavior, and we've only modeled one (presumably) large part of them, which is the grammaticality of the dependency.

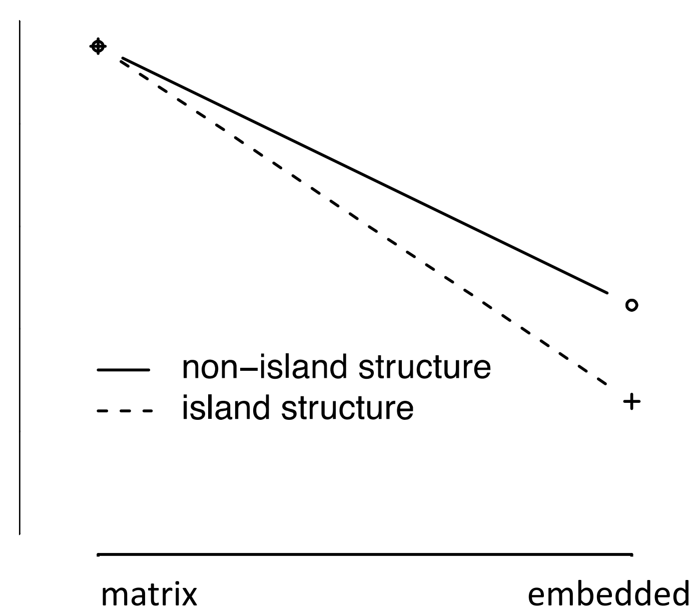
Complex NP



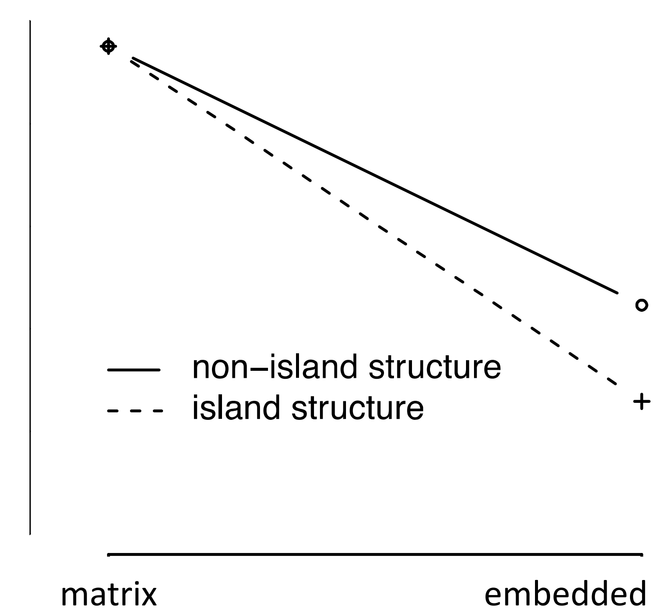
Subject



Whether



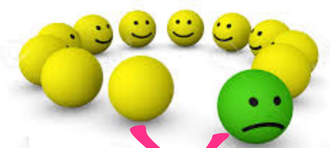
Adjunct



✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island

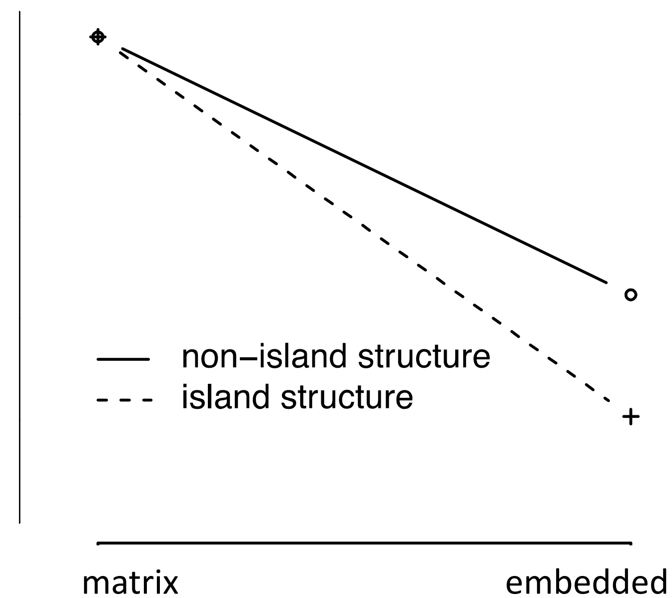


Wh … [$_{CN1}$ … [$_{CN2}$ … ~~[$_{CN3}$~~ … [$_{CN4}$ … [$_{CN5}$ … __]]

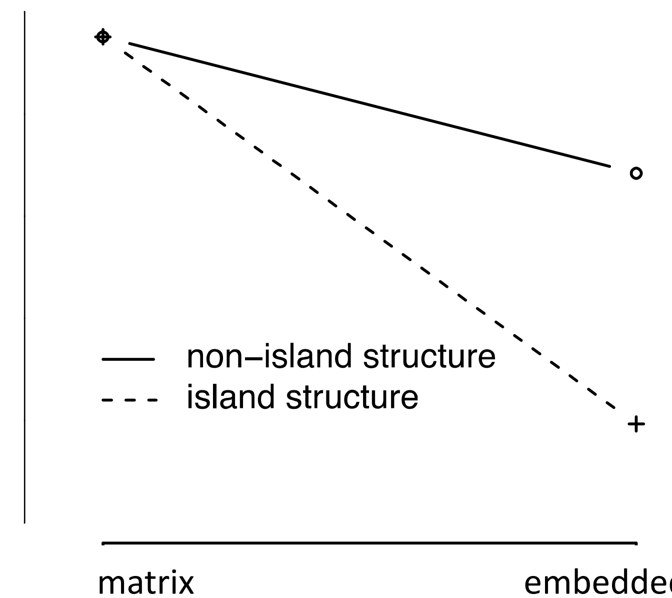
But is this all we can say?

No! One useful aspect of models is that we can look inside the modeled child to see *why* it's behaving the way that it is. (This is something that's harder to do with real children — that is, opening up their minds and seeing how they work.)

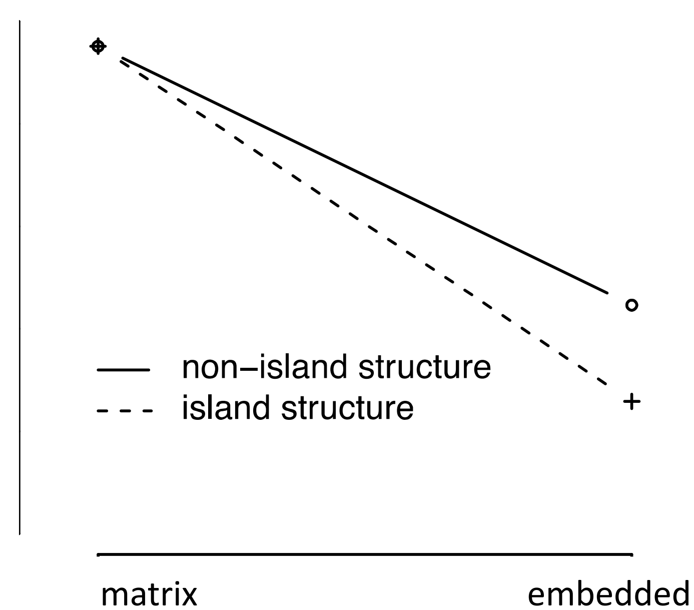
Complex NP



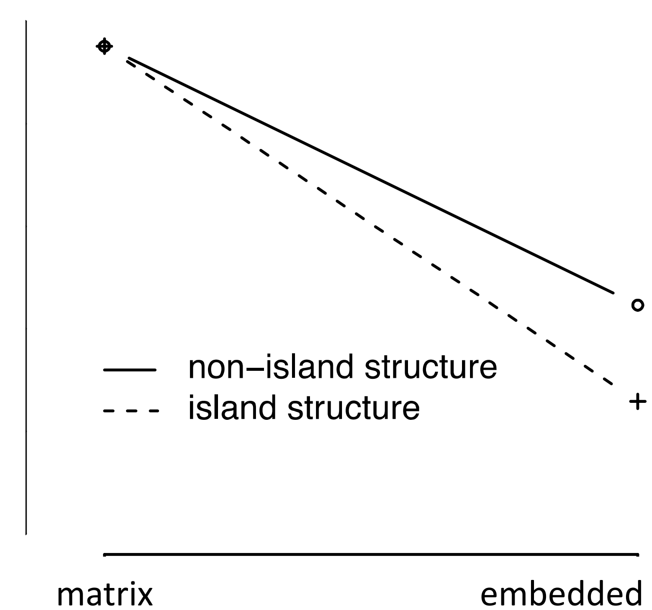
Subject



Whether



Adjunct

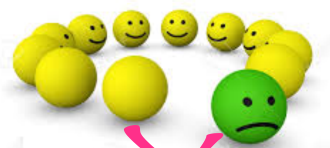


✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $\cancel{[_{CN3}}$ … $[_{CN4}$ … $[_{CN5}$ … ___]]

What's going on?

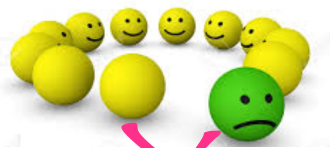
Why are the island-spanning dependencies so much worse than the grammatical ones?

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … ___]]

What's going on?

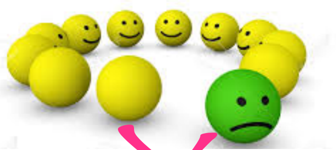
Why are the island-spanning dependencies so much worse than the grammatical ones?

Let's look inside them and see!

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

Let's look inside them and see!

It turns out that each island-spanning dependency contains at least one very low probability container node trigram. So these are the relevant "island" representations.

a. Complex NP

(i) * What did [$_{IP}$ the teacher [$_{VP}$ make [$_{NP}$ the claim $_{CP_{that}}$ that [$_{IP}$ Lily $_{VP}$ forgot __]]]]]?

(ii) *start*-IP-VP-NP-CP$_{that}$-IP-VP-*end*

(iii) Low probability:

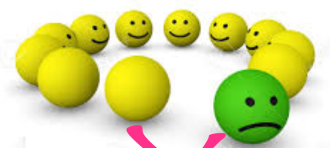
VP-NP-CP$_{that}$

NP-CP$_{that}$-IP

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $\cancel{[_{CN3}}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Let's look inside them and see!

It turns out that each island-spanning dependency contains at least one very low probability container node trigram. So these are the relevant "island" representations.

b. Subject

(i) * Who does $[_{IP}$ Jack $[_{VP}$ think $[_{CP_{null}}$ $[_{IP}$ $[_{NP}$ the necklace $[_{PP}$ for __]] is expensive]]]]?

(ii) *start*-IP-VP-CP_{null}-IP-NP-PP-*end*

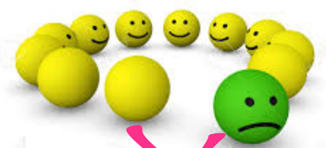
(iii) Low probability:

CP_{null}-IP-NP

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Let's look inside them and see!

It turns out that each island-spanning dependency contains at least one very low probability container node trigram. So these are the relevant "island" representations.

c. Whether

(i) * What does $[_{IP}$ the teacher $[_{VP}$ wonder $[_{CP_{whether}}$ whether $[_{IP}$ Jack $[_{VP}$ stole __]]]]]?

(ii) *start*-IP-VP-CP$_{whether}$-IP-VP-*end*

(iii) Low probability:

IP-VP-CP$_{whether}$

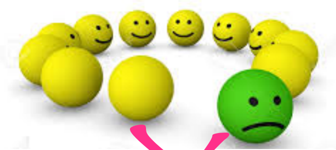
VP-CP$_{whether}$-IP

CP$_{whether}$-IP-VP

✓ Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)


syntax

syntactic island



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Let's look inside them and see!

It turns out that each island-spanning dependency contains at least one very low probability container node trigram. So these are the relevant "island" representations.

d. Adjunct

(i) * What does $[_{IP}$ the teacher $[_{VP}$ worry $[_{CP_{if}}$ if $[_{IP}$ Lily $[_{VP}$ forgot __]]]]]?

(ii) *start*-IP-VP-CP_{if}-IP-VP-*end*

(iii) Low probability:

IP-VP-CP_{if}

VP-CP_{if}-IP

CP_{if}-IP-VP

Learning strategies

syntax

syntactic island

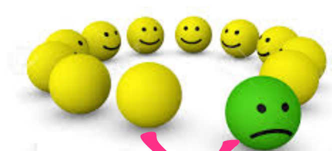
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

can't cross 2+ bounding nodes
from a fixed set (CP, IP, and/or NP)

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

A dependency can't cross a very low probability sequence of container nodes



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

In common: Local structural anomaly is the problem

The way Subjacency-ish implements this local structural anomaly can allow the development of syntactic island knowledge without relying on prior knowledge about bounding nodes and how many a dependency is limited to crossing.

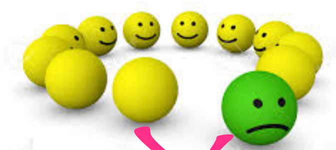
Less reliance on island-specific prior knowledge

Learning strategies


syntax

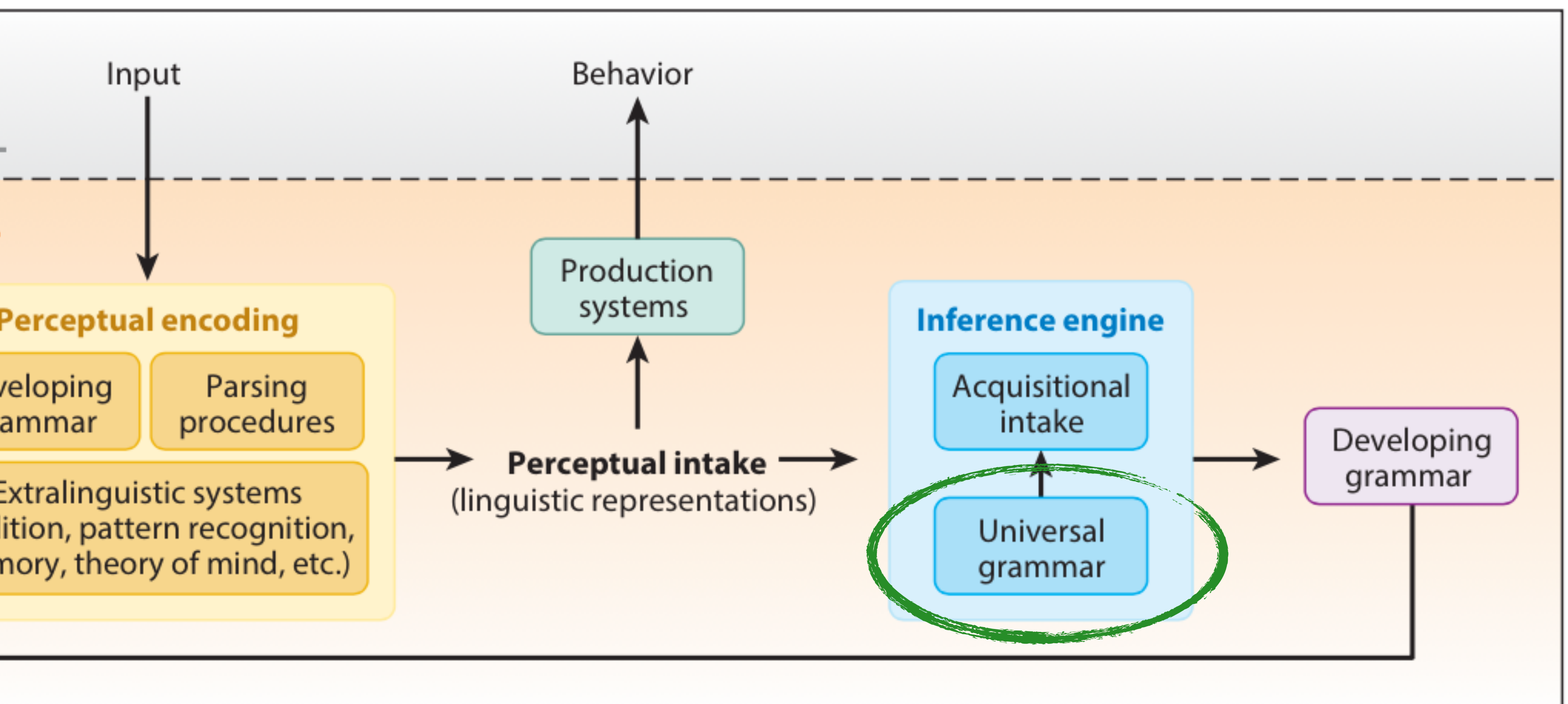
syntactic island

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)



Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Less reliance on island-specific prior knowledge



Today's Plan:
Computational models of syntactic acquisition

I. Some non-parametric examples

Who does … is pretty?

syntax

another one

syntax, semantics

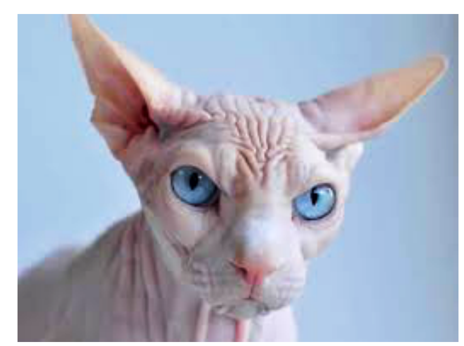
Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Look — there's another one!"



Pronoun interpretation

syntax, semantics

another one

antecedent

"Oh look — a pretty kitty!"

"Look — there's another one!"

Interpretation: another pretty kitty

same
syntactic category
as antecedent
???

Pronoun interpretation

syntax, semantics

another one

antecedent

"Oh look — a pretty kitty!"

"Look — there's another one!"

Interpretation: another

same
syntactic category
as antecedent

???

bigger than a plain **Noun**

Noun
|
pretty **kitty**

Pronoun interpretation

syntax, semantics

another one

antecedent

"Oh look — a pretty kitty!"

"Look — there's another one!"

Interpretation: another ~~the~~ pretty kitty

same
syntactic category
as antecedent
???

smaller than a full **Noun Phrase**

Noun Phrase

the

Noun

pretty kitty

Pronoun interpretation

syntax, semantics

another one

antecedent

"Oh look — a pretty kitty!"

"Look — there's another one!"

Interpretation: another

same
syntactic category
as antecedent
???

In-between category **Noun'**
that includes strings with nouns
and modifiers+nouns

Noun Phrase

the **Noun'**

Noun'

Noun

pretty kitty

Pronoun interpretation

syntax, semantics

another one

antecedent

"Oh look — a pretty kitty!"

"Look — there's another one!"

Interpretation: another

same
syntactic category
as antecedent

Noun Phrase

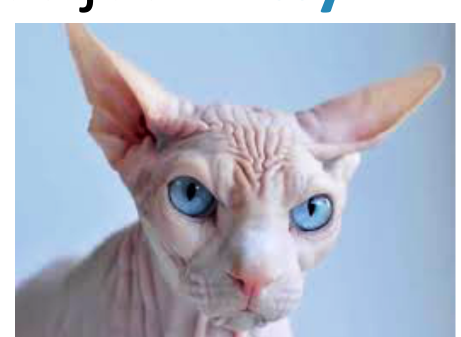
the Noun'

Noun'

Noun

pretty **kitty**

This is why we can also interpret one as just **kitty.**



Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another one?"

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another one?"

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

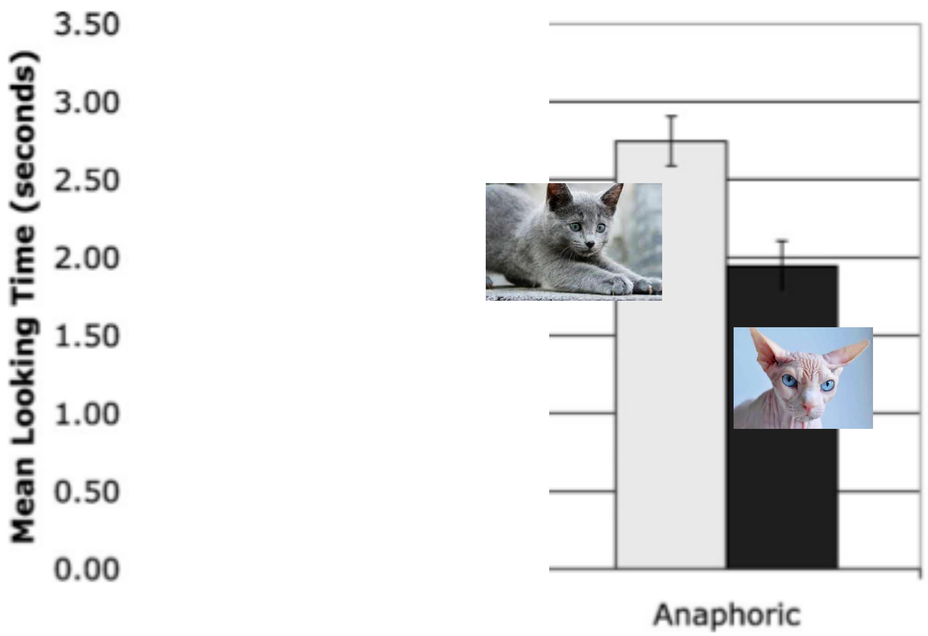
"Oh look — a pretty kitty!"

"Do you see another one?"

pretty kitty

Noun'

J. Lidz et al. / Cognition 89 (2003) B65–B73



Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"What do you see now?"

another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"What do you see now?"

another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

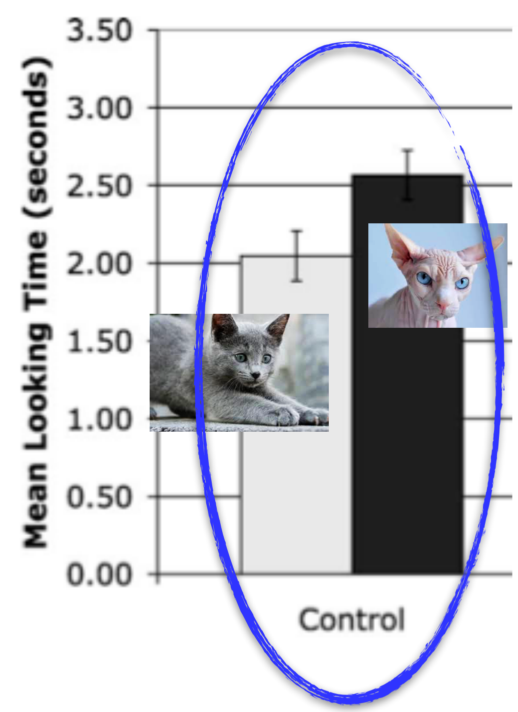
syntax, semantics

another one

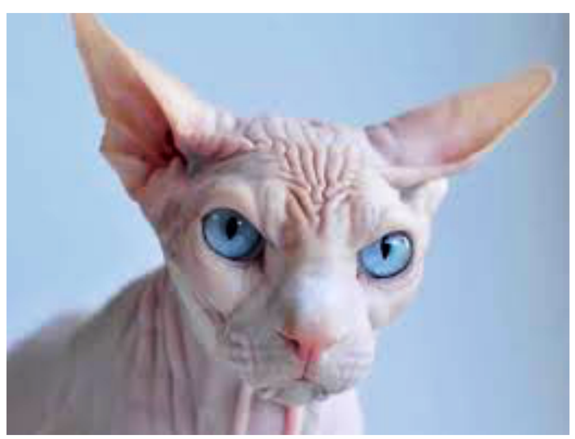
"Oh look — a pretty kitty!"

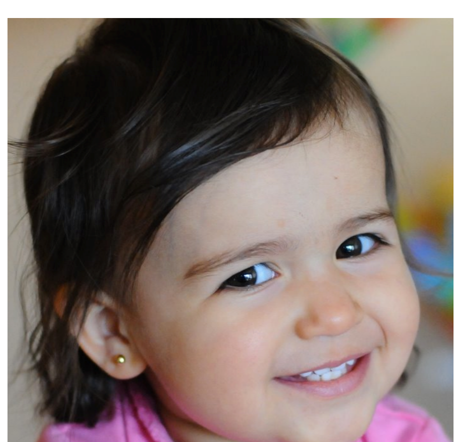
Shows baseline looking preference

J. Lidz et al. / Cognition 89 (2003) B65–B73



"What do you see now?"





another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

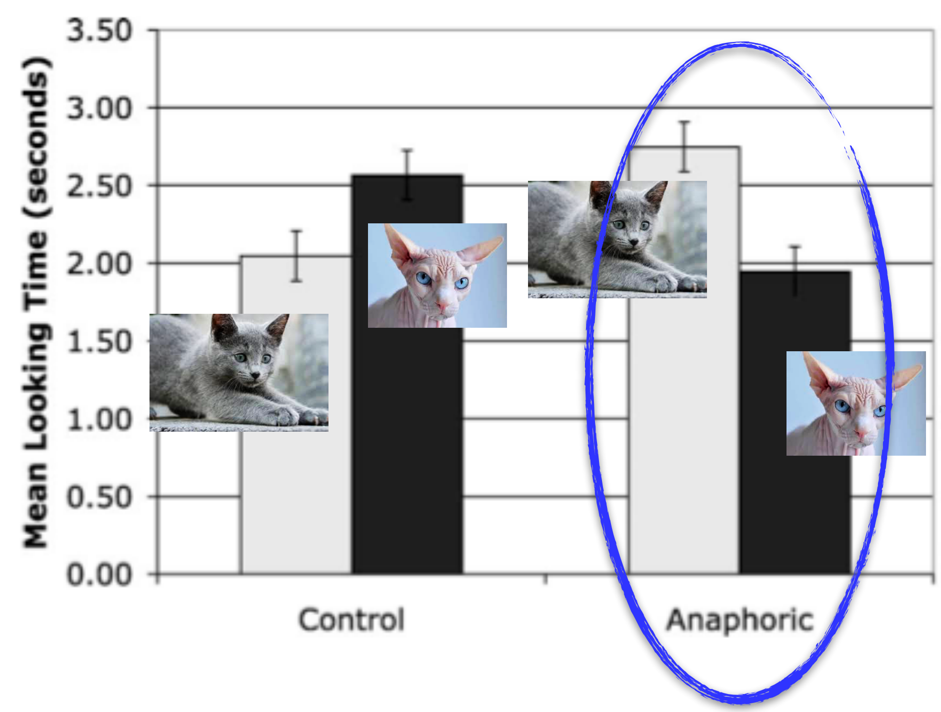
another one

"Oh look — a pretty kitty!"

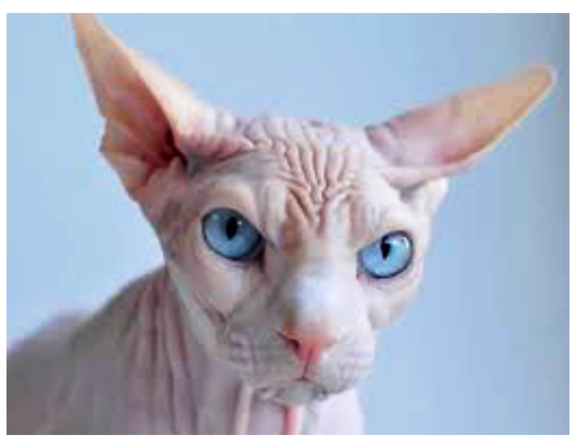
Shows baseline looking preference

which is counteracted with "Do you see another one?"

J. Lidz et al. / Cognition 89 (2003) B65–B73



"What do you see now?"



another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another kitty?"

another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another kitty?"

another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

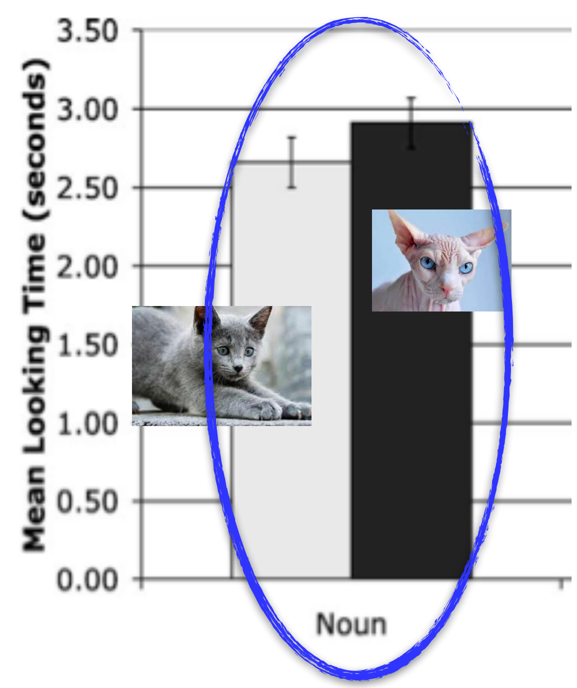
syntax, semantics

another one

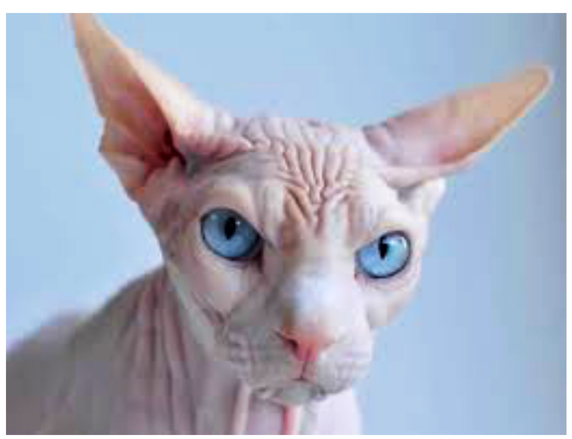
"Oh look — a pretty kitty!"

Shows baseline looking preference

J. Lidz et al. / Cognition 89 (2003) B65–B73



"Do you see another kitty?"



another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another pretty kitty?"

another one

pretty kitty

Noun'

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

"Oh look — a pretty kitty!"

"Do you see another pretty kitty?"

another one

pretty kitty

Noun'

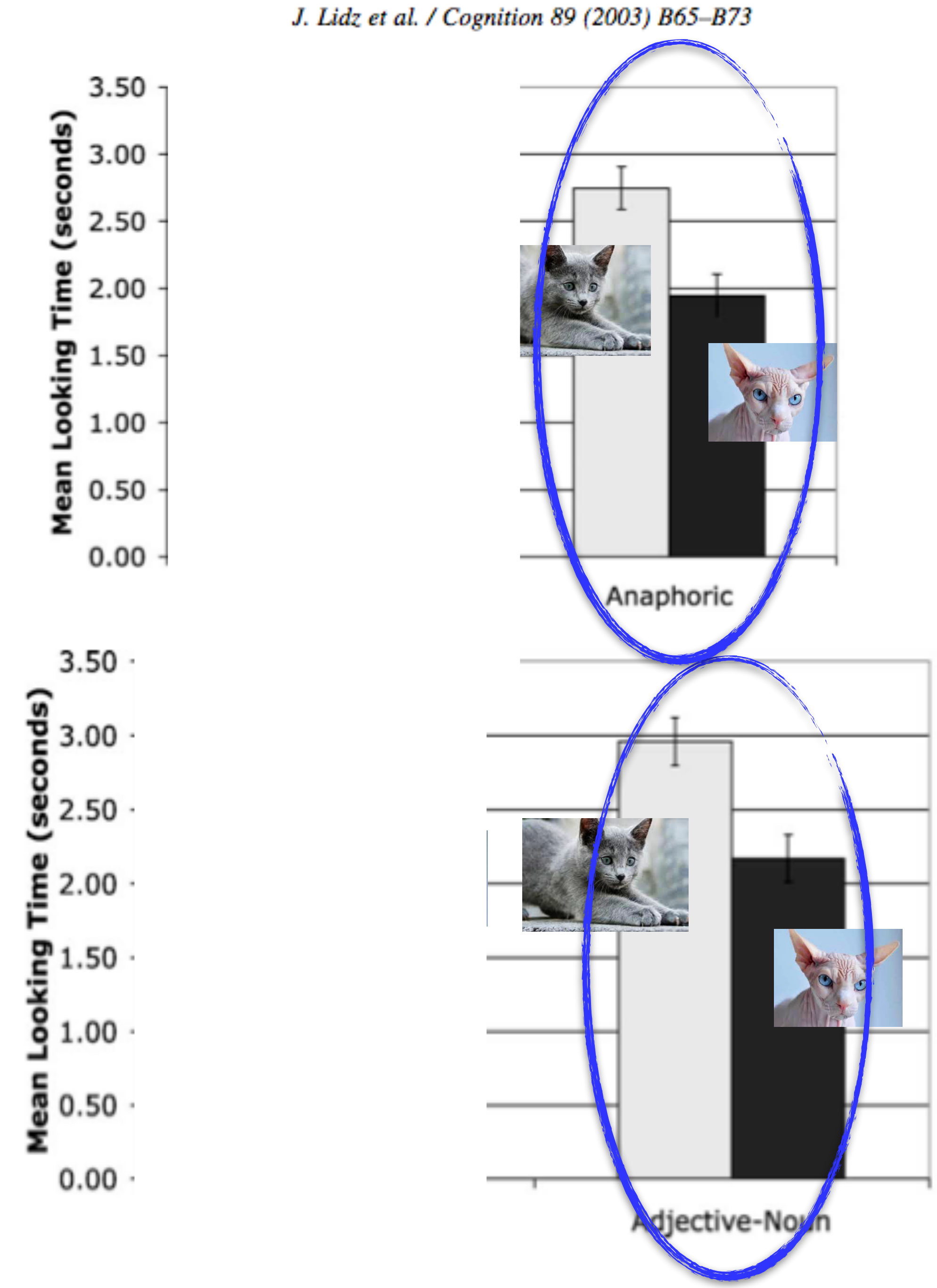
Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

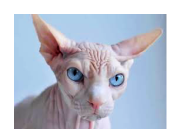
another one

“Oh look — a pretty kitty!”



Same looking pattern as “another one”

“Do you see another pretty kitty?”



another one

pretty kitty

Noun’

Lidz, Waxman, & Freedman 2003:
18-month-old interpretations

Pronoun interpretation

syntax, semantics

another one

“Oh look — a pretty kitty!”

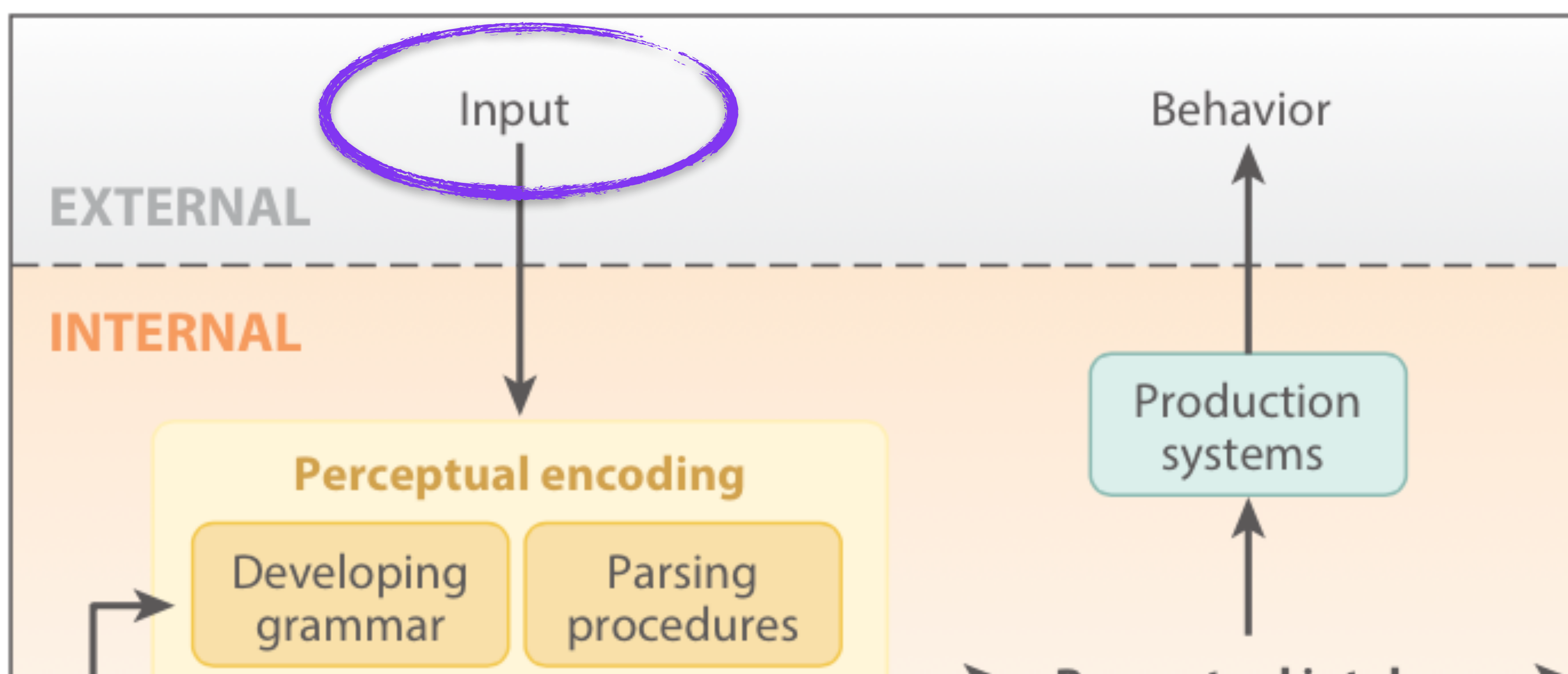
Noun’

pretty kitty

“Do you see another one ?”

Several learning strategies implemented with **algorithmic-level** modeled learners, given realistic samples of English child-directed speech.

Pearl & Mis 2016



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

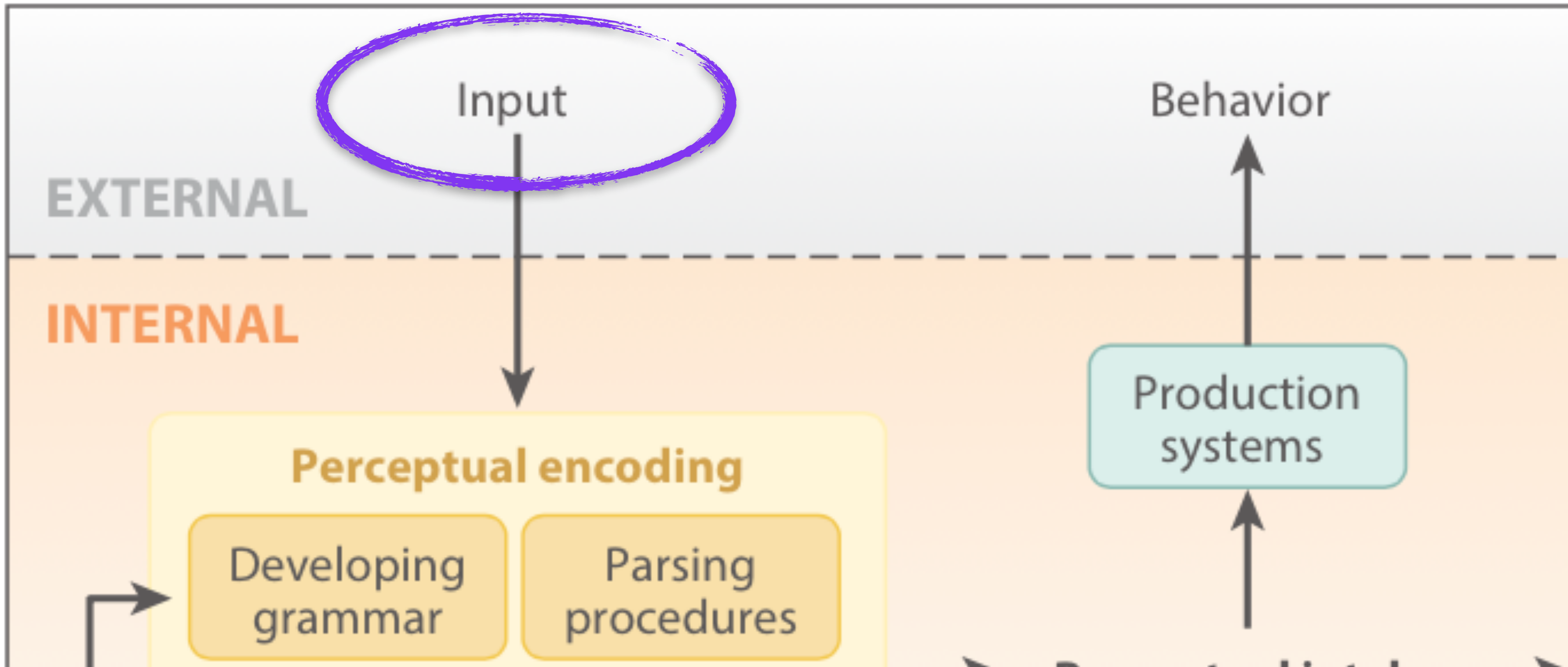
English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Syntactically (SYN) ambiguous data

(92% according to corpus study by Pearl & Mis 2011, 2016):

"Look – a kitty! Oh, look – another one."



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

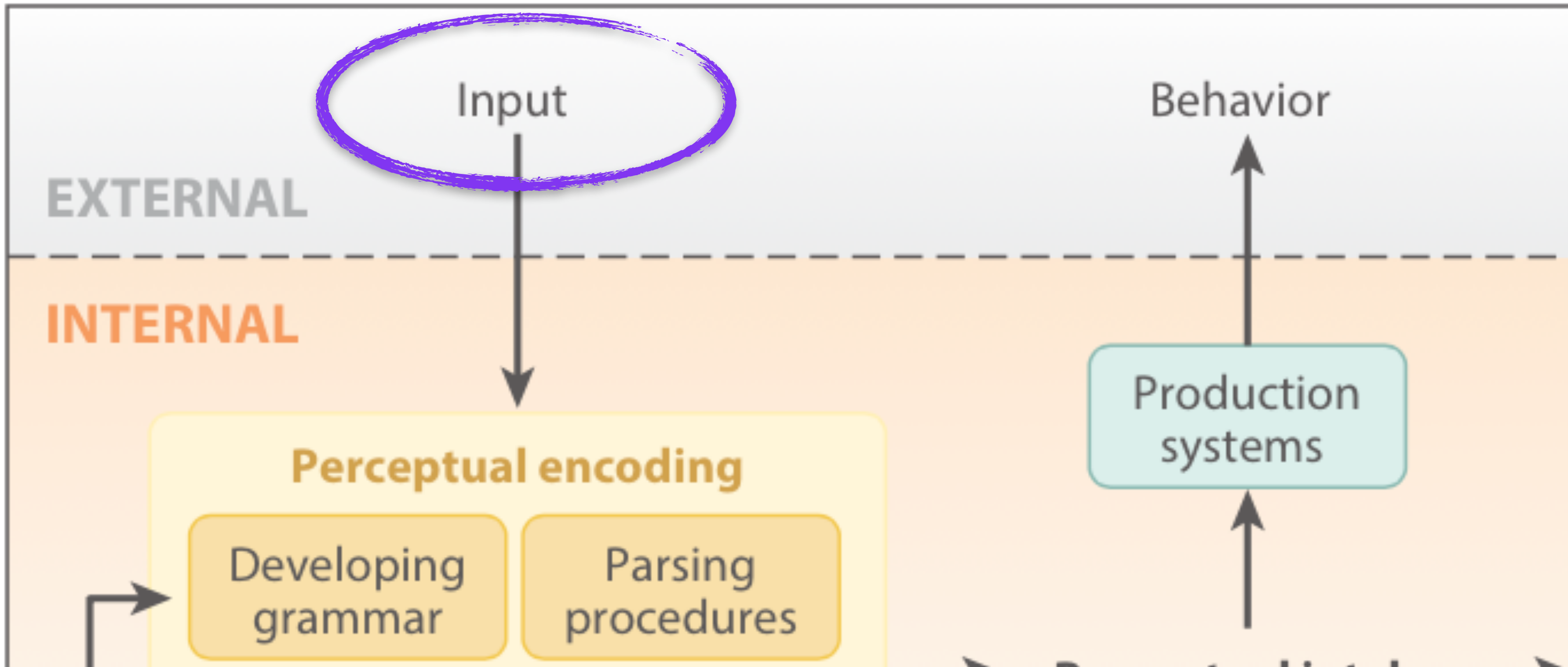
Syntactically (SYN) ambiguous data

(92% according to corpus study by Pearl & Mis 2011, 2016):

"Look – a **kitty**! Oh, look – another **one**."

Antecedent = "kitty"

Referent



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Syntactically (SYN) ambiguous data

(92% according to corpus study by Pearl & Mis 2011, 2016):

"Look – a **kitty**! Oh, look – another **one**."

Antecedent = "kitty"

Referent

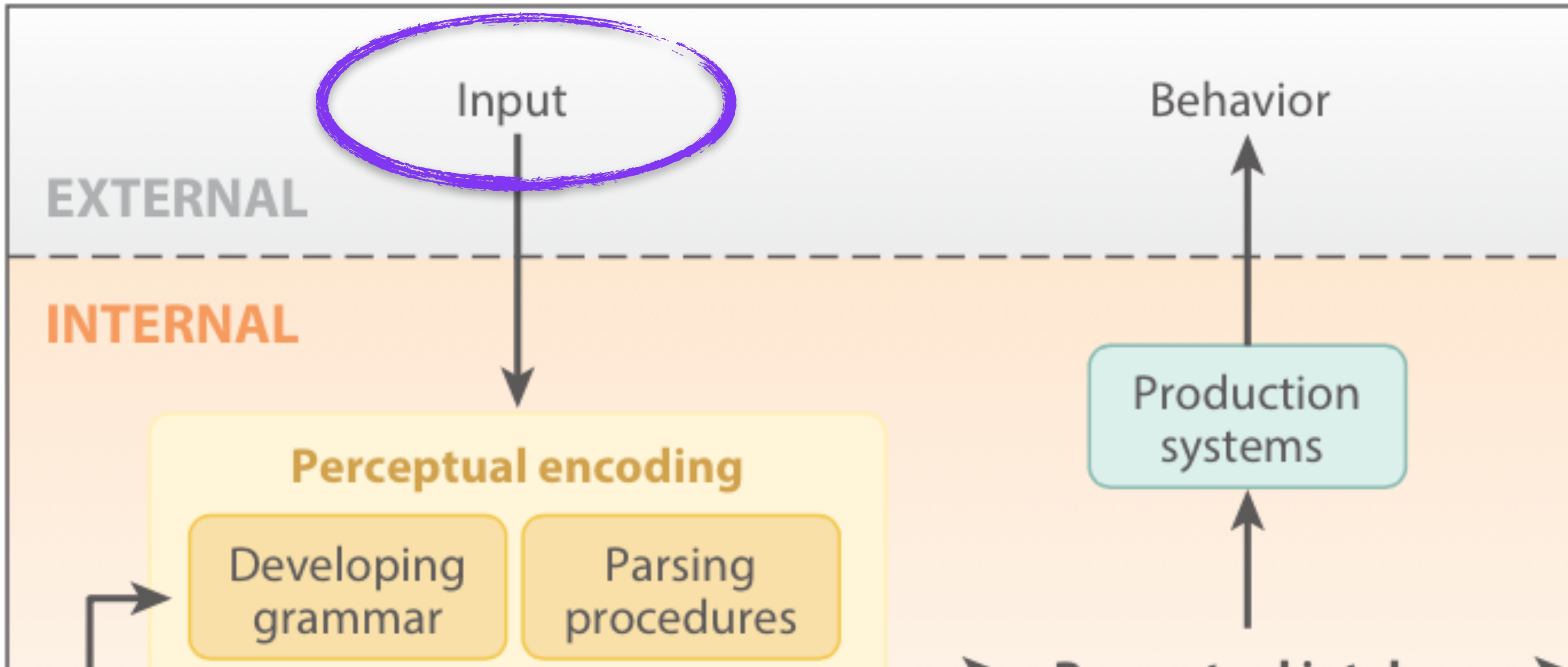
Syntactic category?

???

Noun'

Noun

kitty



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

Noun'

pretty kitty

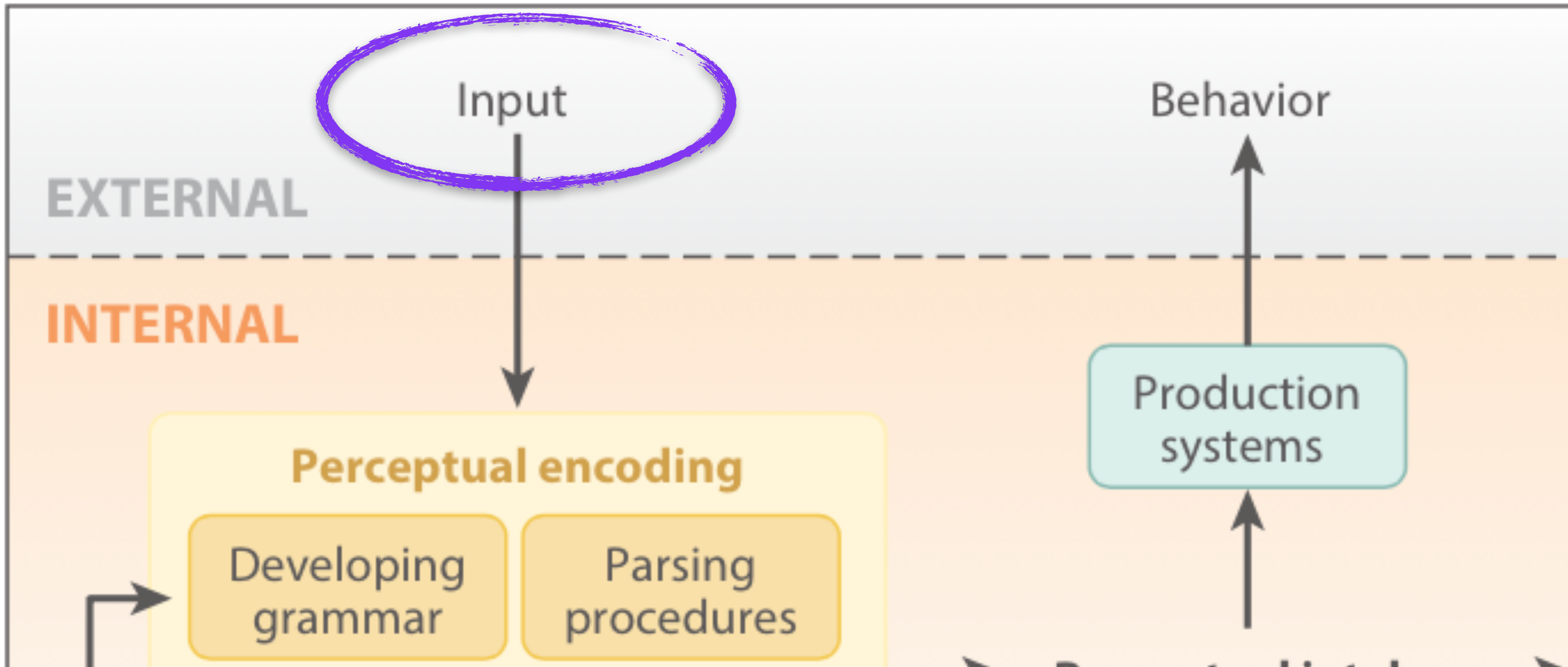
English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Referentially and syntactically (REF-SYN) ambiguous

(8% according to corpus study by Pearl & Mis 2011, 2016)

"Look – a pretty kitty! Oh, look – another one."



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

Noun'

pretty kitty

English child-directed speech

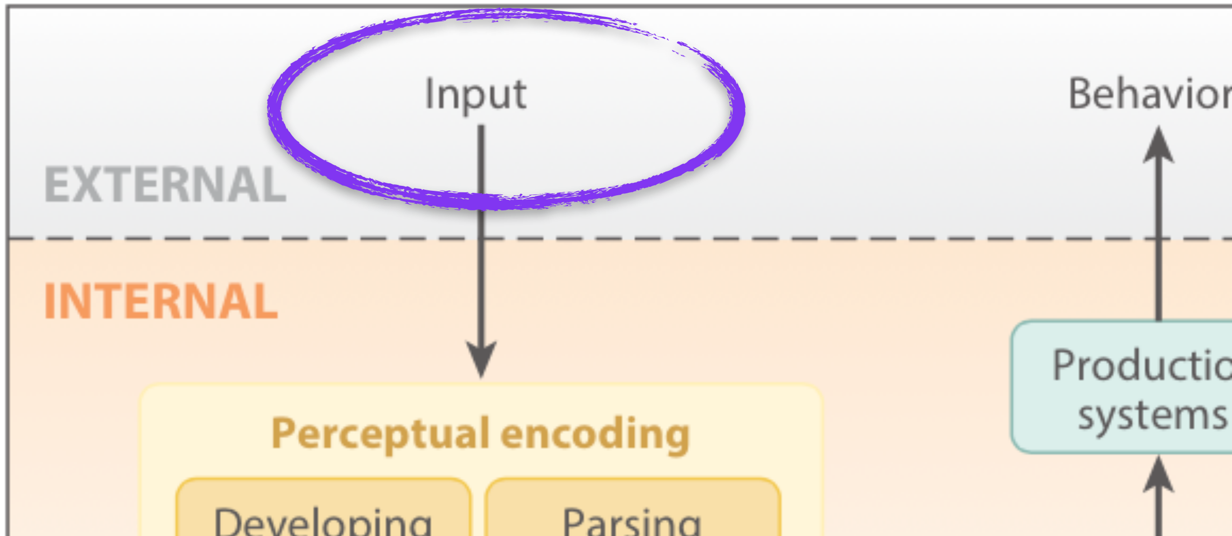
Problem: Most direct evidence children encounter is ambiguous.

Referentially and syntactically (REF-SYN) ambiguous

(8% according to corpus study by Pearl & Mis 2011, 2016)

"Look – a pretty **kitty**! Oh, look – another **one**."

Referent



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Referentially and syntactically (REF-SYN) ambiguous

(8% according to corpus study by Pearl & Mis 2011, 2016)

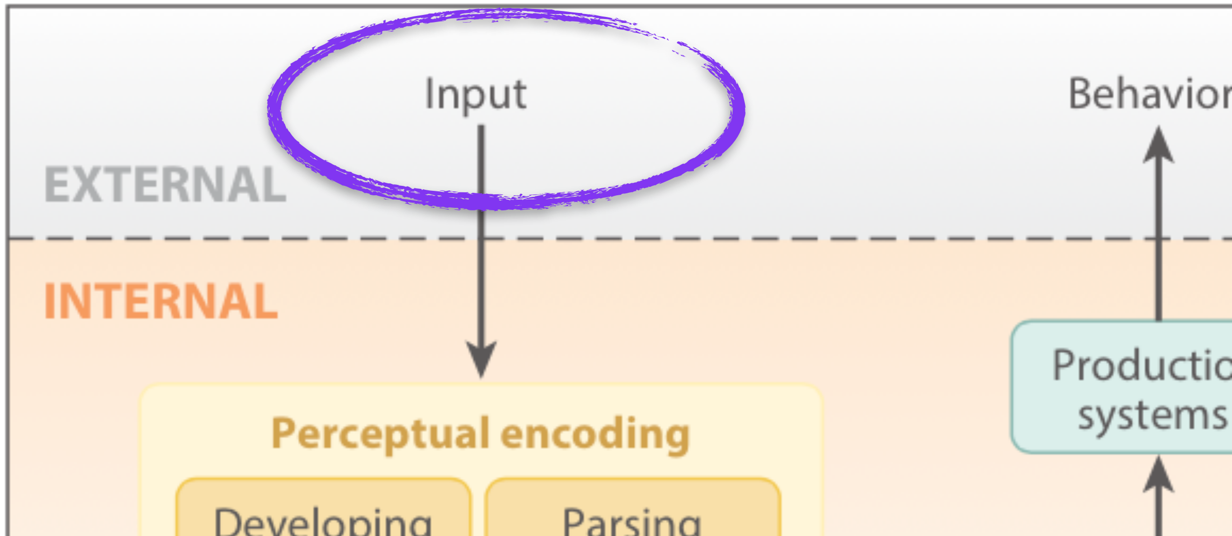
"Look – a **pretty kitty**! Oh, look – another **one**."

Antecedent = "pretty kitty"

OR

Antecedent = "kitty"

Referent



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Referentially and syntactically (REF-SYN) ambiguous

(8% according to corpus study by Pearl & Mis 2011, 2016)

"Look – a **pretty** **kitty**! Oh, look – another **one**."

Antecedent = "pretty kitty"

???

Antecedent = "kitty"

Referent

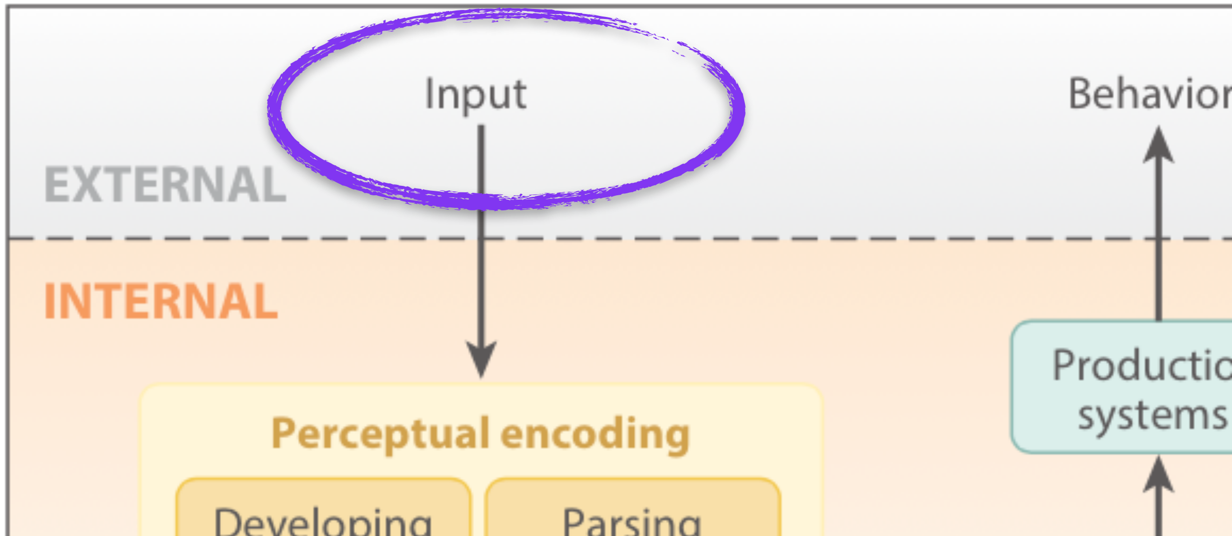
Syntactic category?

Noun'

???

Noun

kitty



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

Noun’

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Referentially and syntactically (REF-SYN) ambiguous

(8% according to corpus study by Pearl & Mis 2011, 2016)

“Look – a **pretty kitty**! Oh, look – another **one**.”

Antecedent = “pretty kitty”

???

Antecedent = “kitty”

Referent

Noun’

Noun’

Noun

pretty **kitty**

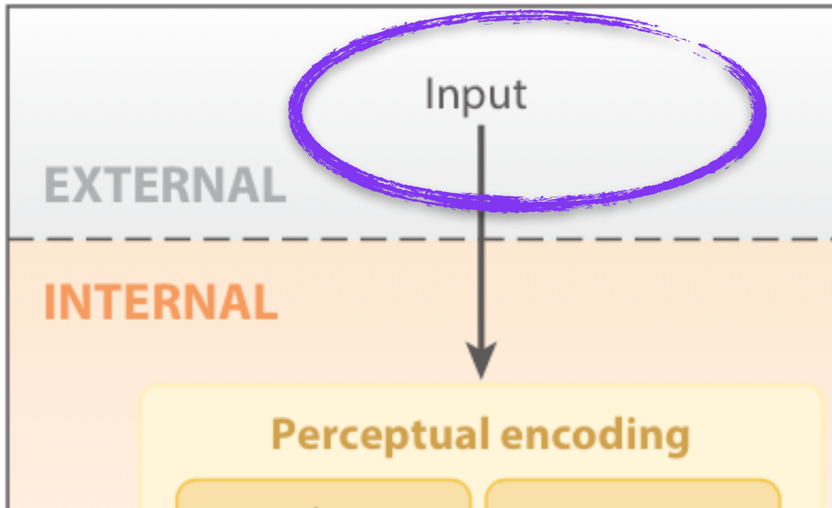
Syntactic category?

???

Noun’

Noun

kitty



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

8% REF-SYN ambiguous

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

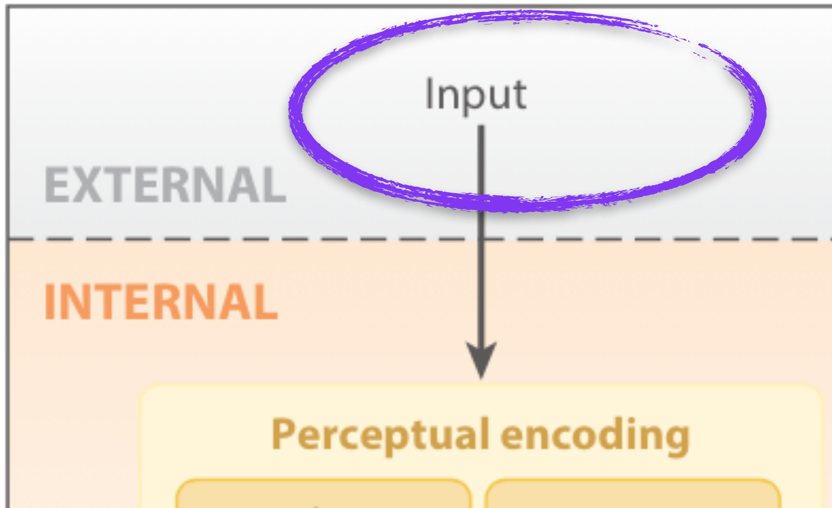
Unambiguous (UNAMB) data

What we wish were there but isn't

(0% according to corpus study by Pearl & Mis 2011, 2016)

"Look – a pretty kitty!

Hmmm - there doesn't seem to be another one here, though."



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous
8% REF-SYN ambiguous

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Unambiguous (UNAMB) data

What we wish were there but isn't

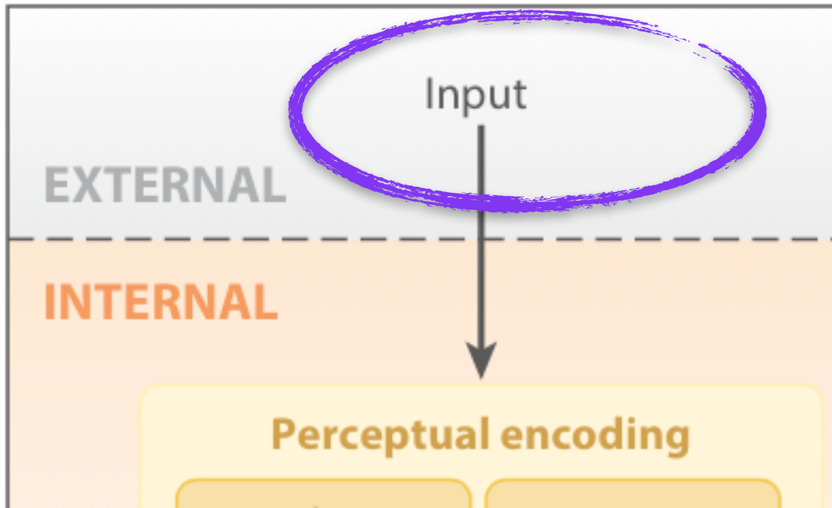
(0% according to corpus study by Pearl & Mis 2011, 2016)

Can't have "**kitty**" as its antecedent, because there *is* another kitty here. This would be a false thing to say.

"Look – a pretty **kitty**!

Hmmm - there doesn't seem to be another **one** here, though."

~~**kitty**~~



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

8% REF-SYN ambiguous

Noun'

pretty kitty

Referent

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Unambiguous (UNAMB) data

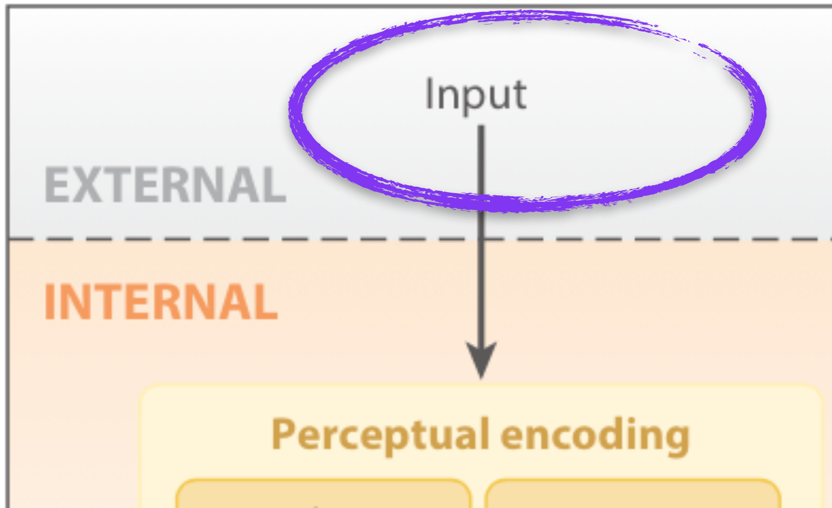
What we wish were there but isn't

(0% according to corpus study by Pearl & Mis 2011, 2016)

"Look – a **pretty kitty**!

Hmmm - there doesn't seem to be another **one** here, though."

Must have "pretty kitty" as its antecedent.



Pronoun interpretation

syntax, semantics

another one

92% SYN ambiguous

8% REF-SYN ambiguous

Noun’

pretty kitty

Referent

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

Unambiguous (UNAMB) data

What we wish were there but isn’t

(0% according to corpus study by Pearl & Mis 2011, 2016)

“Look – a **pretty kitty**!

Hmmm - there doesn’t seem to be another **one** here, though.”

Must have “pretty kitty” as its antecedent.

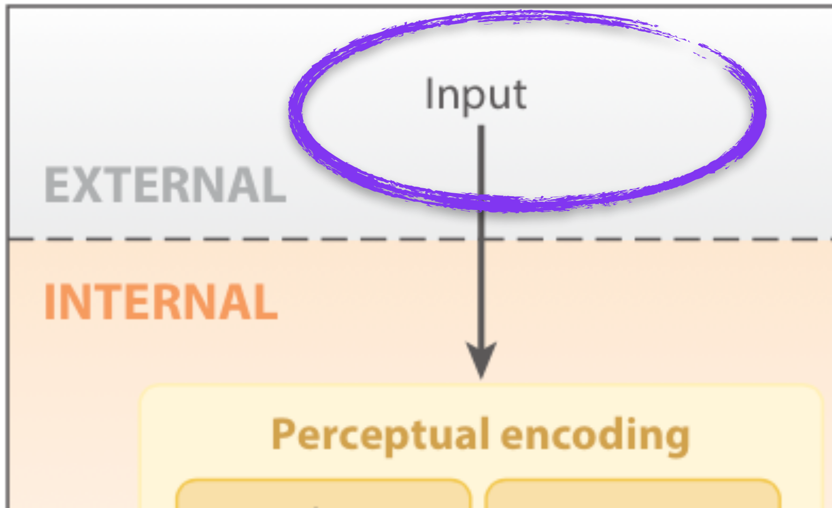
Noun’

Noun’

Noun

pretty **kitty**

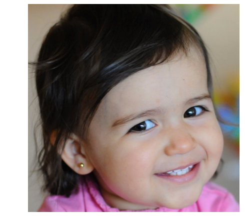
and be a Noun’ category.



Pronoun interpretation

syntax, semantics

another one



English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

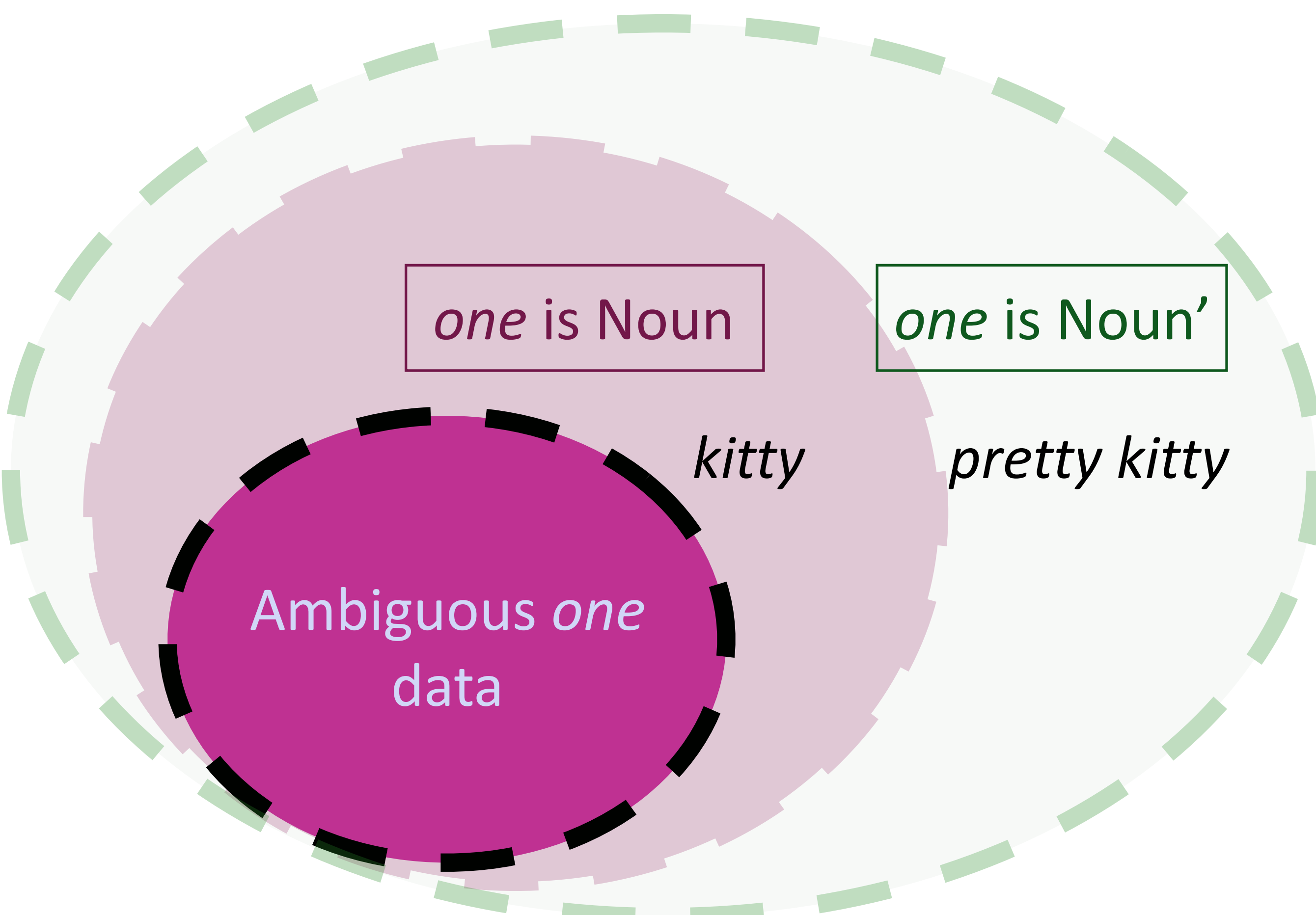
8% REF-SYN ambiguous

Noun'

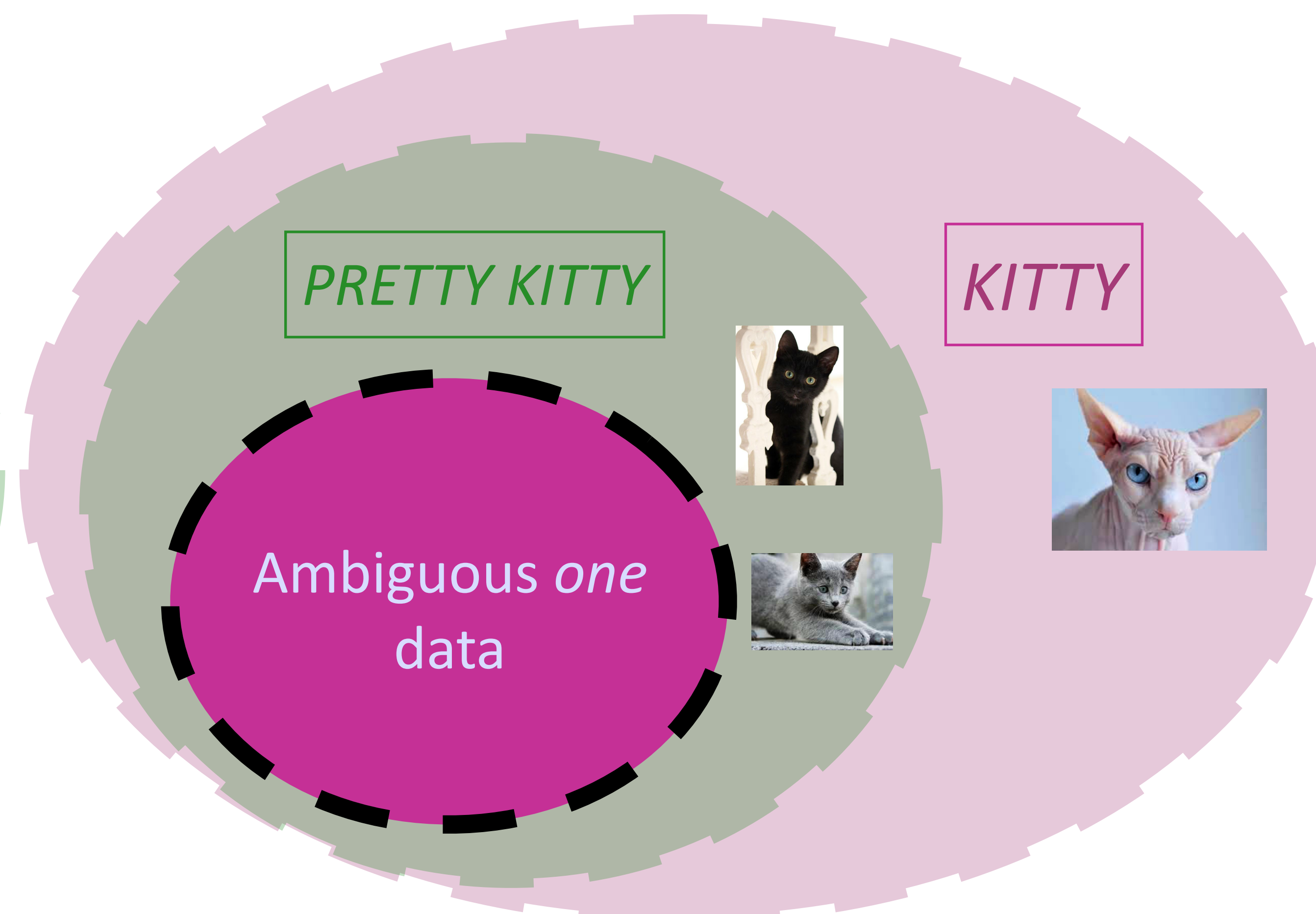
pretty kitty

How do children learn the right generalizations for interpreting *one*?

syntactic category



referent in context



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

How do children learn the right generalizations for interpreting *one*?

Regier & Gahl (2004), Pearl & Lidz (2009): **Filtering the direct evidence** (being more selective about what you learn from) & learning from it in more sophisticated ways

Pearl & Mis (2016): **Leveraging a broader set of data** to learn from & learning from in it more sophisticated ways

Pronoun interpretation

syntax, semantics

another one

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

Noun'

pretty kitty

How do children learn the right generalizations for interpreting *one*?

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Pearl & Mis (2016):

Leveraging a broader set of data

Learning from it in more sophisticated ways

Pronoun interpretation

syntax, semantics

another one

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

Noun'

pretty kitty

How do children learn the right generalizations for interpreting *one*?

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

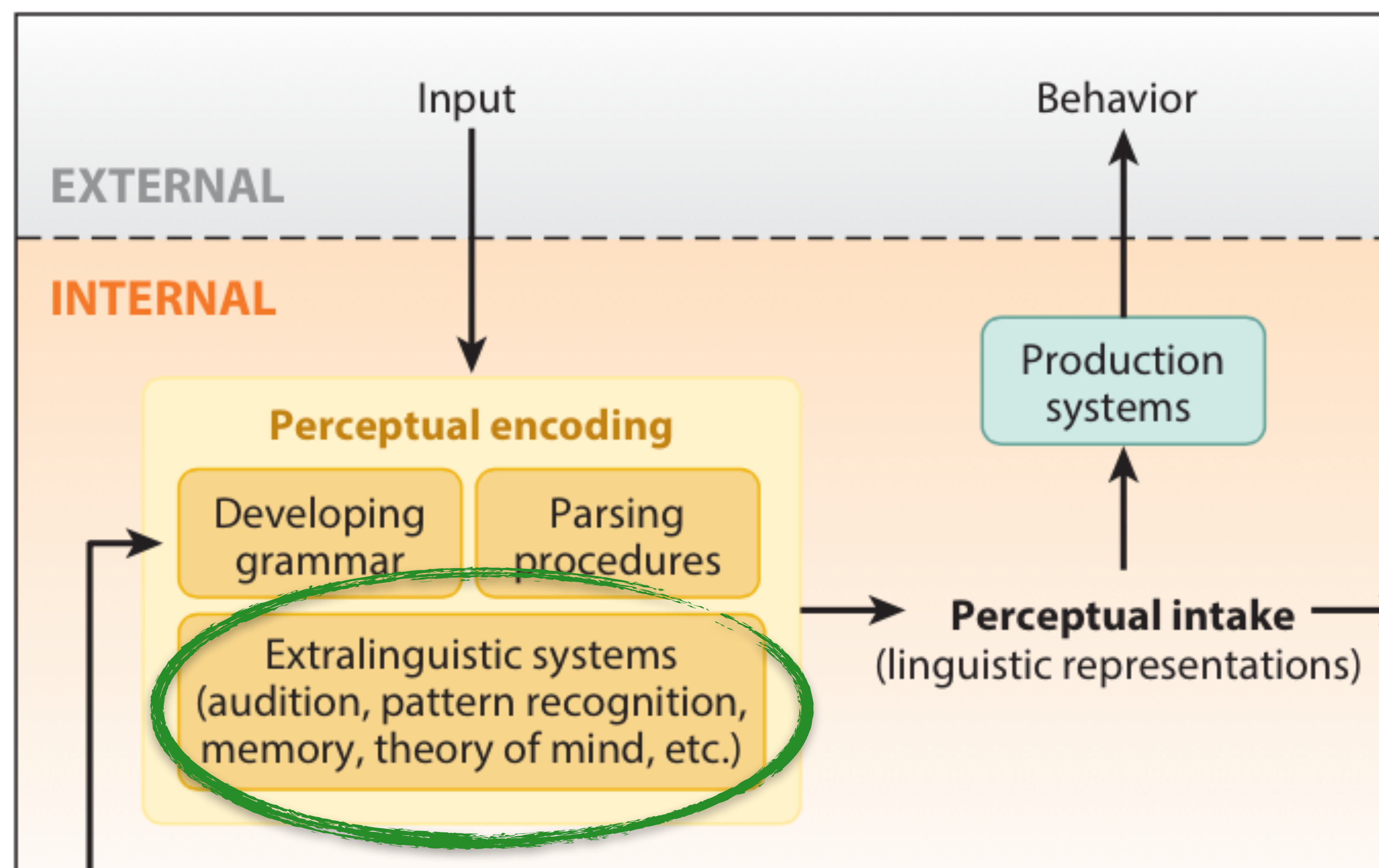
Pearl & Mis (2016):

Leveraging a broader set of data

Learning from it in more sophisticated ways

Probabilistic reasoning about input:

Bayesian inference



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

How do children learn the right generalizations for interpreting *one*?

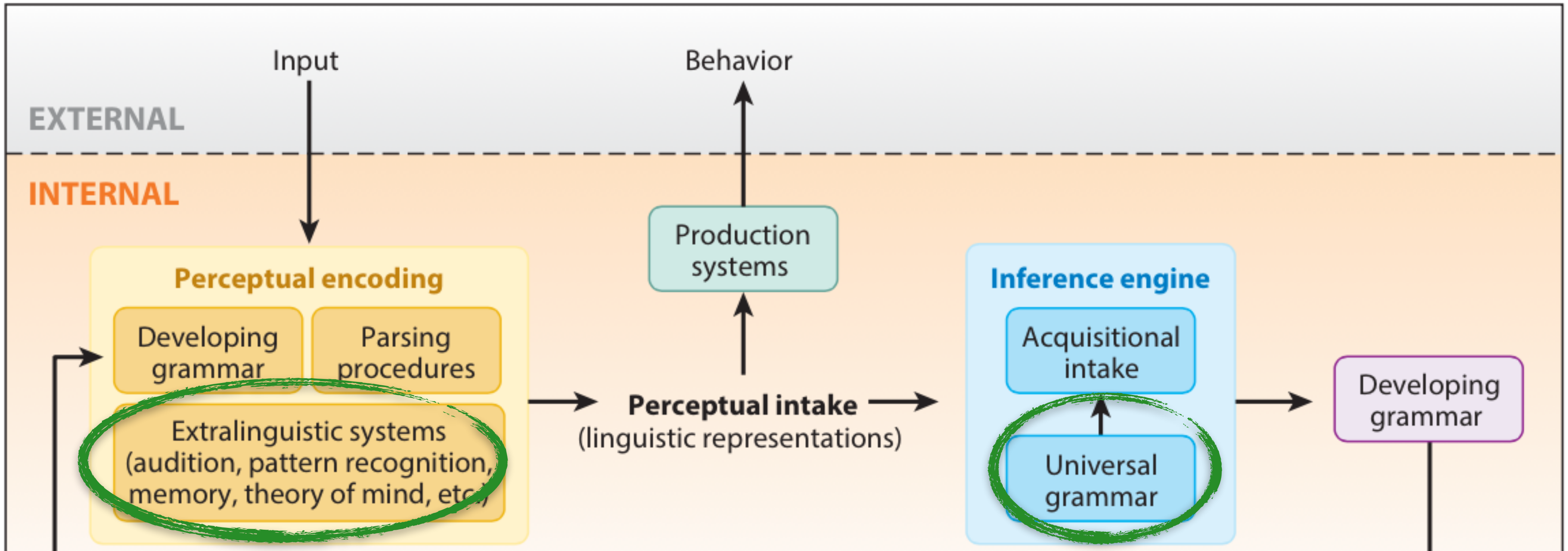
Pearl & Mis (2016):

Leveraging a broader set of data

Learning from it in more sophisticated ways

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence



Pronoun interpretation

syntax, semantics

another one

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

8% REF-SYN ambiguous

Noun'

pretty kitty

How do children learn the right generalizations for interpreting *one*?

Pearl & Mis (2016):

Leveraging a broader set of data

Learning from it in more sophisticated ways

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Ignore these data *92% SYN ambiguous*

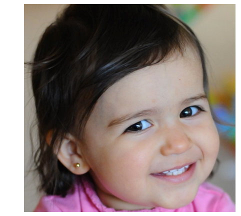
"Look – a **kitty**!

Oh, look – another **one**."

Pronoun interpretation

syntax, semantics

another one



Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

How do children learn the right generalizations for interpreting *one*?

Pearl & Mis (2016):

Leveraging a broader set of data

Learning from it in more sophisticated ways

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Ignore these data *92% SYN ambiguous*

and learn from these data using Bayesian inference 8% REF-SYN ambiguous

"Look – a **pretty kitty**!

Oh, look – another **one**."

Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

How do children learn the right generalizations for interpreting *one*?

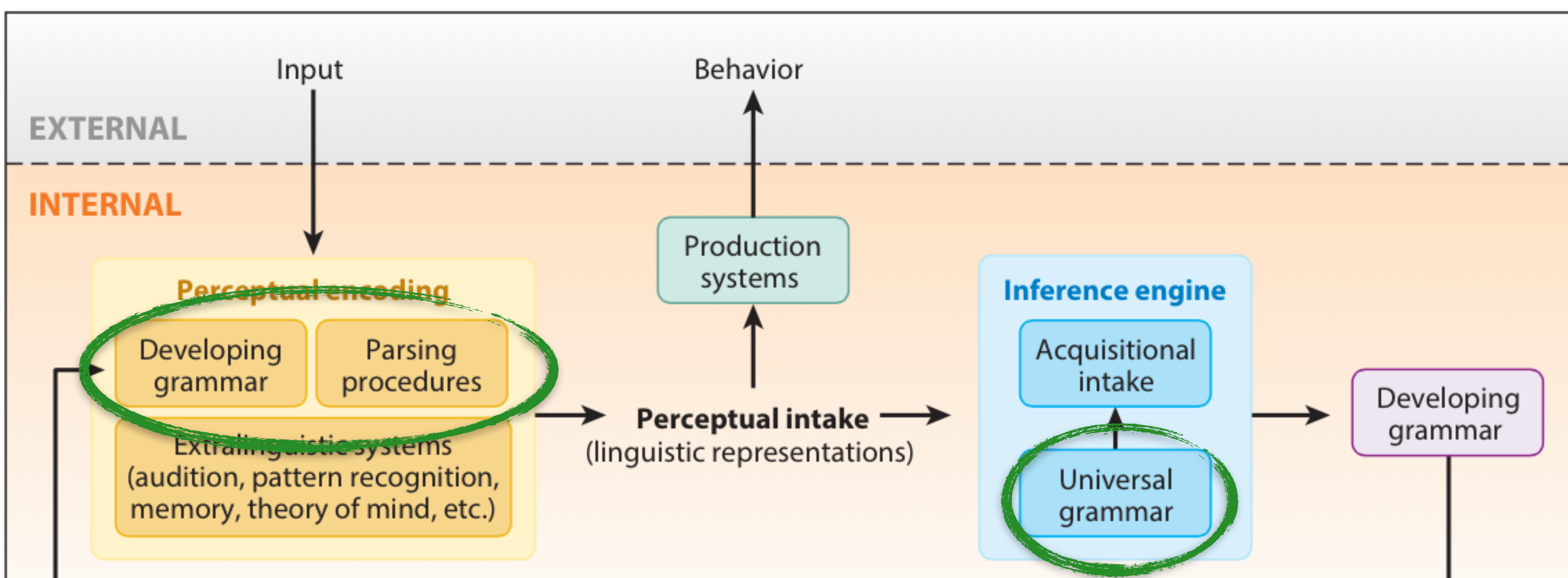
Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Learning from it in more sophisticated ways

Pearl & Mis (2016):

Leveraging a broader set of data



Pronoun interpretation

syntax, semantics

another one

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

Noun'

pretty kitty

How do children learn the right generalizations for interpreting *one*?

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Learning from it in more sophisticated ways

Pearl & Mis (2016):

Leveraging a broader set of data

Learn from data like these that involve other pronouns

"Look – **a pretty kitty**!

I want to pet **it**."

Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

English child-directed speech

Problem: Most direct evidence children encounter is ambiguous.

92% SYN ambiguous

8% REF-SYN ambiguous

How do children learn the right generalizations for interpreting *one*?

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Learning from it in more sophisticated ways

Pearl & Mis (2016):

Leveraging a broader set of data

Learn from data like these that involve other pronouns

"Look – **a pretty kitty**! I want to pet **it**."

Key: modifier is included in antecedent.
Implication: May want to include the modifier whenever it's an option.

one

pretty kitty

Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

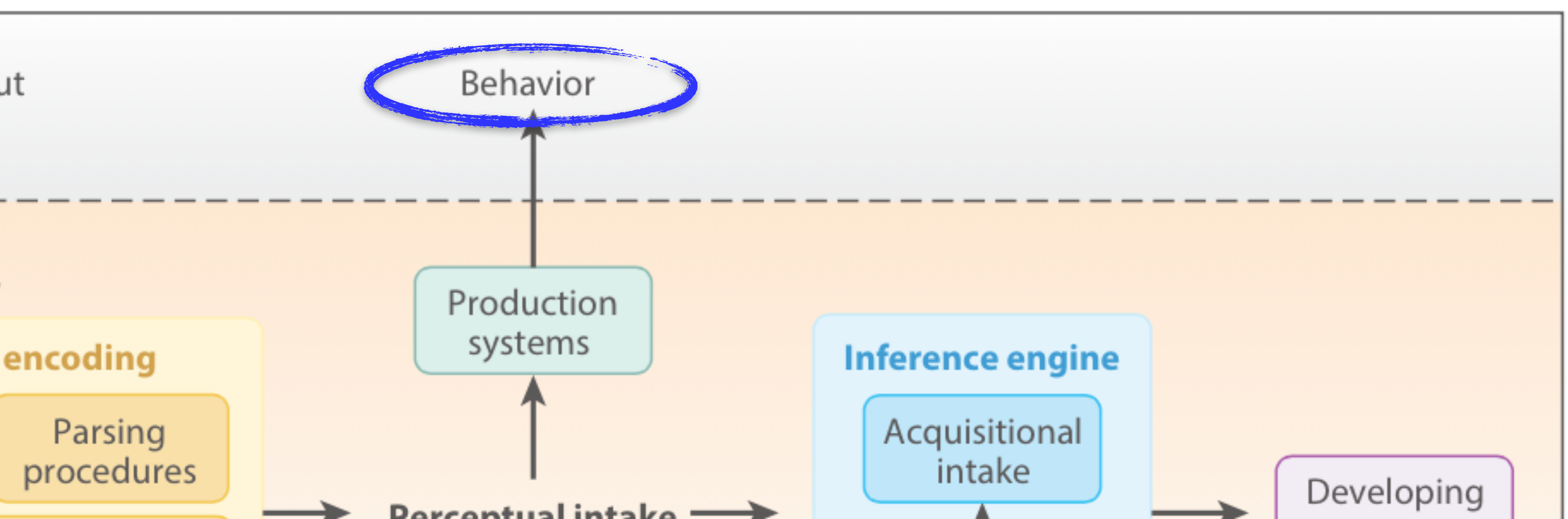
Regier & Gahl (2004), Pearl & Lidz (2009):
Filtering the direct evidence

Learning from it in more sophisticated ways

Pearl & Mis (2016):
Leveraging a broader set of data

Algorithmic-level implementation of these strategies

Evaluated on whether they matched
18-month-old looking preferences.



Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

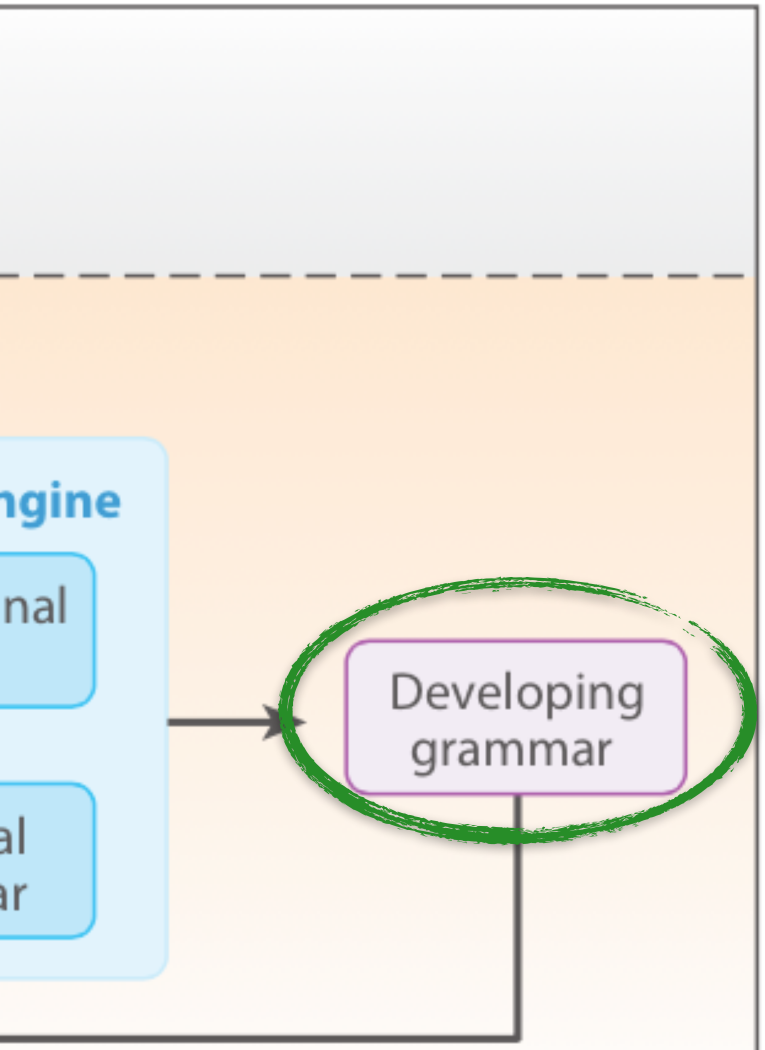
Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence

Learning from it in more sophisticated ways

Pearl & Mis (2016):

Leveraging a broader set of data



Algorithmic-level

Both were successful at generating the 18-month-old behavior. We can then look inside the modeled learners and see what the underlying representations were.

Pronoun interpretation

syntax, semantics

another one

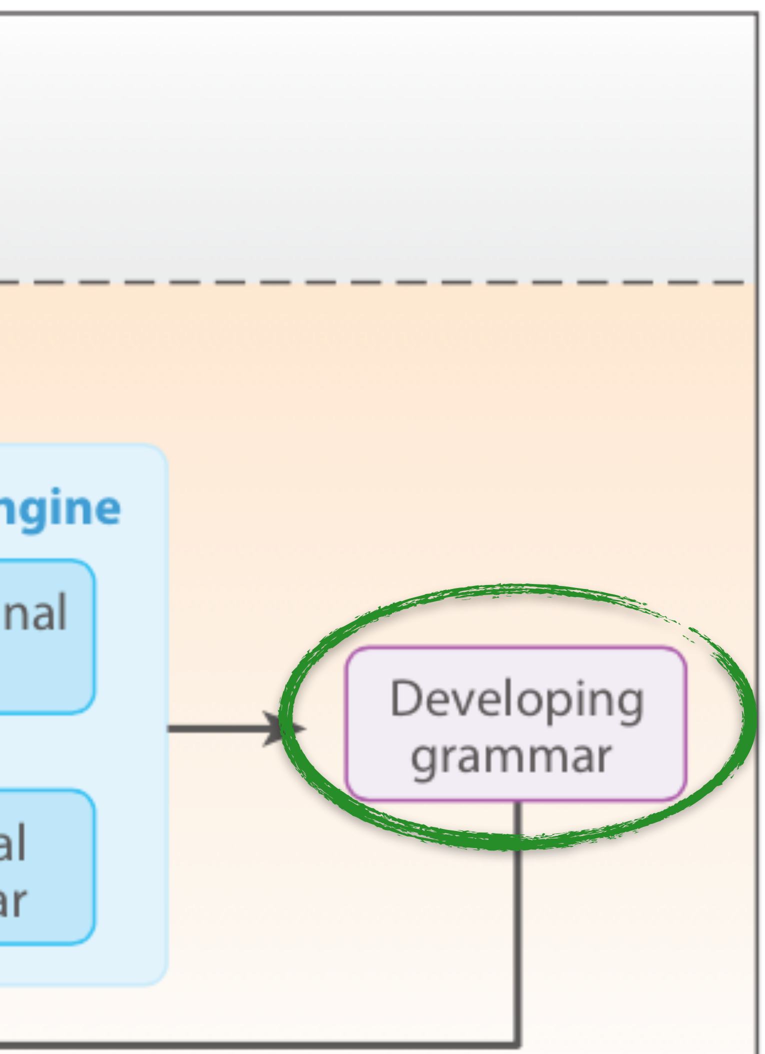
Noun’

pretty kitty

Learning from it in more sophisticated ways

Pearl & Mis (2016):
Leveraging a broader set of data

Algorithmic-level



Regier & Gahl (2004), Pearl & Lidz (2009):
Filtering the direct evidence

Adult representations

✓ **Noun’**
pretty kitty

But…required additional situational context to be present to succeed.

Pronoun interpretation

syntax, semantics

another one

Noun'

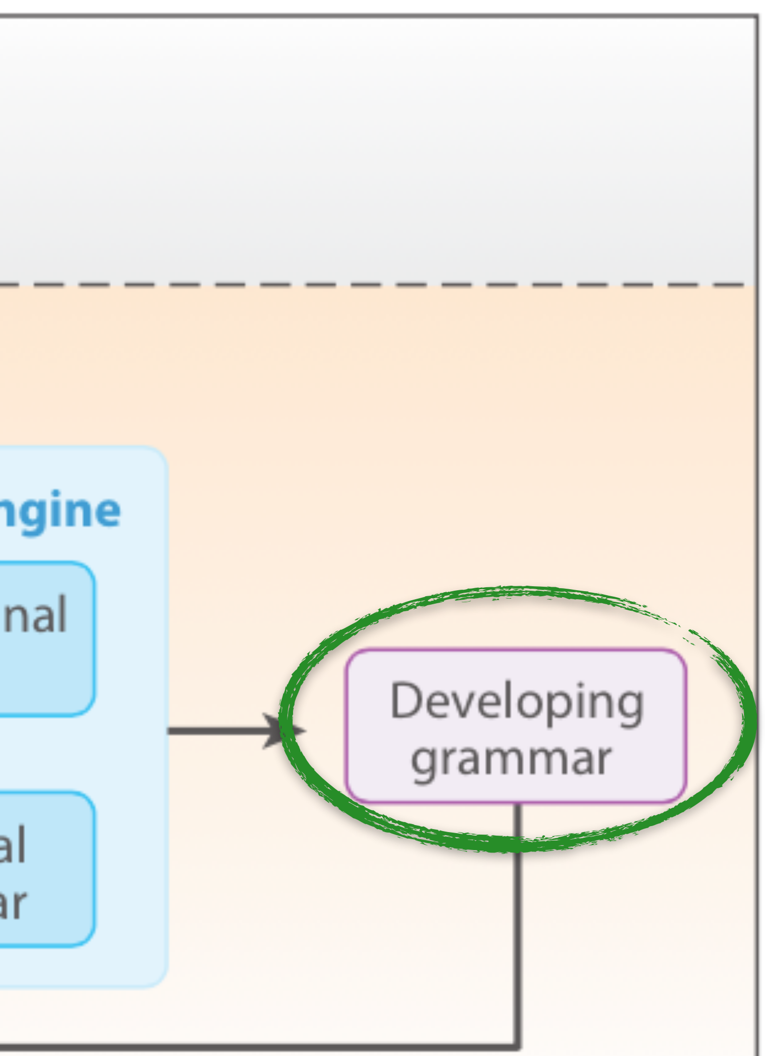
pretty kitty

Learning from it in more sophisticated ways

Pearl & Mis (2016):
Leveraging a broader set of data

Regier & Gahl (2004), Pearl & Lidz (2009):
Filtering the direct evidence

Algorithmic-level

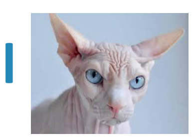


"Look – a **pretty kitty**!
Oh, look – another **one**."

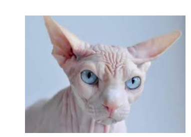
Adult representations

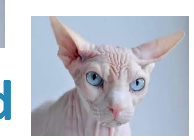
✓ **Noun'**
pretty kitty

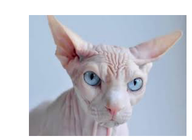
small



furry







light-eyed

big-eared

But…required additional situational context to be present to succeed.

Less robust

Needed to have a lot of alternative options so it's a suspicious coincidence that the referent is pretty if "pretty" wasn't actually included in the antecedent.

Pronoun interpretation

syntax, semantics

another one

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence ✓ **Less robust**

Learning from it in more sophisticated ways

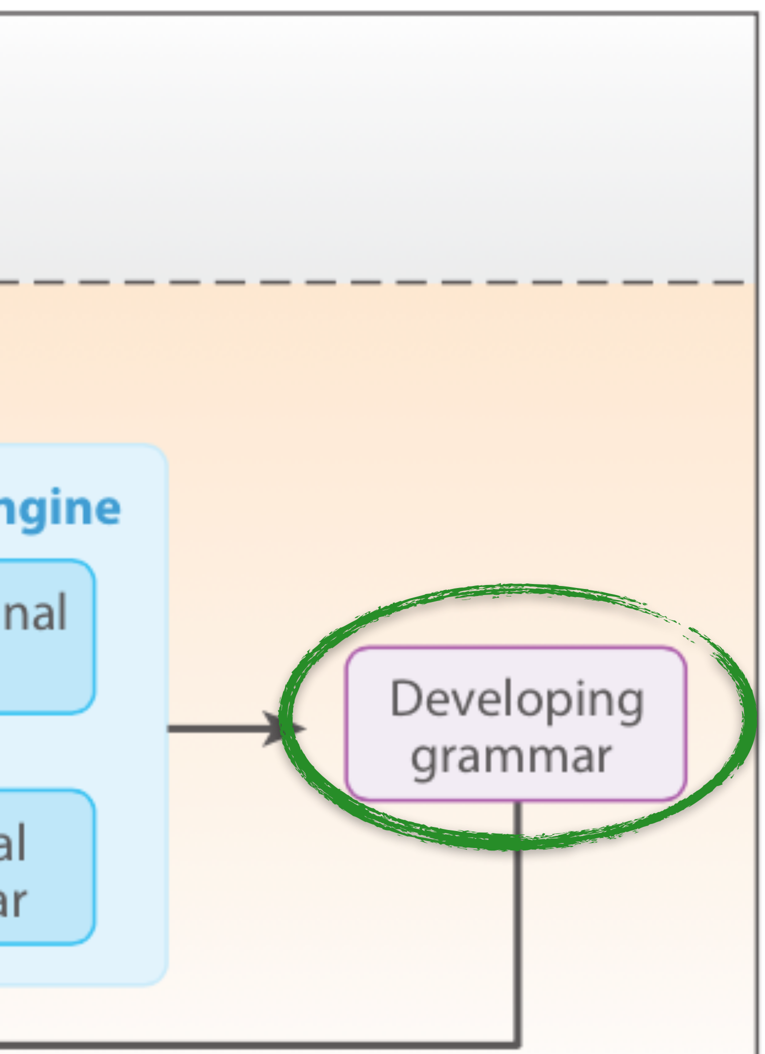
Noun'

pretty kitty

Pearl & Mis (2016):

Leveraging a broader set of data

Algorithmic-level



Immature representations

✓ **Noun'** only in certain linguistic contexts

pretty kitty

Noun'

Noun'

Noun

pretty **kitty**

"Look – a **pretty kitty**!

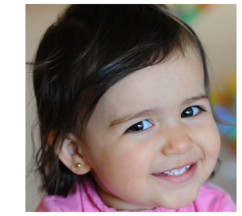
Oh, look – another **one**."

Noun'

Pronoun interpretation

syntax, semantics

another one



Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence ✓ **Less robust**

Learning from it in more sophisticated ways

Noun'

pretty kitty

Pearl & Mis (2016):

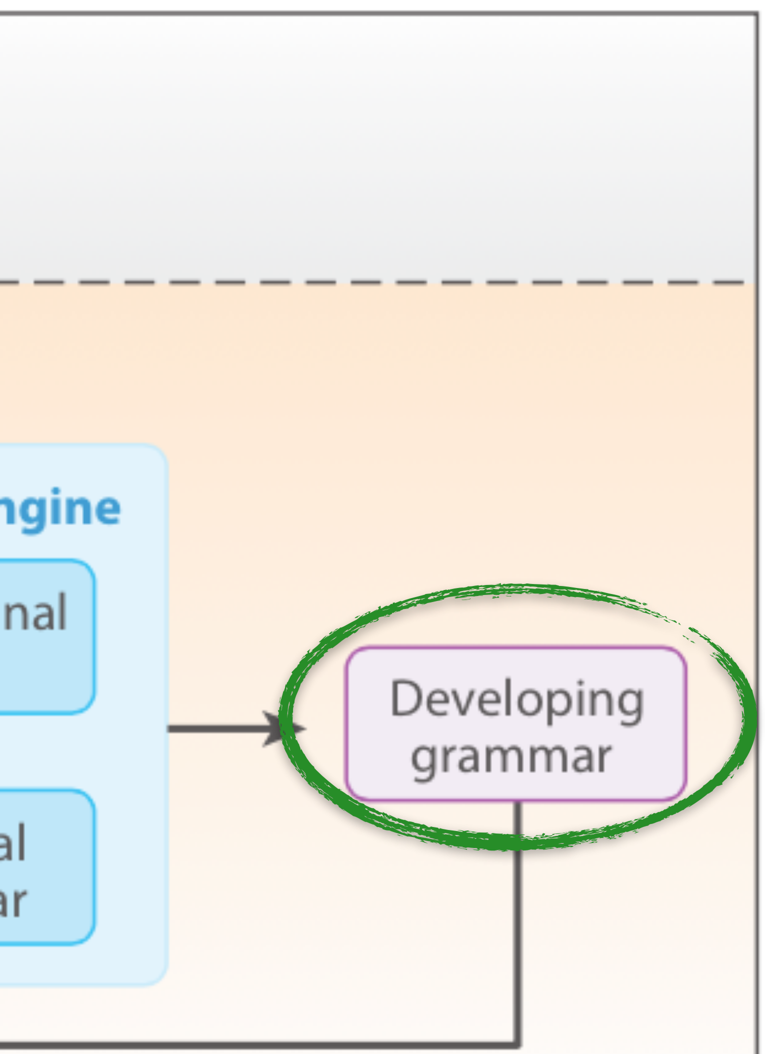
Leveraging a broader set of data

Noun

|

kitty

Algorithmic-level



Immature representations

✓ **Noun'** only in certain linguistic contexts

pretty kitty

✗ otherwise **Noun**

"Look – a **kitty**!

Oh, look – another **one**."

Noun

But…does this for pretty much any situational context.

More robust

Pronoun interpretation

syntax, semantics

another one

Noun'

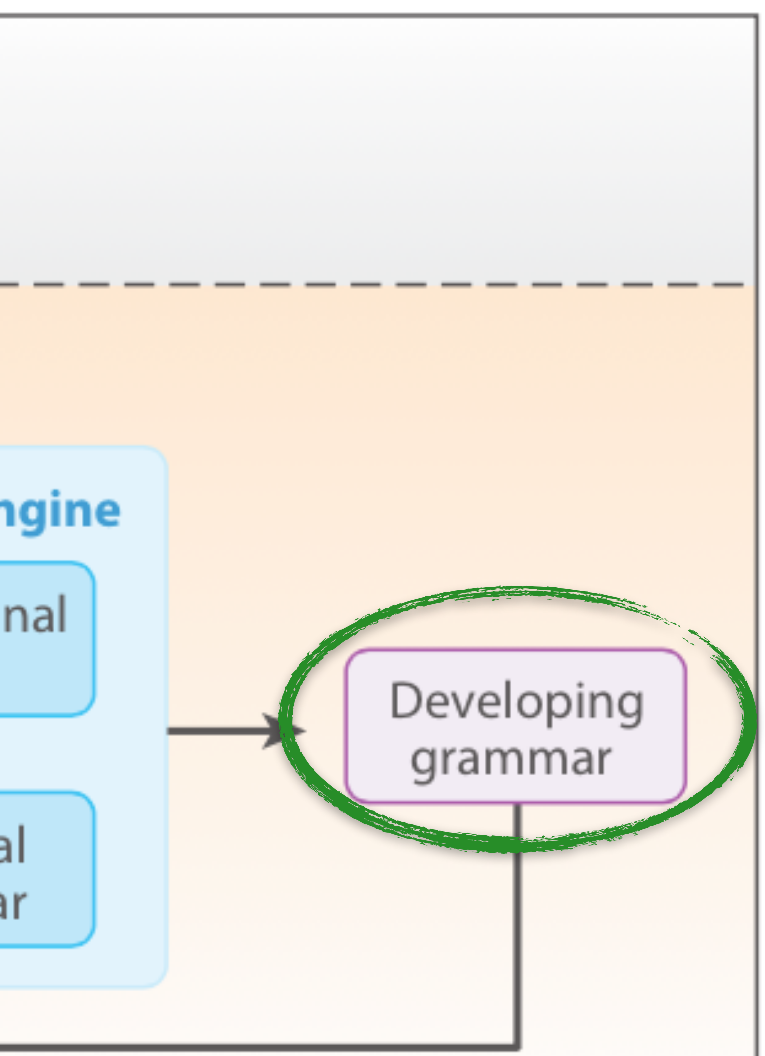
pretty kitty

Regier & Gahl (2004), Pearl & Lidz (2009):
Filtering the direct evidence ✓ **Less robust**

Learning from it in more sophisticated ways

Pearl & Mis (2016):
Leveraging a broader set of data ✓ **More robust**

Algorithmic-level



By modeling, we have two concrete proposals for how children learn the knowledge they do by 18 months.

This also motivates future experimental work to distinguish these two possibilities.

Pronoun interpretation

syntax, semantics

another one

Noun'

pretty kitty

Regier & Gahl (2004), Pearl & Lidz (2009):

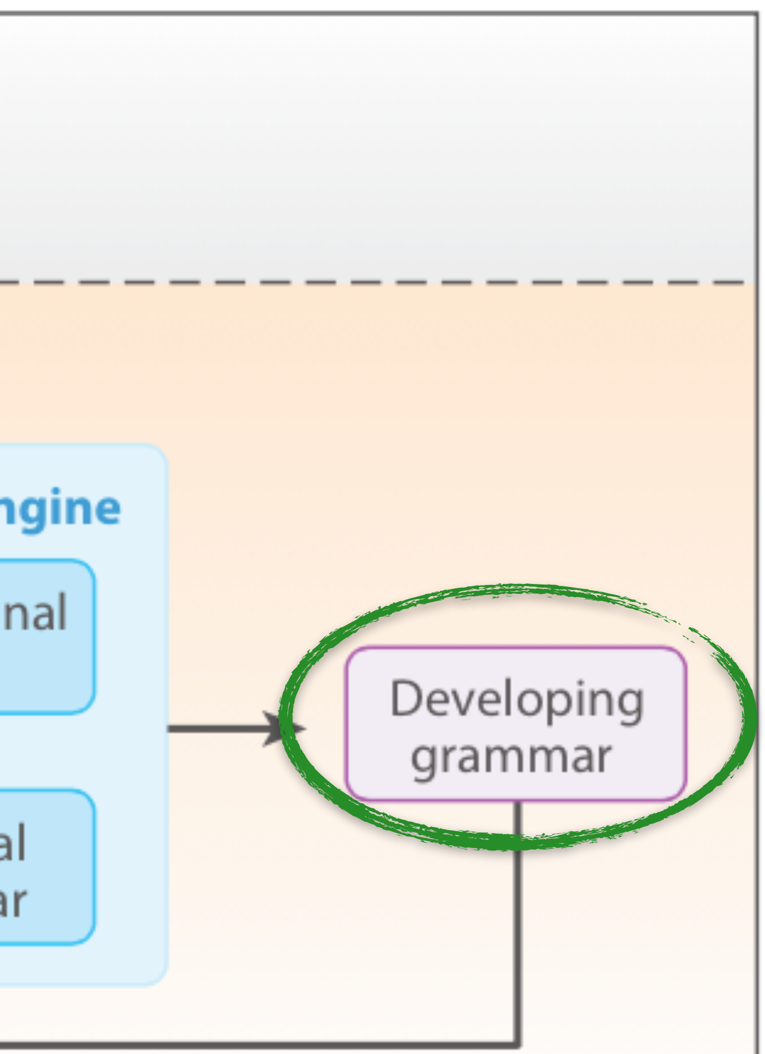
Filtering the direct evidence ✓ **Less robust**

Learning from it in more sophisticated ways

Pearl & Mis (2016):

Leveraging a broader set of data ✓ **More robust**

Algorithmic-level



This also motivates future experimental work to distinguish these two possibilities.

"This kitty likes the **cup** of milk but not the **one** of water."

Adults generally don't like this because it forces *one* to be category Noun.

Pronoun interpretation

syntax, semantics

another one

Regier & Gahl (2004), Pearl & Lidz (2009):

Filtering the direct evidence ✓ **Less robust**

Learning from it in more sophisticated ways

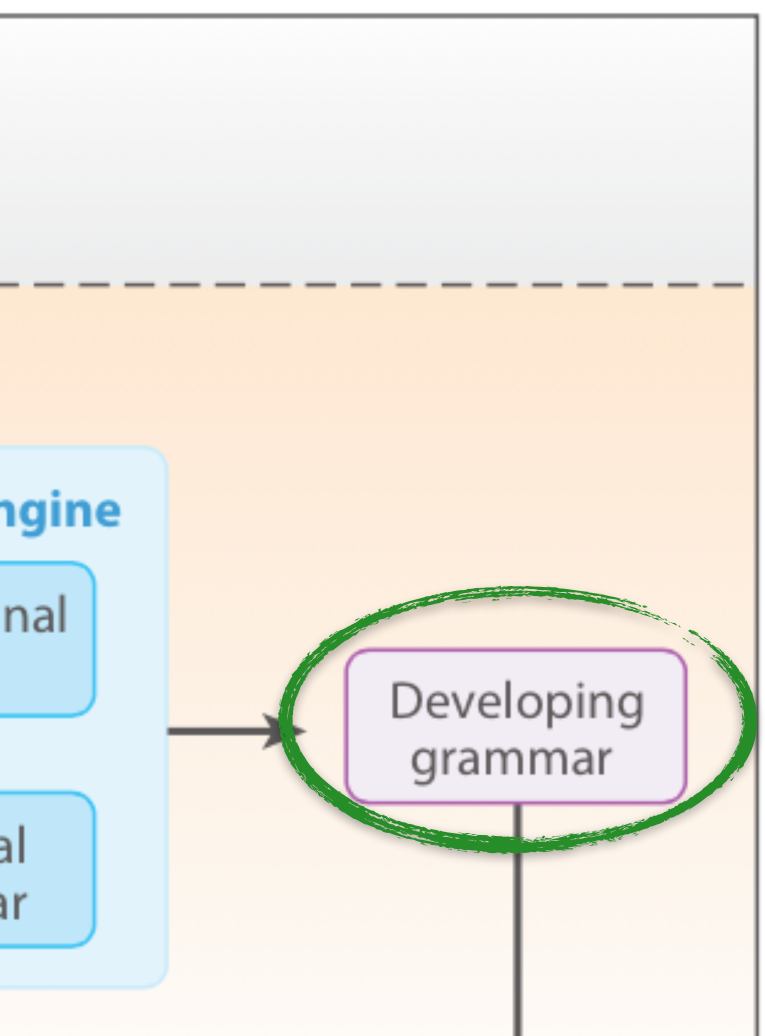
Pearl & Mis (2016):

Leveraging a broader set of data ✓ **More robust**

Noun'

pretty kitty

Algorithmic-level



This also motivates future experimental work to distinguish these two possibilities.

"This kitty likes the **cup** of milk but not the **one** of water."

Noun

When do children have this same judgment? Is it before 18 months?

Pronoun interpretation

syntax, semantics

another one

Noun'

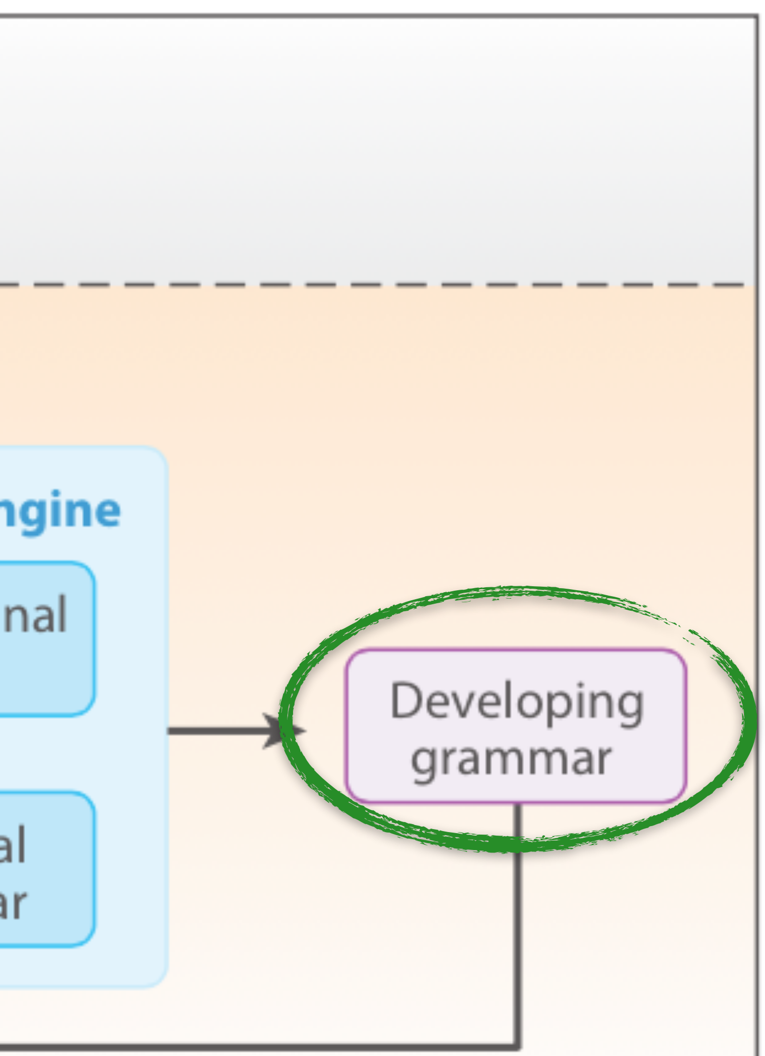
pretty kitty

Learning from it in more sophisticated ways

Pearl & Mis (2016):
Leveraging a broader set of data

More robust

Algorithmic-level



By 18 months

Regier & Gahl (2004),
Pearl & Lidz (2009):
Filtering the direct evidence

✓

"This kitty likes the **cup** of milk but not the **one** of water."

Noun

When do children have this same judgment? Is it before 18 months?

Pronoun interpretation

syntax, semantics

another one

Noun’

pretty kitty

By 18 months

Regier & Gahl (2004),

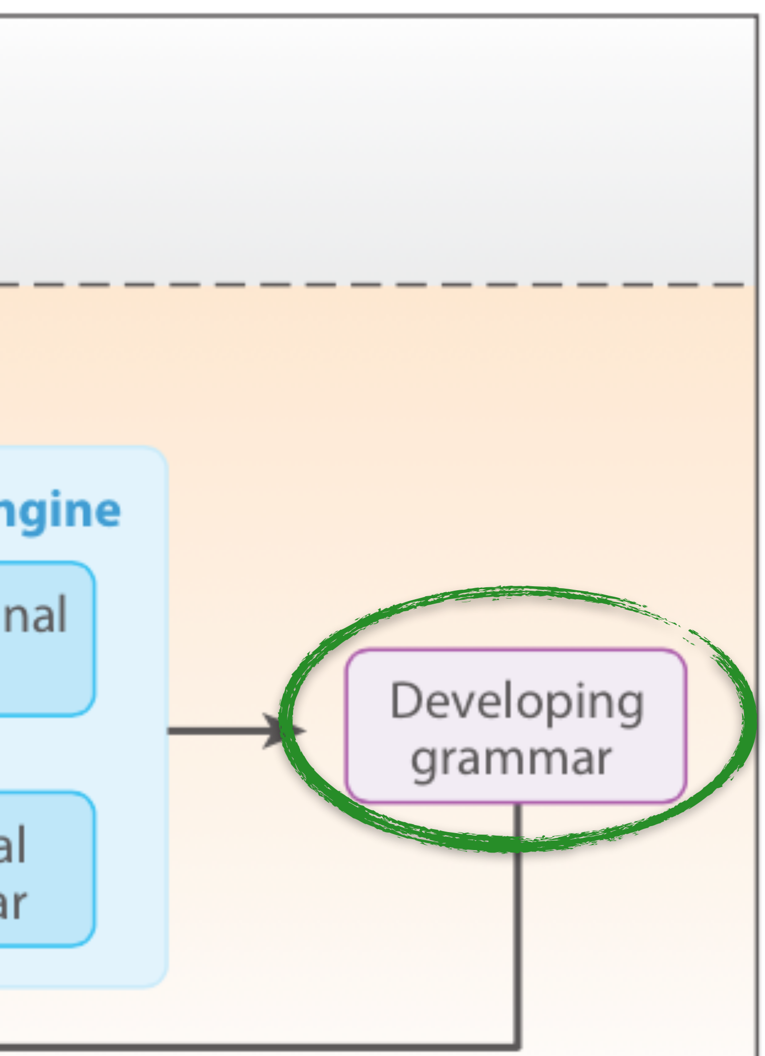
Pearl & Lidz (2009):

Filtering the direct evidence

“This kitty likes the **cup** of milk but not the **one** of water.”

Noun

Algorithmic-level



Not by 18 months

Pearl & Mis (2016):

Leveraging a broader set of data

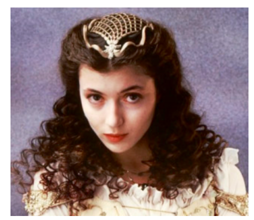
When do children have this same judgment? Is it before 18 months?

Today's Plan:
Computational models of syntactic acquisition

I. Some non-parametric examples


Who does … is pretty?

syntax

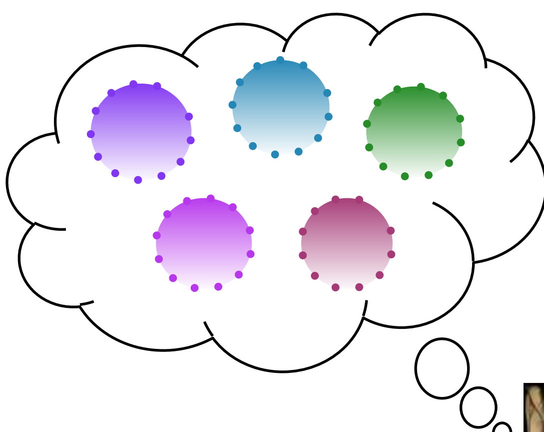


another one



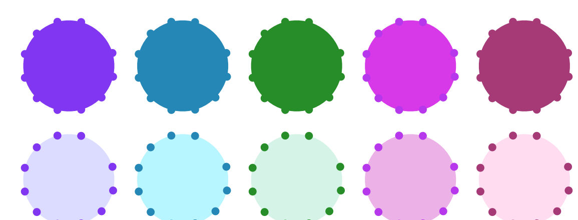
syntax, semantics

II. About linguistic parameters




III. Learning with parameters

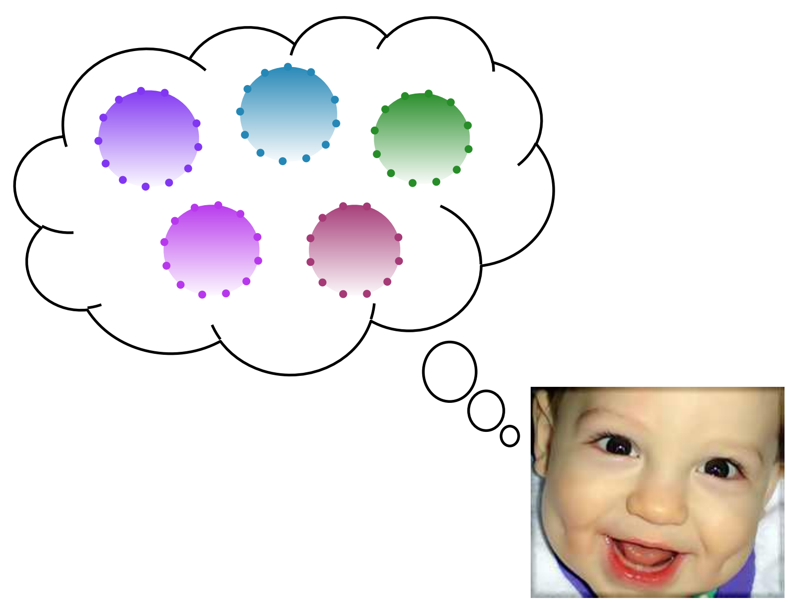
0.2 0.3 0.8 0.7 0.1



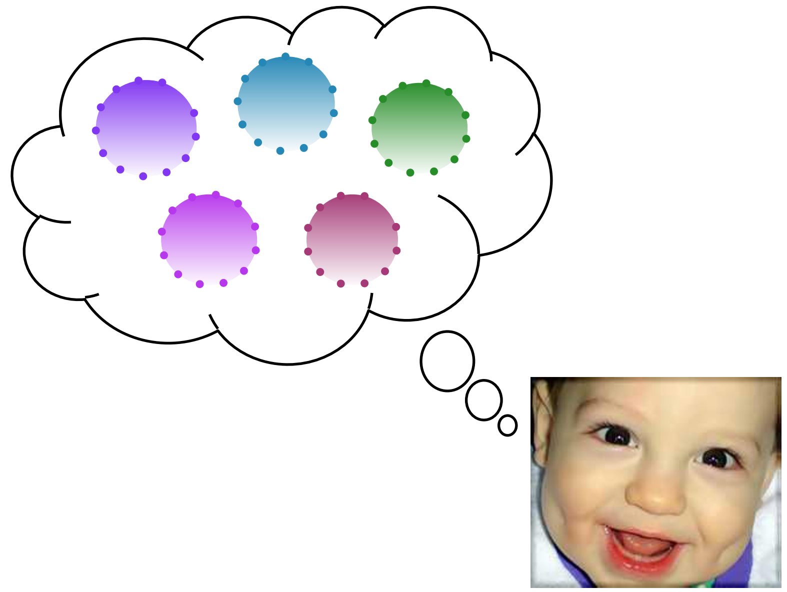
0.8 0.7 0.2 0.3 0.9

Today's Plan:
Computational models of syntactic acquisition

II. About linguistic parameters



About linguistic parameters

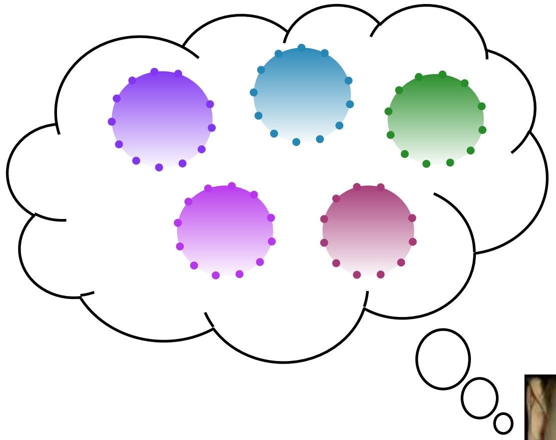


What are linguistic parameters?

How do they work?

What exactly are they supposed to do?

About linguistic parameters

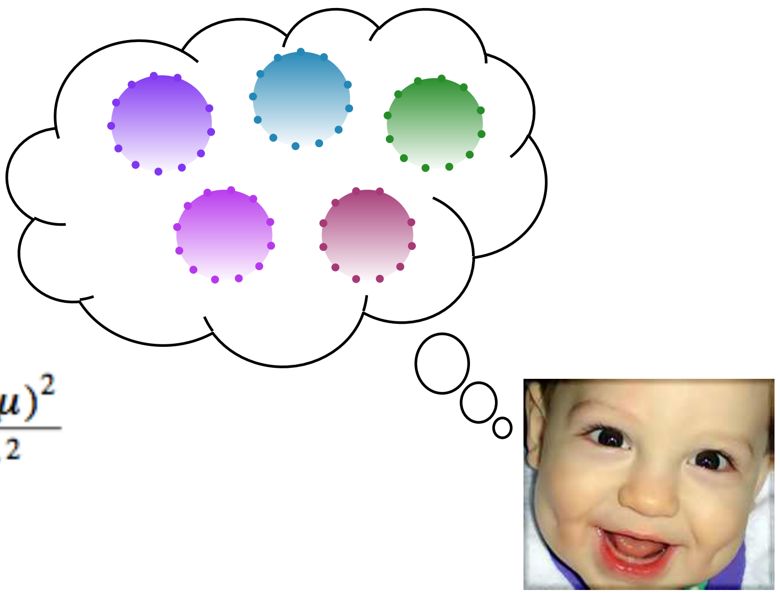


A parameter is meant to be something that can account for multiple observations in some domain.

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

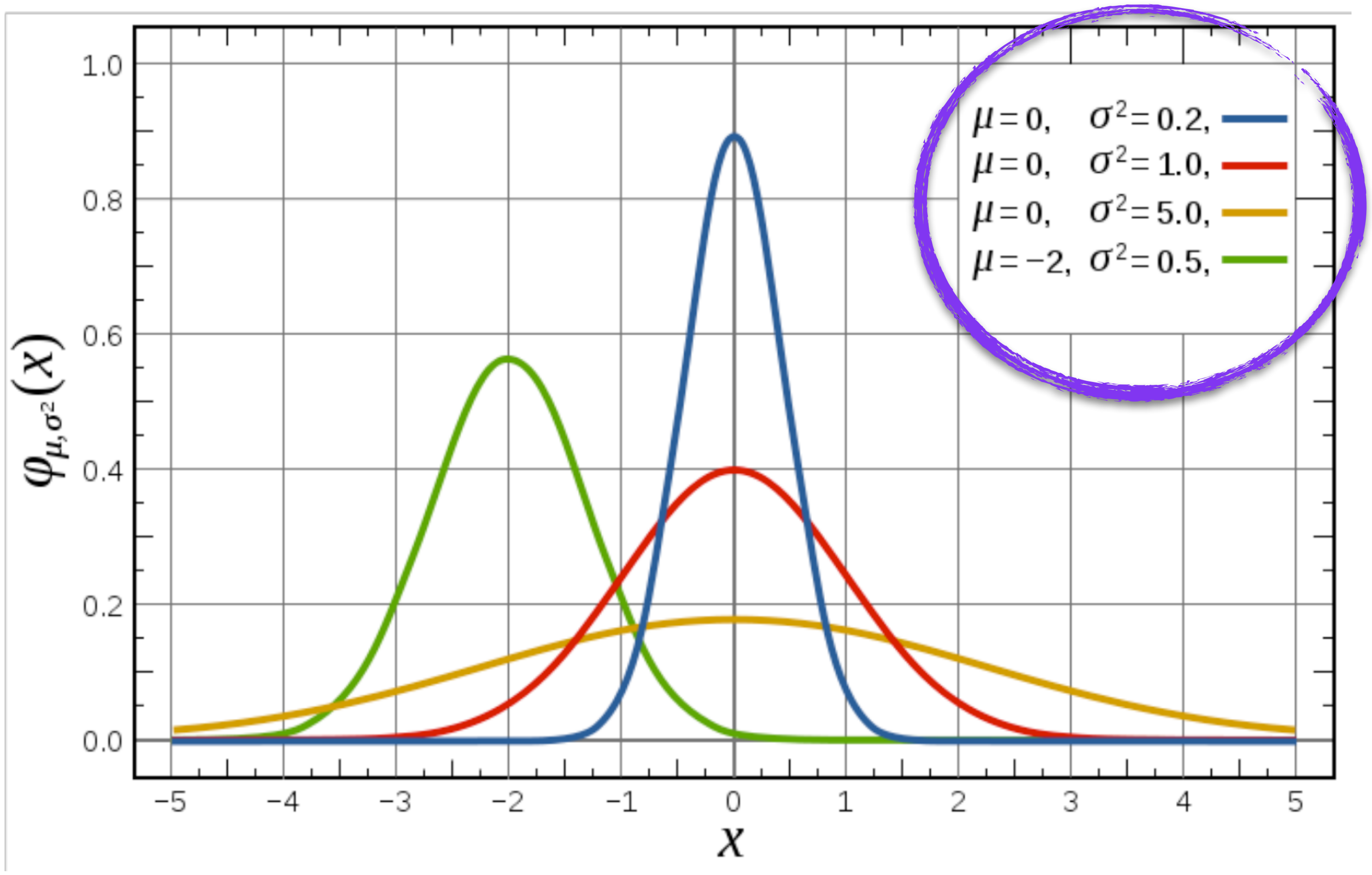


Statistical parameter

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

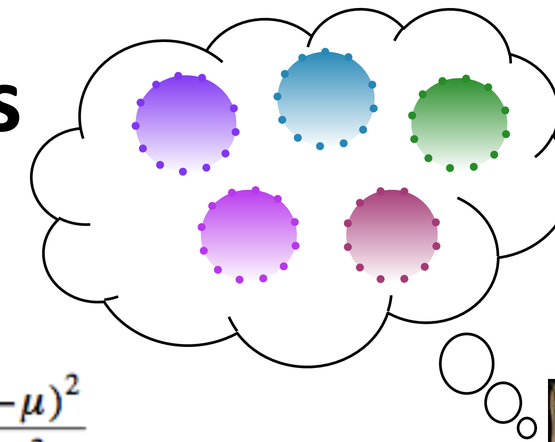
The normal distribution is a statistical model that uses **two parameters**:

- 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

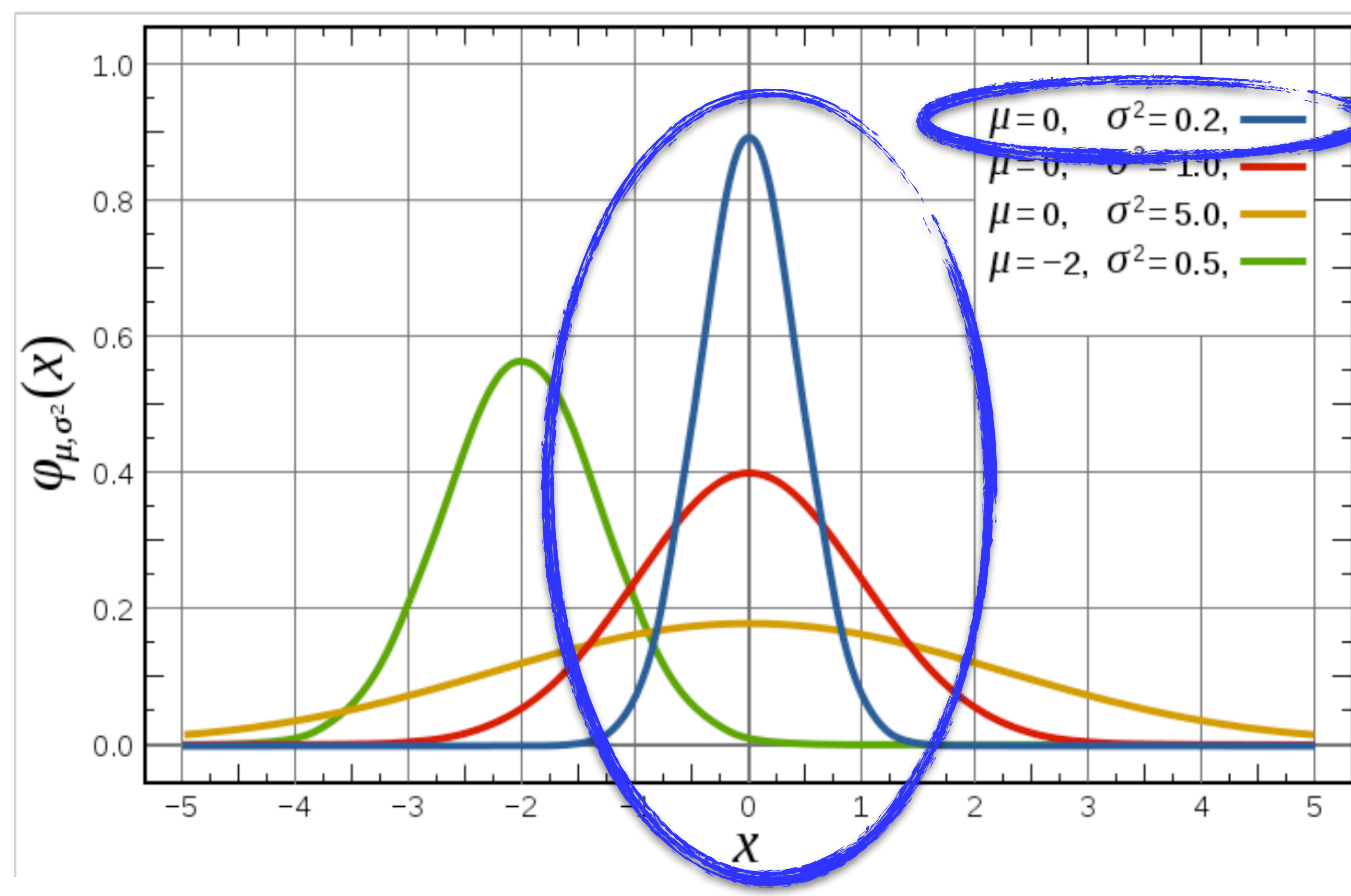


$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Statistical parameter

μ for the mean

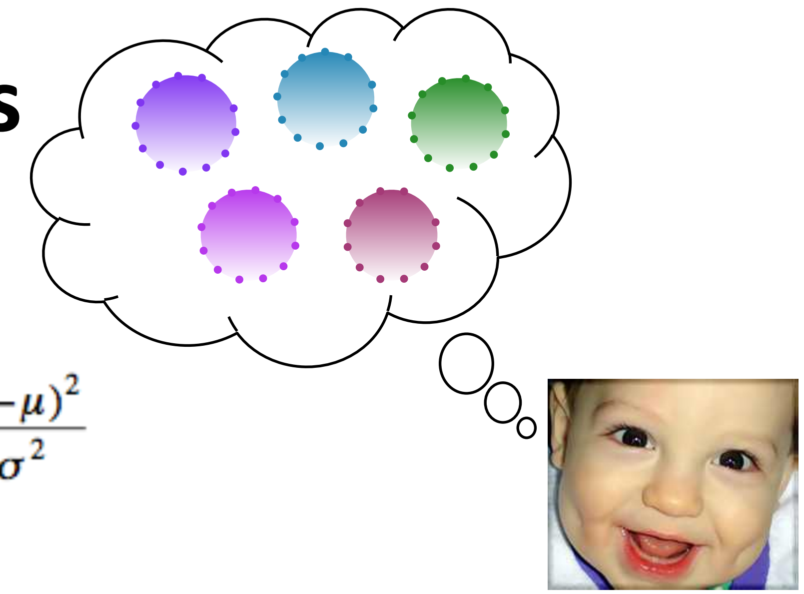
σ for the standard deviation



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



Statistical parameter

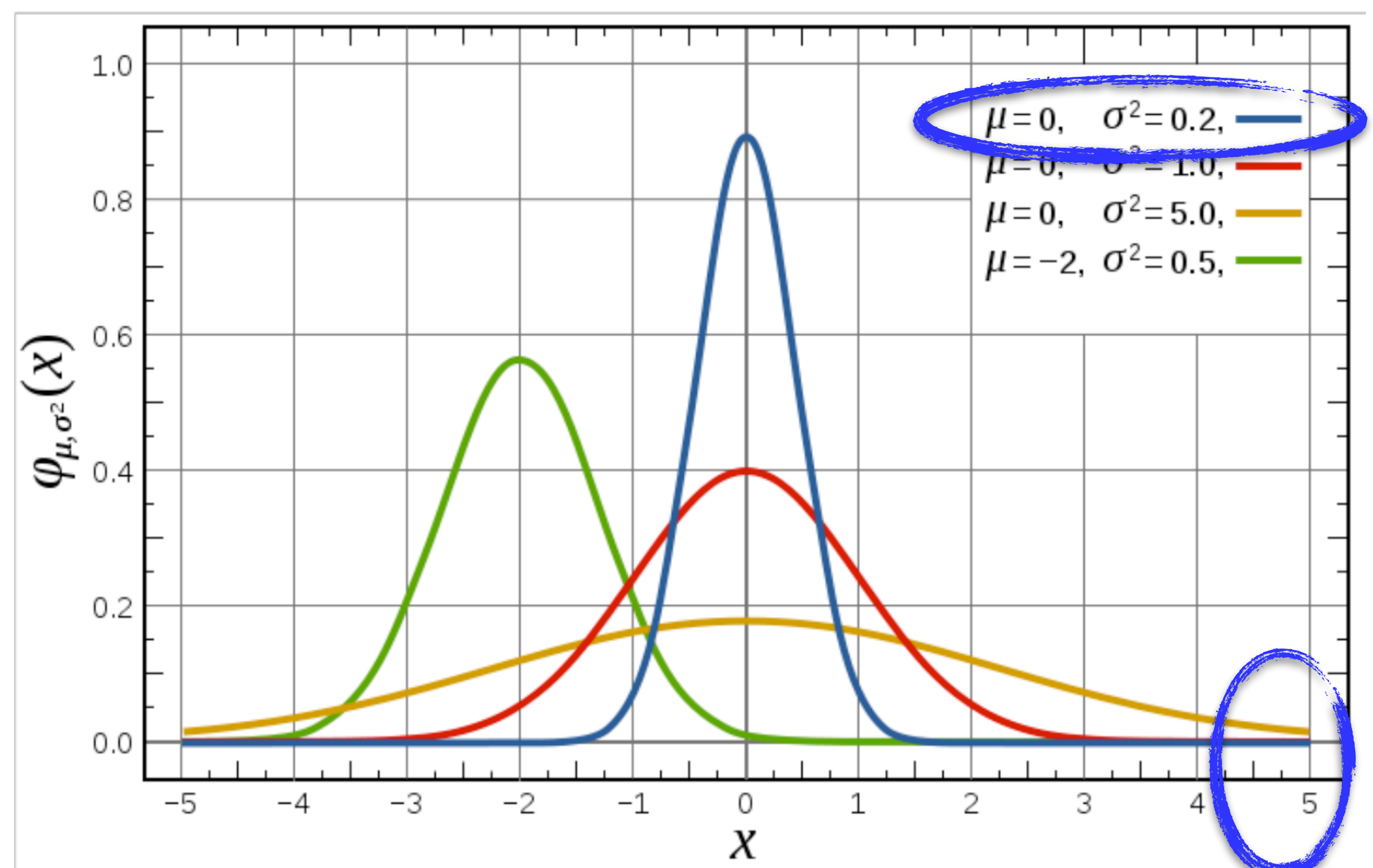
μ for the mean

σ for the standard deviation

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

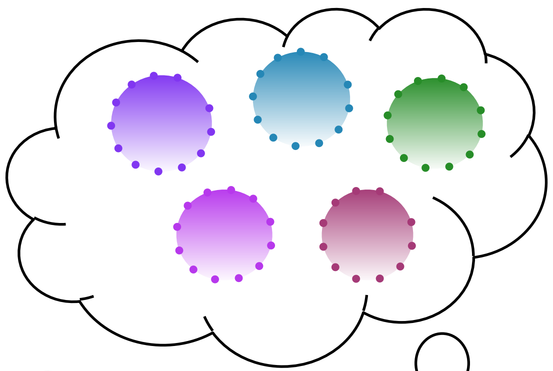
Let's use the model with
μ = 0 and σ^2 = 0.2.

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

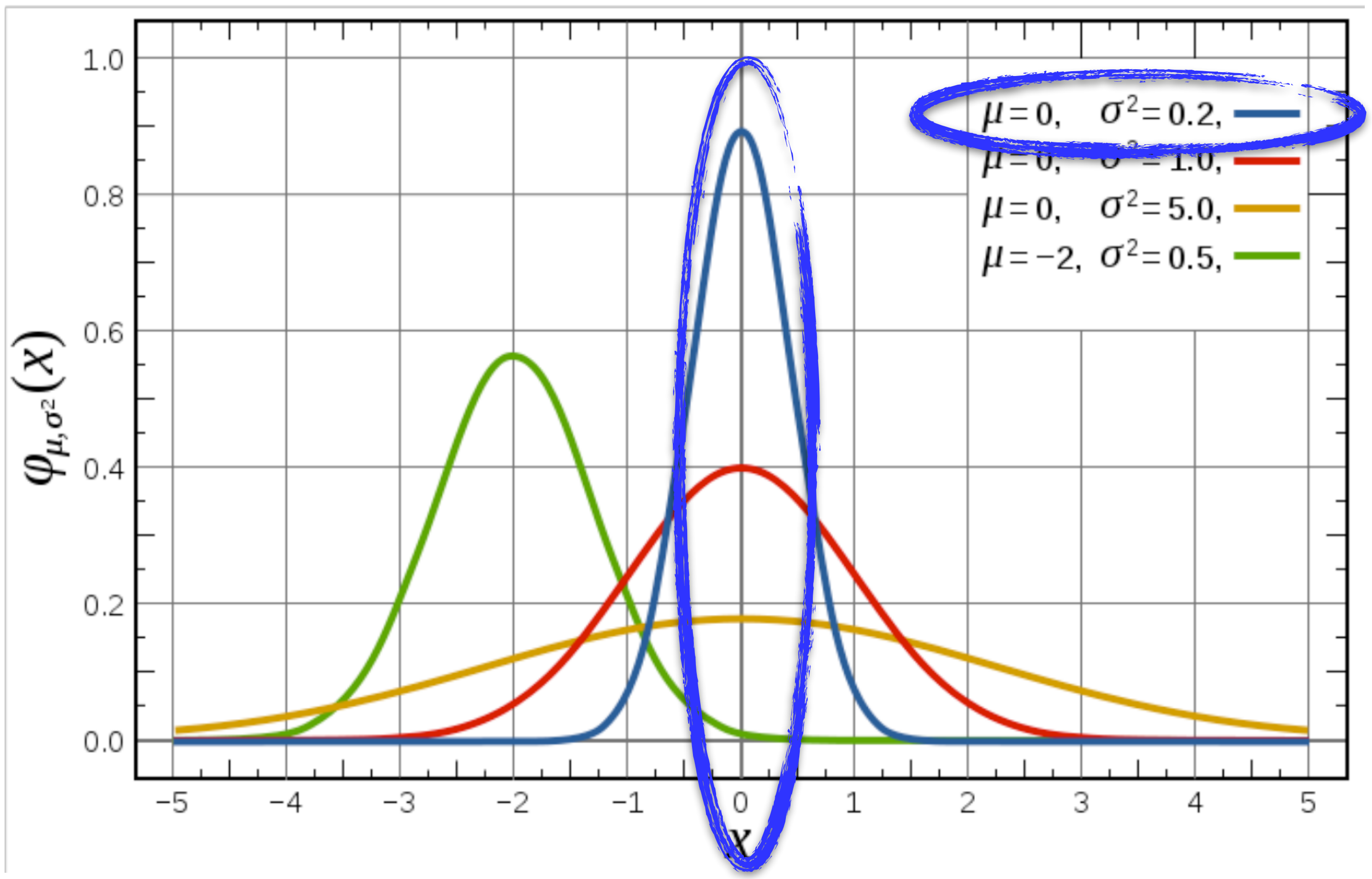
σ for the standard deviation

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

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

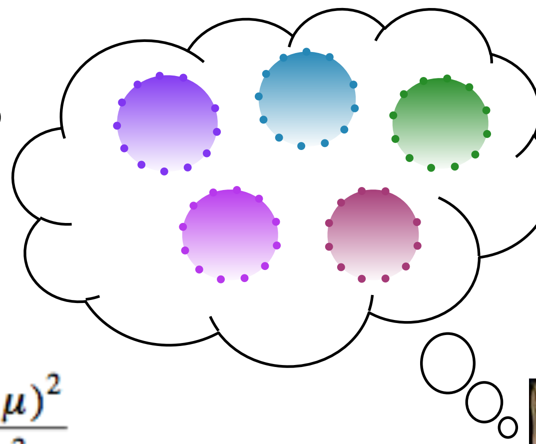
5 minutes late? ✗

What about right on time? ✔



Much more probable!

About linguistic parameters



Statistical parameter

μ for the mean

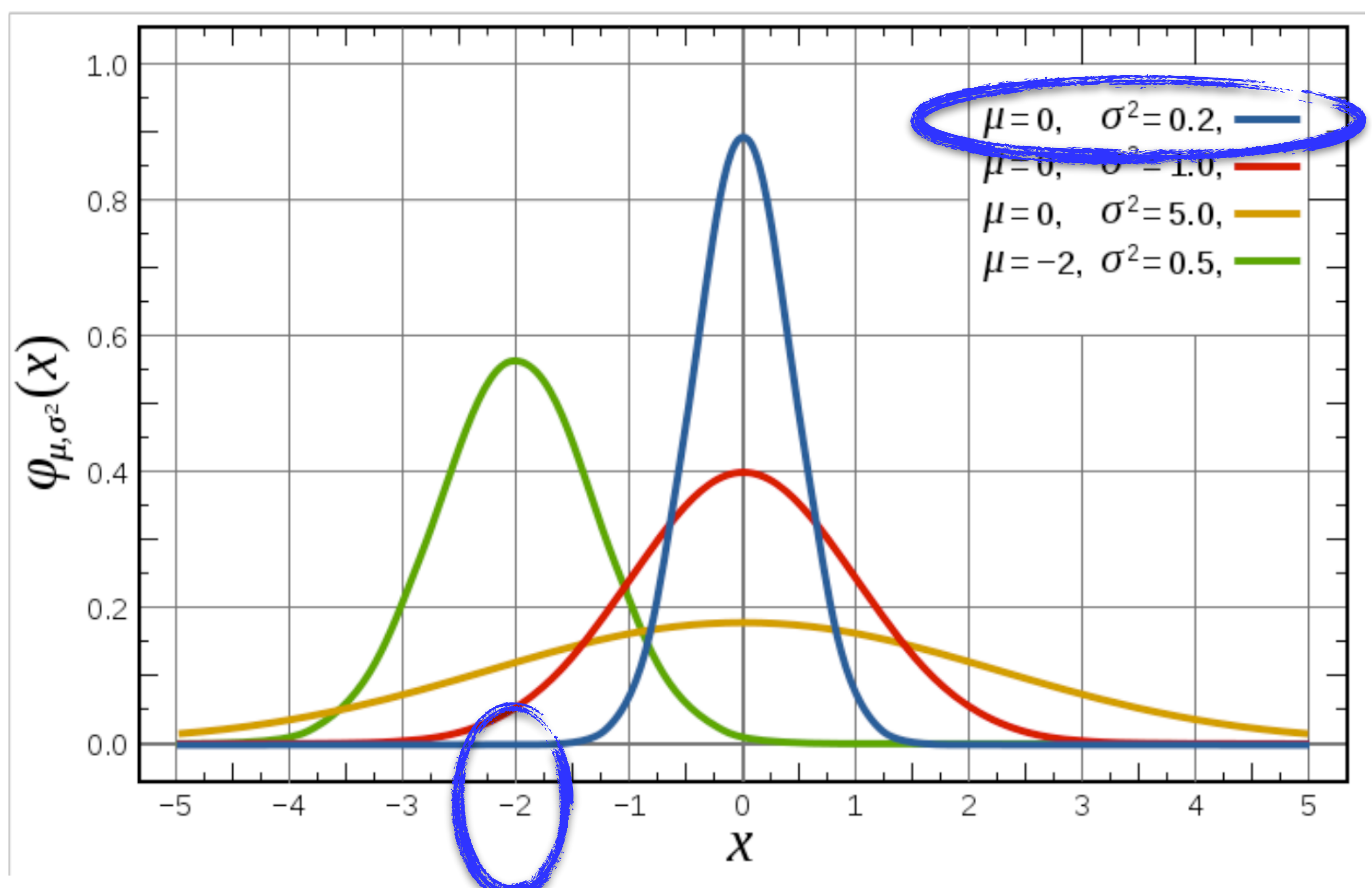
σ for the standard deviation

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

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

5 minutes late? X
On time? ✔

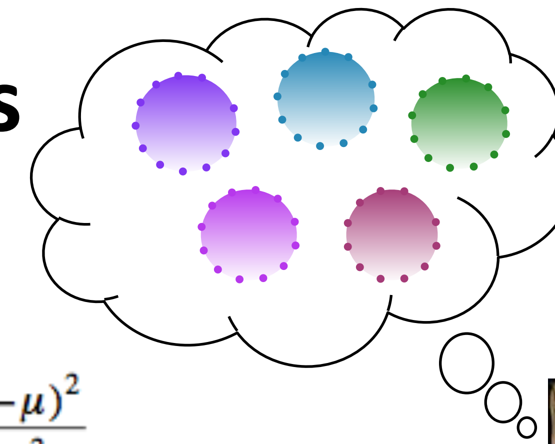
What about 2 minutes early? X



Not very probable!

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.

About linguistic parameters



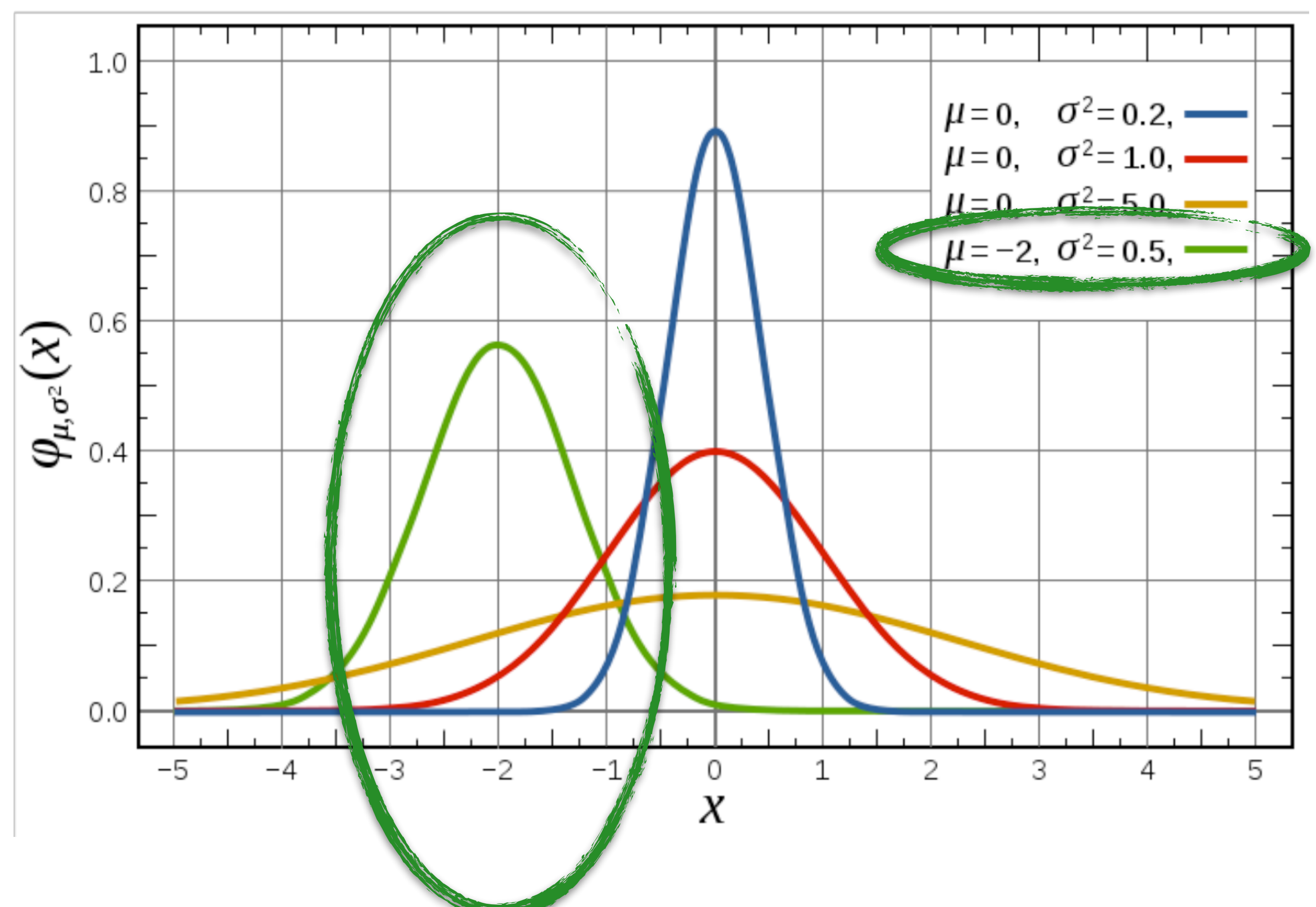
Statistical parameter

μ for the mean

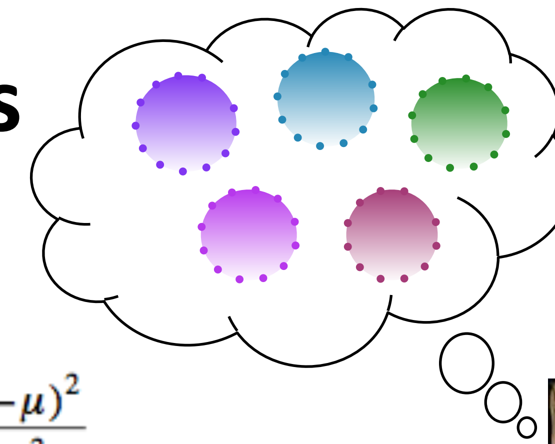
σ for the standard deviation

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

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



About linguistic parameters



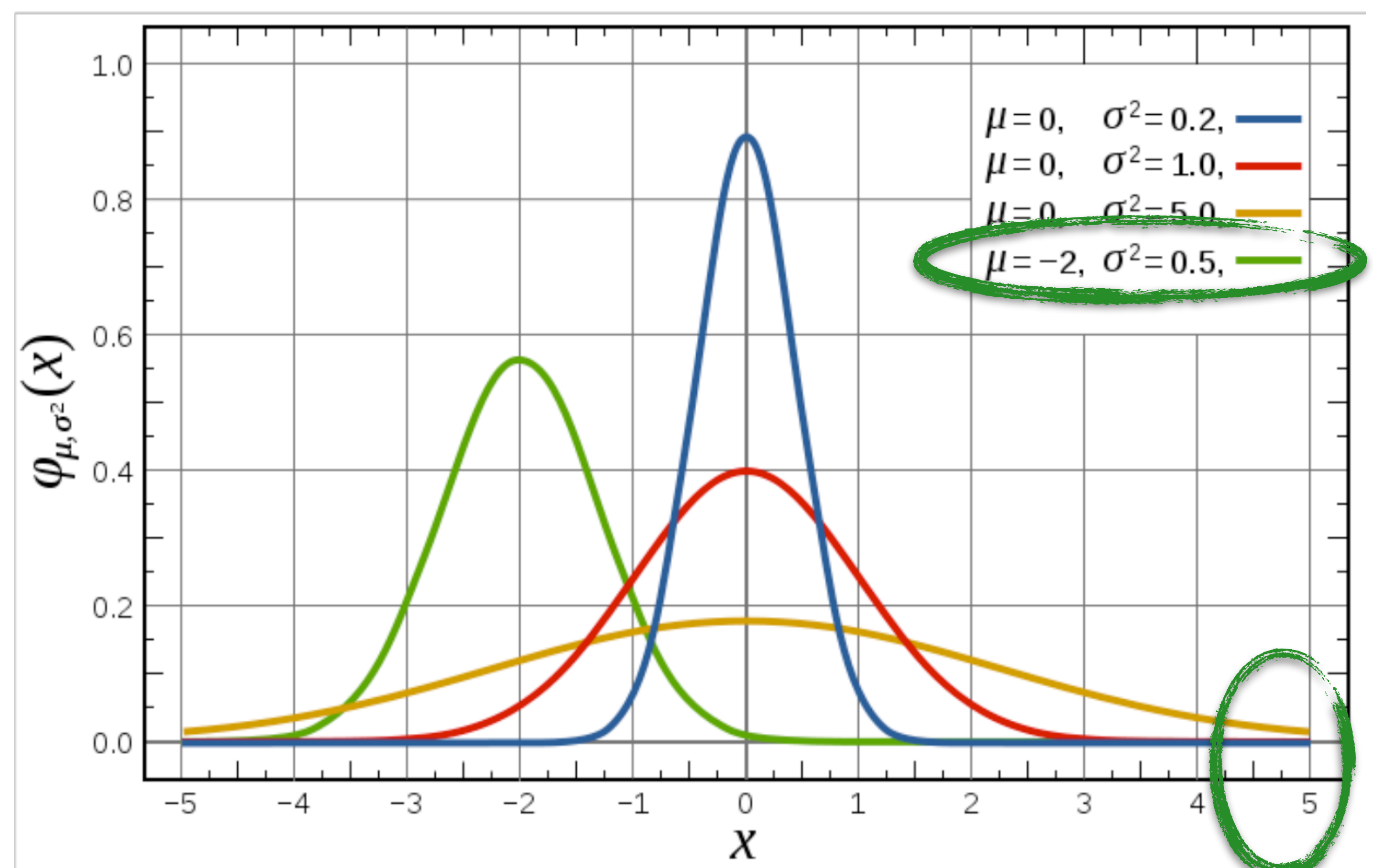
Statistical parameter

μ for the mean

σ for the standard deviation

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

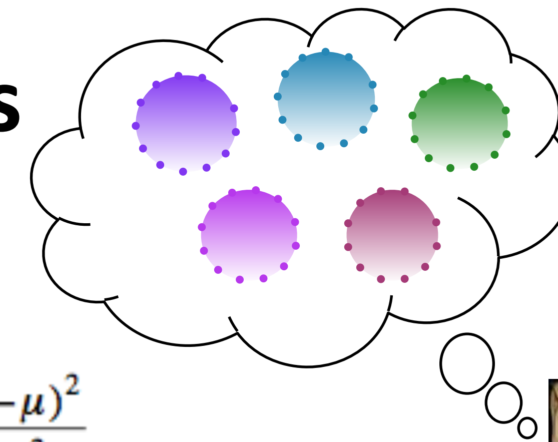
Let's shift to the model with **μ = -2 and σ^2 = 0.5**.



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

Not very probable!

About linguistic parameters



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

Statistical parameter

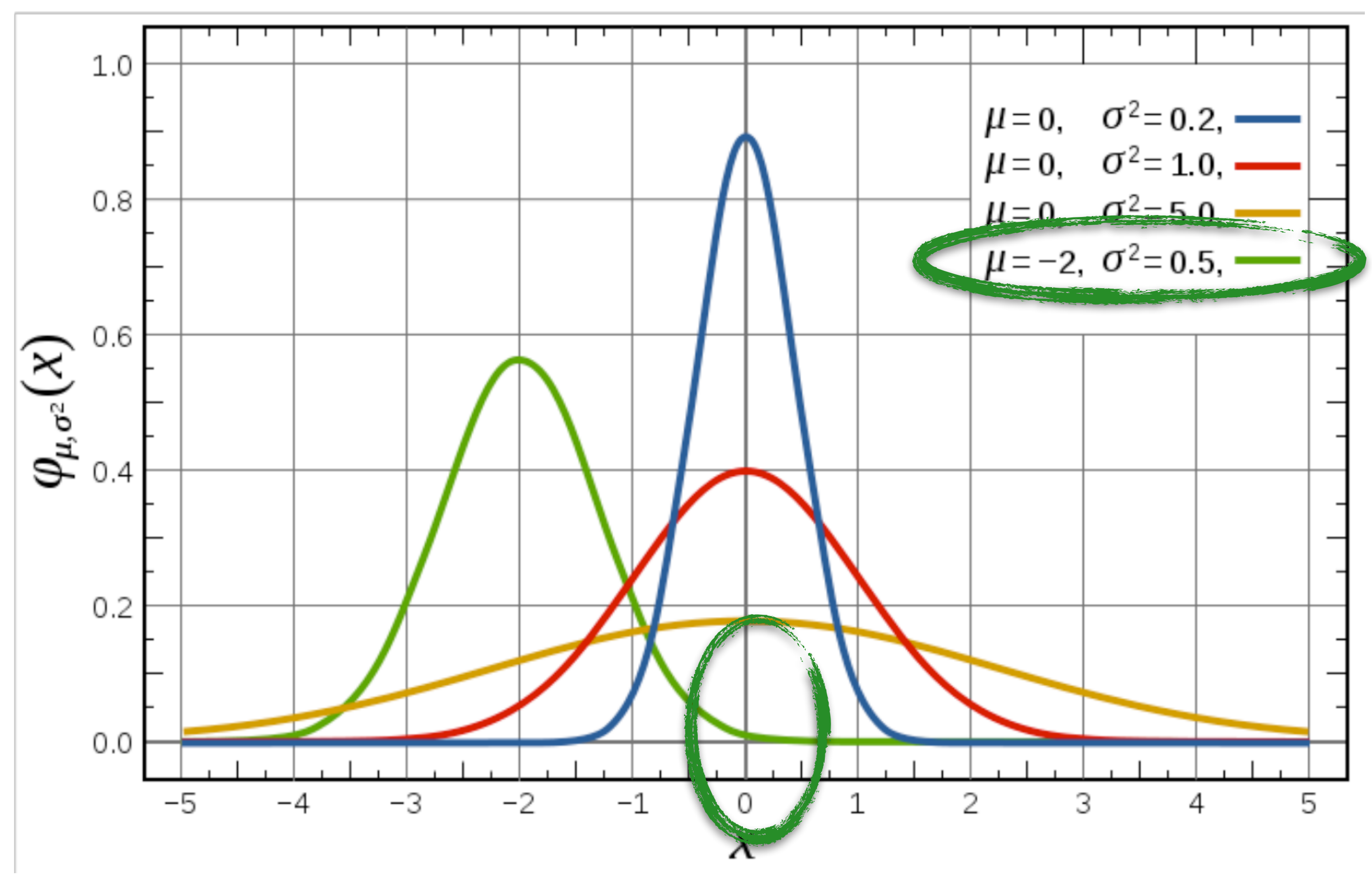
μ for the mean

σ for the standard deviation

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

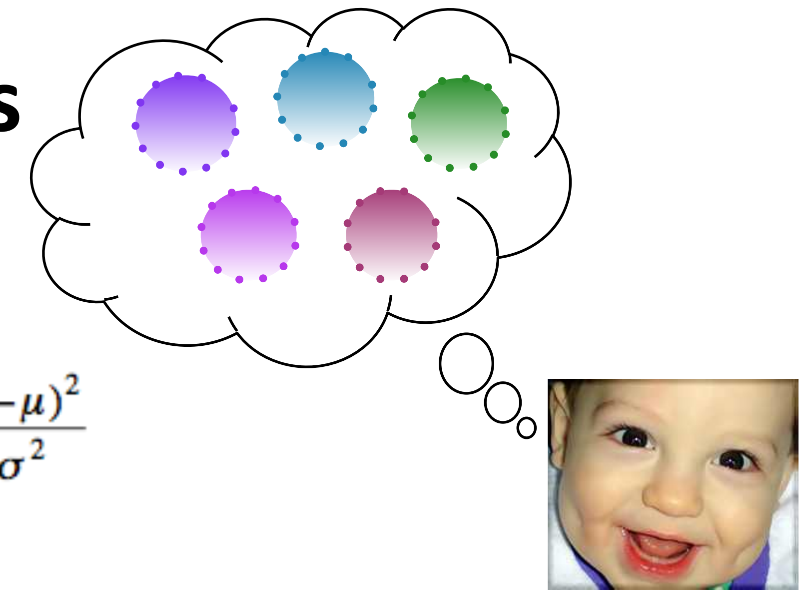
5 minutes late?

What about right on time?



Not very probable!

About linguistic parameters

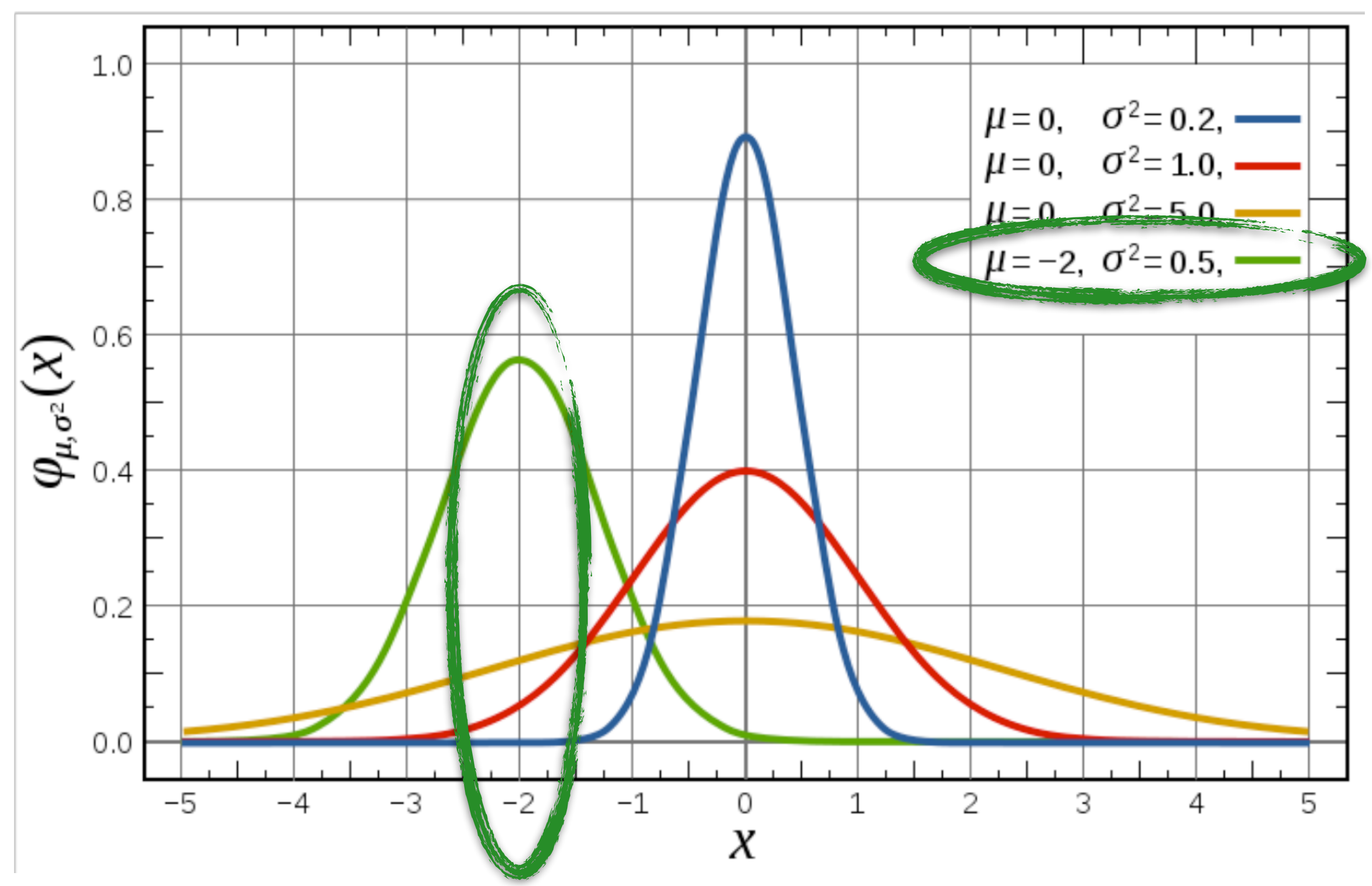


Statistical parameter

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? ✗

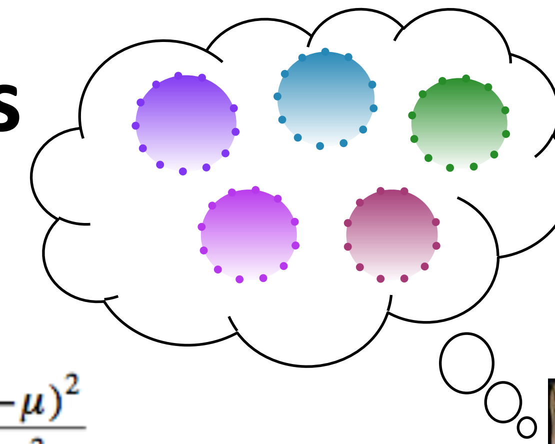
On time? ✗

What about 2 minutes early? ✔

Much more probable!

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

About linguistic parameters



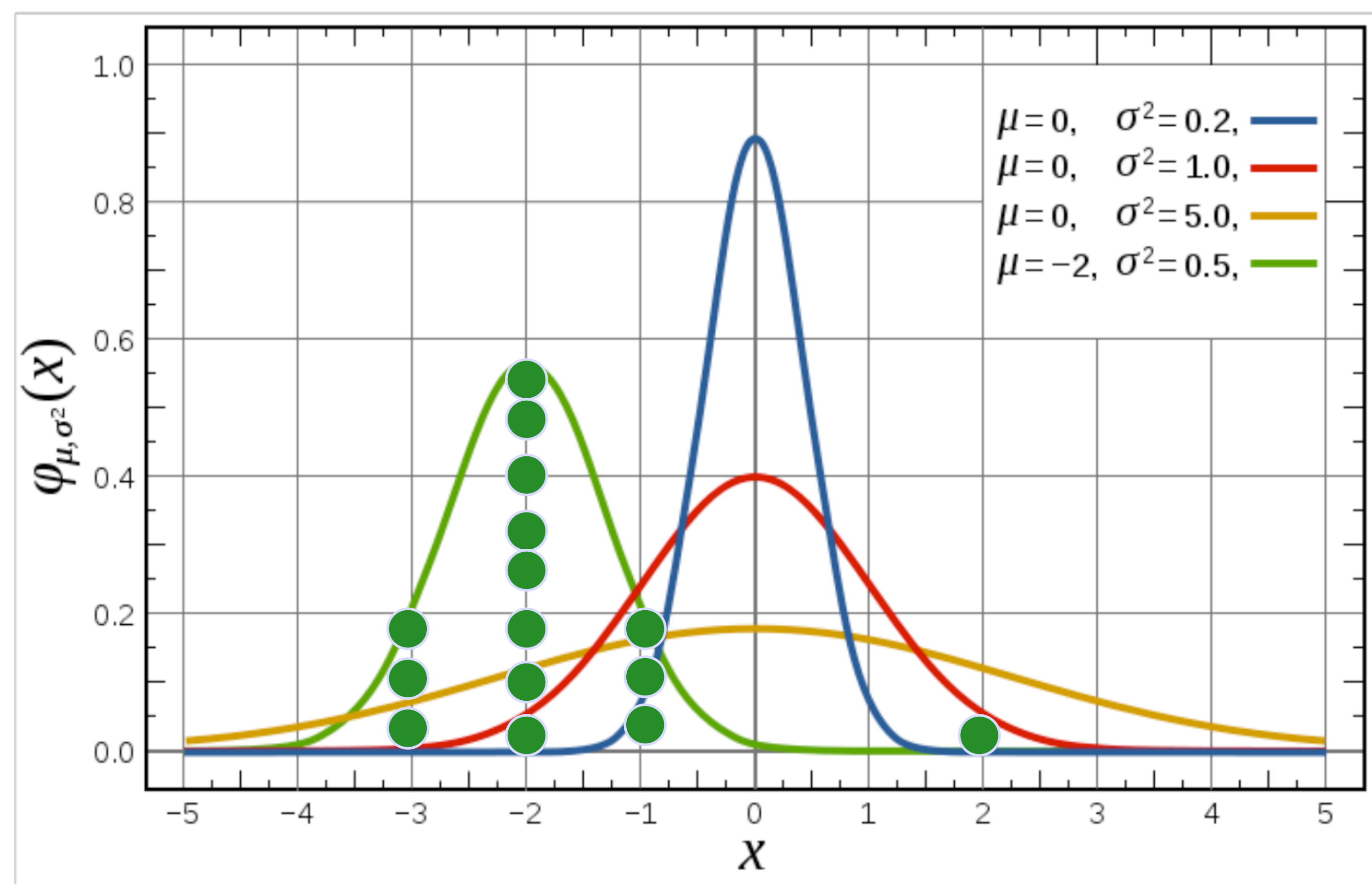
Statistical parameter

μ for the mean

σ for the standard deviation

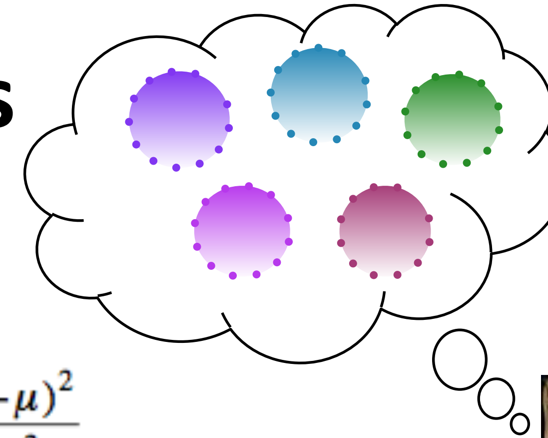
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

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

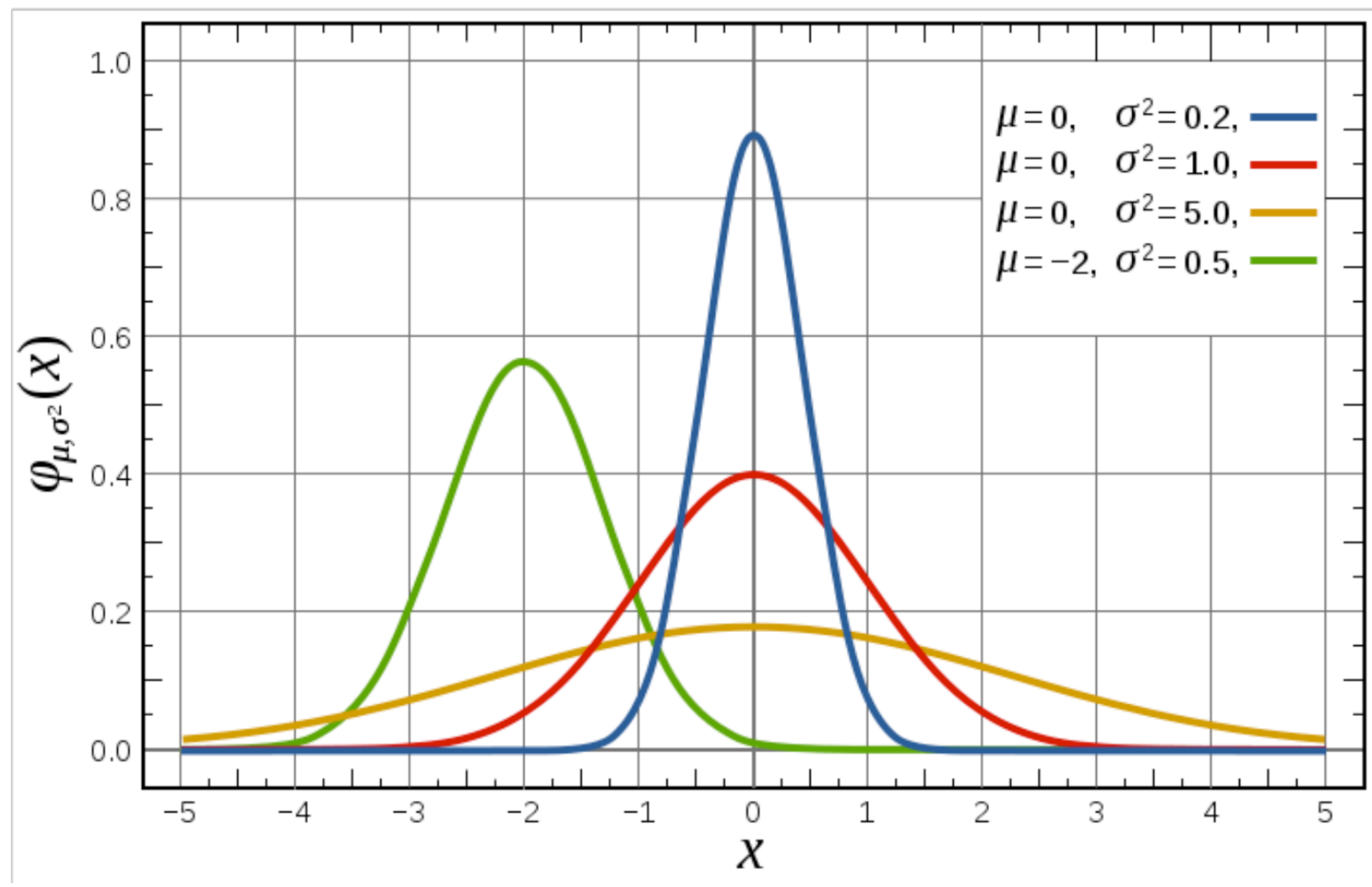


Statistical parameter

μ for the mean

σ for the standard deviation

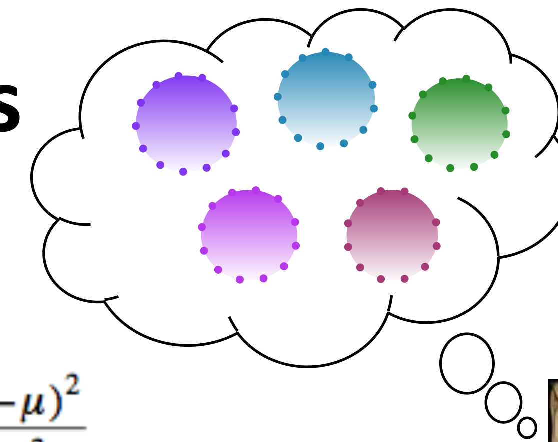
$$\varphi_{\mu,\sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$



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

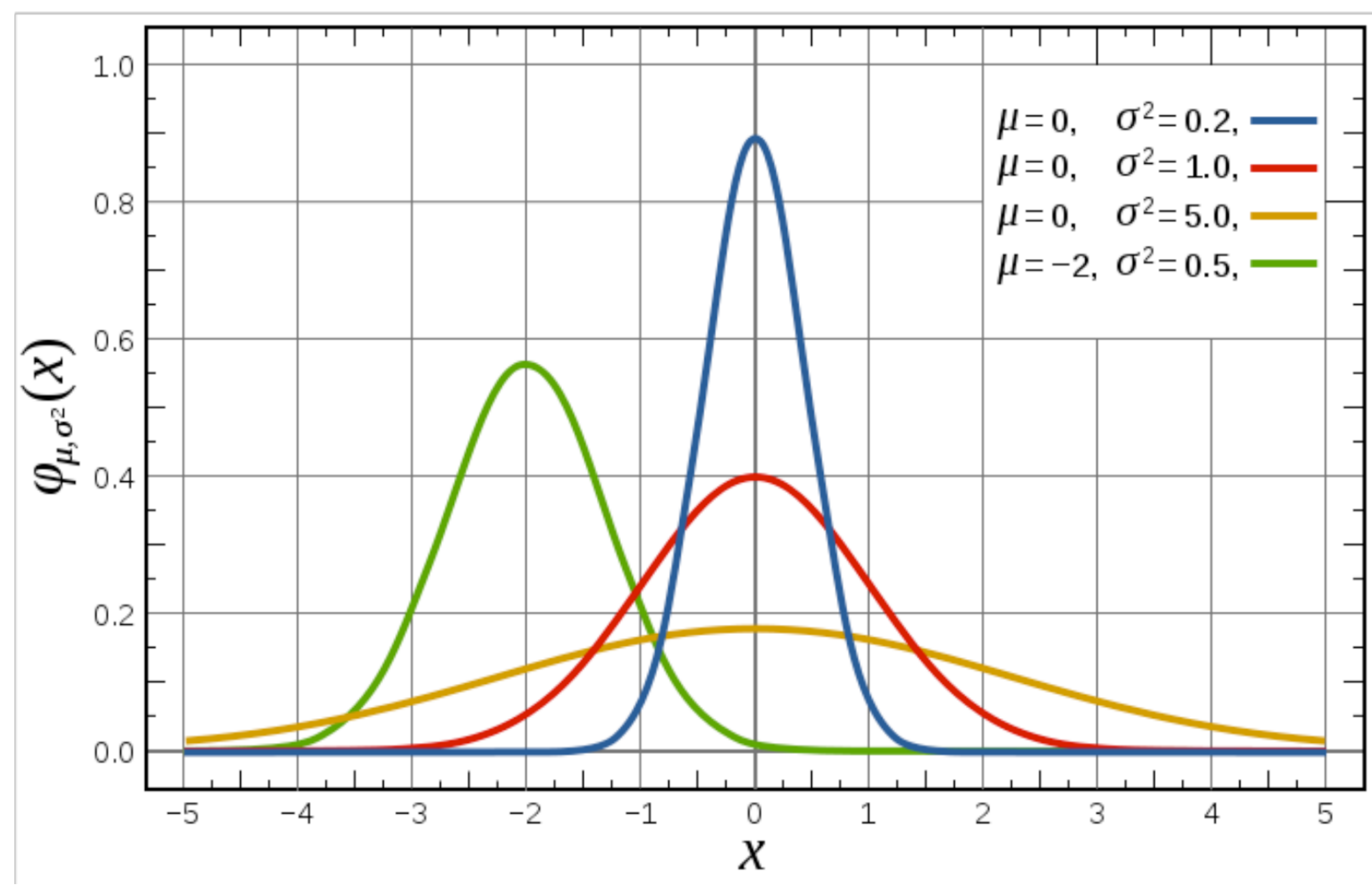


Statistical parameter

μ for the mean

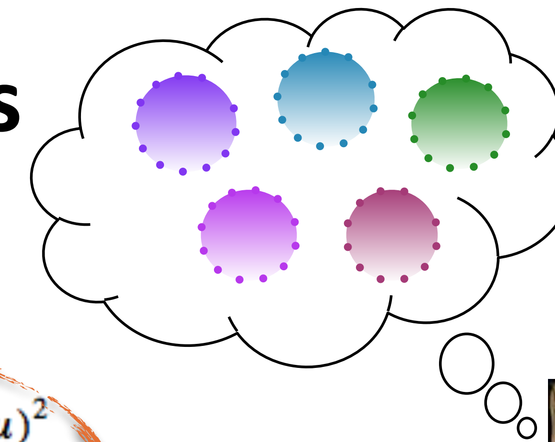
σ for the standard deviation

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



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

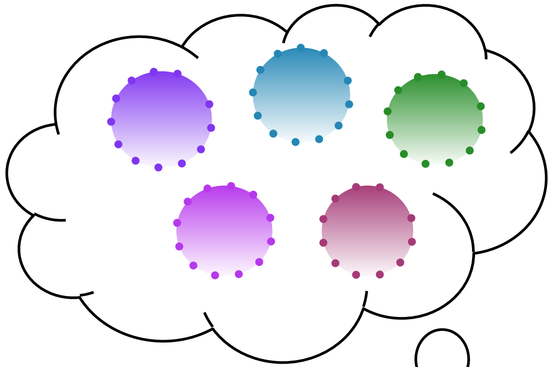
Comparison: μ and σ^2 determine the exact form of the curve that represents the probability of observing certain data. While different values for these parameters can produce many different curves, these curves share their underlying form due to the common invariant function.

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

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

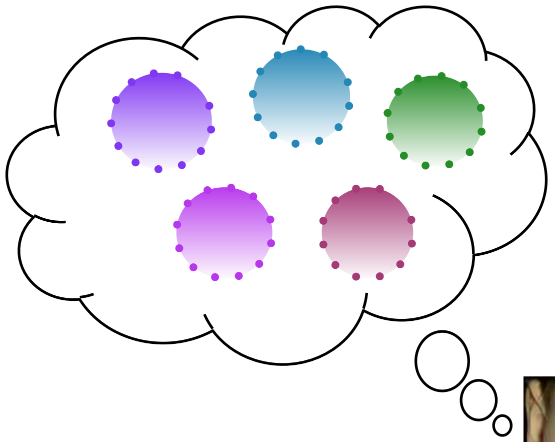
Linguistic parameters correspond to the properties that vary across human languages

About linguistic parameters for language acquisition



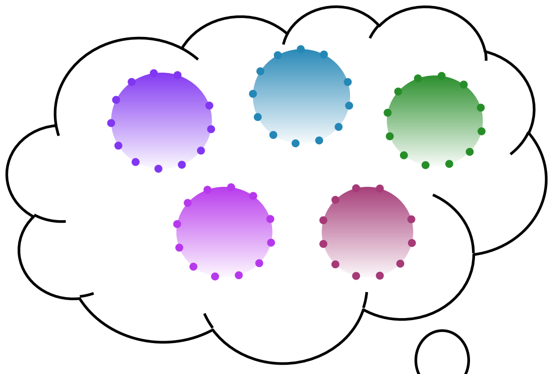
Parameters connecting to multiple structural properties is a very good thing from the perspective of someone trying to acquire language (like a child). This is because a child can learn about a parameter's value by observing many different kinds of examples in the language.

About linguistic parameters for language acquisition



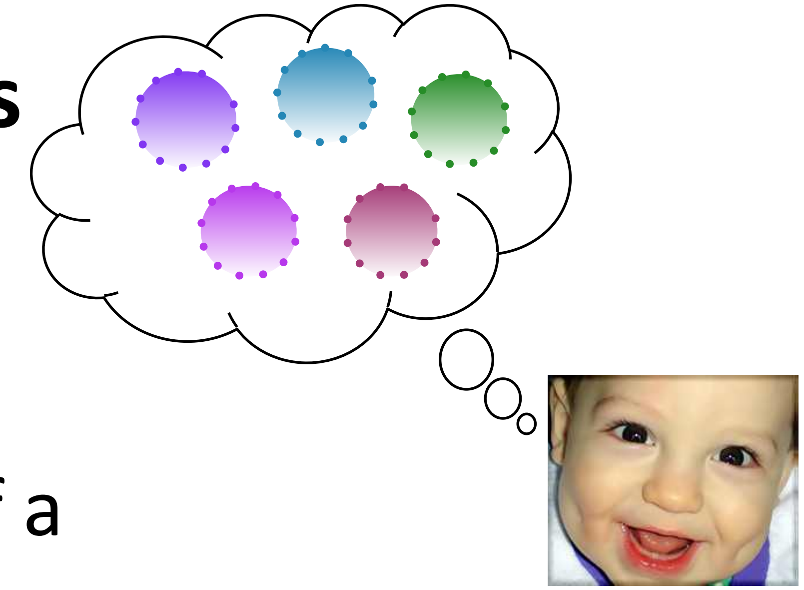
"The richer the deductive structure associated with a particular parameter, the greater the range of potential 'triggering' data which will be available to the child for the 'fixing' of the particular parameter" – Hyams (1987)

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.

About linguistic parameters for language acquisition



An issue: The observable data are often the result of a combination of interacting parameters.

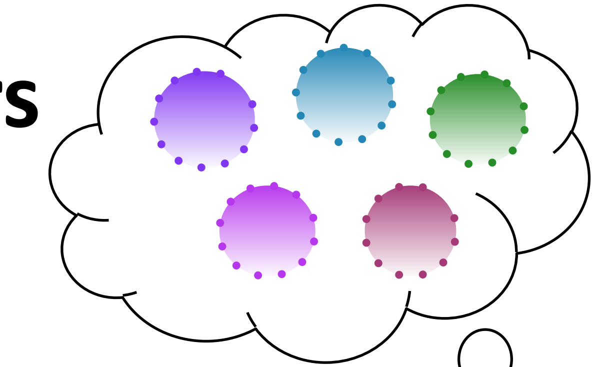
This can make it hard to figure out what parameter values might have produced the observable data - even if the child already knows what the parameters are.

Observable data can be ambiguous for which parameter values they signal.

Observable data

Subject Verb Object

About linguistic parameters for language acquisition



An issue: The observable data are often the result of a combination of interacting parameters.

Observable data can be ambiguous for which parameter values they signal.

Subject Verb Object

German

Subject Verb *Subject* Object *Verb*

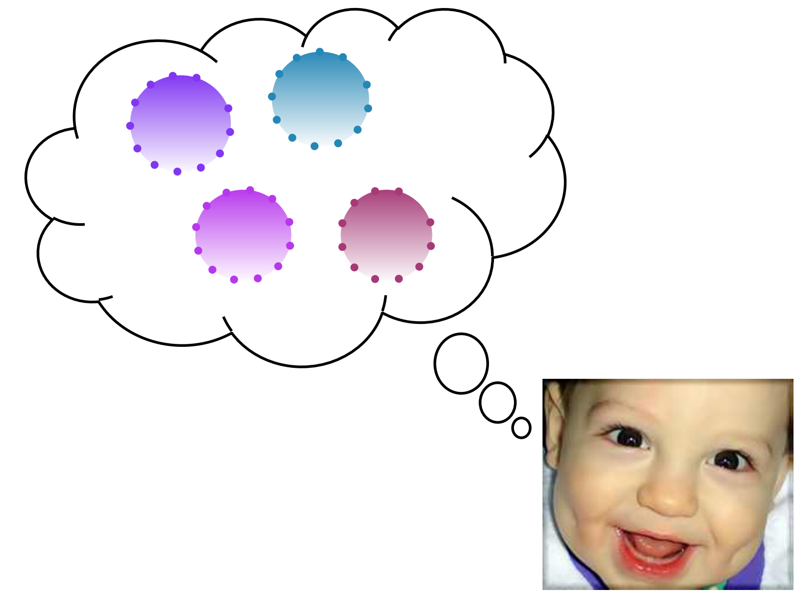
Kannada

Subject *Object* Verb Object

English

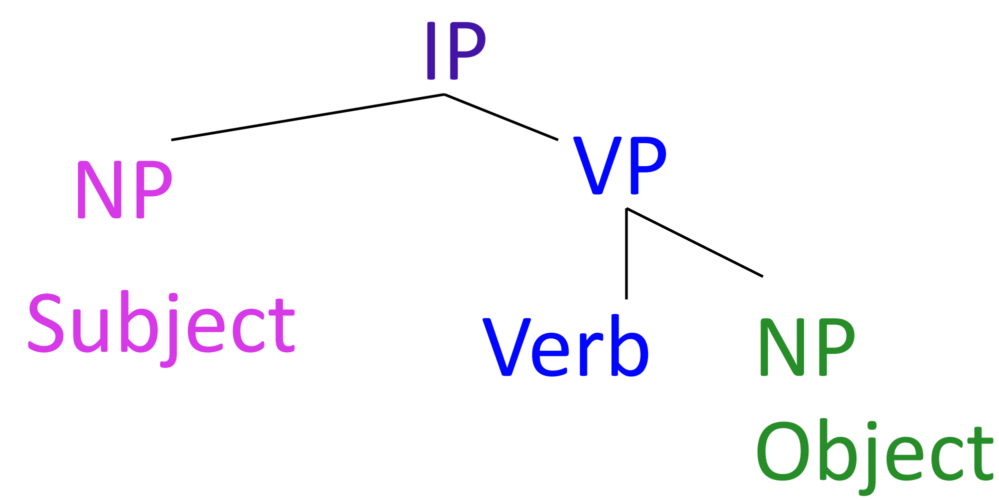
Subject Verb Object

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

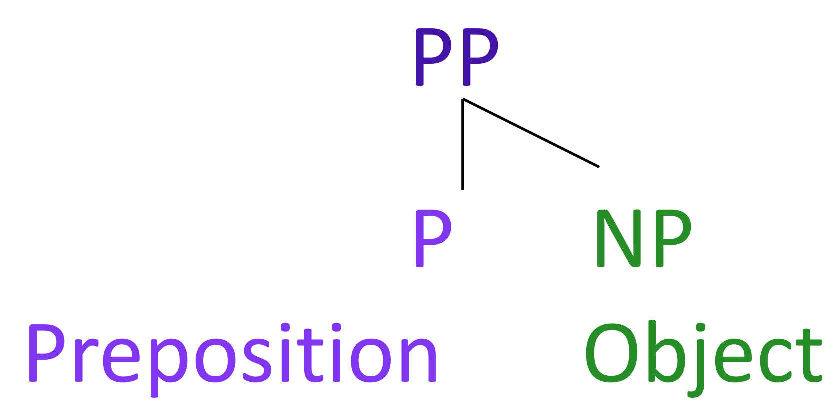


Basic word order:

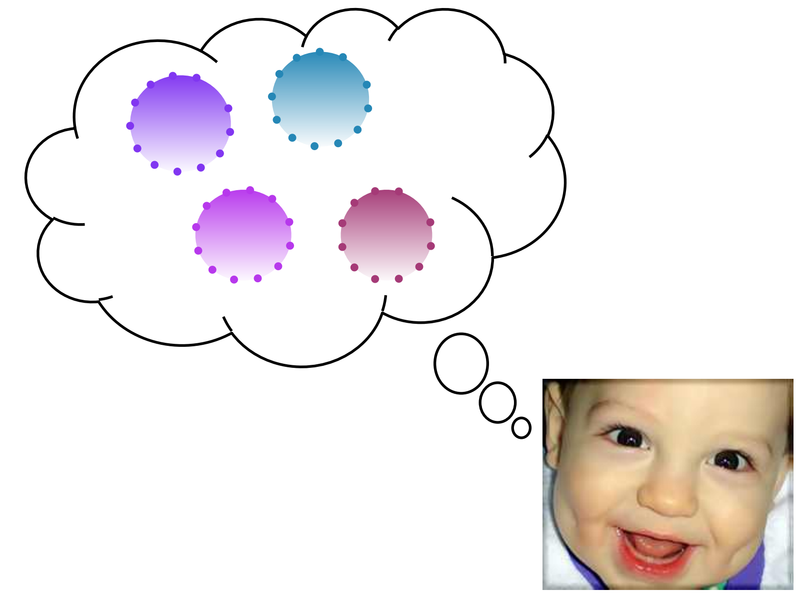
Subject Verb Object [SVO]

Prepositions:

Preposition Noun Phrase



Interacting parameters



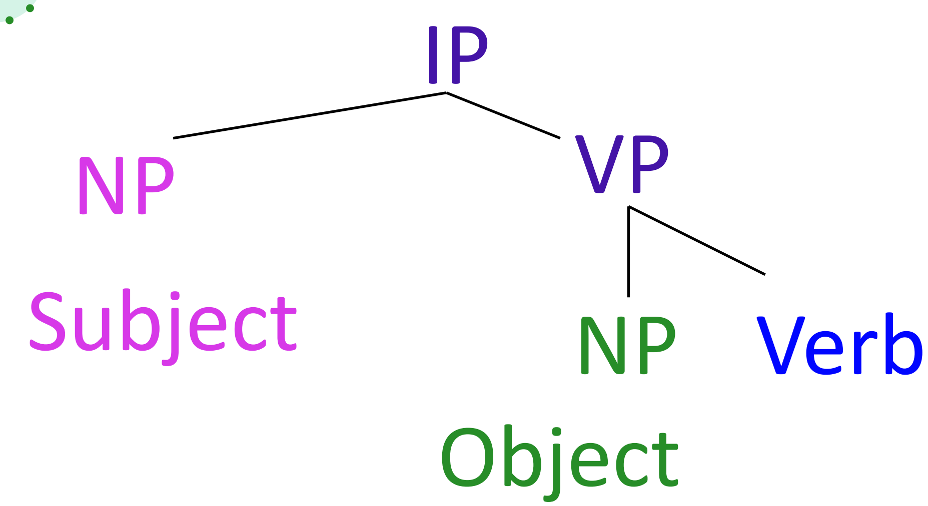
Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

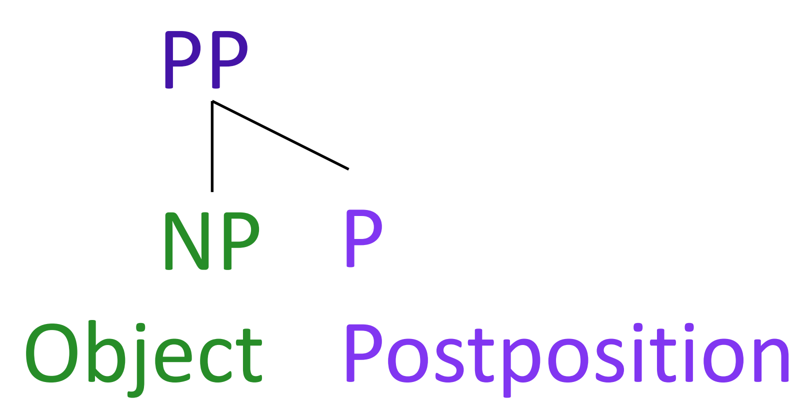
Basic word order:

Subject Object Verb [SOV]

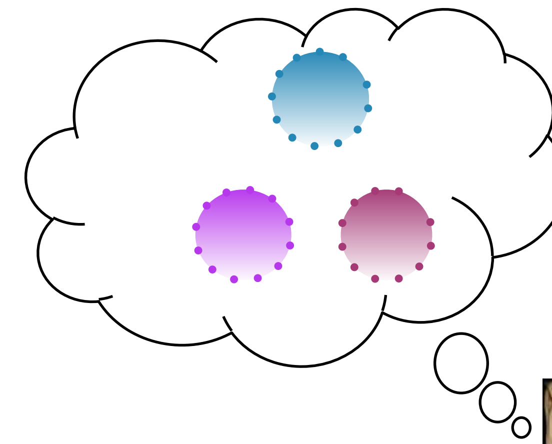


Postpositions:

Noun Phrase Postposition



Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

German: +V2

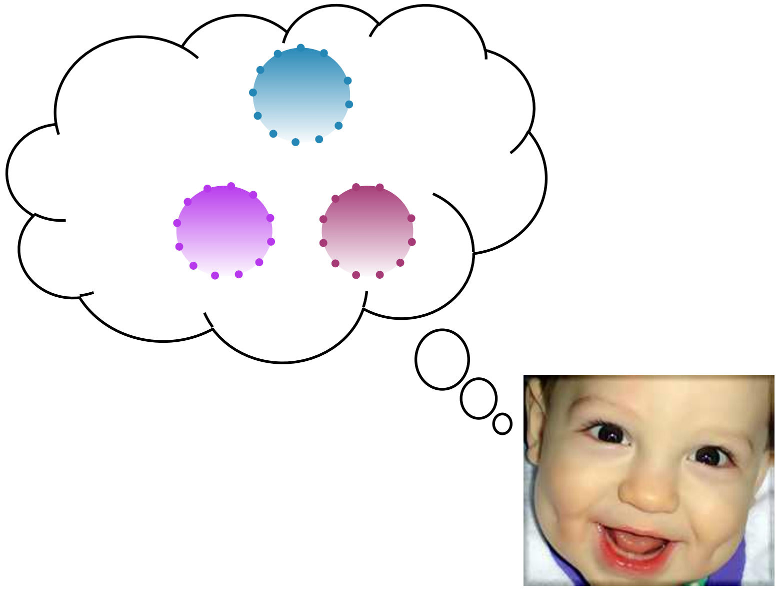
Verb moves to second phrasal position, some other phrase moves to the first position

Sarah das Buch liest

Sarah the book reads

Underlying form of the sentence

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

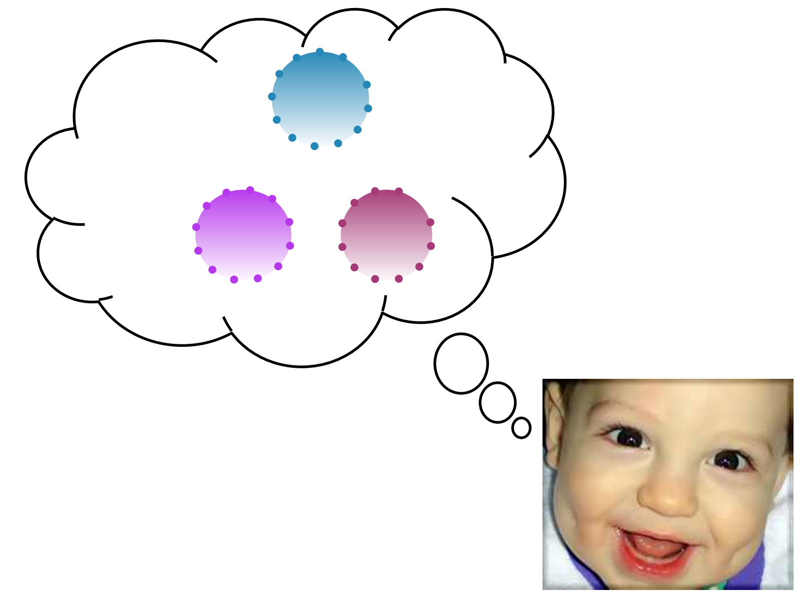
German: +V2

Verb moves to second phrasal position, some other phrase moves to the first position

Sarah	liest	Sarah	das Buch	liest
Sarah	*reads*		*the book*	

Observable form of the sentence

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

German: +V2

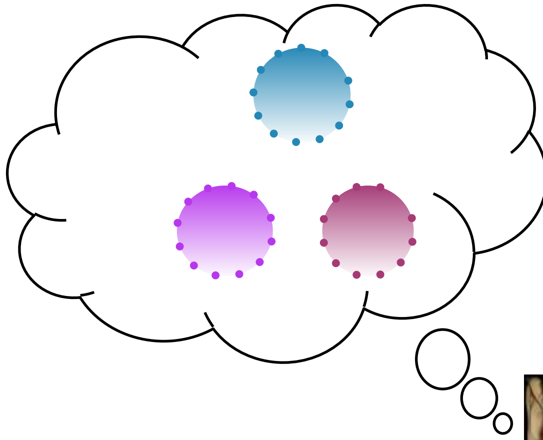
Verb moves to second phrasal position, some other phrase moves to the first position

Sarah das Buch liest

Sarah the book reads

Underlying form of the sentence

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

German: +V2

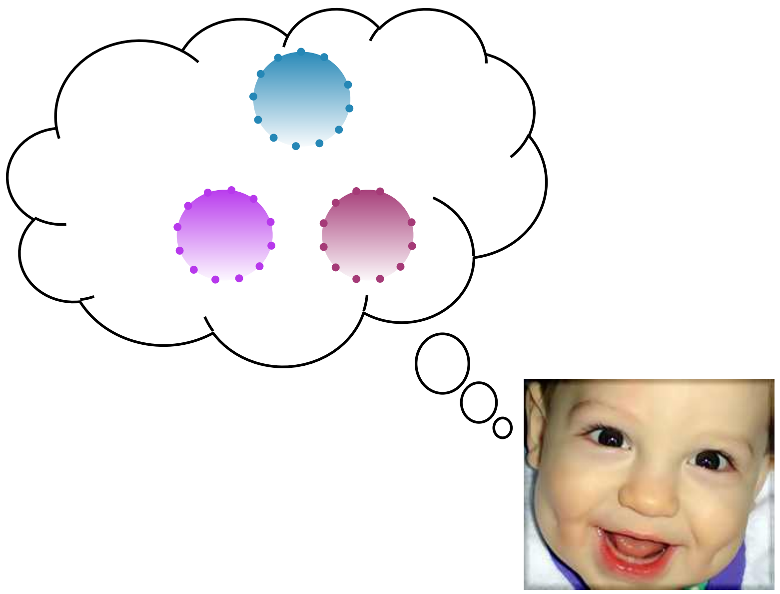
Verb moves to second phrasal position, some other phrase moves to the first position

Das Buch liest Sarah das Buch liest

The book reads Sarah

Observable form of the sentence

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

German: +V2

English: -V2

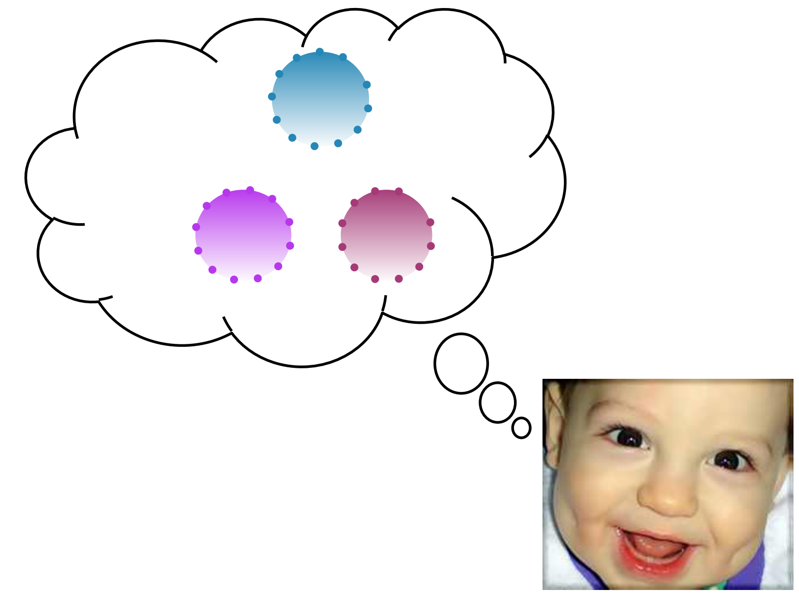
Verb doesn't move.

Sarah reads the book

Underlying form of the sentence

Observable form of the sentence

Interacting parameters



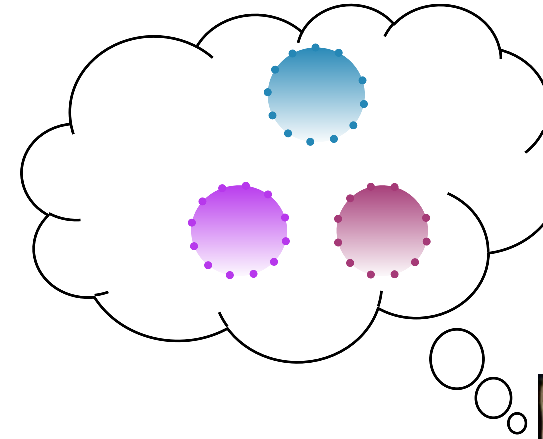
Head-directionality

Verb Second (V2)

Grammars available

G1	Head-first	+V2
G2	Head-final	+V2
G3	Head-first	-V2
G4	Head-final	-V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Data point

Subject Verb Object

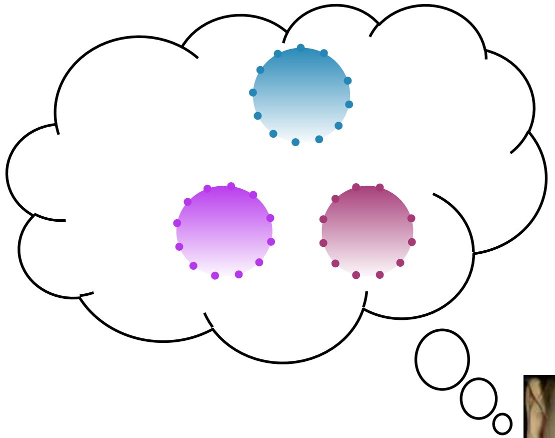
G1 Head-first +V2

G2 Head-final +V2

G3 Head-first -V2

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

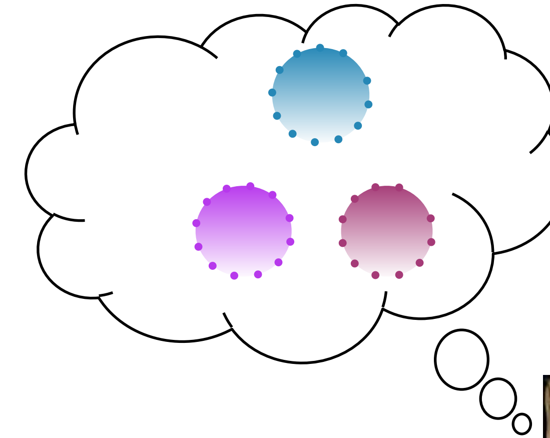
"I love kitties."

Subject **Verb** **Object**

Which grammars can analyze this data point?

G1	Head-first	+V2
G2	Head-final	+V2
G3	Head-first	-V2
G4	Head-final	-V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject **Verb** ***Verb*** **Object**

G1 Head-first +V2

✔+head-first predicts SVO

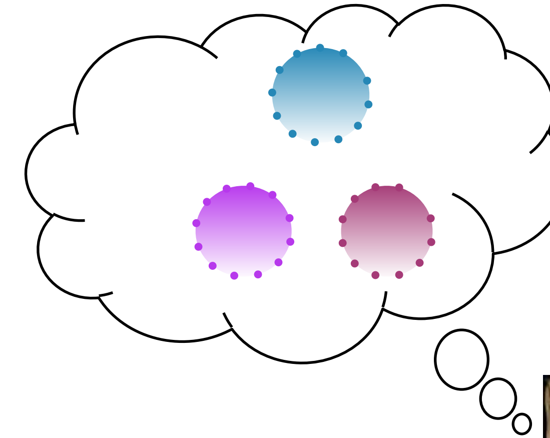
✔+V2 predicts Verb moved to second position

G2 Head-final +V2

G3 Head-first -V2

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject Verb *Subject* Object *Verb*

G2 Head-final +V2

✔ head-final predicts SOV

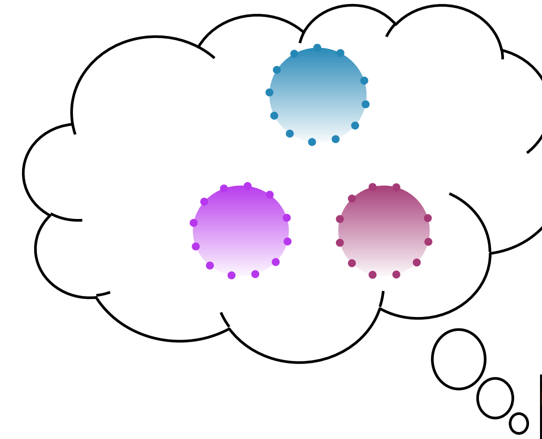
✔ +V2 predicts Verb moved to second position

✔ **G1** Head-first +V2

G3 Head-first -V2

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject   Verb   Object

G3 Head-first, -V2

✔ head-first predicts SVO

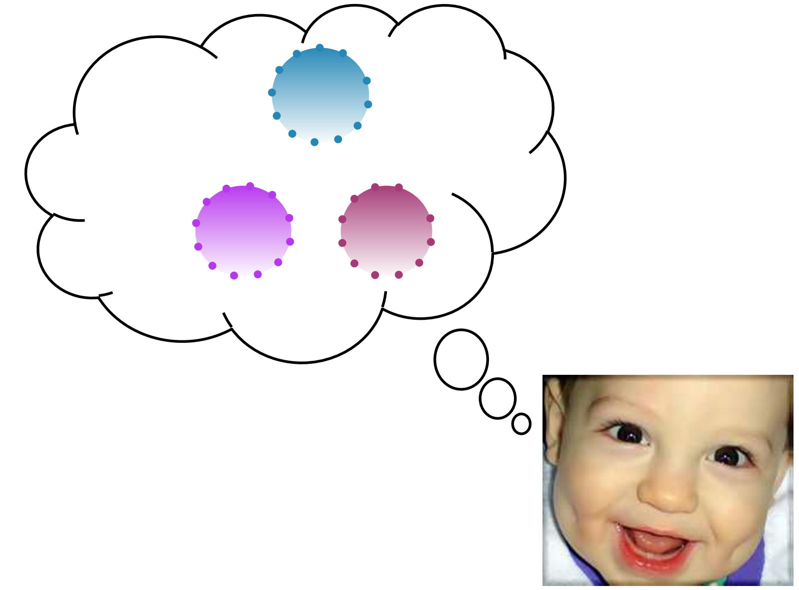
✔ -V2 predicts Verb doesn't move

✔ **G1** Head-first, +V2

✔ **G2** Head-final, +V2

G4 Head-final, -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject **Verb** **Object**

G4 Head-final -V2

✗ head-final predicts SOV

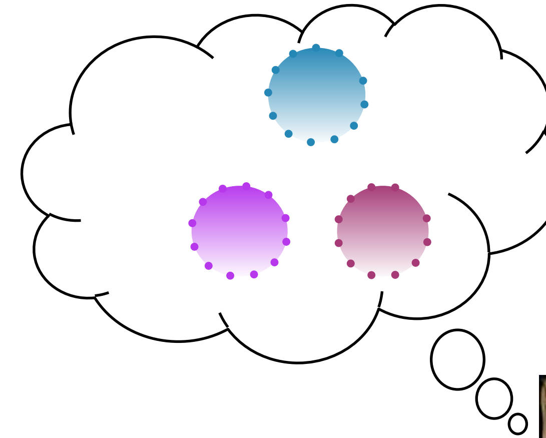
✔ -V2 predicts Verb doesn't move

✔ **G1** Head-first +V2

✔ **G2** Head-final +V2

✔ **G3** Head-first -V2

Interacting parameters



Head-directionality Verb Second (V2)

“I love kitties.”

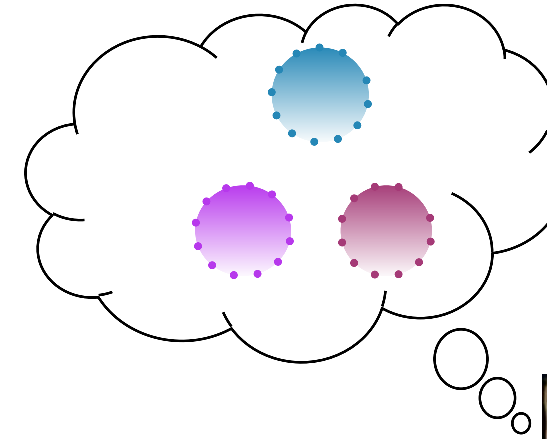
Subject Verb Object

✔ **G1** Head-first +V2	✔ **G2** Head-final +V2	✔ **G3** Head-first -V2

What do the grammars that can analyze this data point have in common?

✗ **G4** Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

“I love kitties.”

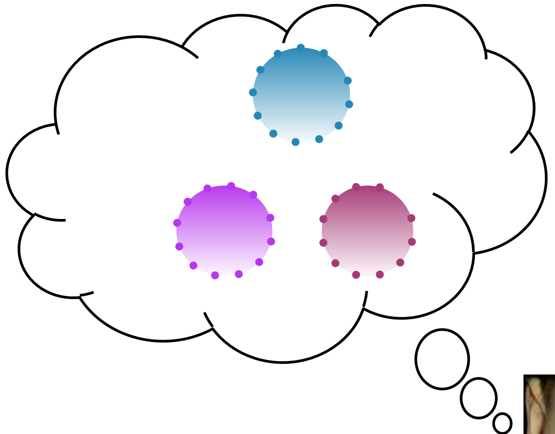
Subject Verb Object

G1 Head-first +V2	**G2** Head-final +V2	**G3** Head-first -V2

We don’t know whether the true grammar is head-first or head-final since there’s a grammar of each kind.

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject **Verb** **Object**

✔ **G1** Head-first +V2

✔ **G2** Head-final +V2

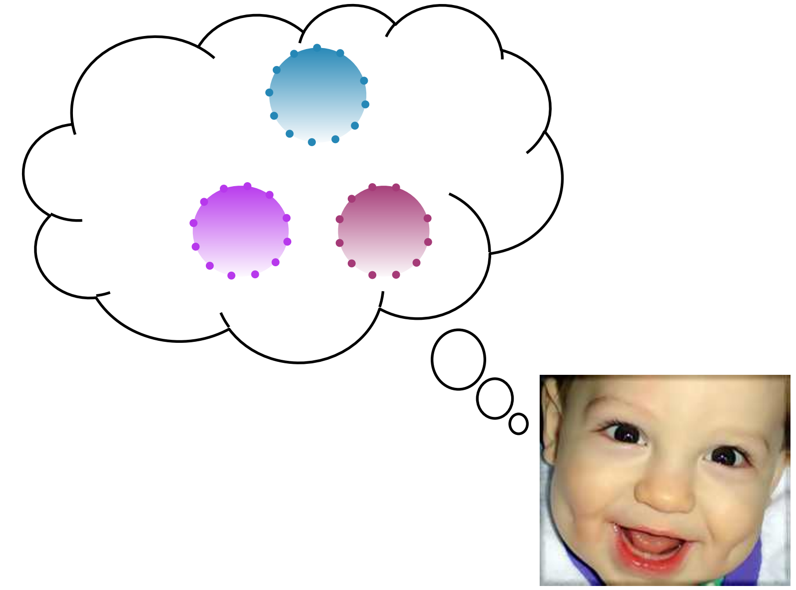
✔ **G3** Head-first -V2

We don't know whether the true grammar is head-first or head-final since there's a grammar of each kind.

(though there are more head-first)

✗ **G4** Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

Subject   Verb   Object

G1 Head-first +V2

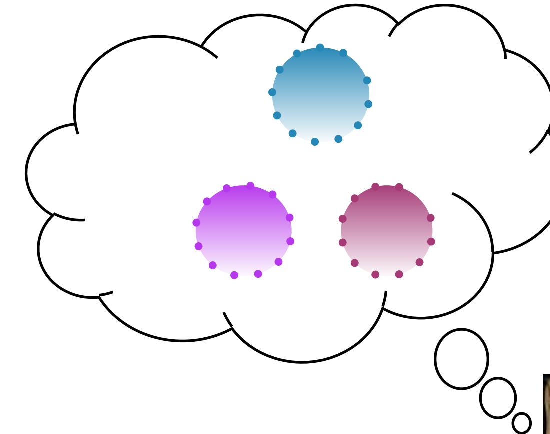
G2 Head-final +V2

G3 Head-first -V2

We don't know whether the true grammar is +V2 or -V2 since there's a grammar of each kind.

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject Verb Object

G1 Head-first +V2

G2 Head-final +V2

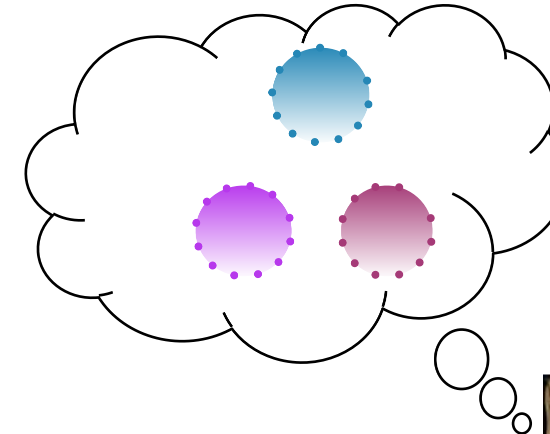
G3 Head-first -V2

We don't know whether the true grammar is +V2 or -V2 since there's a grammar of each kind.

(though there are more +V2)

G4 Head-final -V2

Interacting parameters



Head-directionality   Verb Second (V2)

"I love kitties."

Subject **Verb** **Object**

G1 Head-first +V2

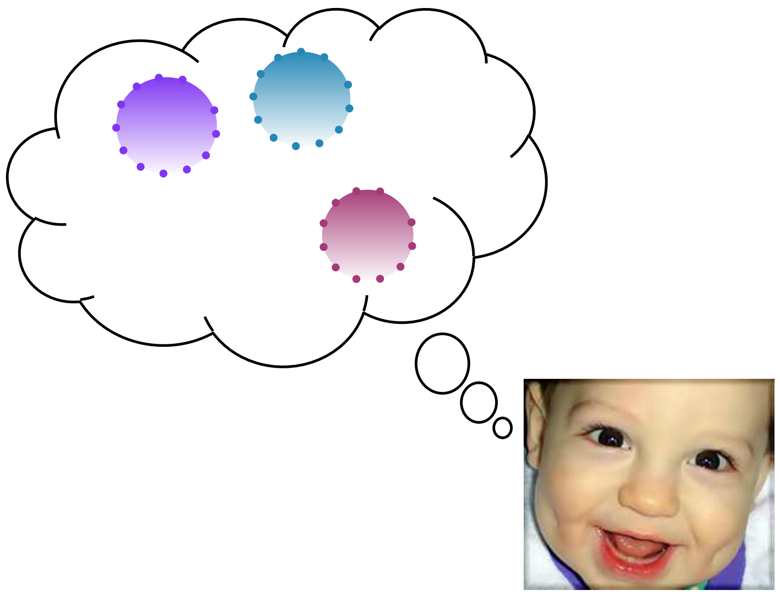
G2 Head-final +V2

G3 Head-first -V2

This data point isn't unambiguous for any of the parameters we're interested in because **the parameters interact**…even though we feel like it might be somewhat informative for head-first and +V2 because these occur in more grammars that are compatible.

G4 Head-final -V2

Interacting parameters



Head-directionality
Edo/English: Head-first
Japanese/Navajo: Head-final

Example Parameter 3: Subject drop

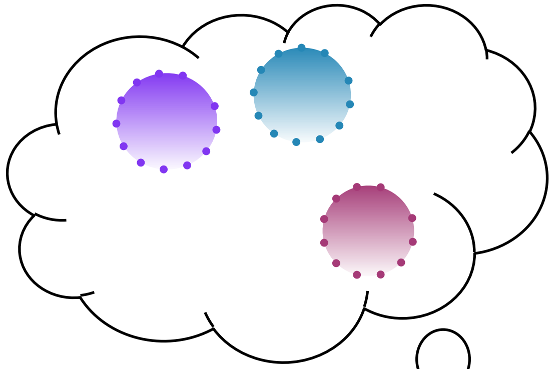
Spanish: +subj-drop
Allows Subject to be overt or dropped

✔ Ellos beben
they drink-3rd-pl

"They drink"

✔ Beben
drink-3rd-pl

Interacting parameters



Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 3: Subject drop

Spanish: +subj-drop

English: -subj-drop

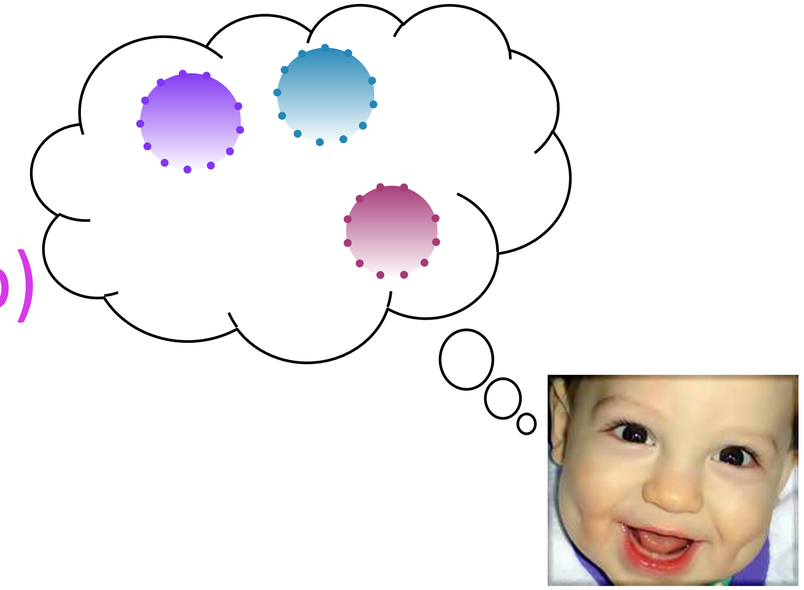
Subject must be overt

✔ They drink

✘ Drink

"They drink"

Interacting parameters



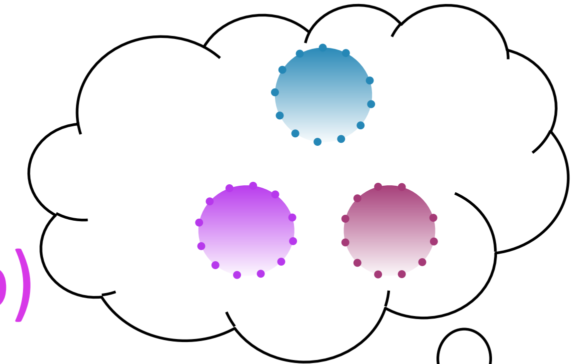
Head-directionality

Subject drop (subj-drop)

Grammars available

G1	Head-first	+subj-drop
G2	Head-final	+subj-drop
G3	Head-first	-subj-drop
G4	Head-final	-subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

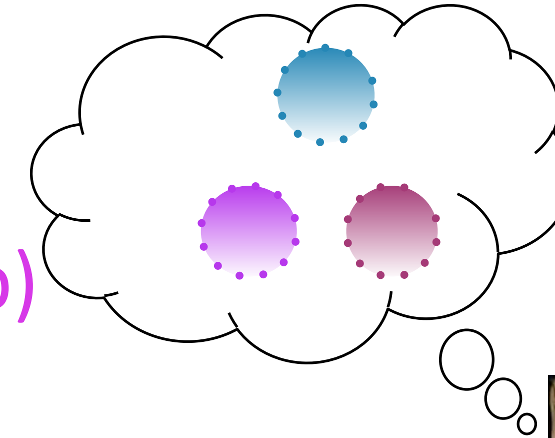
...that I Kitties love

Subject Object Verb

Which grammars can analyze this data point?

G1	Head-first	+subj-drop
G2	Head-final	+subj-drop
G3	Head-first	-subj-drop
G4	Head-final	-subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject Object Verb

✗ head-first predicts SVO

✔ +subj-drop allows subject to be overt

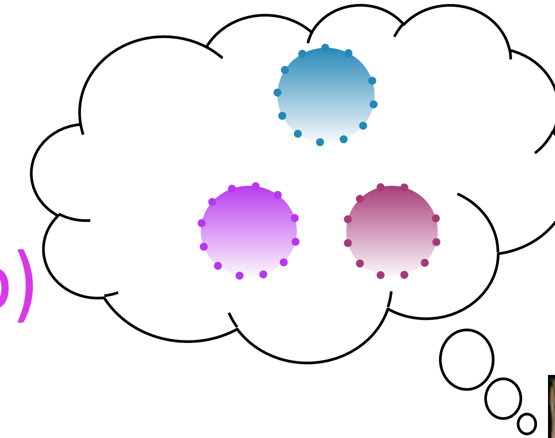
G1 Head-first
+subj-drop

G2 Head-final
+subj-drop

G3 Head-first
-subj-drop

G4 Head-final
-subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass **ich** **Kätzchen** **liebe**."
...that I Kitties love

Subject **Object** **Verb**

✔ head-final predicts SOV
✔+subj-drop allows subject to be overt

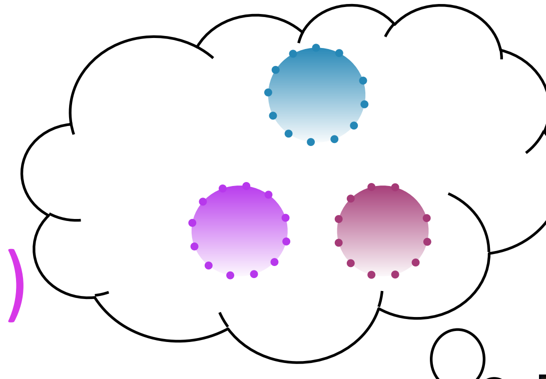
G2 Head-final
+subj-drop

G3 Head-first
-subj-drop

G1 Head-first
+subj-drop

G4 Head-final
-subj-drop

Interacting parameters



Head-directionality   Subject drop (subj-drop)

"…dass ich Kätzchen liebe."

…that I Kitties love

Subject   Object   Verb

✗ head-first predicts SVO

✔ -subj-drop requires subject to be overt

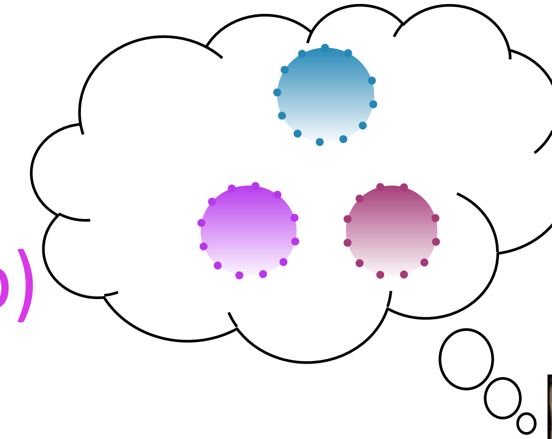
G3 Head-first -subj-drop

✔ **G2** Head-final +subj-drop

✗ **G1** Head-first +subj-drop

G4 Head-final -subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass ich
Kätzchen liebe."
...that I Kitties love

Subject **Object** **Verb**

✔ head-final predicts SOV
✔ -subj-drop requires subject to be overt

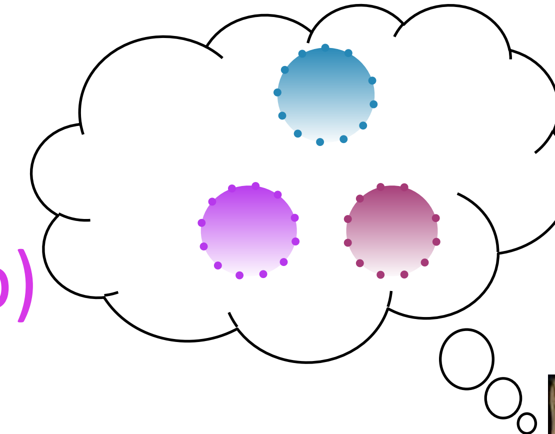
G4 Head-final
-subj-drop

✔ **G2** Head-final
+subj-drop

✗ **G1** Head-first
+subj-drop

✗ **G3** Head-first
-subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

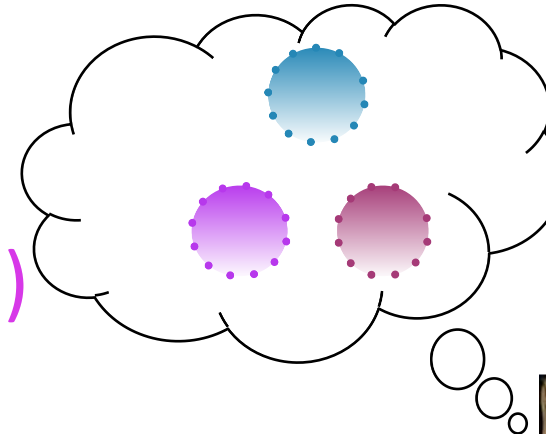
Subject Object Verb

G2	Head-final	+subj-drop
G4	Head-final	-subj-drop

There's more than one grammar compatible with this data point...even though we feel like it ***should definitely*** be informative for head-final (since that's the only value in the compatible grammars).

G1	Head-first	+subj-drop
G3	Head-first	-subj-drop

Interacting parameters



Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject Object Verb

✔ **G2** Head-final +subj-drop

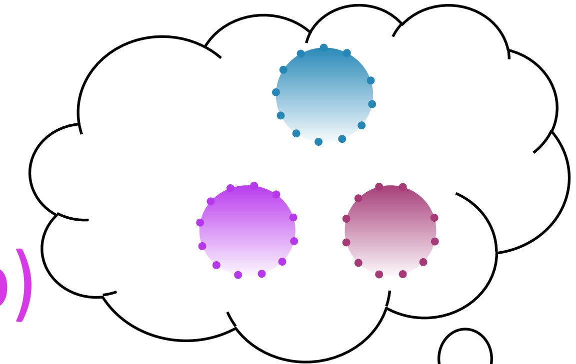
✔ **G4** Head-final -subj-drop

But technically, this is still an ambiguous data point because more than one grammar will work....

✗ **G1** Head-first +subj-drop

✗ **G3** Head-first -subj-drop

Interacting parameters



Head-directionality   Subject drop (subj-drop)

"...dass ich
Kätzchen liebe."
...that I Kitties love

Subject   Object   Verb

G2 Head-final
+subj-drop

G4 Head-final
-subj-drop

So what can we do?

G1 Head-first
+subj-drop

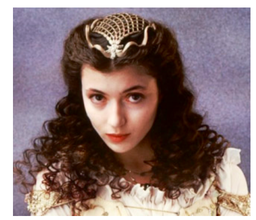
G3 Head-first
-subj-drop

Today's Plan:
Computational models of syntactic acquisition

I. Some non-parametric examples


Who does … is pretty?

syntax

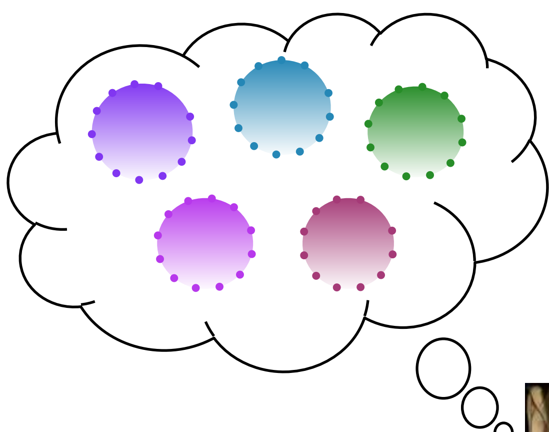


another one



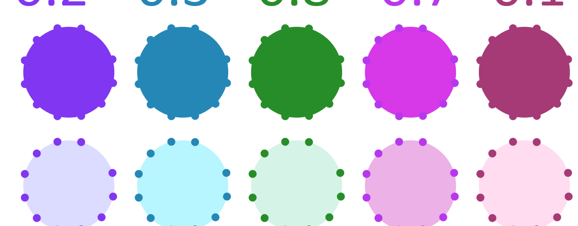
syntax, semantics

II. About linguistic parameters




III. Learning with parameters

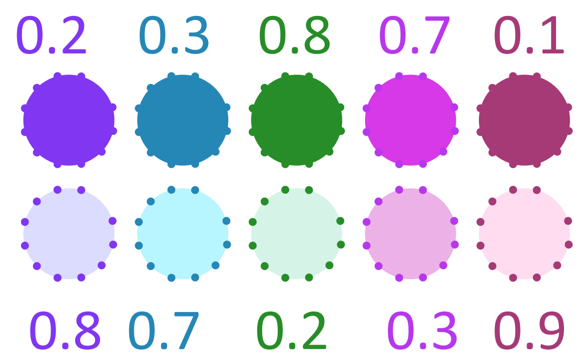
0.2 0.3 0.8 0.7 0.1



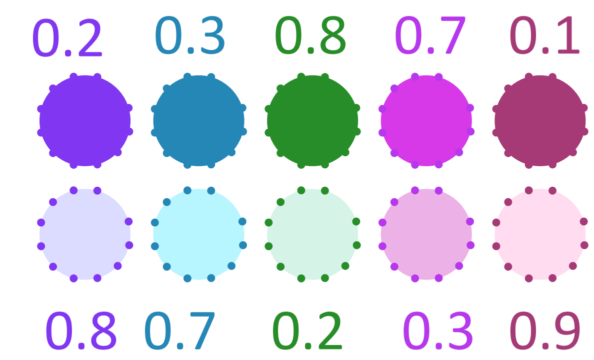
0.8 0.7 0.2 0.3 0.9

Today's Plan:
Computational models of syntactic acquisition

III. Learning with parameters



Learning with parameters

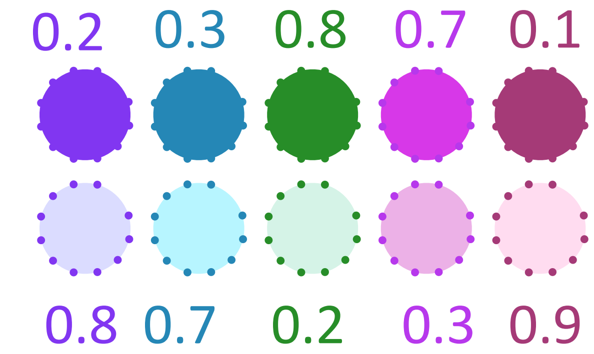


A language's grammar = combination of parameter values

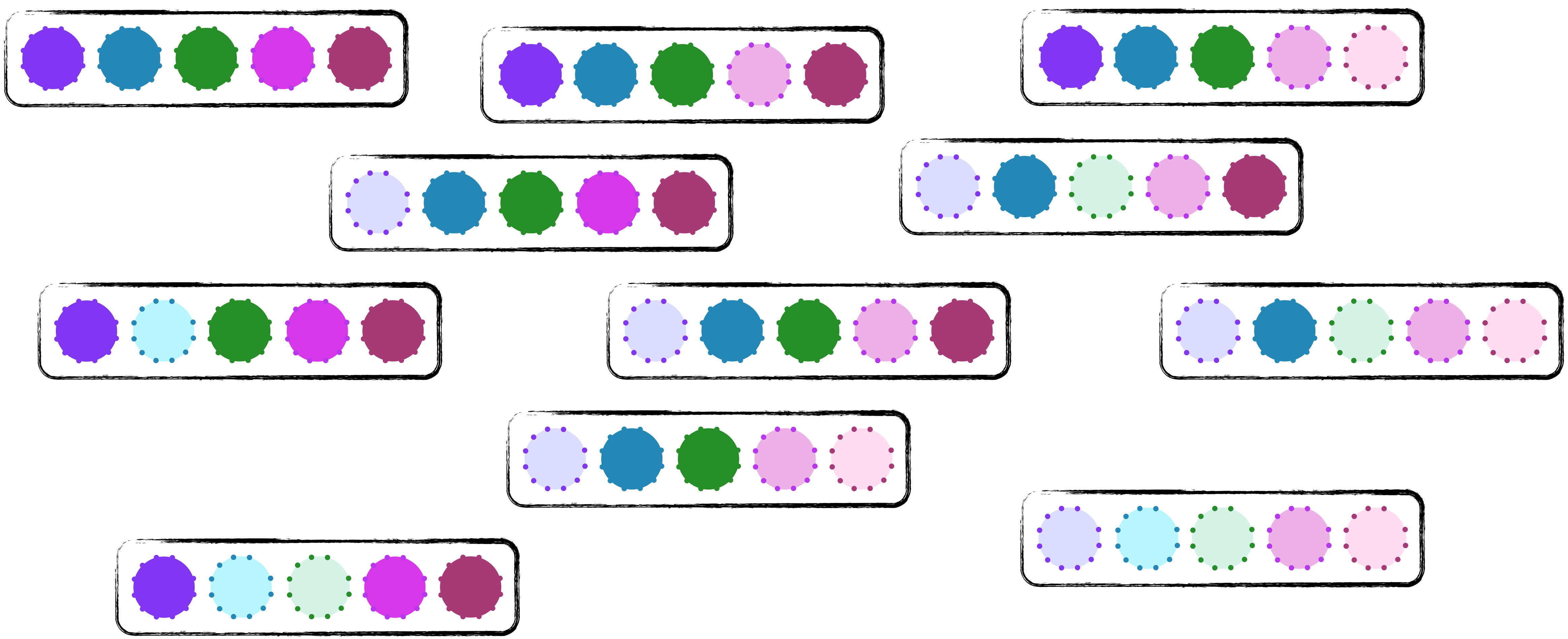
G2 Head-final
+subj-drop

G4 Head-final
-subj-drop

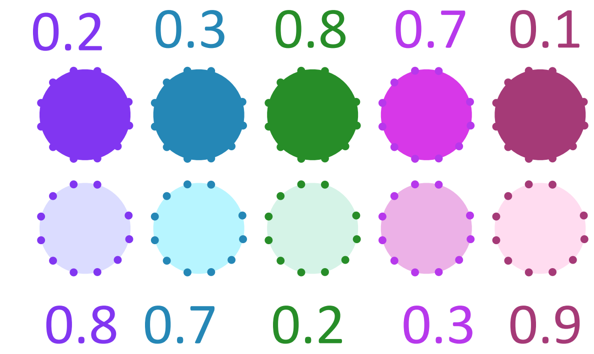
Learning with parameters

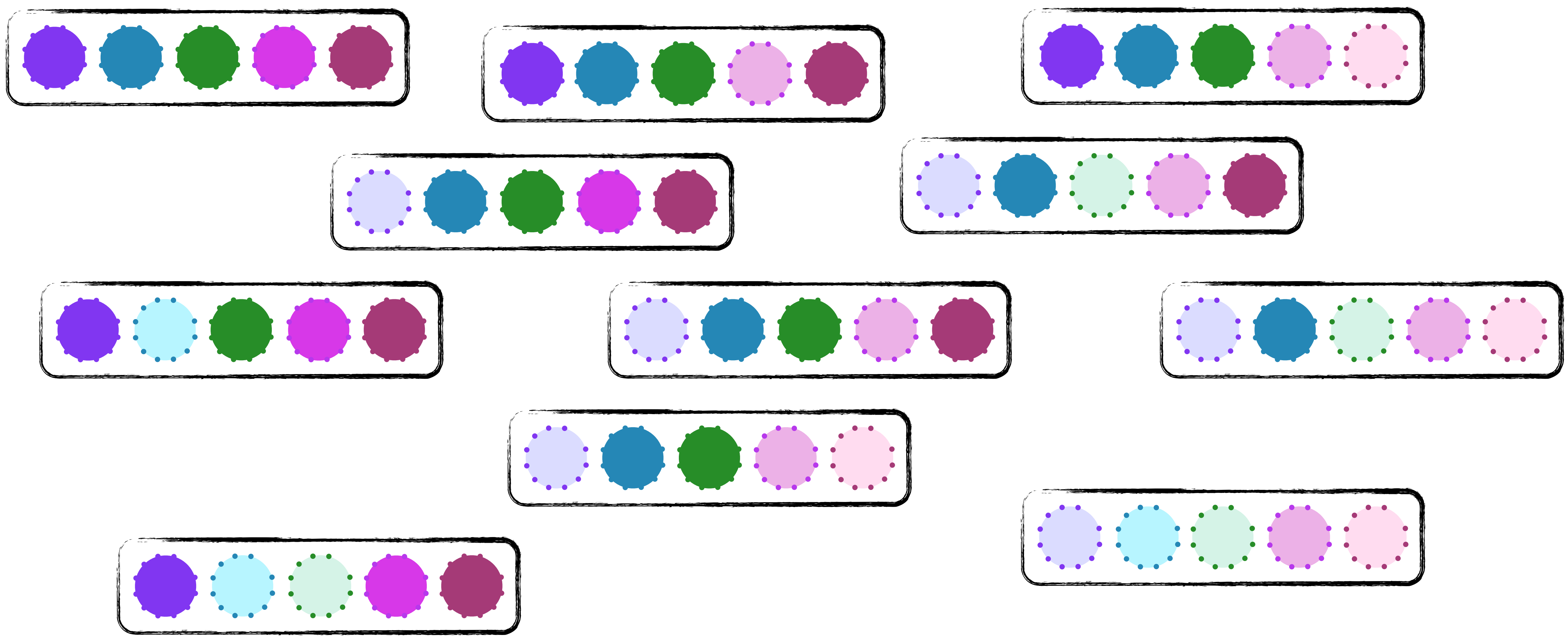


A language's grammar = combination of parameter values



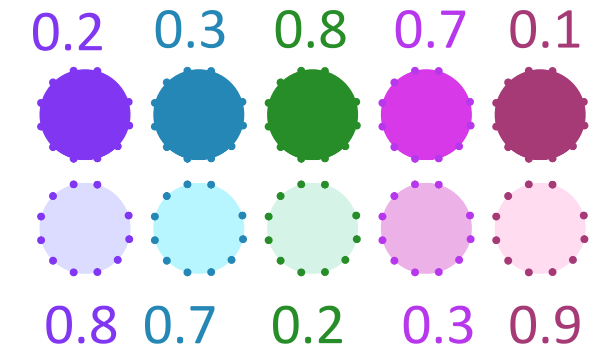
Learning with parameters





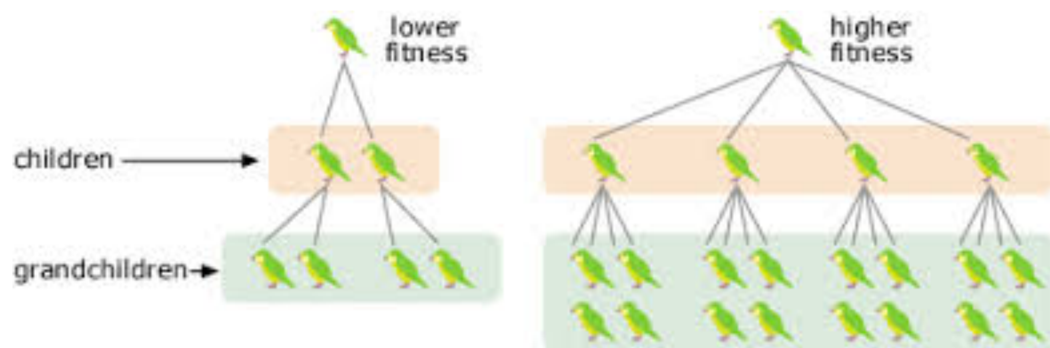
Variational learning (Yang 2002, 2004, 2012): use reinforcement learning to learn which value (for each parameter) that the native language uses for its grammar. This is a combination of using linguistic knowledge & statistical learning.

Learning with parameters



Variational learning

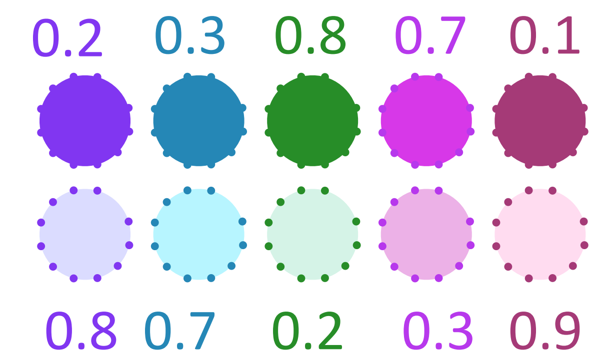
Idea taken from evolutionary biology:
In a population, individuals compete against each other. The fittest individuals survive while the others die out.



How do we translate this to learning with parameters?

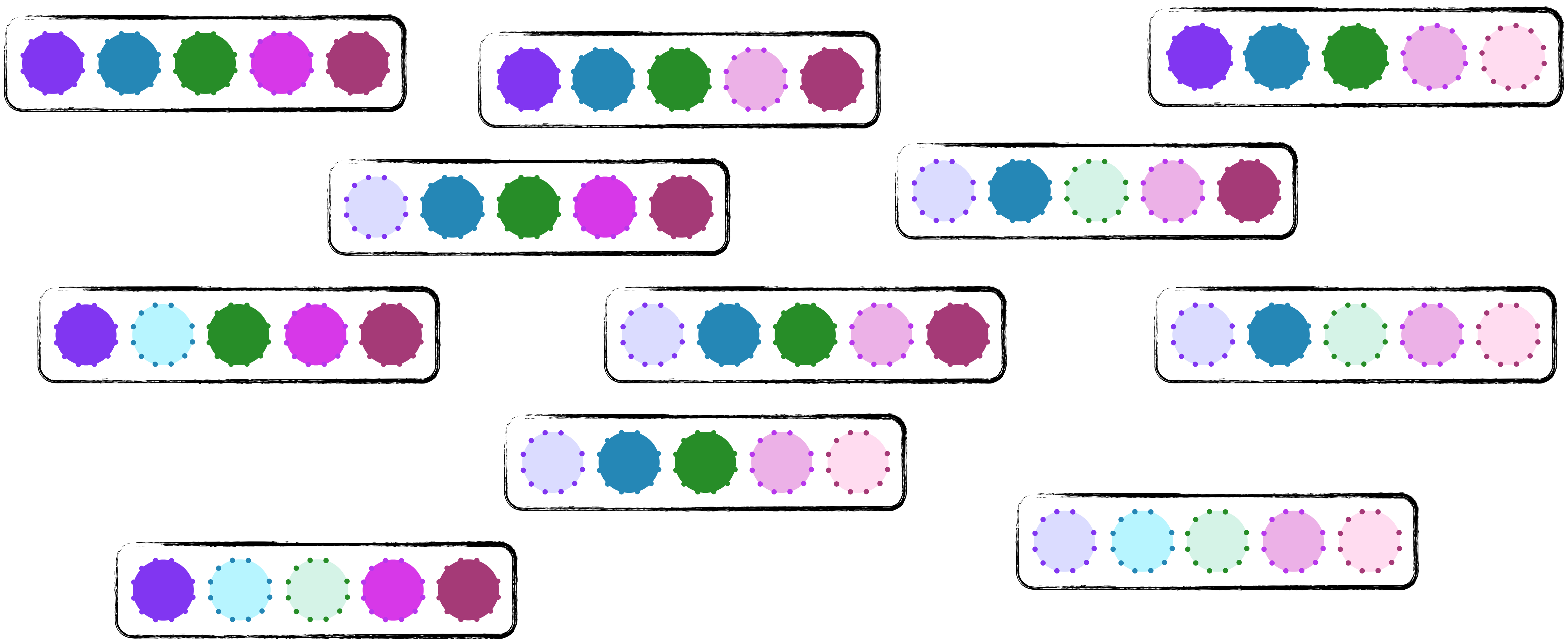
Learning with parameters

Variational learning

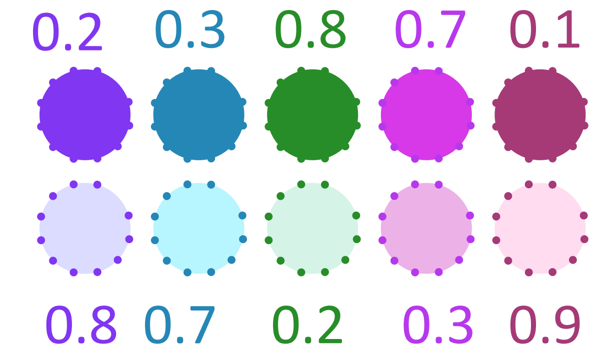


The fittest **individuals** survive while the others die out.

Individual = grammar (combination of parameter values that represents the structural properties of a language)



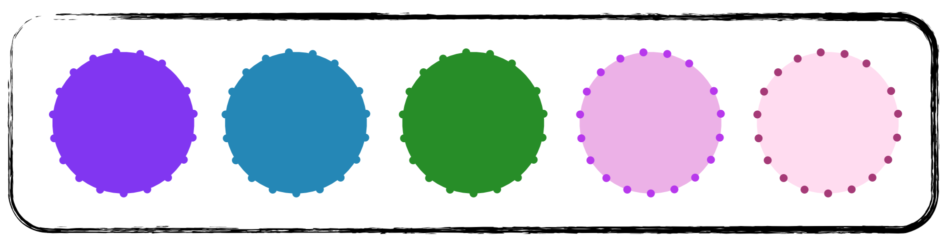
Learning with parameters



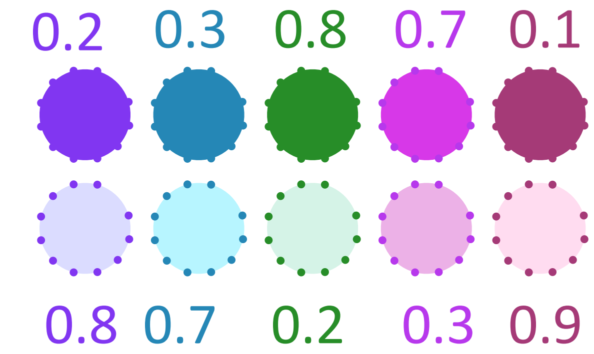
Variational learning

The **fittest** individuals survive while the others die out.

Fitness = how well a grammar can analyze the data the child encounters

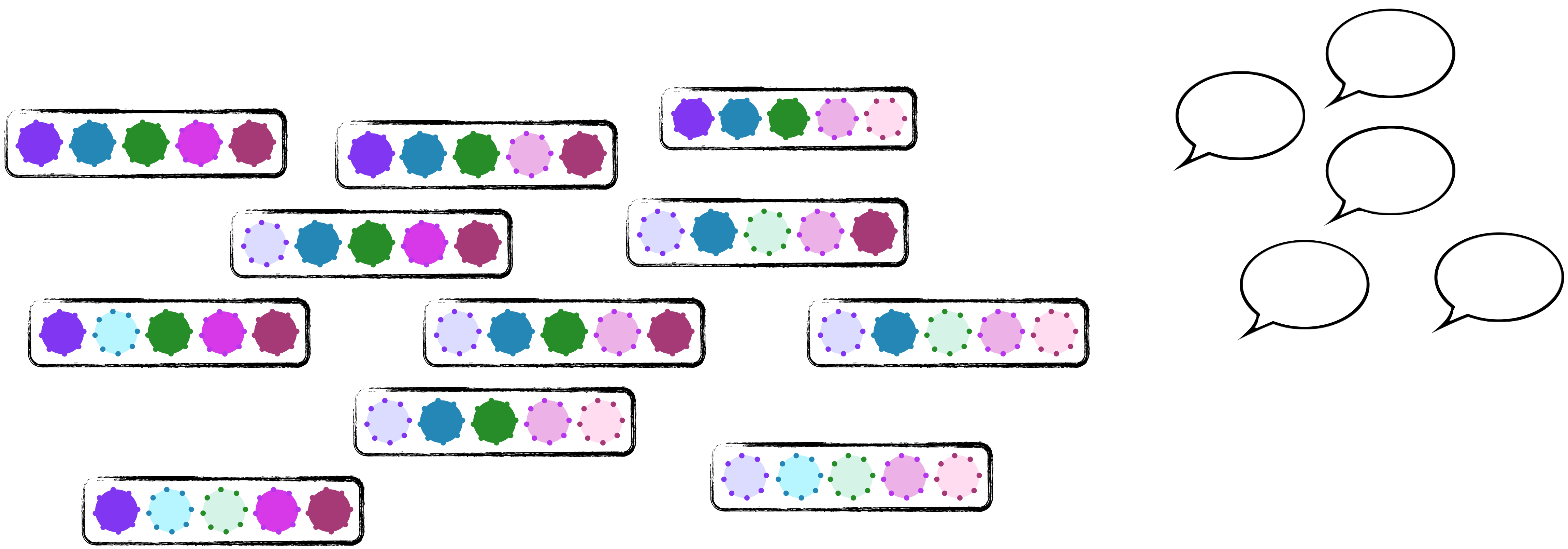


Learning with parameters

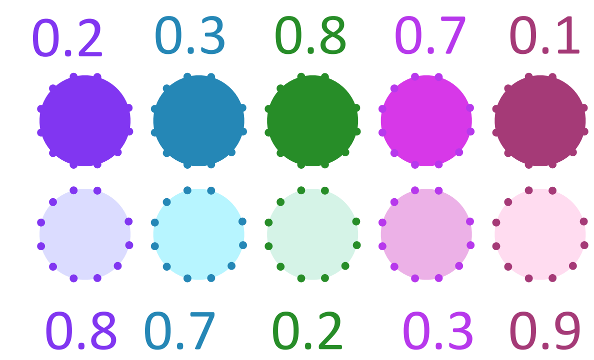


Variational learning

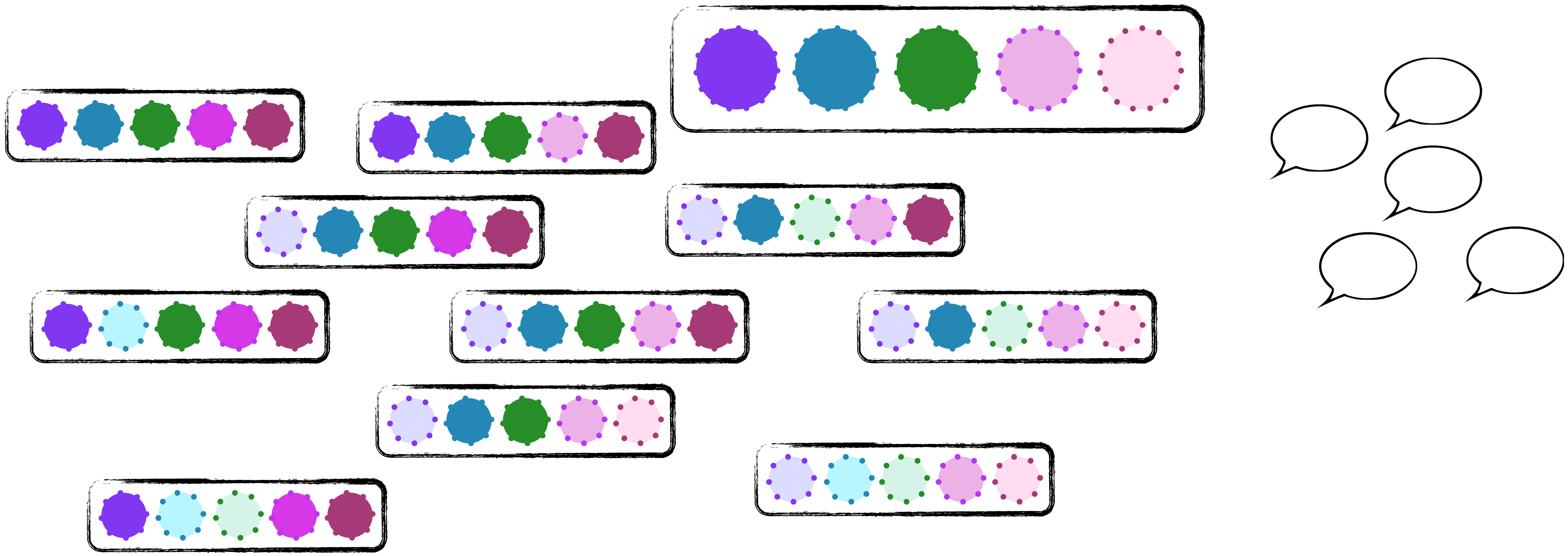
A child's mind consists of a population of grammars that are competing to analyze the data in the child's native language.



Learning with parameters



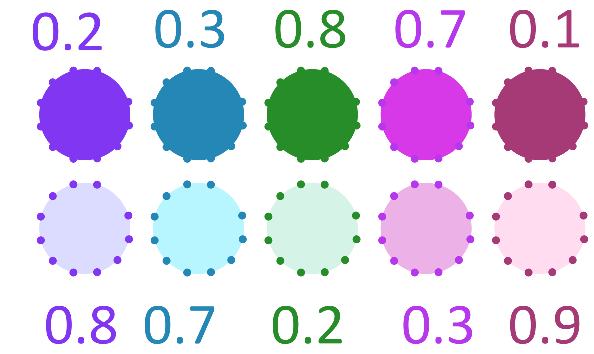
Variational learning

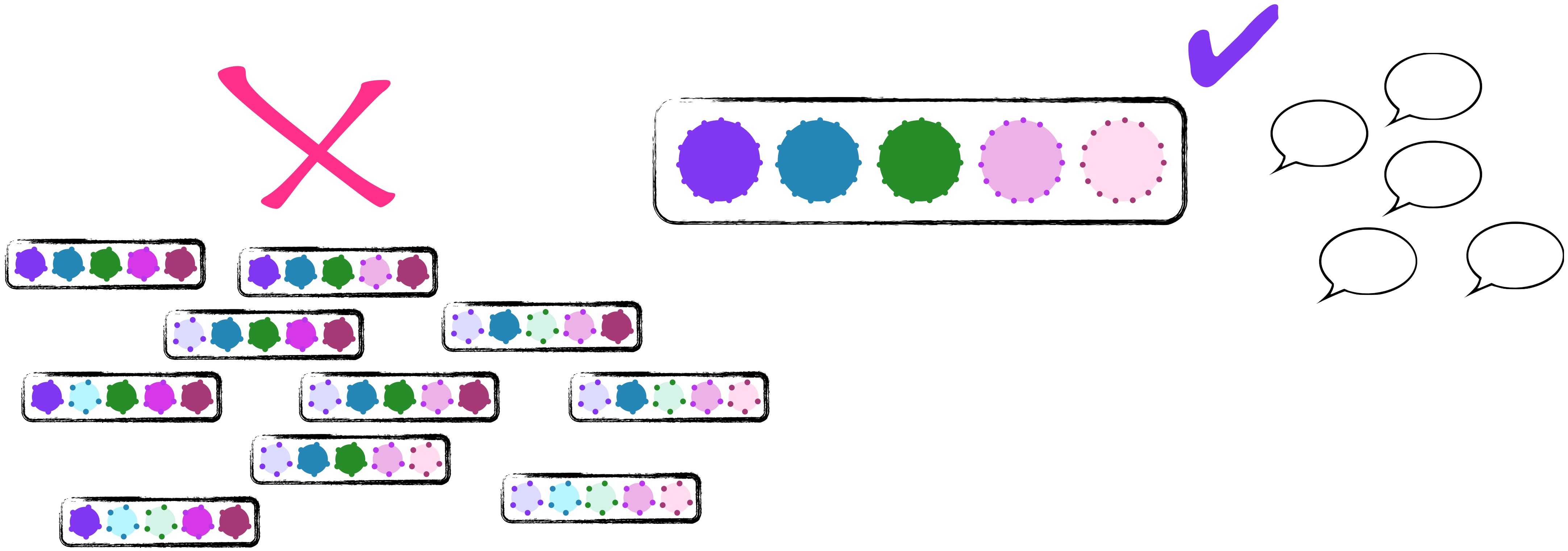


Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will “win”, once the child encounters enough native language data. This is because none of the other competing grammars can analyze all the data.

Learning with parameters

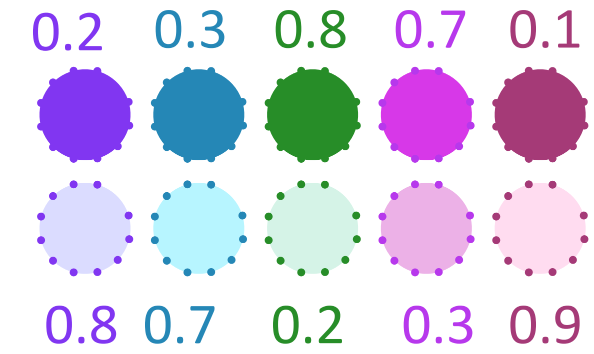
Variational learning



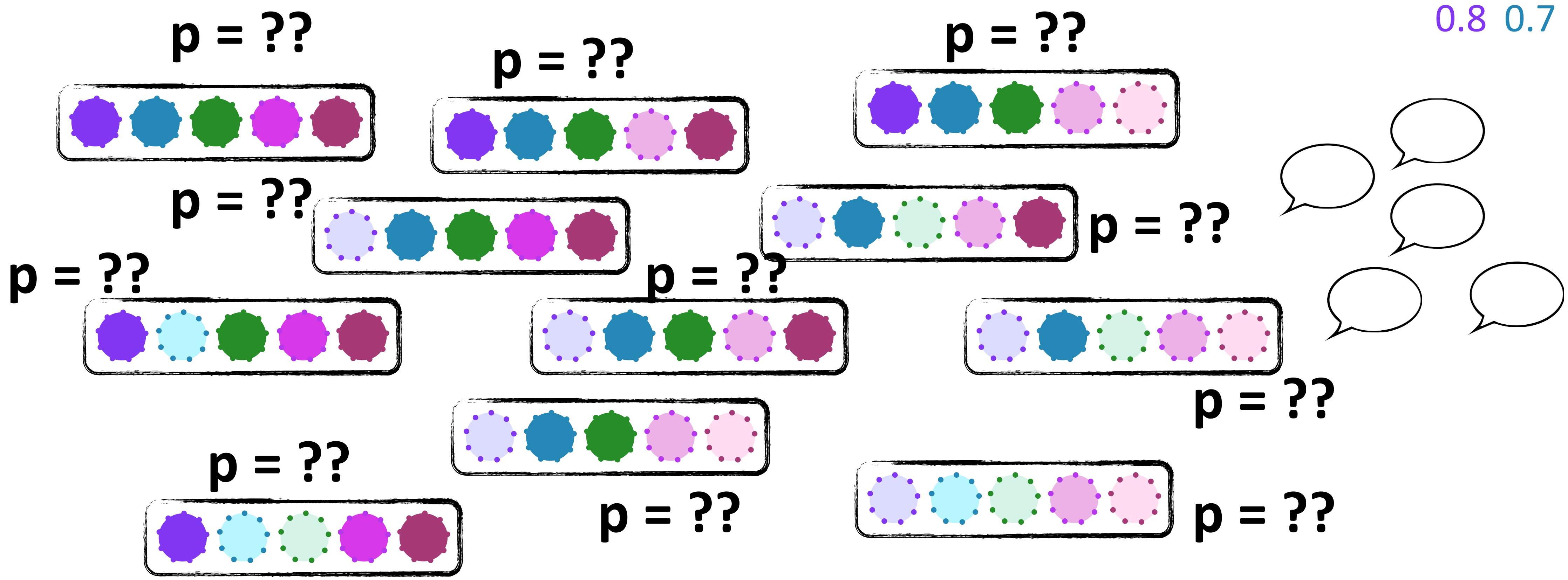


If this is the native language grammar, this grammar can analyze all the intake while the others can’t.

Learning with parameters

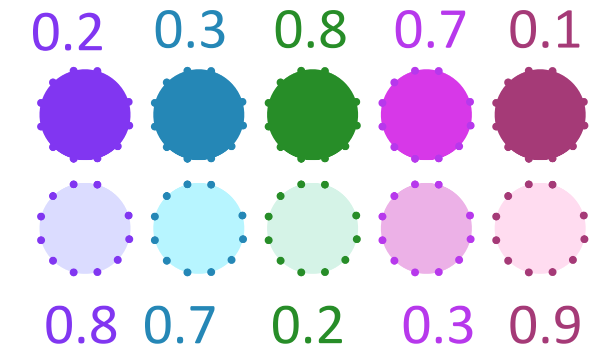


Variational learning

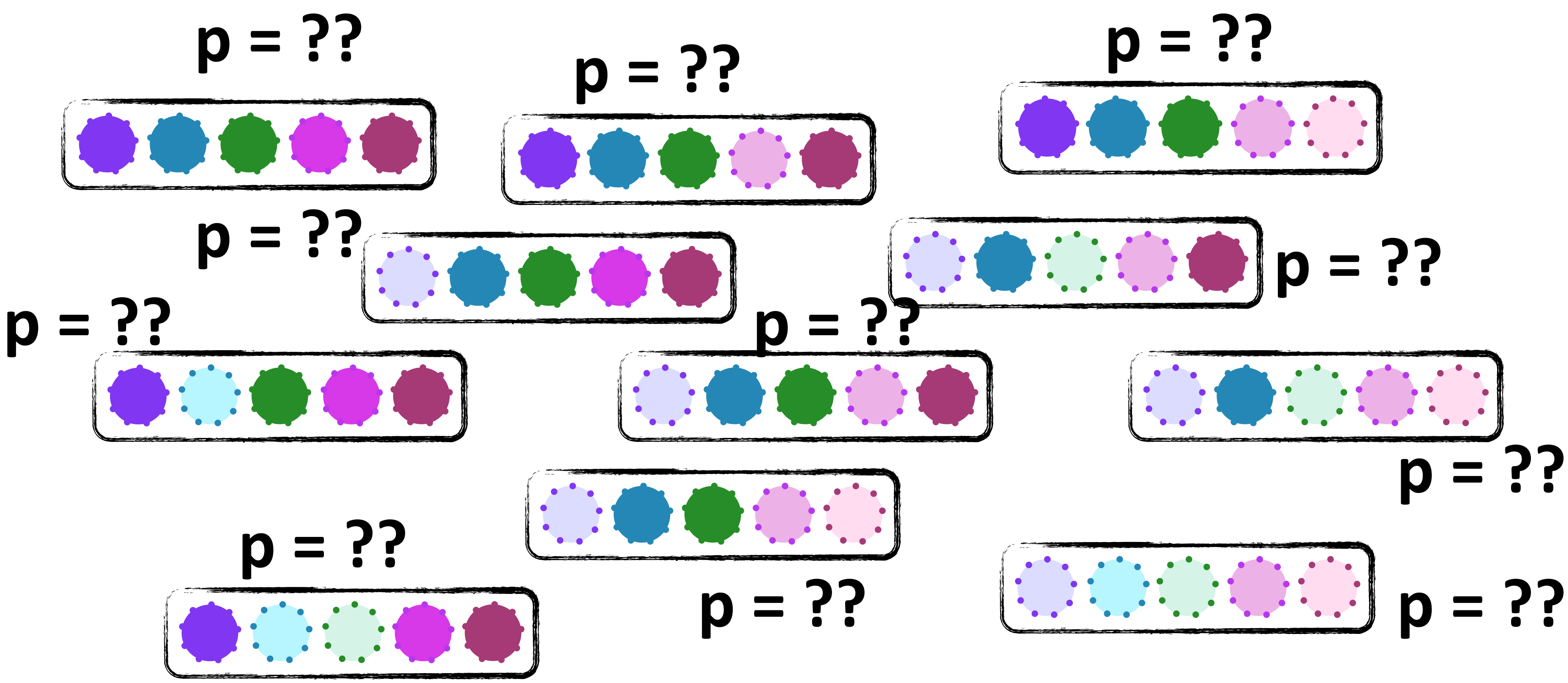


At any point in time, a grammar in the population will have a **probability** associated with it. This represents the child's belief that this grammar is the correct grammar for the native language.

Learning with parameters

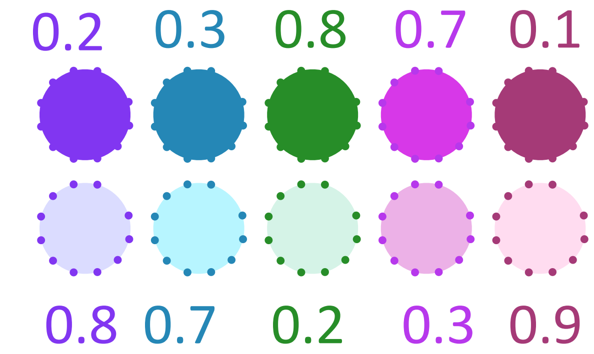


Variational learning

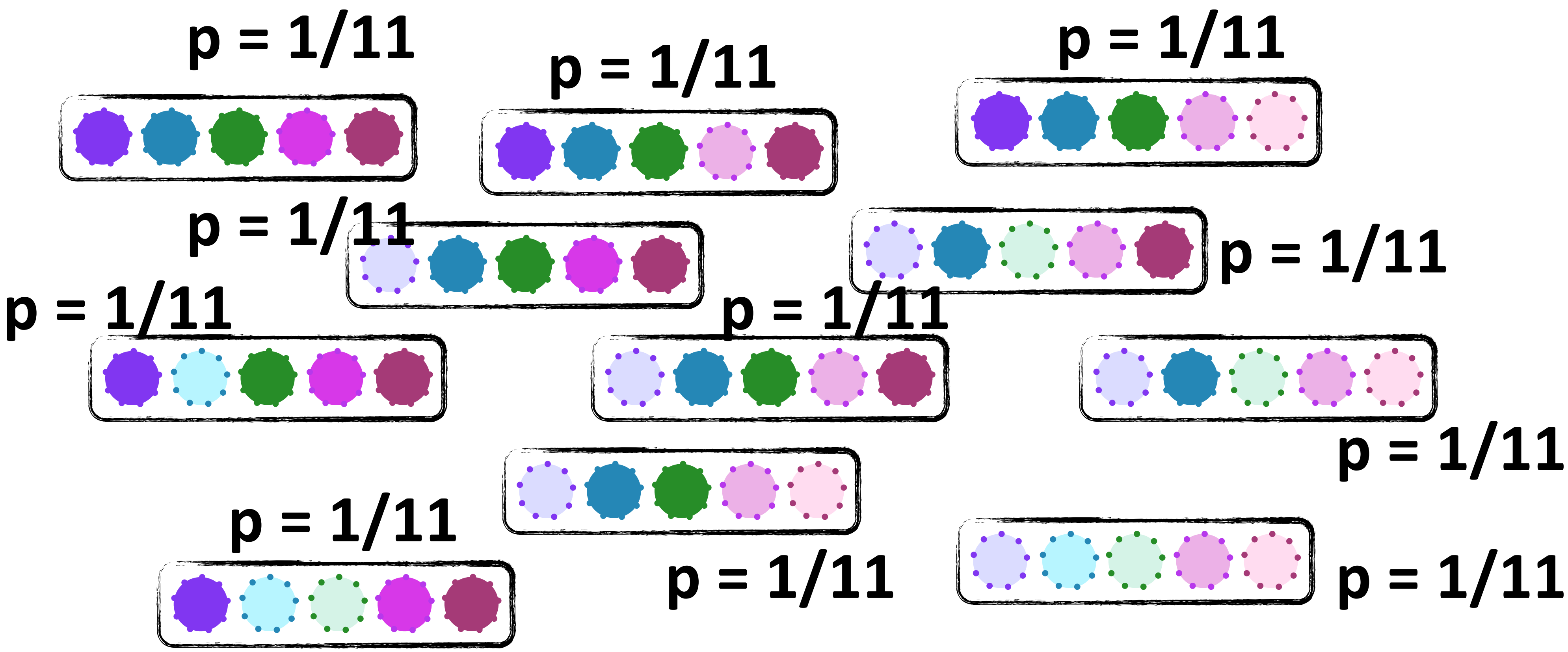


Before the child has encountered any native language data, all grammars are **equally likely**. So, initially all grammars have the same probability, which is 1 divided the number of grammars available.

Learning with parameters

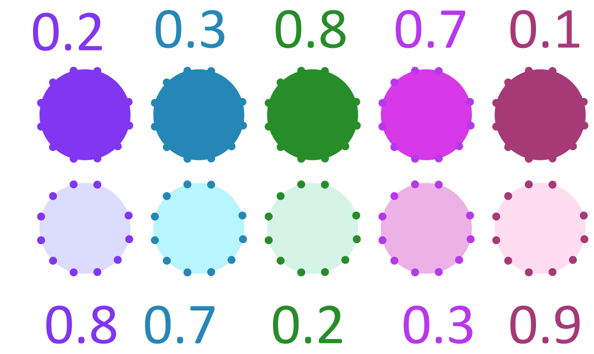


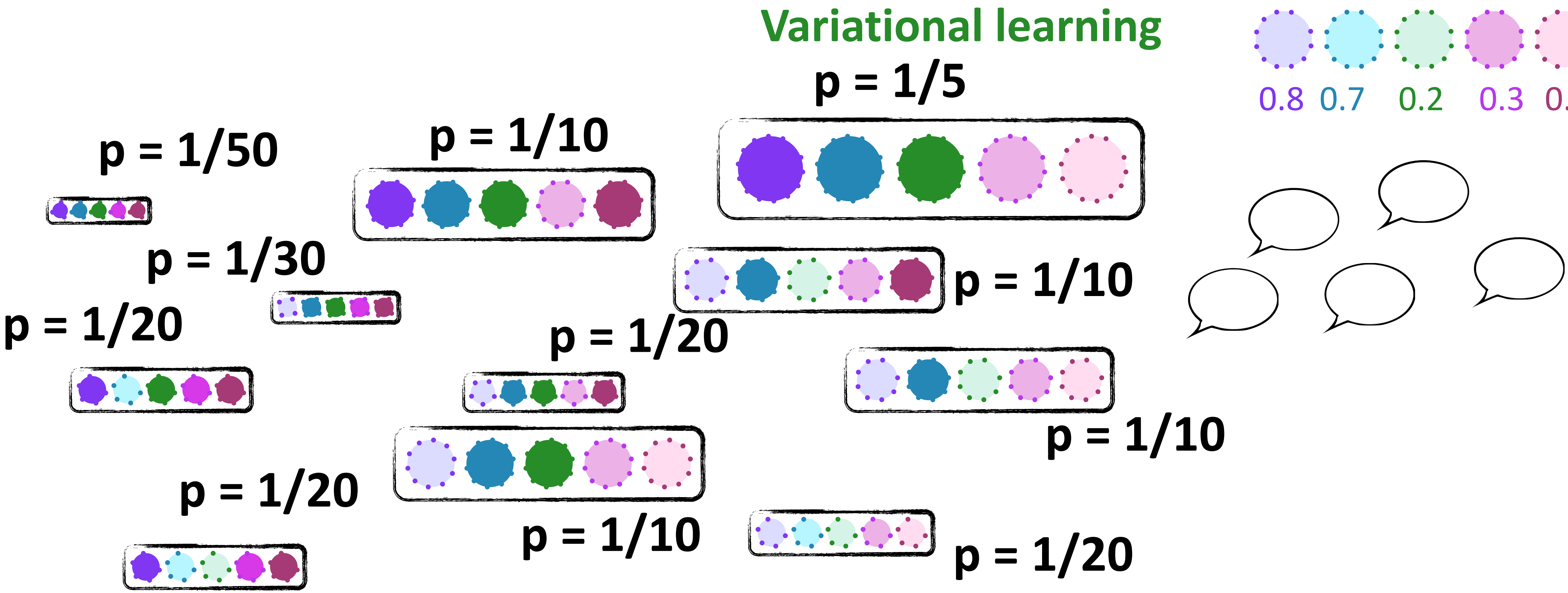
Variational learning



Since there are 11 grammars here, each begins with probability 1/11.

Learning with parameters

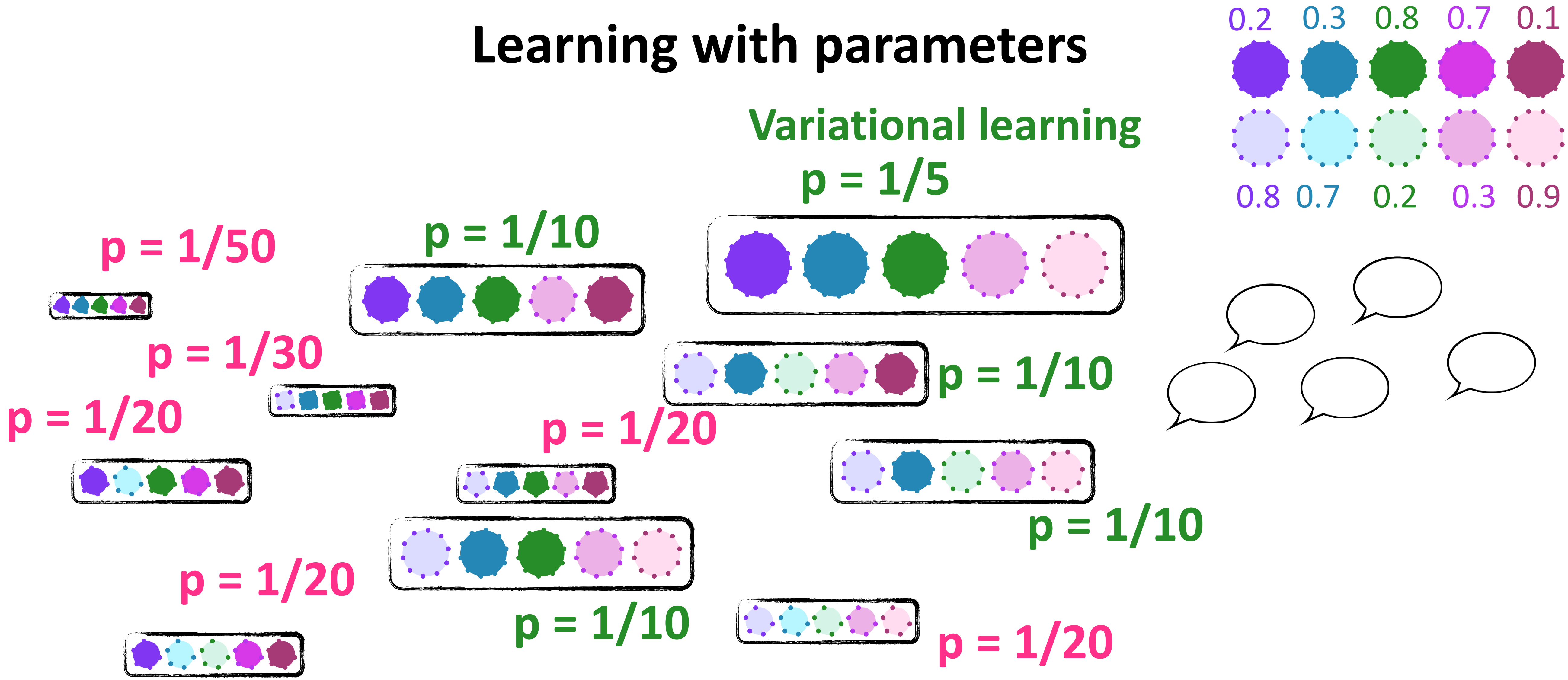




As the child encounters data from the native language, some of the grammars will be more fit because they are better able to account for the syntactic properties of the intake.

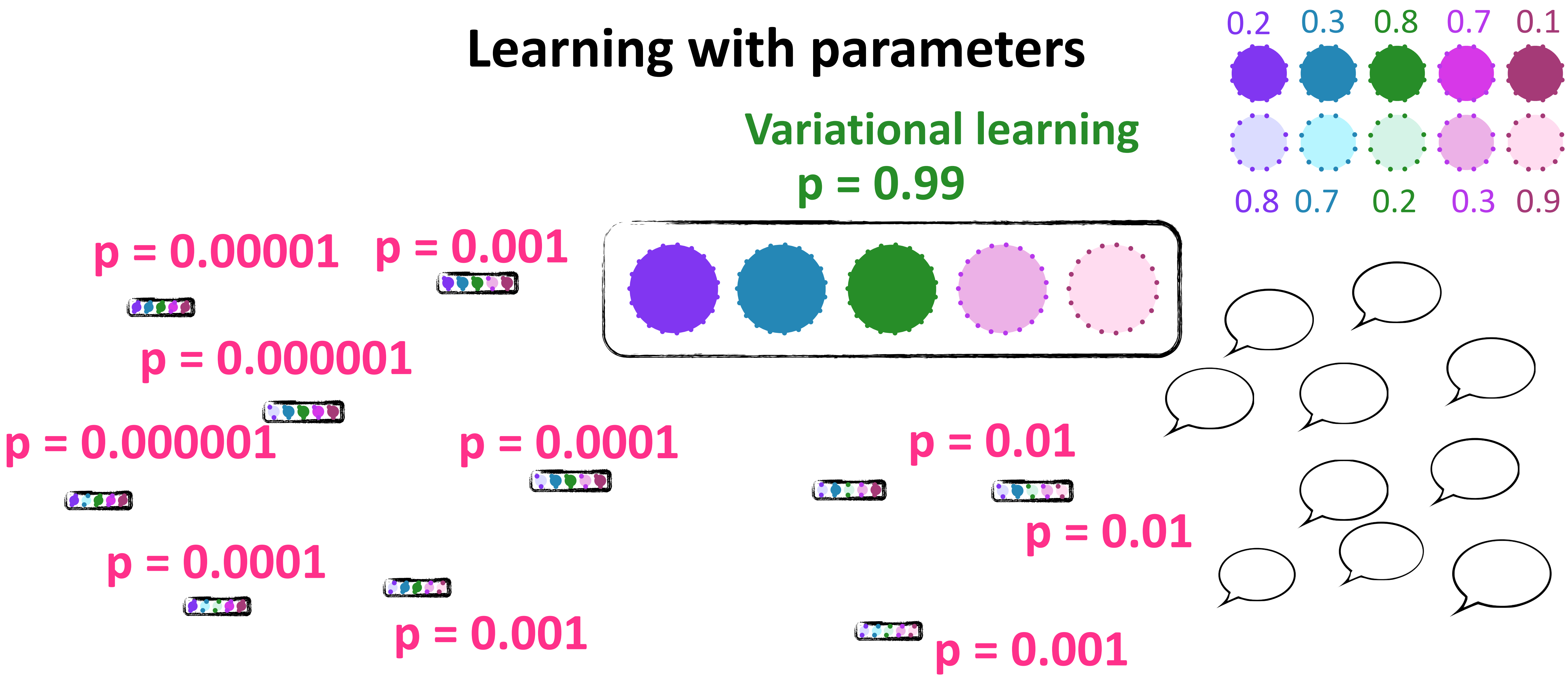
Other grammars will be less fit because they cannot account for some of the data encountered.

Learning with parameters



Grammars that are more compatible with the native language data intake will have their **probabilities increased** while grammars that are less compatible will have their **probabilities decreased** over time.

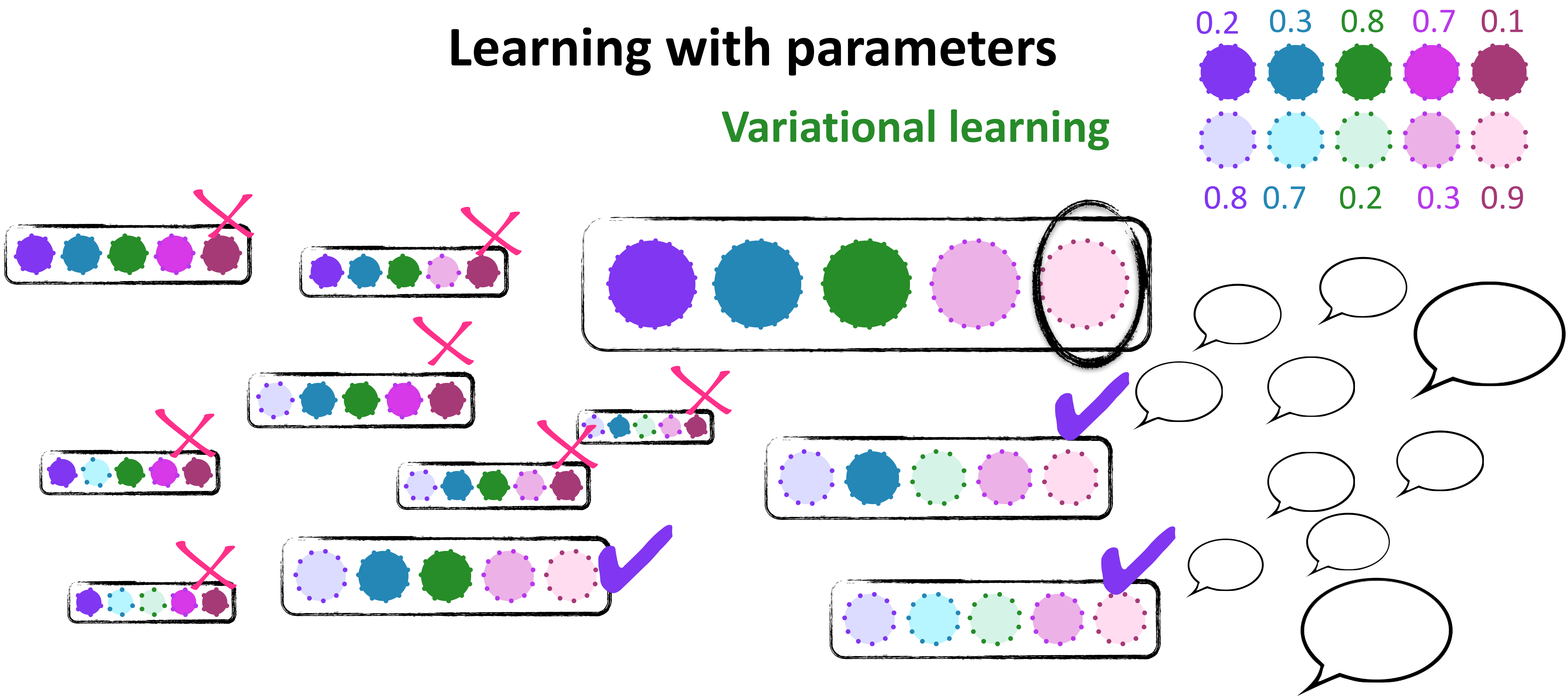
Learning with parameters



After the child has encountered enough data from the native language, the native language grammar should have a probability near 1.0 while the other grammars have a probability near 0.0.

Learning with parameters

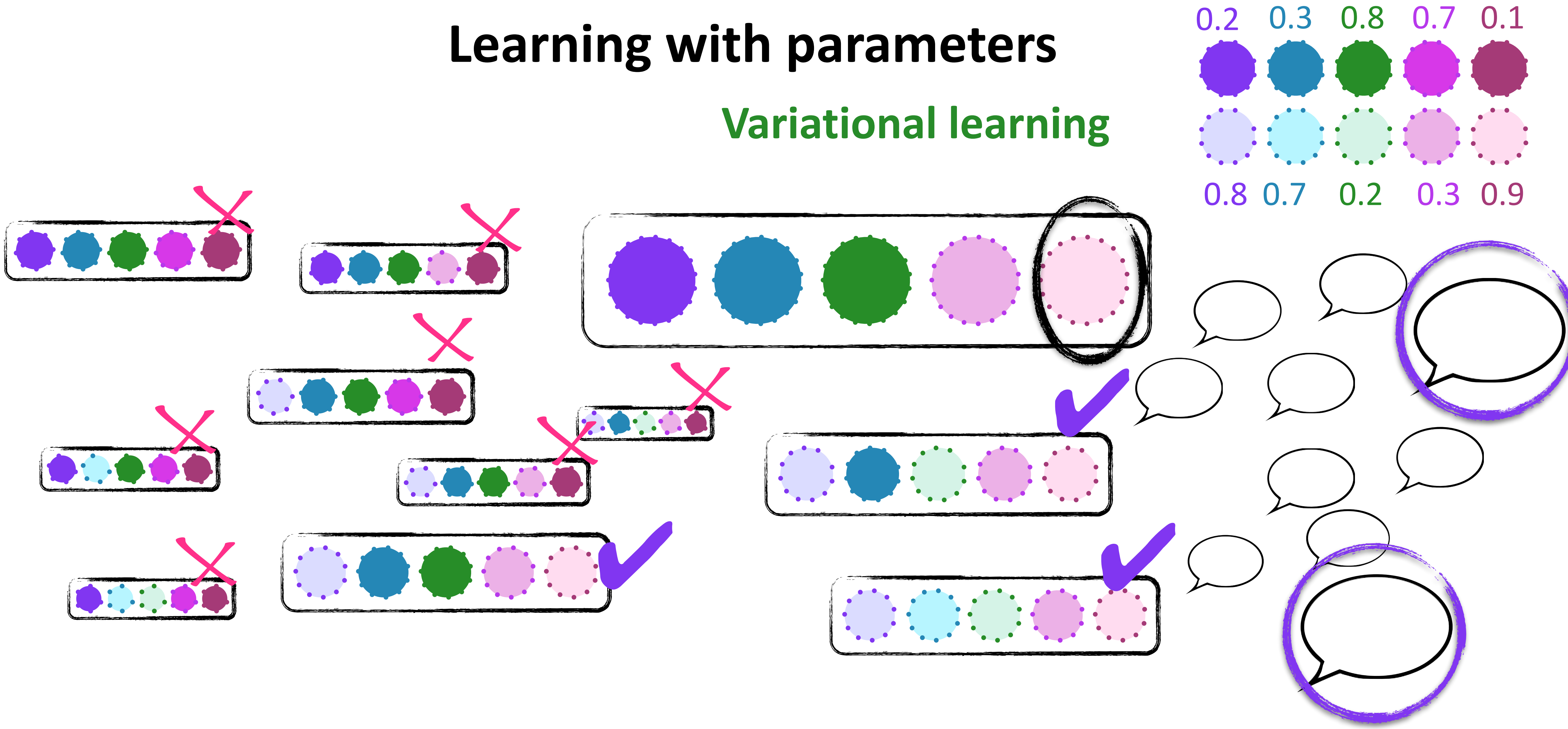
Variational learning



The power of **unambiguous data**:
Unambiguous data from the native language can only be analyzed by grammars that use the **native language's parameter value**.

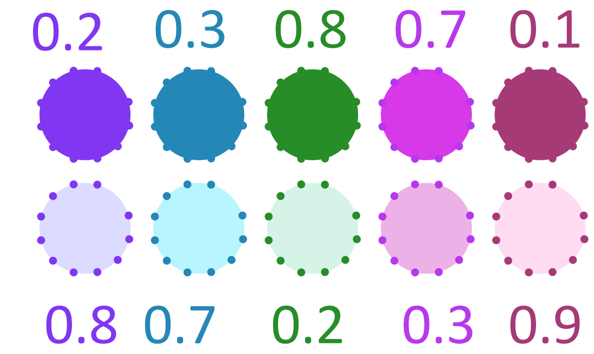
Learning with parameters

Variational learning

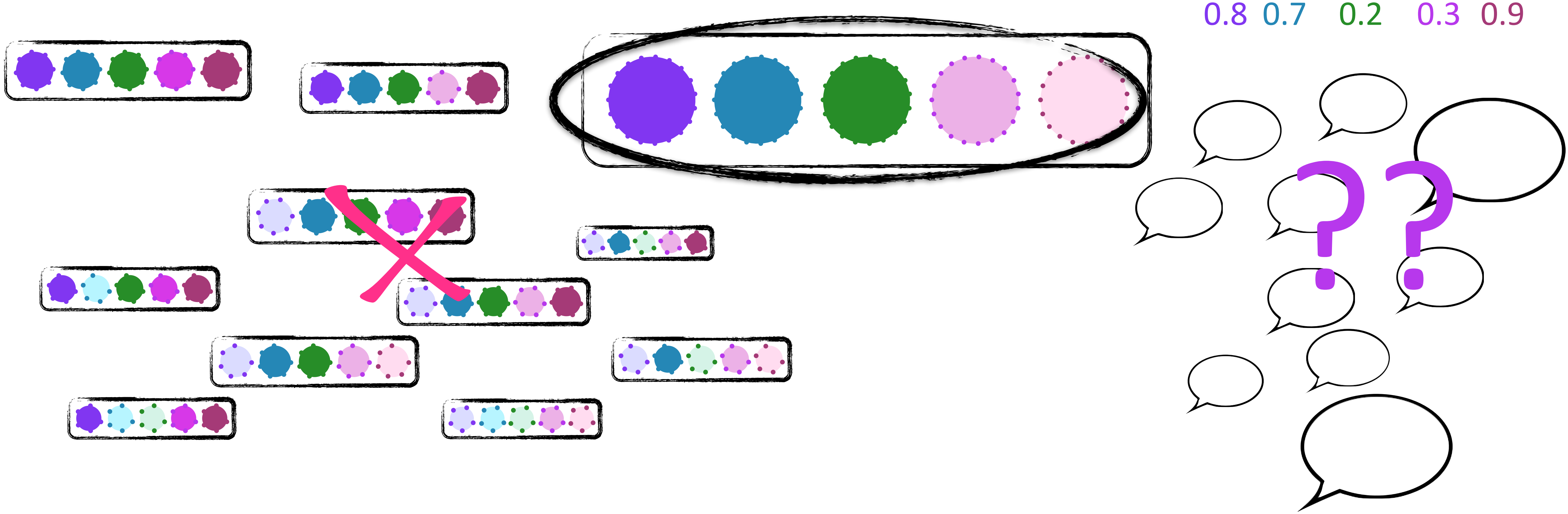


This makes unambiguous data very influential data for the child to encounter, since these data are only compatible with the parameter value that is correct for the native language.

Learning with parameters



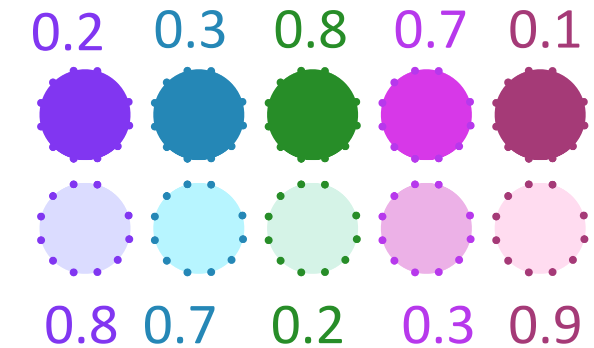
Variational learning



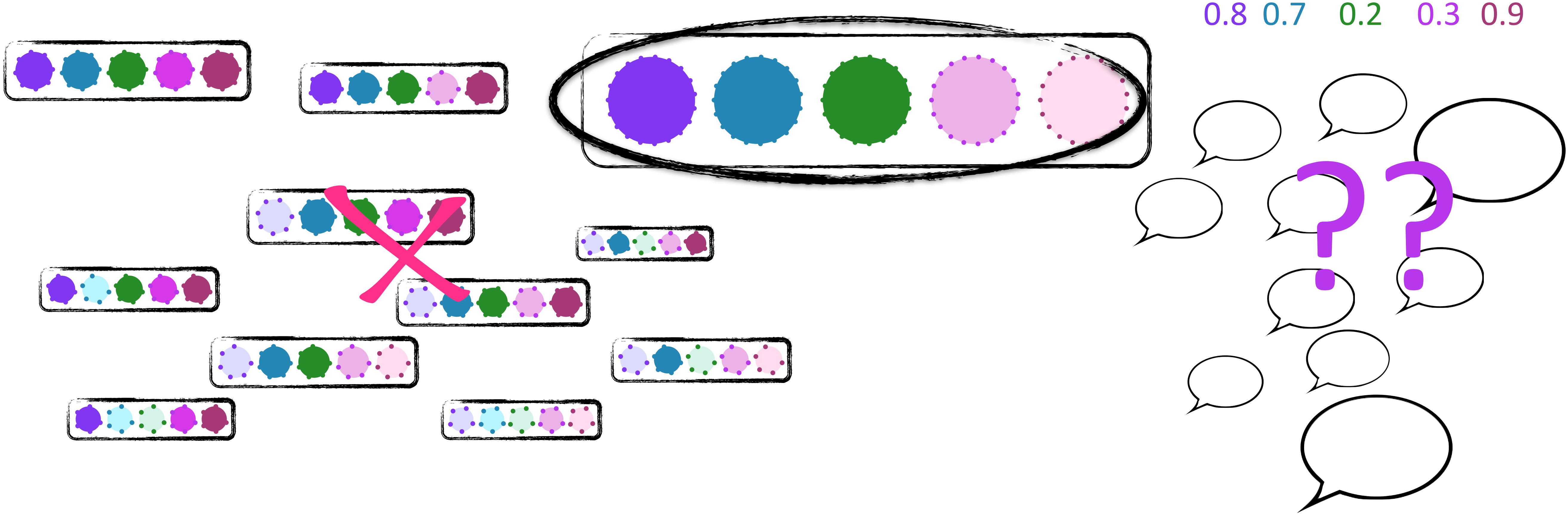
Problem: Do unambiguous data exist for entire grammars?

This requires data that are incompatible with every other possible parameter value of every other possible grammar….

Learning with parameters

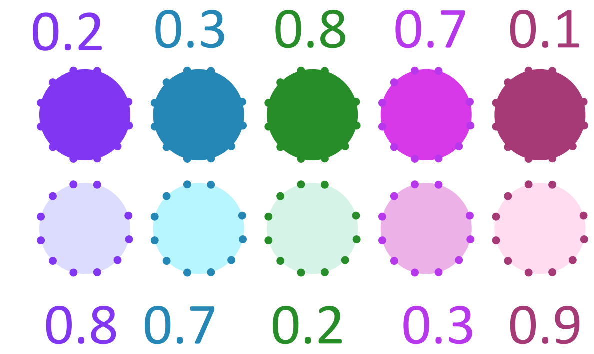


Variational learning

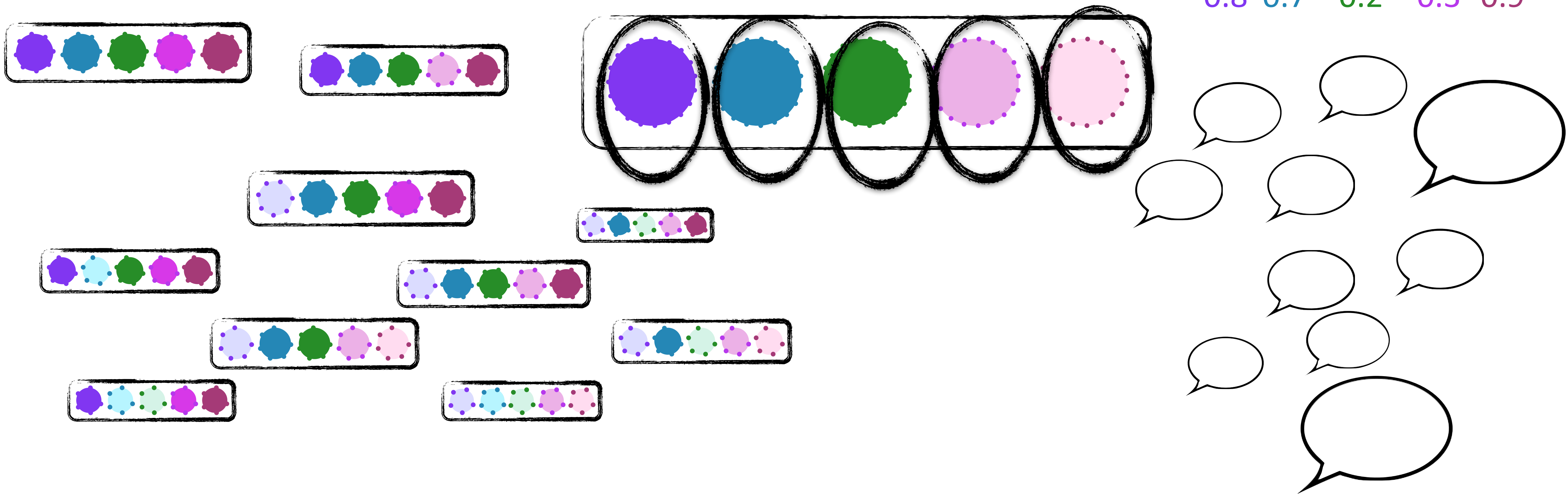


This seems unlikely for real language data because linguistic parameters connect with different types of patterns, which may have nothing to do with each other, or parameters may interact with each other.

Learning with parameters



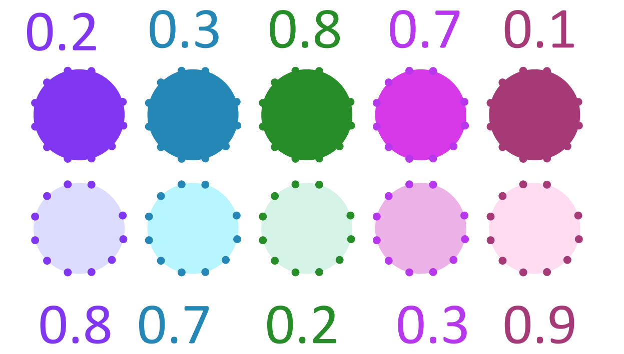
Variational learning

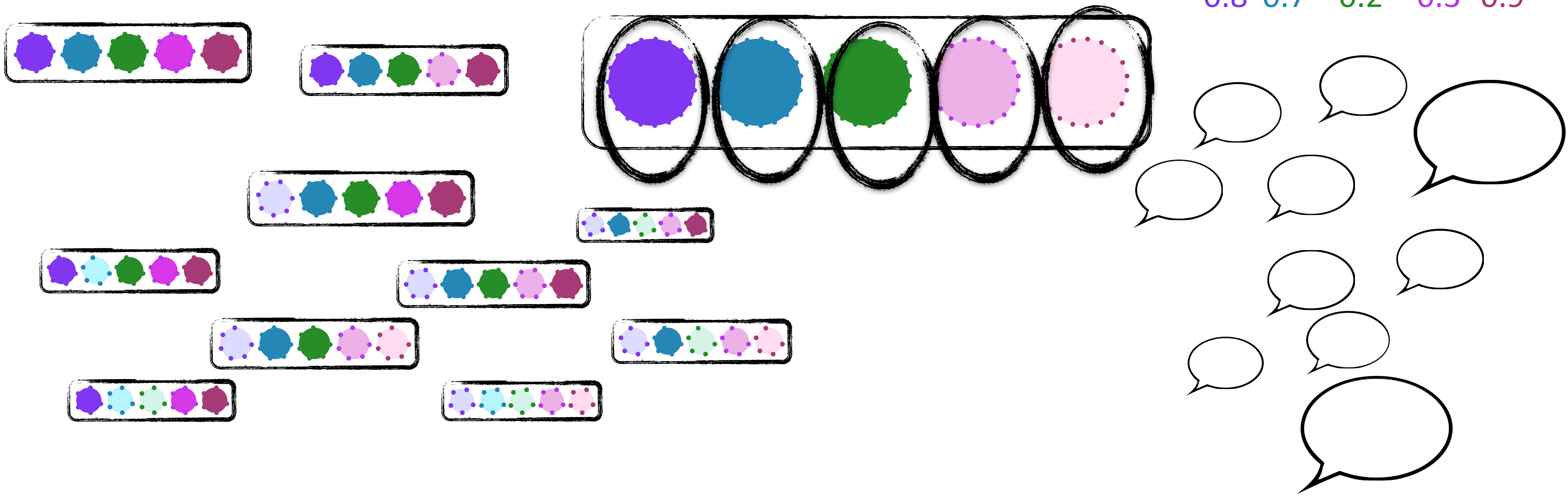


Key: Parameters are separable components of grammars

Learning with parameters

Variational learning

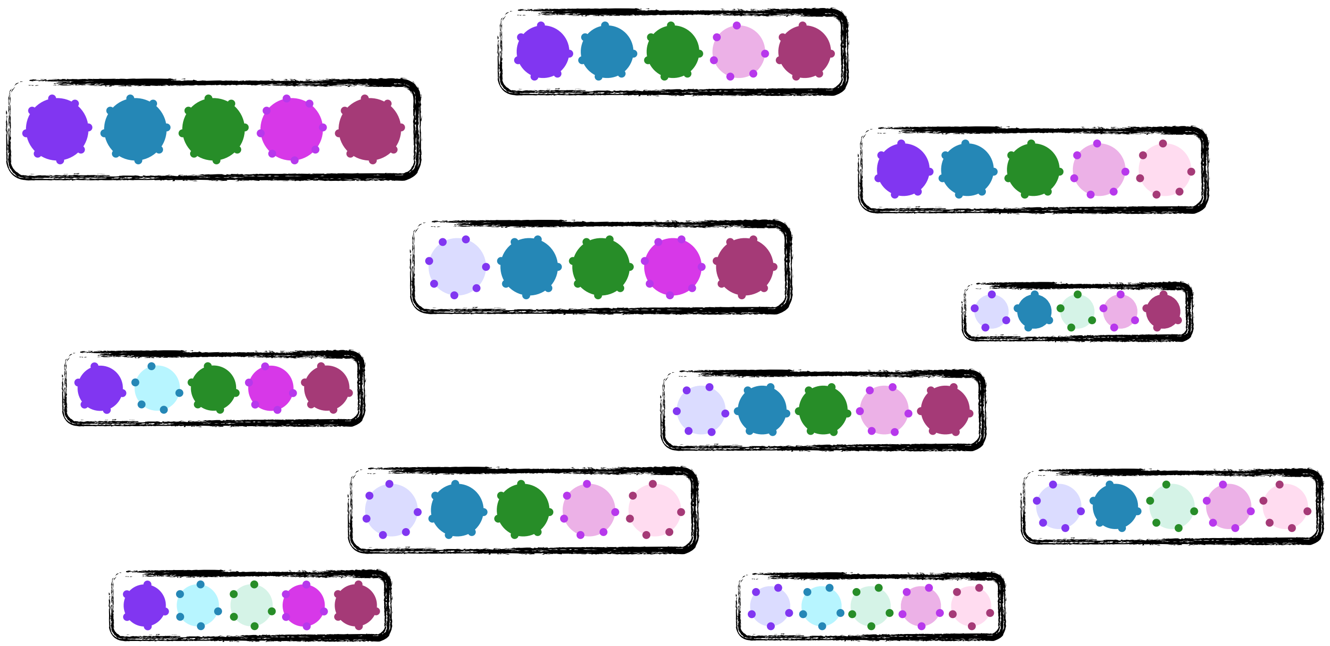


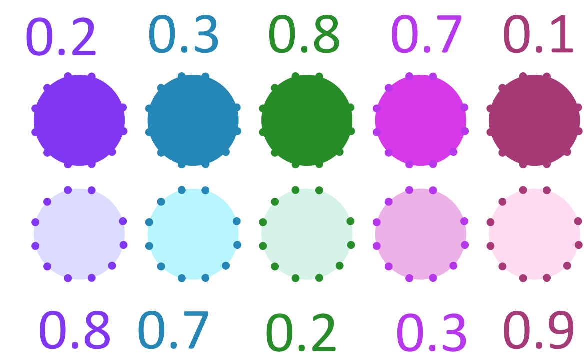


A variational learner can take advantage of the fact that grammars are really sets of parameter values.

Learning with parameters

Variational learning

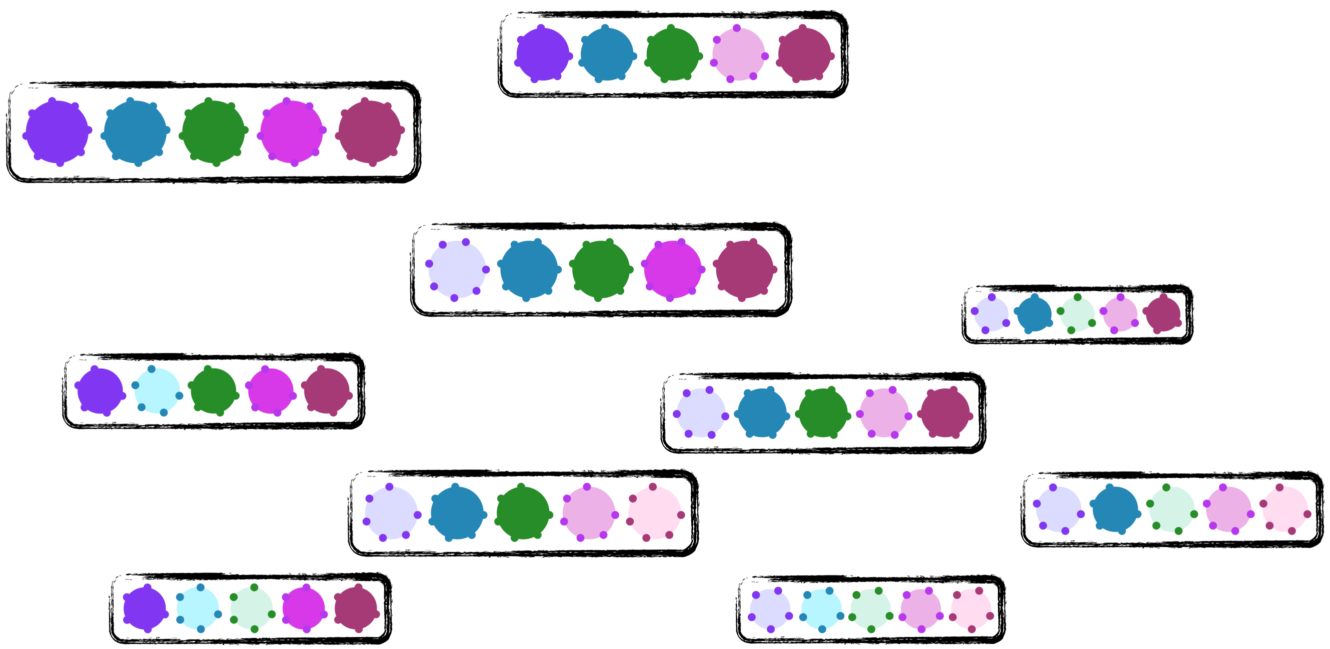


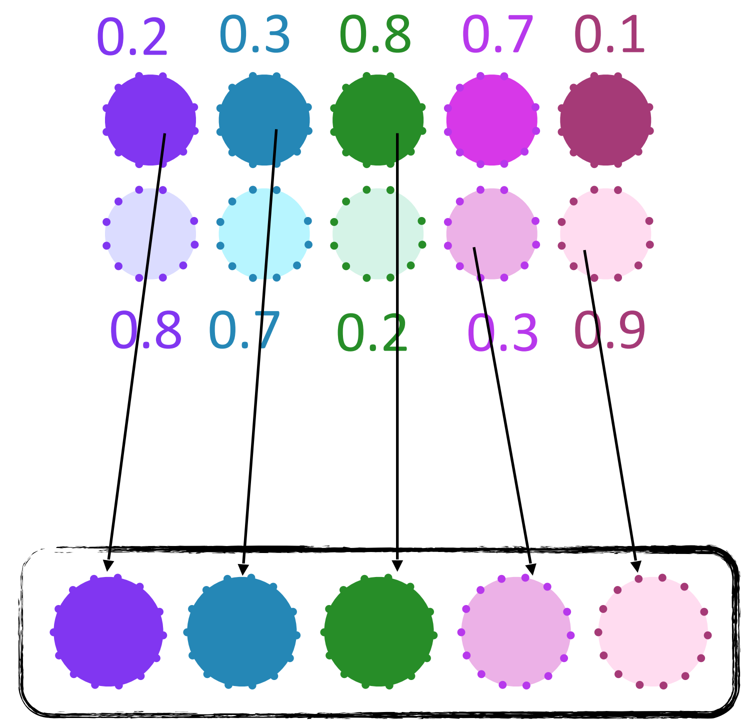


Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning



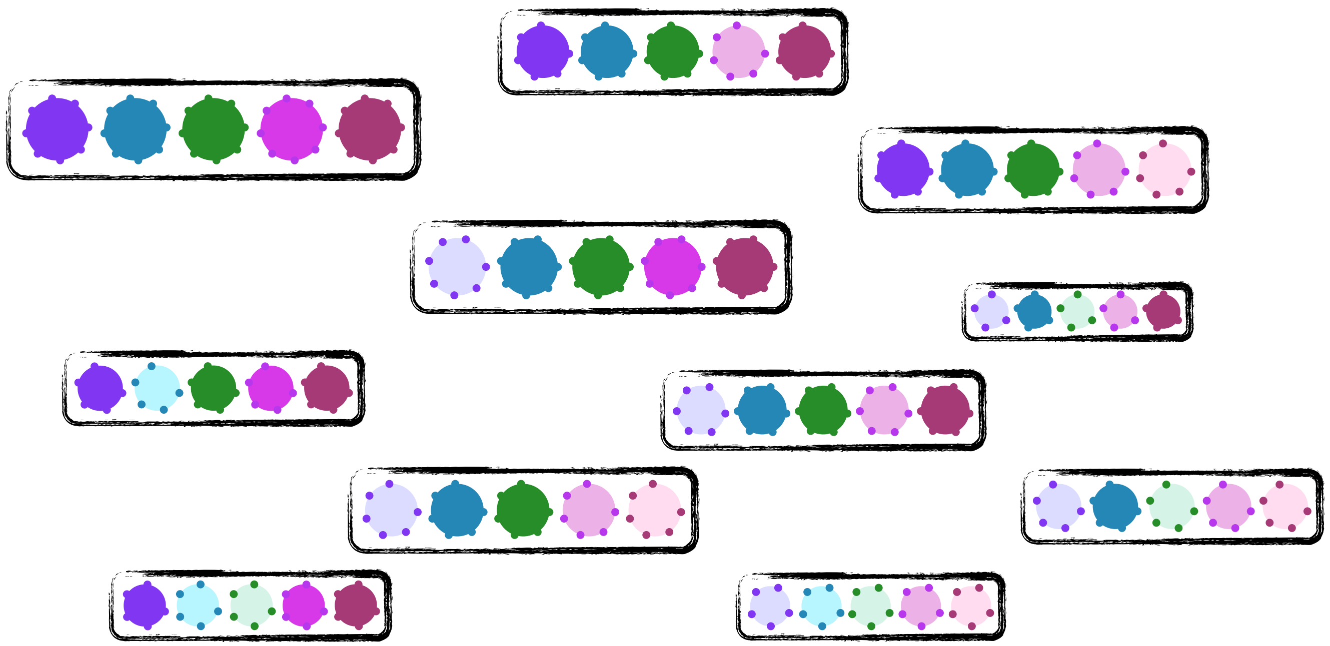


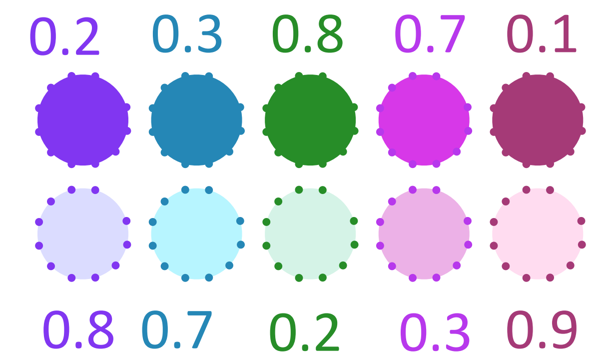
$p = .2 * .3 * .8 * .3 * .9$

Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning

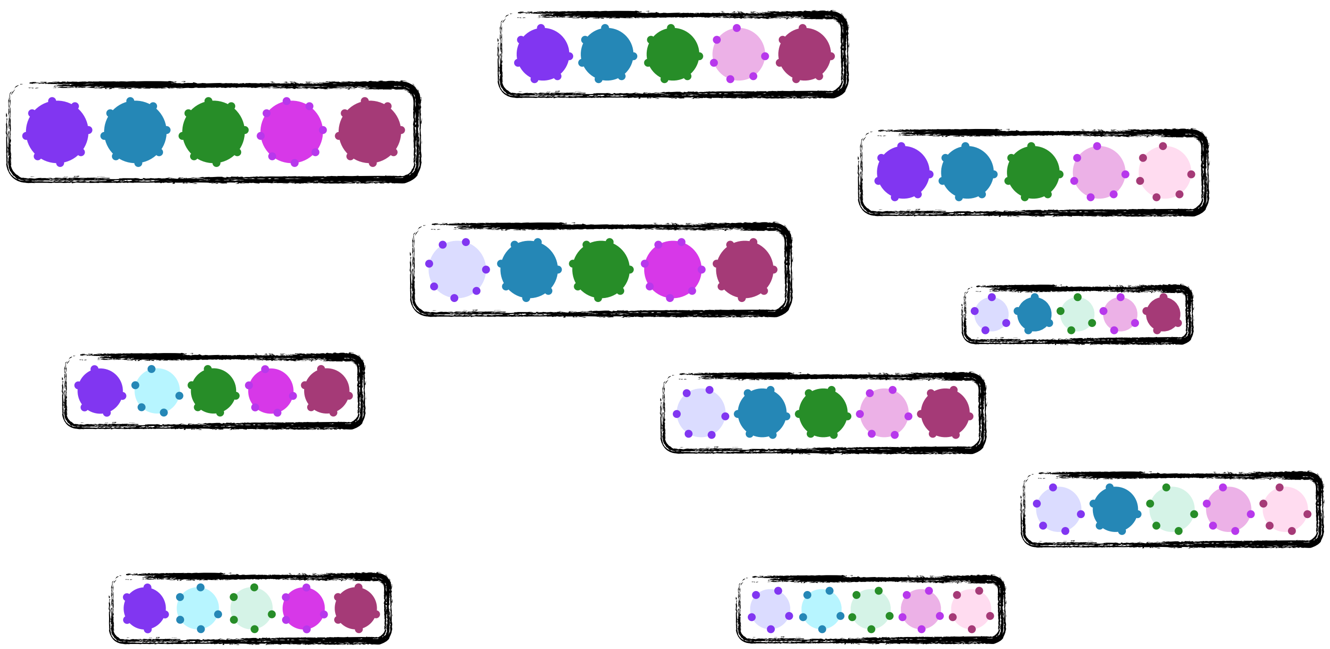


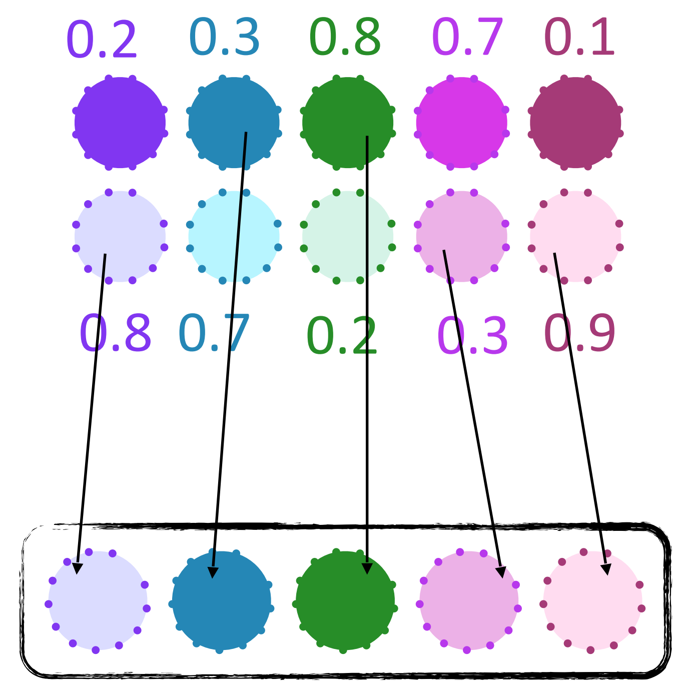


Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning



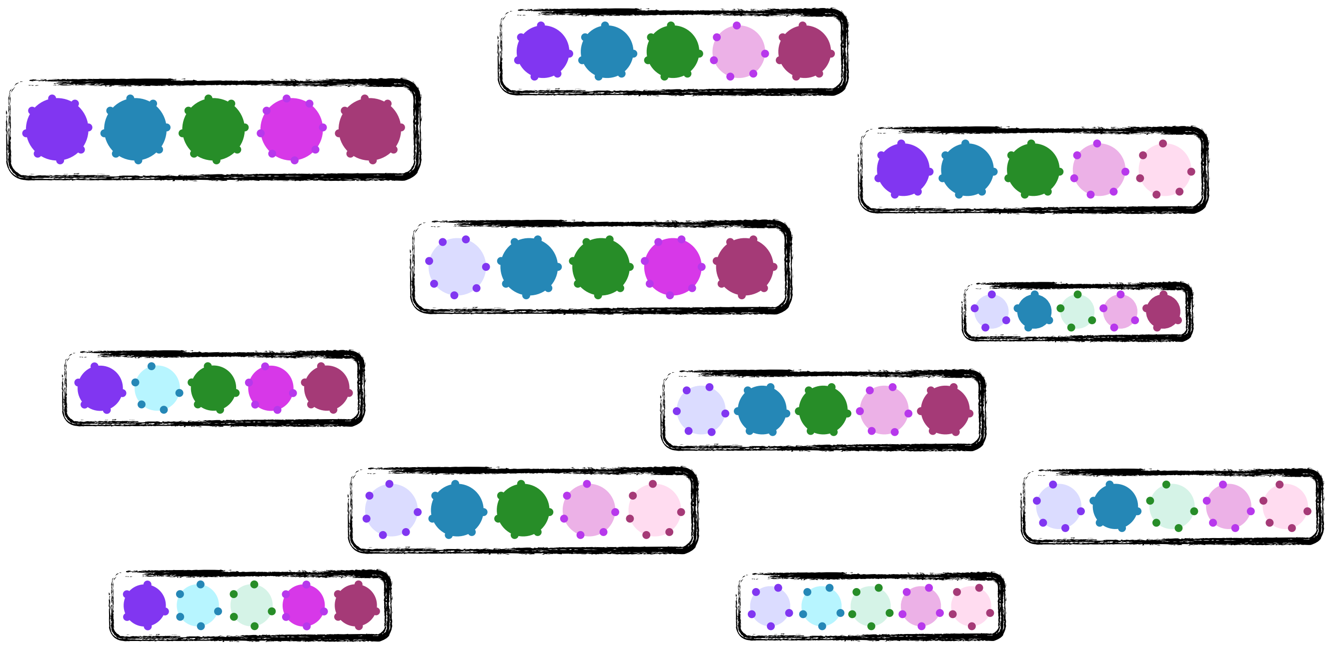


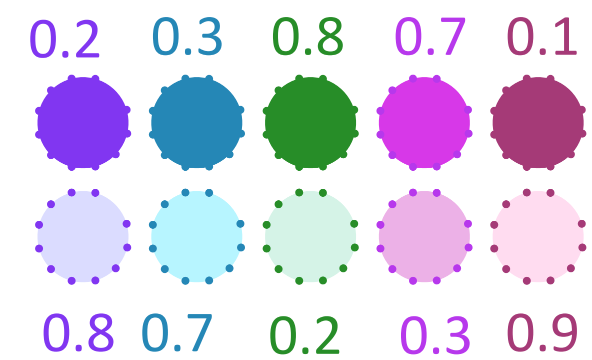
p = .8*.3*.8*.3*.9

Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning

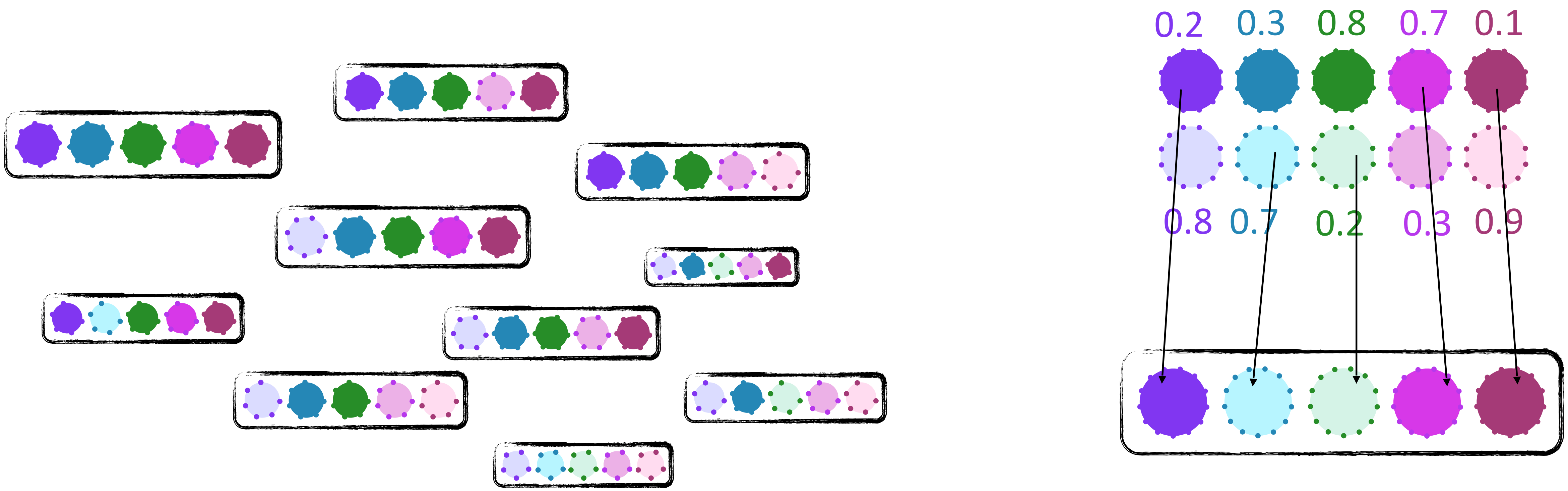




Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning

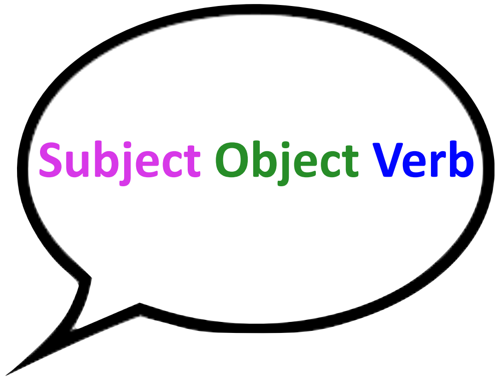


p = .2*.7*.2*.7*.1

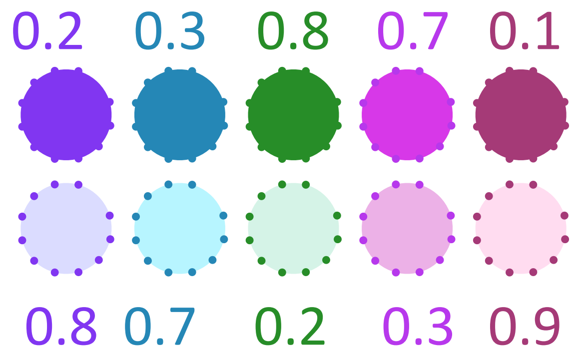
Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters
The learning algorithm

Variational learning



For each data point encountered in the input…

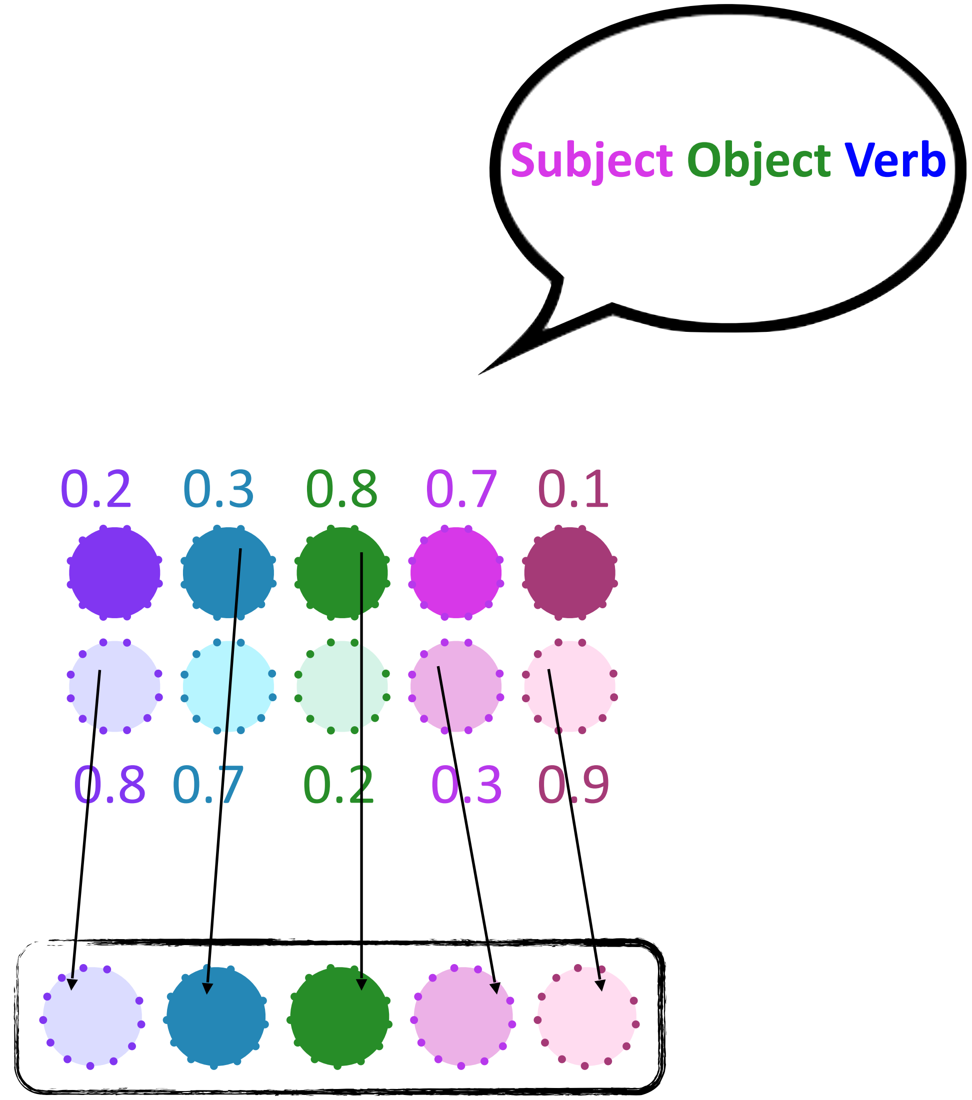


Learning with parameters
The learning algorithm

Variational learning

For each data point encountered in the input…

(1) Choose a grammar to test out on a particular data point. Select a grammar by choosing a set of parameter values, based on the probabilities associated with each parameter value.

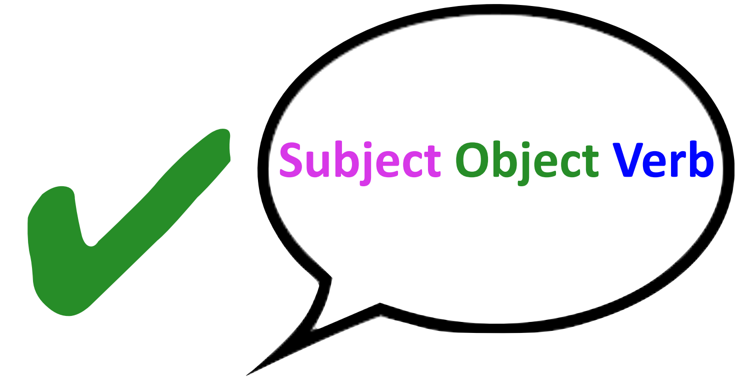


p = .8*.3*.8*.3*.9

Denison, Bonawitz, Gopnik, & Griffiths 2013: Experimental evidence from 4 and 5-year-olds suggests that children are sensitive to the probabilities of complex representations (which parameters are), and so this kind of sampling is not unrealistic.

Learning with parameters
The learning algorithm

Variational learning

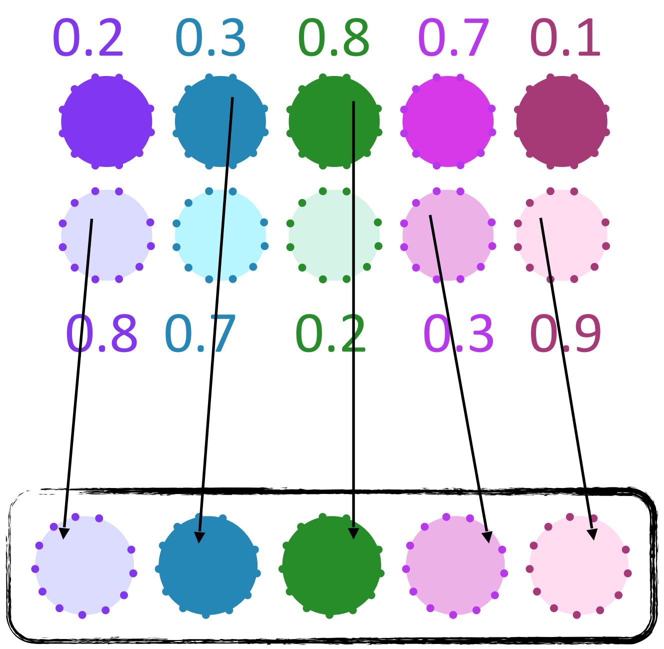


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

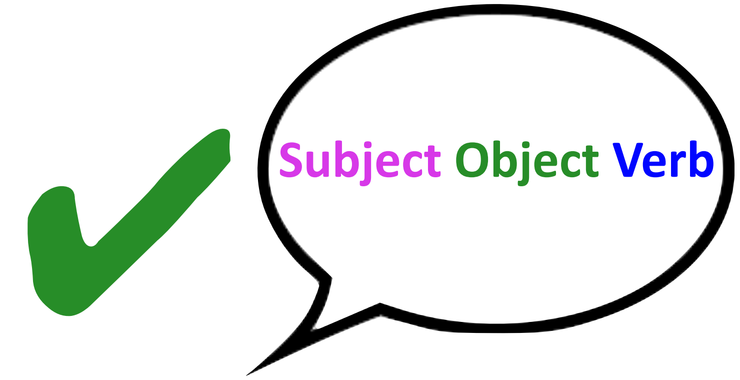
If this grammar can analyze the data point, increase the probability of **all participating parameter** values slightly (reward each value).



p = .8*.3*.8*.3*.9

Learning with parameters
The learning algorithm

Variational learning

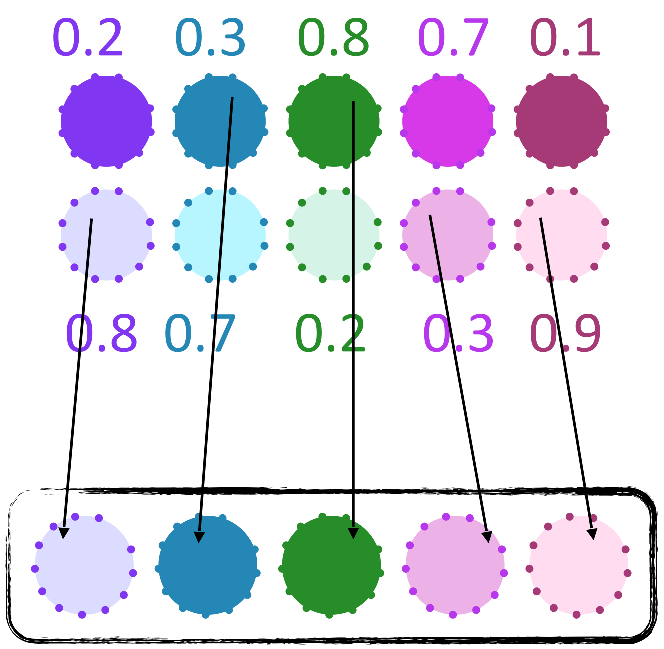


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

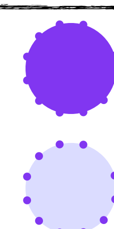
(3) Update parameter value probabilities.



p = .8*.3*.8*.3*.9

1st parameter

Actual update equation for reward:



= .2

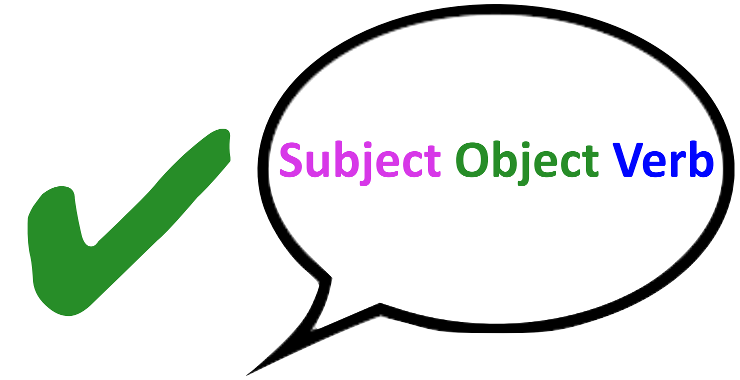
= .8

p_v = previous value of successful parameter value

p_o = previous value of opposing parameter value

Learning with parameters
The learning algorithm

Variational learning

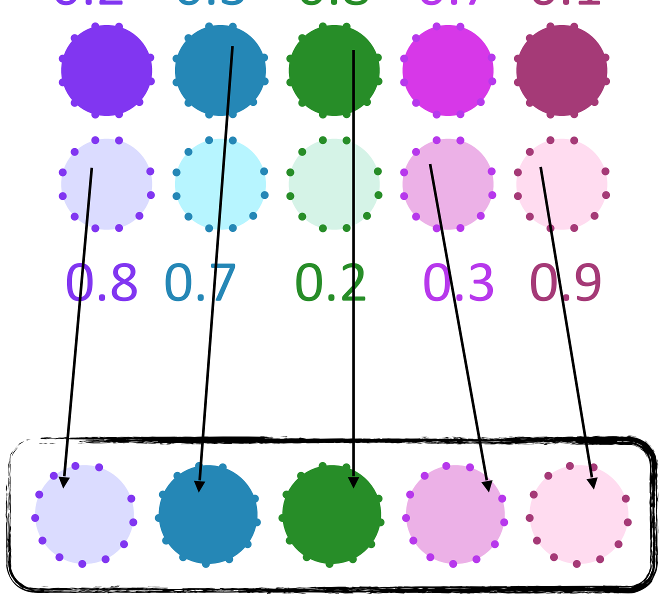


For each data point encountered in the input…

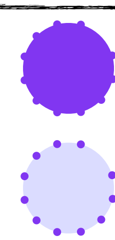
(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter


= .2

= .8

p = .8*.3*.8*.3*.9

Actual update equation for reward:

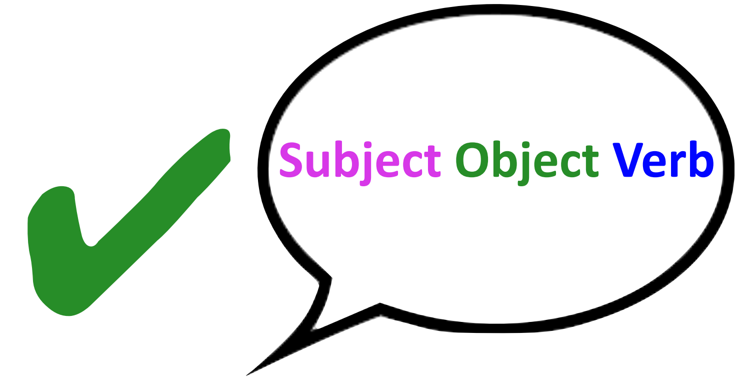
$p_v = 0.8$

$p_o = 0.2$

Learning with parameters

The learning algorithm

Variational learning

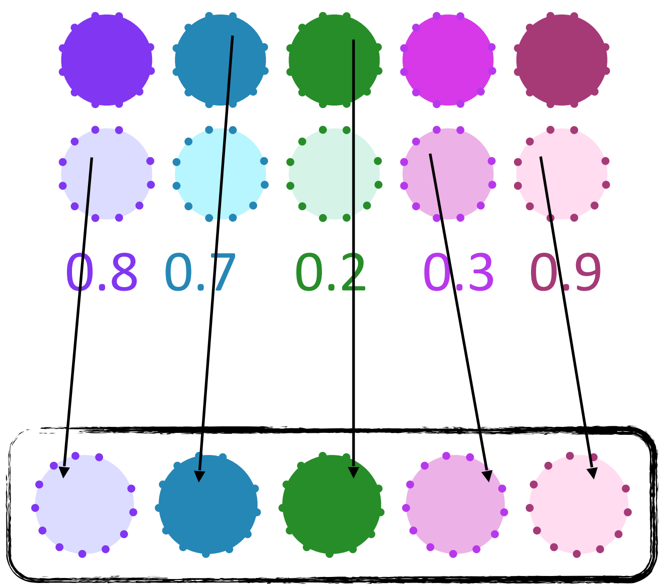


For each data point encountered in the input…

(1) Choose a grammar.

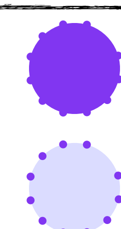
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



p = .8*.3*.8*.3*.9

1st parameter



= .2

= .8

Actual update equation for reward:

$p_v = 0.8$

$p_o = 0.2$

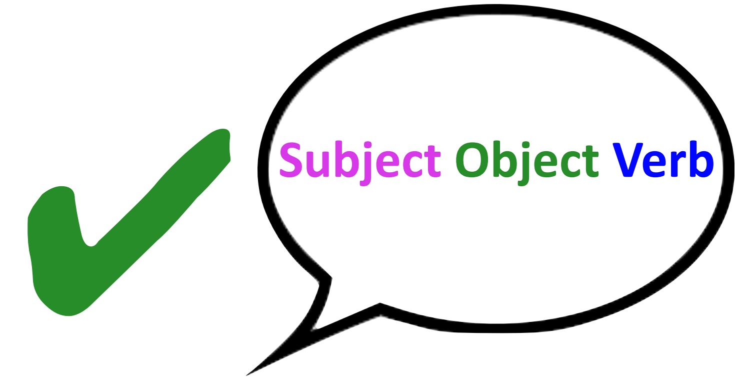
$p_{v\_updated} = p_v + \gamma(1- p_v)$

$p_{o\_updated} = (1-\gamma)p_o$

γ = learning rate (ex: γ = .125)

Learning with parameters
The learning algorithm

Variational learning

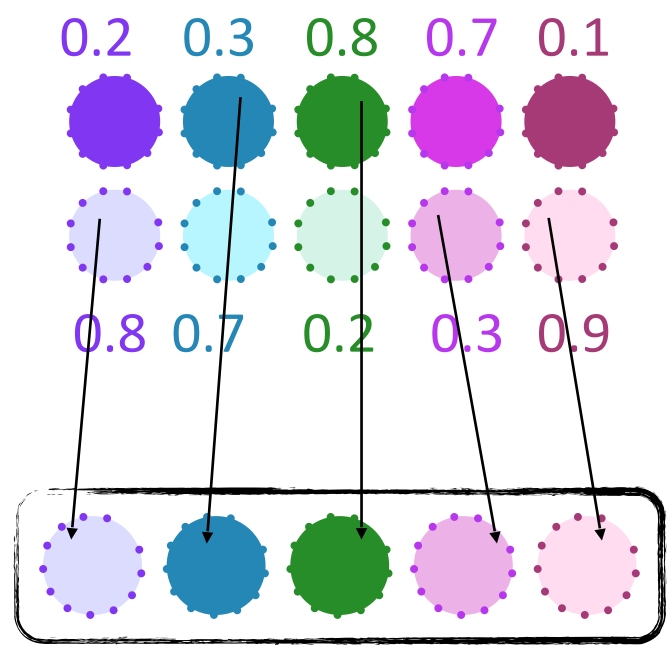


For each data point encountered in the input…

(1) Choose a grammar.

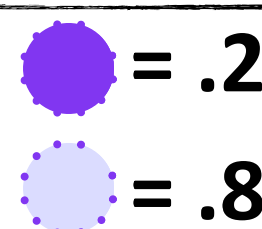
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



$p = .8 * .3 * .8 * .3 * .9$

1st parameter



Actual update equation for reward:

$p_v = 0.8$

$p_o = 0.2$

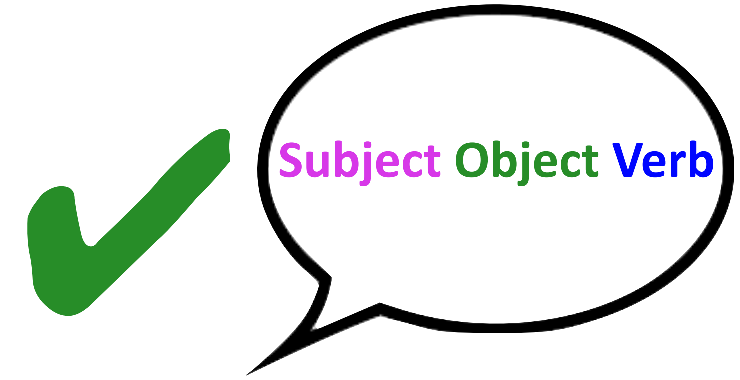
$p_{v\_updated} = 0.8 + 0.125(1- 0.8)$

$p_{o\_updated} = (1-0.125)0.2$

γ = learning rate (ex: $\gamma = .125$)

Learning with parameters
The learning algorithm

Variational learning

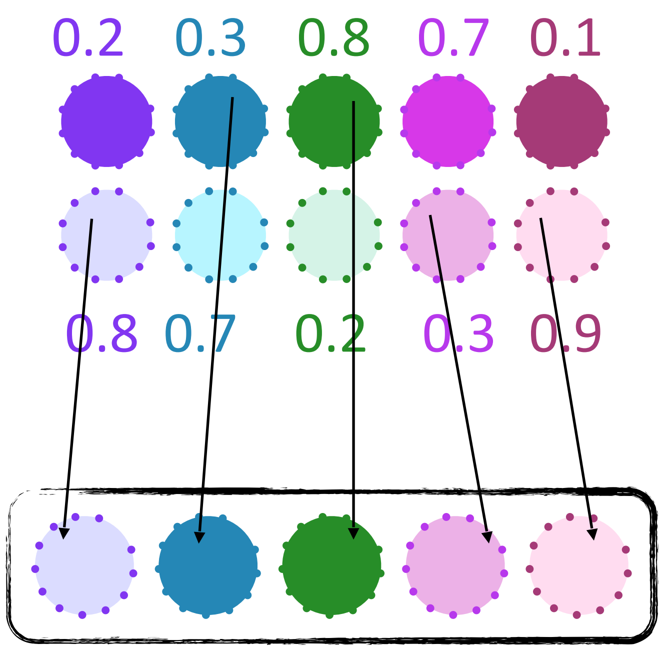


For each data point encountered in the input…

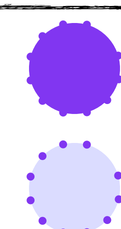
(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter



p = .8*.3*.8*.3*.9

Actual update equation for reward:

$p_v = 0.8$

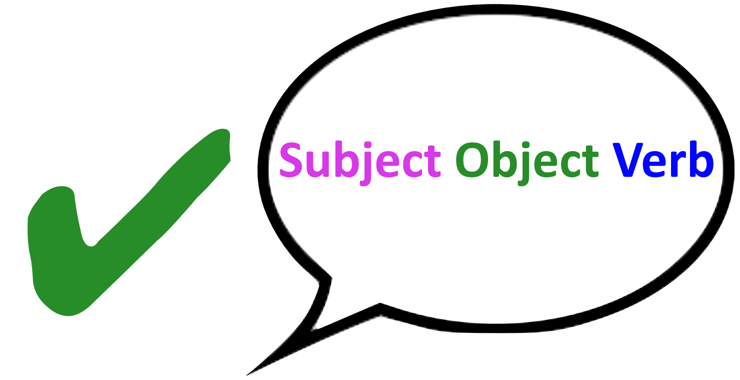
$p_o = 0.2$

$p_{v\_updated} = 0.825$

$p_{o\_updated} = 0.175$

Learning with parameters
The learning algorithm

Variational learning

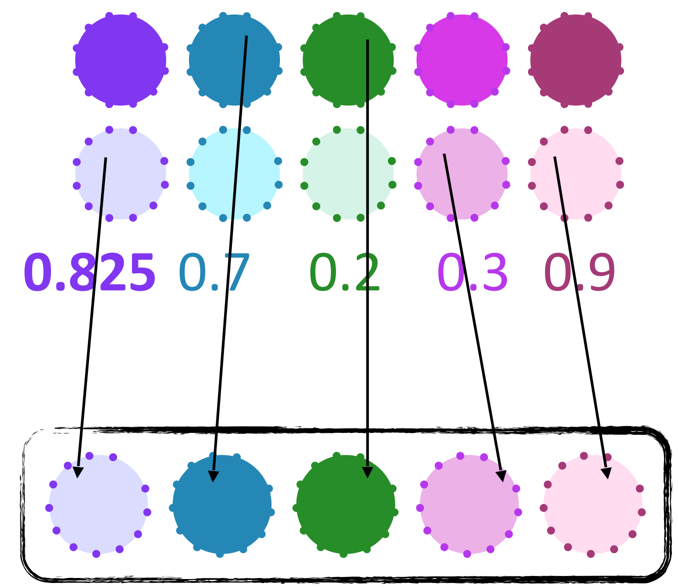


For each data point encountered in the input…

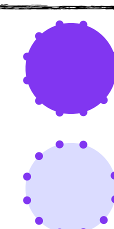
(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter


= .2
= .8

p = .8*.3*.8*.3*.9

Actual update equation for reward:

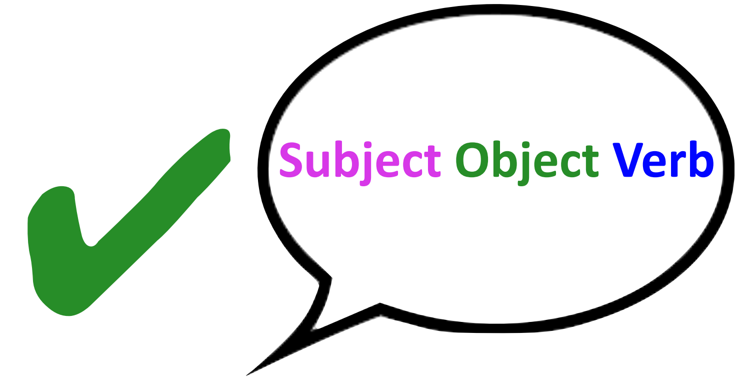
$p_v = 0.8$
$p_o = 0.2$

$p_{v\_updated} = 0.825$
$p_{o\_updated} = 0.175$

Do this for all the other parameters, too.

Learning with parameters
The learning algorithm

Variational learning

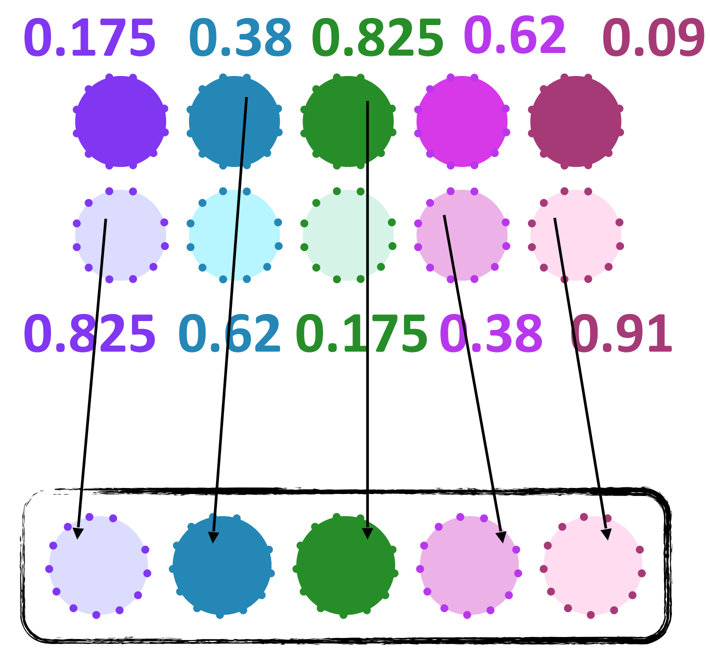


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

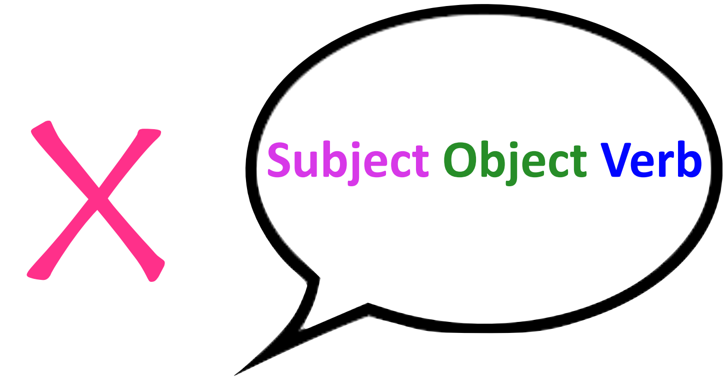
(3) Update parameter value probabilities.



p = .8*.3*.8*.3*.9

Learning with parameters
The learning algorithm

Variational learning

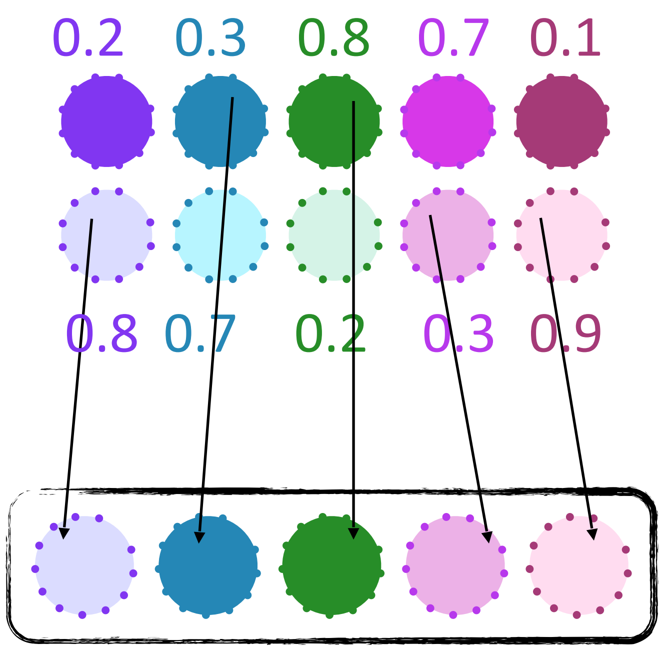


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



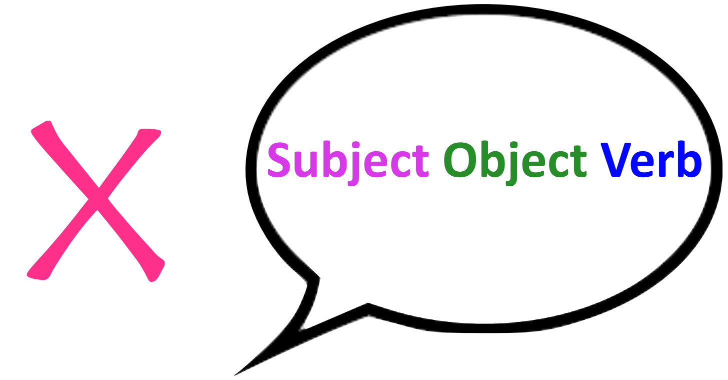
p = .8*.3*.8*.3*.9

But what happens if the selected grammar can't account for the data point?

Then all the participating parameter values are punished.

Learning with parameters
The learning algorithm

Variational learning

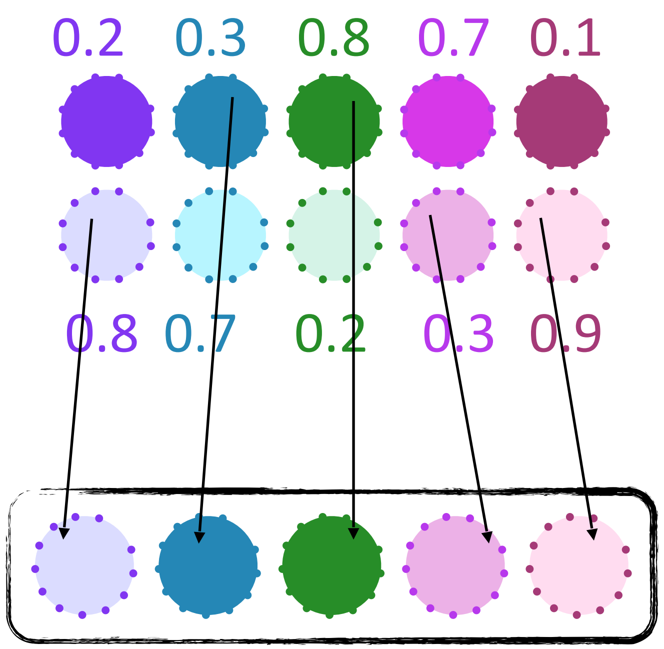


For each data point encountered in the input…

(1) Choose a grammar.

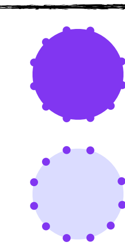
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter

Actual update equation for punishment:



= .2

= .8

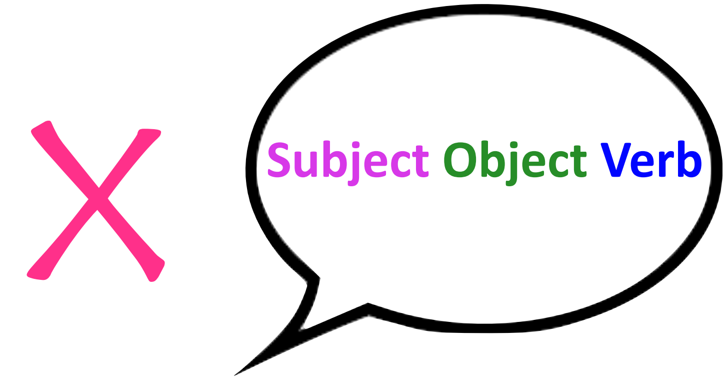
p = .8*.3*.8*.3*.9

p_v = previous value of unsuccessful parameter value

p_o = previous value of opposing parameter value

Learning with parameters
The learning algorithm

Variational learning

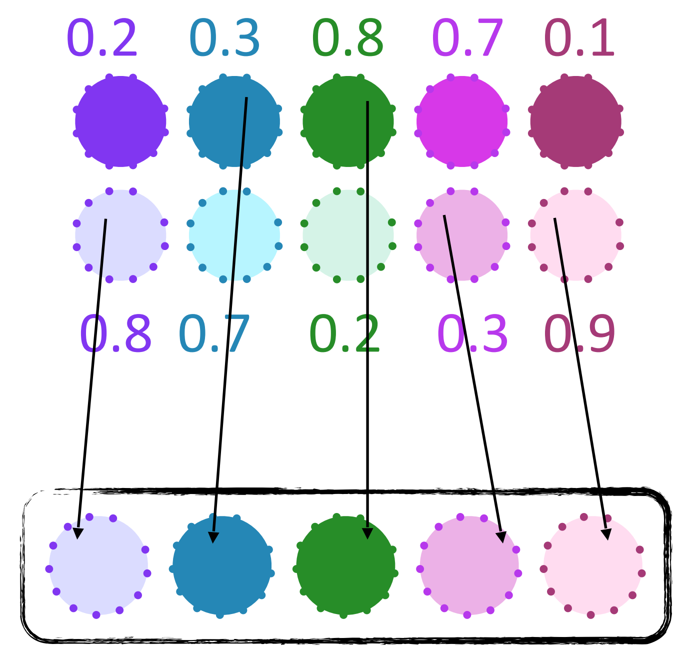


For each data point encountered in the input…

(1) Choose a grammar.

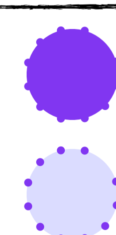
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter

Actual update equation for punishment:



= .2

= .8

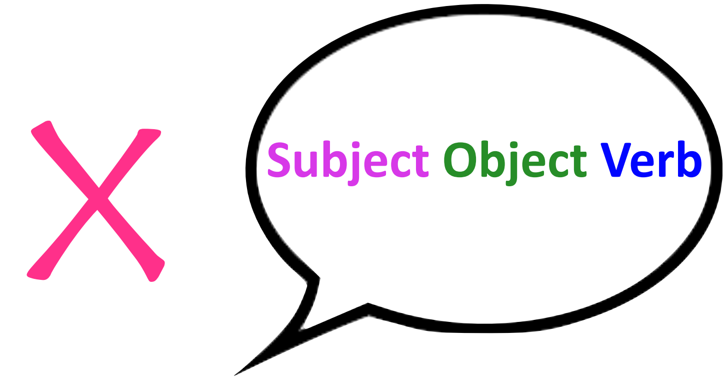
$p_v = 0.8$

$p_o = 0.2$

p = .8*.3*.8*.3*.9

Learning with parameters
The learning algorithm

Variational learning

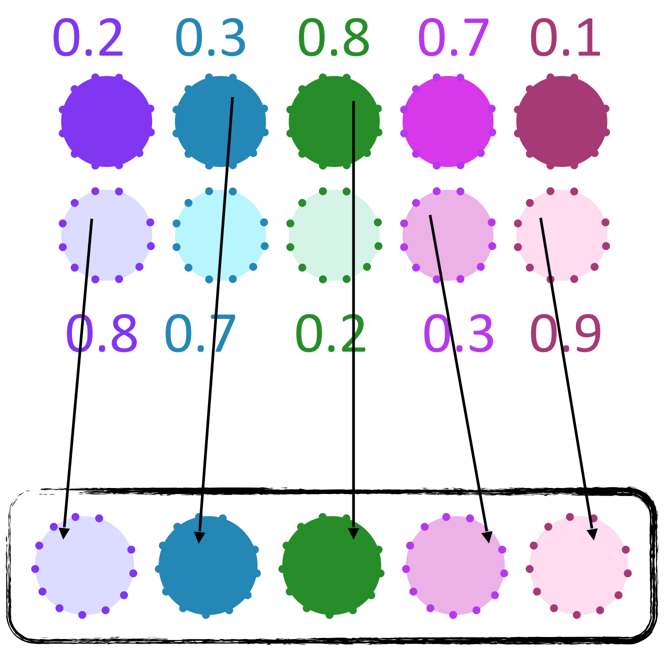


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

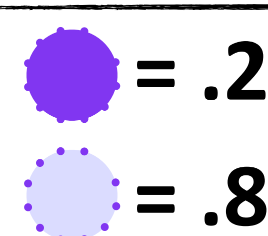
(3) Update parameter value probabilities.



p = .8*.3*.8*.3*.9

1st parameter

Actual update equation for punishment:



$p_v = 0.8$

$p_o = 0.2$

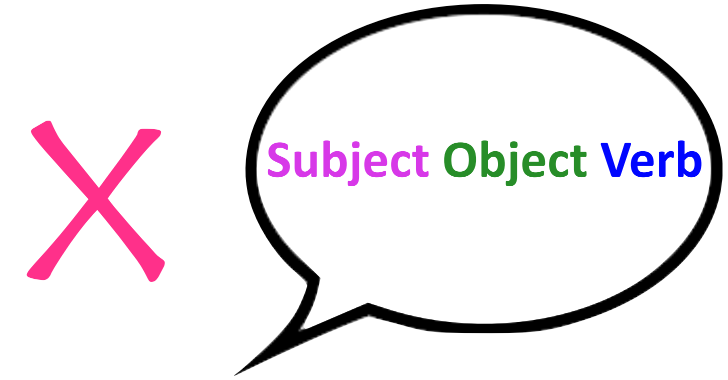
$p_{v\_updated} = (1-\gamma)p_v$

$p_{o\_updated} = \gamma + (1-\gamma)p_o$

γ = learning rate (ex: $\gamma = .125$)

Learning with parameters
The learning algorithm

Variational learning

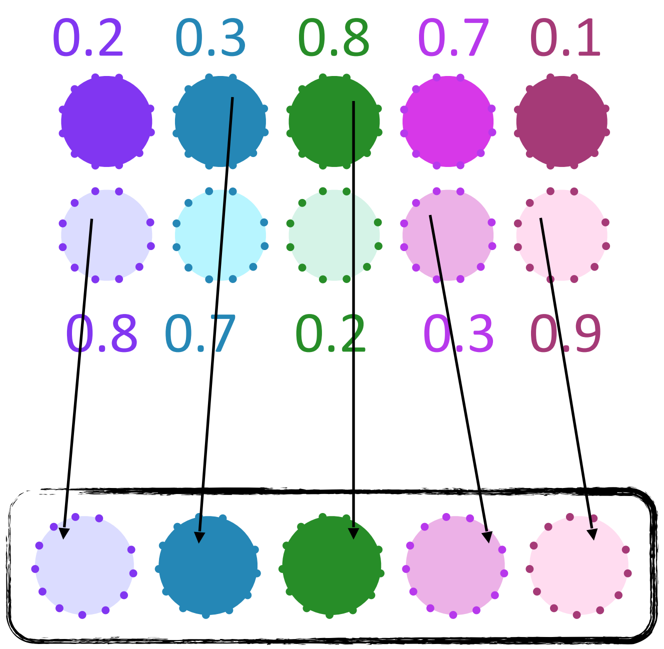


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

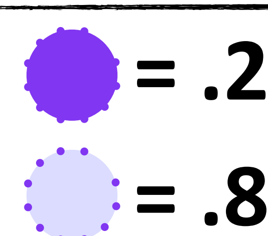
(3) Update parameter value probabilities.



$p = .8 * .3 * .8 * .3 * .9$

1st parameter

Actual update equation for punishment:



$p_v = 0.8$

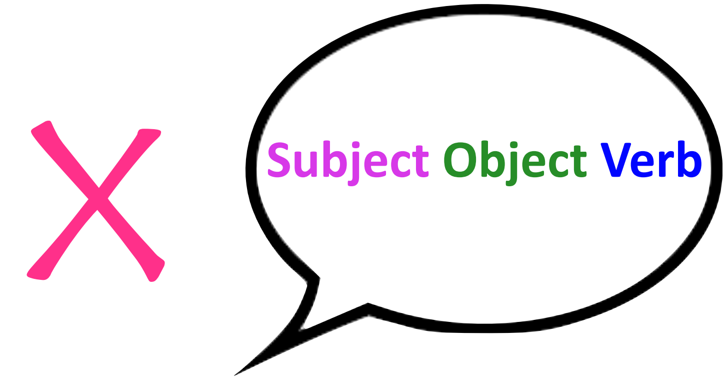
$p_o = 0.2$

$p_{v\_updated} = (1-0.125)0.8$

$p_{o\_updated} = 0.125 + (1-0.125)0.2$

Learning with parameters
The learning algorithm

Variational learning

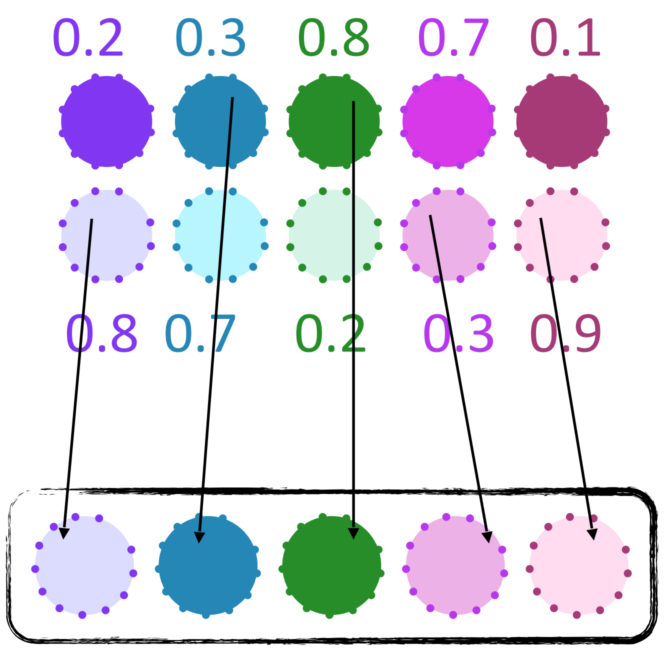


For each data point encountered in the input…

(1) Choose a grammar.

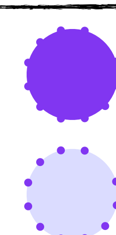
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter

Actual update equation for punishment:


= .2
= .8

p = .8*.3*.8*.3*.9

$p_v = 0.8$

$p_o = 0.2$

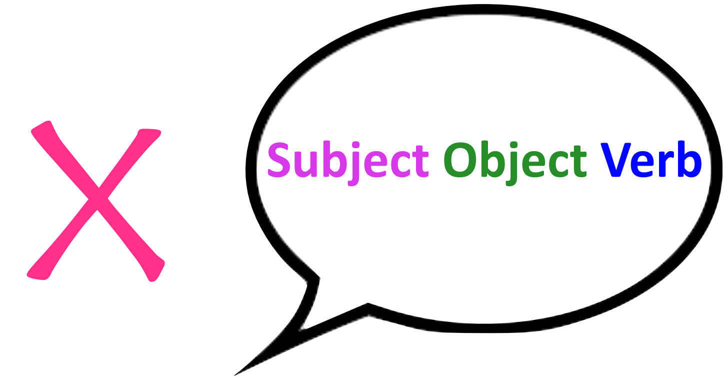
$p_{v\_updated} = 0.70$

$p_{o\_updated} = 0.30$

Learning with parameters

The learning algorithm

Variational learning

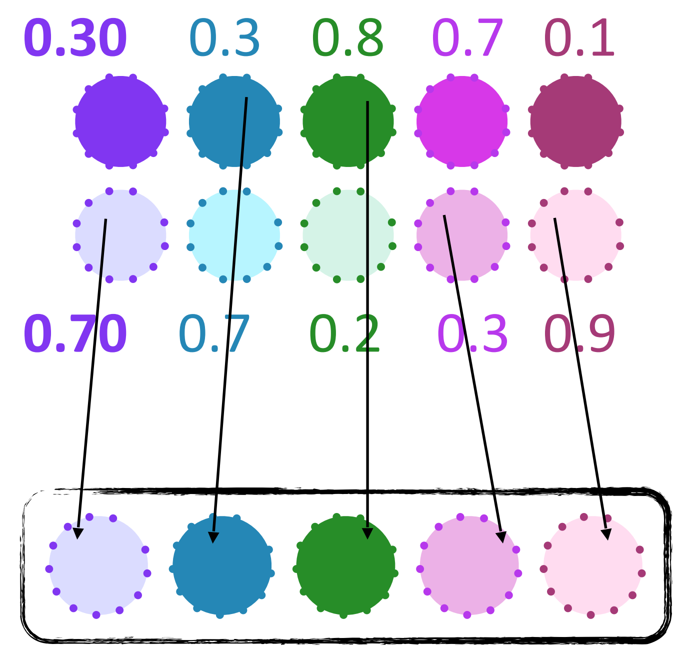


For each data point encountered in the input…

(1) Choose a grammar.

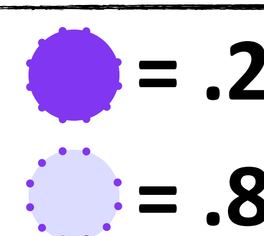
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



1st parameter

Actual update equation for punishment:



p = .8*.3*.8*.3*.9

$p_v = 0.8$

$p_o = 0.2$

$p_{v\_updated} = 0.70$

$p_{o\_updated} = 0.30$

Do this for all the other parameters, too.

Learning with parameters
The learning algorithm

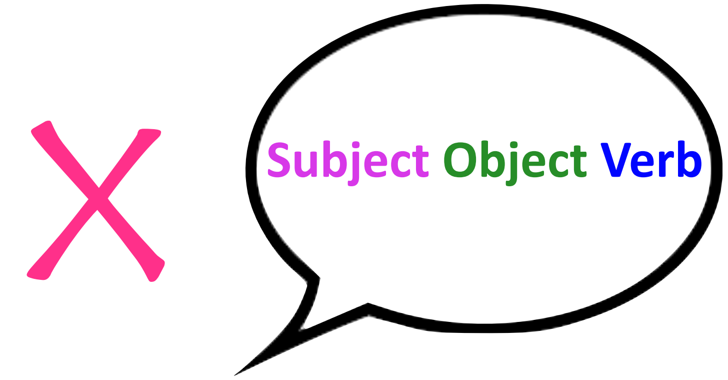
Variational learning

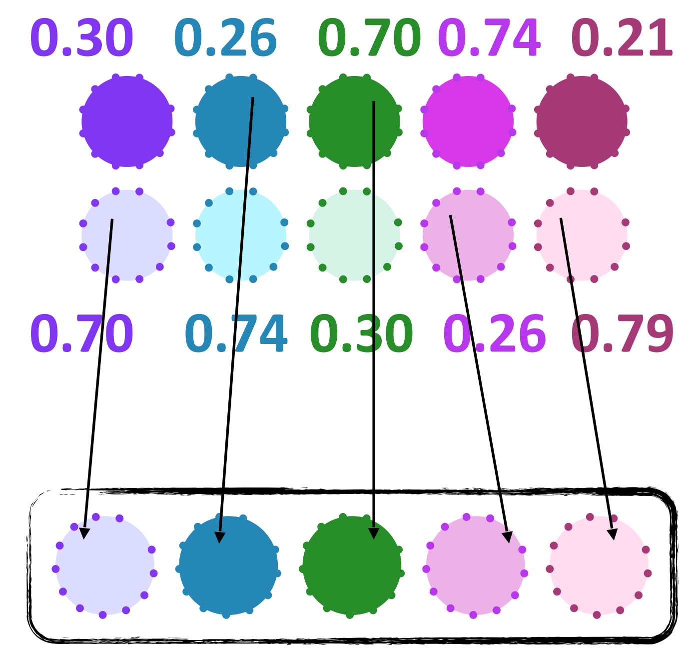
For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.

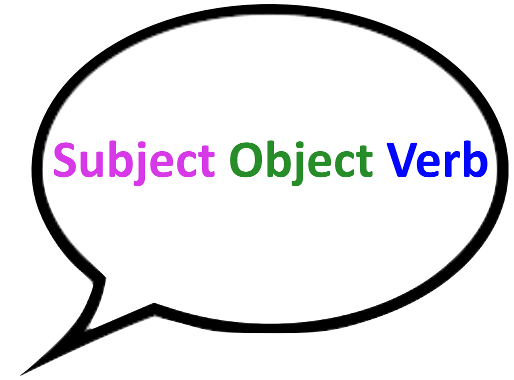




$p = .8 * .3 * .8 * .3 * .9$

Learning with parameters
The learning algorithm

Variational learning

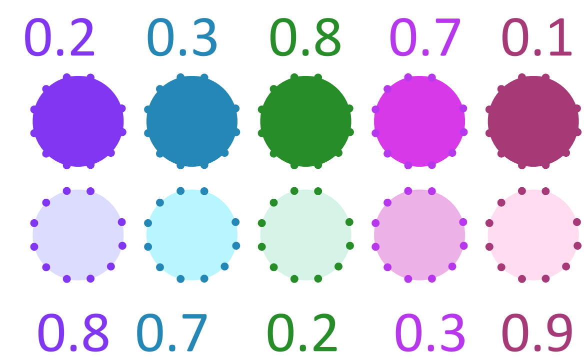


For each data point encountered in the input…

(1) Choose a grammar.

(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.

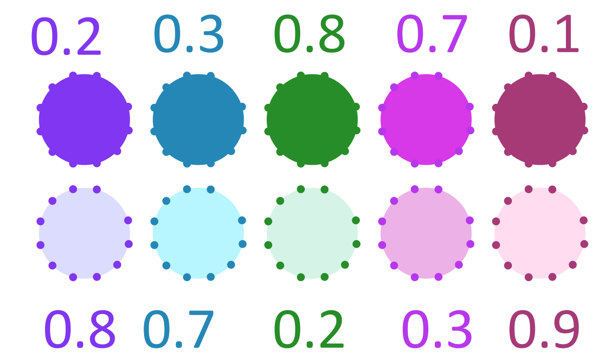


Problem ameliorated!
Unambiguous data are much more likely to exist for **individual parameter values** instead of entire grammars.

Learning with parameters

The learning algorithm

Variational learning



Unambiguous data are much more likely to exist for **individual parameter values** instead of entire grammars.

Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject Object Verb

G2 Head-final +subj-drop

G4 Head-final -subj-drop

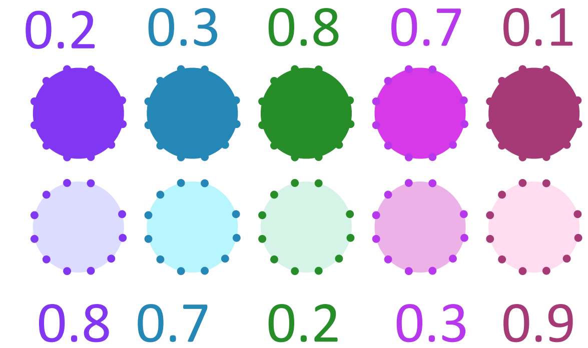
G1 Head-first +subj-drop

G3 Head-first -subj-drop

Learning with parameters

The learning algorithm

Variational learning



In this case, if either G2 or G4 were selected, head-final would be rewarded (in addition to whichever subj-drop value was used).

Head-directionality   Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject   Object   Verb

✔ **G2** Head-final +subj-drop

✔ **G4** Head-final -subj-drop

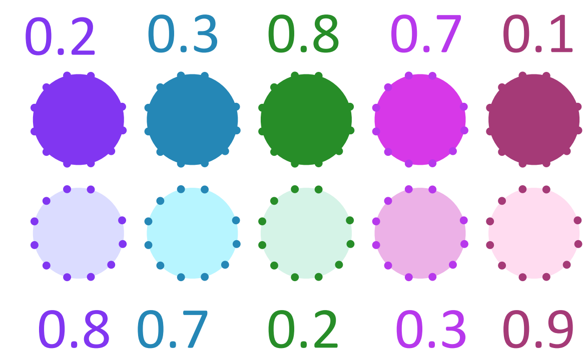
✗ **G1** Head-first +subj-drop

✗ **G3** Head-first -subj-drop

Learning with parameters

The learning algorithm

Variational learning



In this case, if either G1 or G3 were selected, head-first would be punished (in addition to whichever subj-drop value was used).

Head-directionality   Subject drop (subj-drop)

"...dass **ich** **Kätzchen** **liebe**."

...that I Kitties love

Subject **Object** **Verb**

G1 Head-first +subj-drop

G3 Head-first -subj-drop

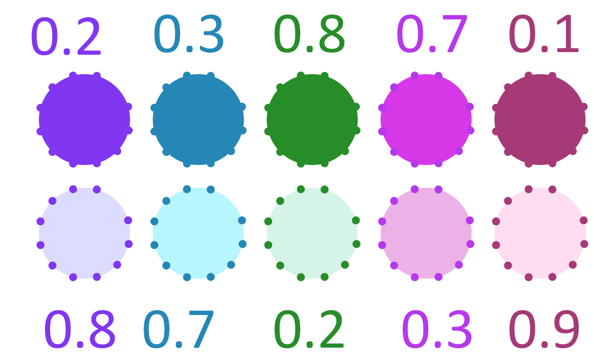
G2 Head-final +subj-drop

G4 Head-final -subj-drop

Learning with parameters

The learning algorithm

Variational learning



Because this data point is unambiguous for head-final, grammars using that value would be rewarded and its probability as a parameter value would become 1.0 over time.

Head-directionality Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject Object Verb

G2 Head-final +subj-drop

G4 Head-final -subj-drop

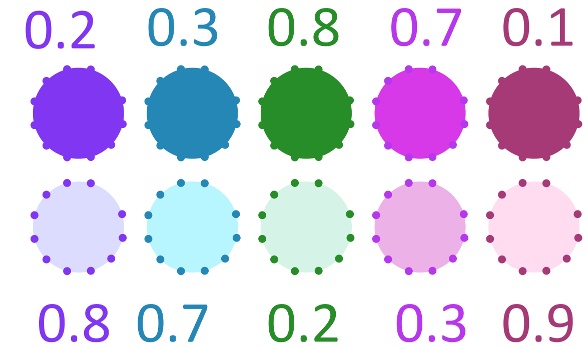
G1 Head-first +subj-drop

G3 Head-first -subj-drop

Learning with parameters

The learning algorithm

Variational learning



Meanwhile, grammars using head-first would be punished every time, and its probability as a parameter value would approach 0.0 over time.

Head-directionality   Subject drop (subj-drop)

"...dass ich Kätzchen liebe."

...that I Kitties love

Subject Object Verb

G1 Head-first +subj-drop ✗

G3 Head-first -subj-drop ✗

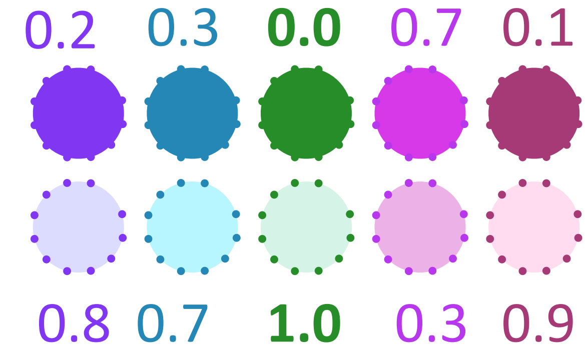
G2 Head-final +subj-drop ✓

G4 Head-final -subj-drop ✓

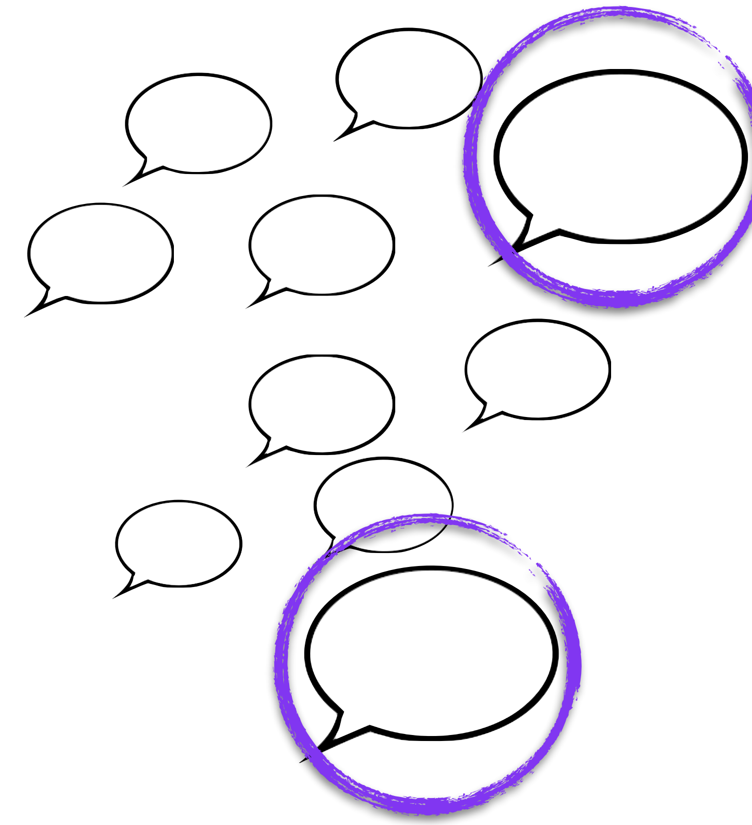
Learning with parameters

The learning algorithm

Variational learning



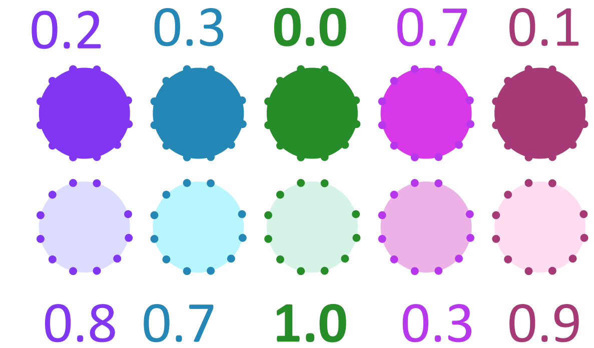
Implication: The more unambiguous data there are, the faster the native language's parameter value will "win" (reach a probability near 1.0). This means that the child will learn the associated structural pattern faster.



Learning with parameters

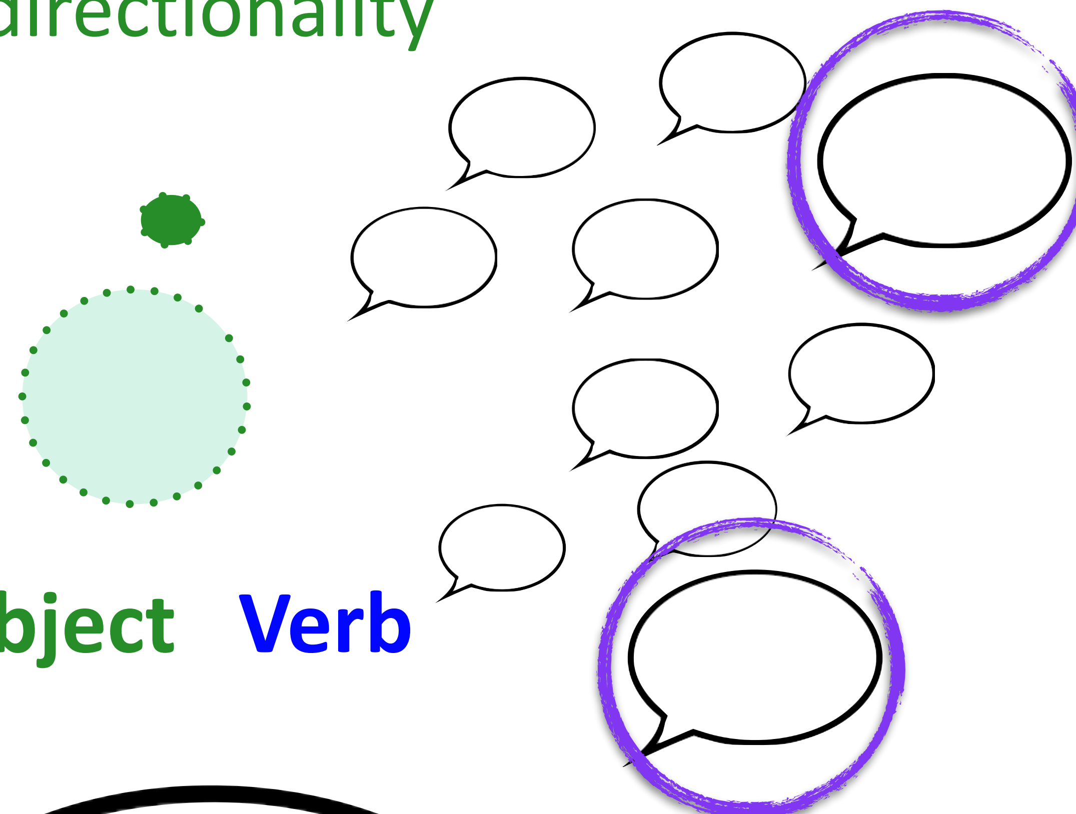
The learning algorithm

Variational learning



Head-directionality

Example: the more unambiguous head-final data the child encounters, the faster a child should learn that the native language prefers objects before verbs as the basic order.



Subject **Object** **Verb**

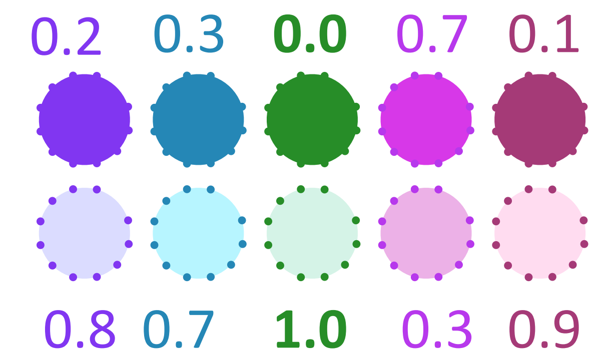
“…dass ich Kätzchen liebe.”

…that I Kitties love

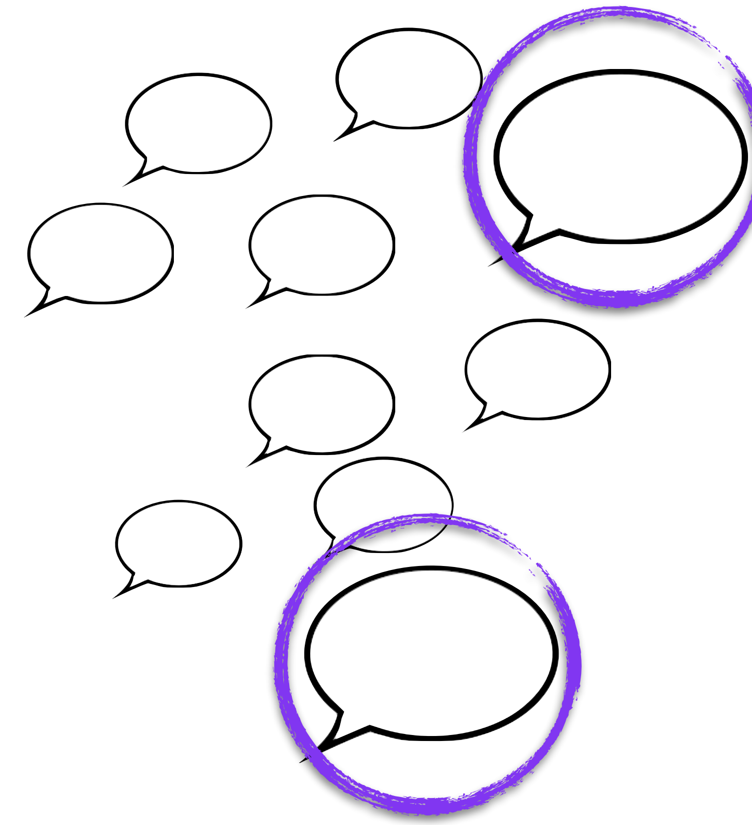
Learning with parameters

The learning algorithm

Variational learning

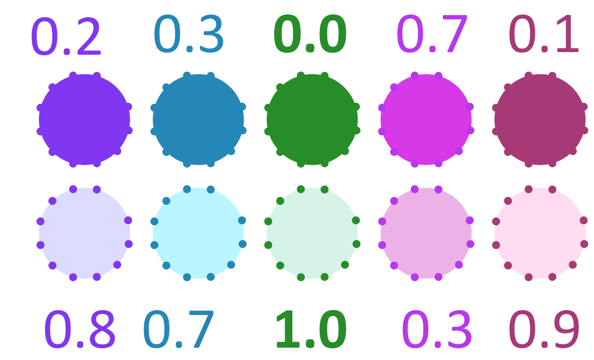


Is it true that the amount of unambiguous data the child encounters for a particular parameter strongly impacts when the child learns that structural property of the language?



Learning with parameters
The learning algorithm

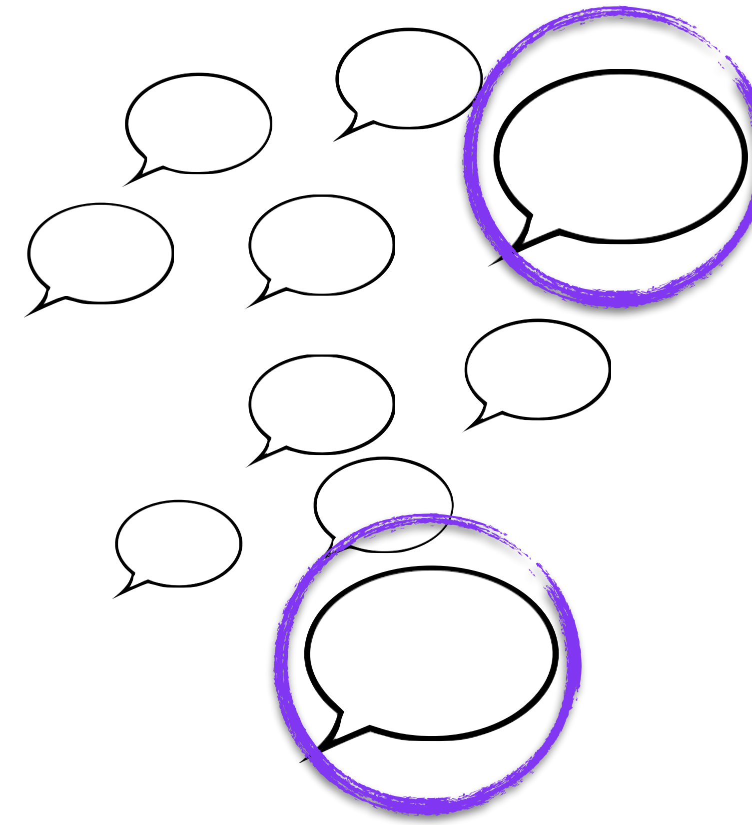
Variational learning



Striking evidence that this is true

Table 1: The qualitative fit Yang discovered between the unambiguous data advantage (Adv) perceived by a VarLearner in its acquisitional intake and the observed age of acquisition (AoA) in children for six parameter values across different languages.

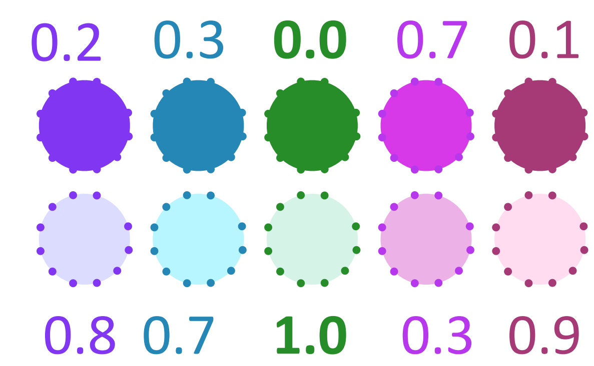
Param Value	Language	Unambiguous Form	Unambiguous Ex	Adv	AoA
+*wh*-fronting	English	*wh*-fronting in questions	*Who did you see?*	25%	<1;8
+topic-drop	Chinese	null objects	*Wǒ méi kànjiàn* *I not see* "I didn't see (him)"	12%	<1;8
+pro-drop	Italian	null subjects in questions	*Chi hai visto* *who have seen* "Who have you seen?"	10%	<1;8
+verb-raising	French	*Verb Adverb*	*Jean voit souvent Marie* *Jean sees often Marie* "Jean often sees Marie"	7%	1;8
-pro-drop	English	expletive subjects	There's a penguin on the ice.	1.2%	3;0
+verb-second	German Dutch	*Object Verb Subject*	*Pinguine liebe ich.* *penguins like I* "I like penguins"	1.2%	3;0-3;2
-scope-marking	English	long-distance *wh* questions without medial-*wh*	*Who do you think is on the ice?*	0.2%	>4.0



Learning with parameters

The learning algorithm

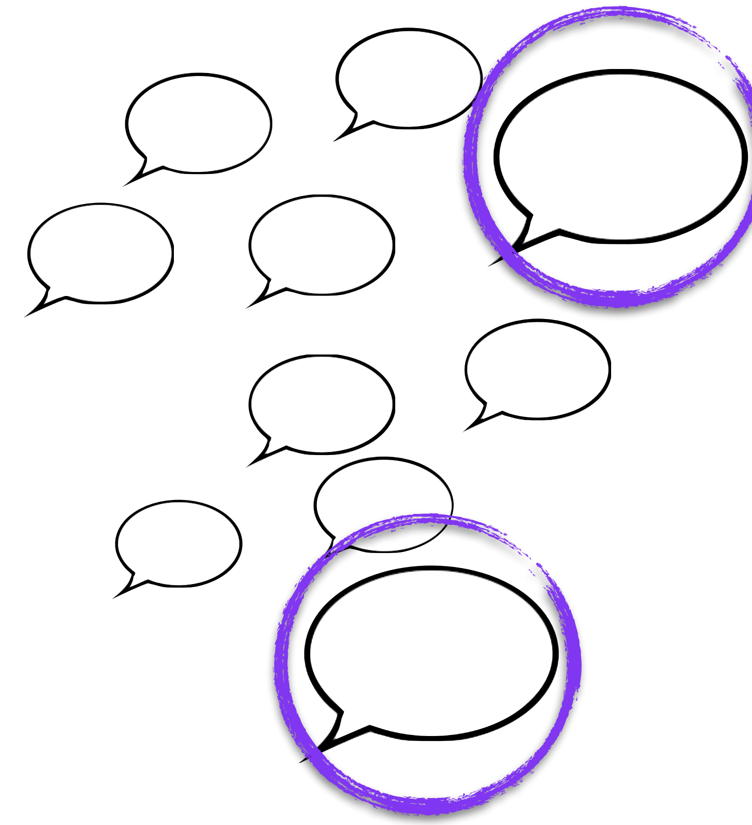
Variational learning



Striking evidence that this is true

Table 1: The qualitative fit Yang discovered between the unambiguous data advantage (Adv) perceived by a VarLearner in its acquisitional intake and the observed age of acquisition (AoA) in children for six parameter values across different languages.

Param Value	Language	Unambiguous Form	Unambiguous Ex	Adv	AoA
+*wh*-fronting	English	*wh*-fronting in questions	*Who did you see?*	25%	<1;8
+topic-drop	Chinese	null objects	*Wǒ méi kànjiàn* *I not see* "I didn't see (him)"	12%	<1;8
+pro-drop	Italian	null subjects in questions	*Chi hai visto* *who have seen* "Who have you seen?"	10%	<1;8
+verb-raising	French	*Verb Adverb*	*Jean voit souvent Marie* *Jean sees often Marie* "Jean often sees Marie"	7%	1;8
-pro-drop	English	expletive subjects	There's a penguin on the ice.	1.2%	3;0
+verb-second	German Dutch	*Object Verb Subject*	*Pinguine liebe ich.* *penguins like I* "I like penguins"	1.2%	3;0-3;2
-scope-marking	English	long-distance *wh* questions without medial-*wh*	*Who do you think is on the ice?*	0.2%	>4.0

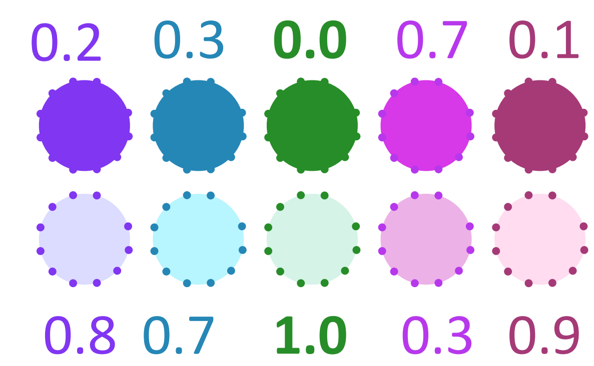


The more unambiguous data there are for one value over another (its advantage)…

Learning with parameters

The learning algorithm

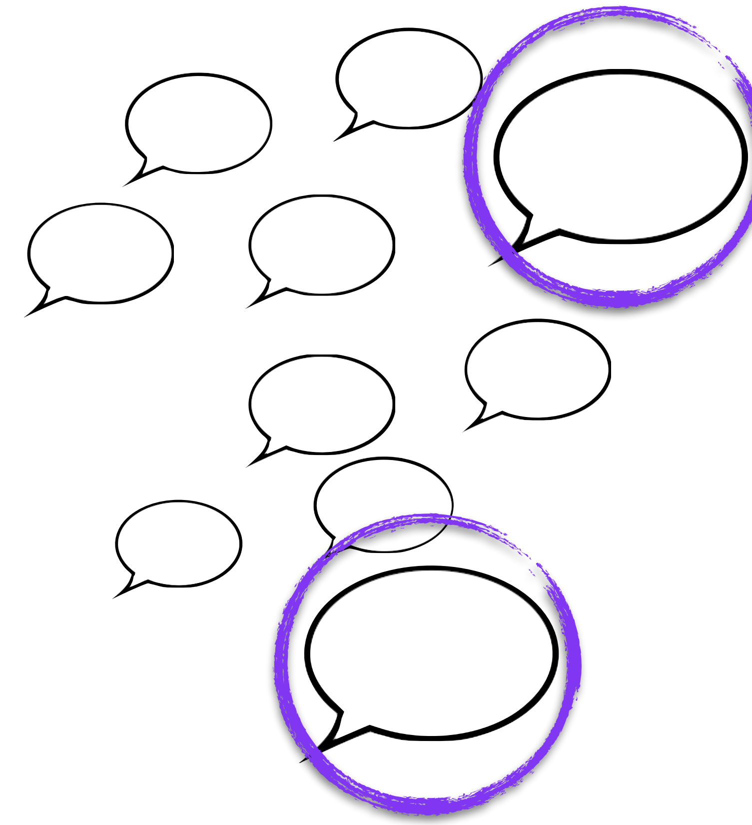
Variational learning



Striking evidence that this is true

Table 1: The qualitative fit Yang discovered between the unambiguous data advantage (Adv) perceived by a VarLearner in its acquisitional intake and the observed age of acquisition (AoA) in children for six parameter values across different languages.

Param Value	Language	Unambiguous Form	Unambiguous Ex	Adv	AoA
+*wh*-fronting	English	*wh*-fronting in questions	*Who did you see?*	25%	<1;8
+topic-drop	Chinese	null objects	*Wǒ méi kànjiàn* *I not see* "I didn't see (him)"	12%	<1;8
+pro-drop	Italian	null subjects in questions	*Chi hai visto* *who have seen* "Who have you seen?"	10%	<1;8
+verb-raising	French	*Verb Adverb*	*Jean voit souvent Marie* *Jean sees often Marie* "Jean often sees Marie"	7%	1;8
-pro-drop	English	expletive subjects	There's a penguin on the ice.	1.2%	3;0
+verb-second	German Dutch	*Object Verb Subject*	*Pinguine liebe ich.* *penguins like I* "I like penguins"	1.2%	3;0-3;2
-scope-marking	English	long-distance *wh* questions without medial-*wh*	*Who do you think is on the ice?*	0.2%	>4.0



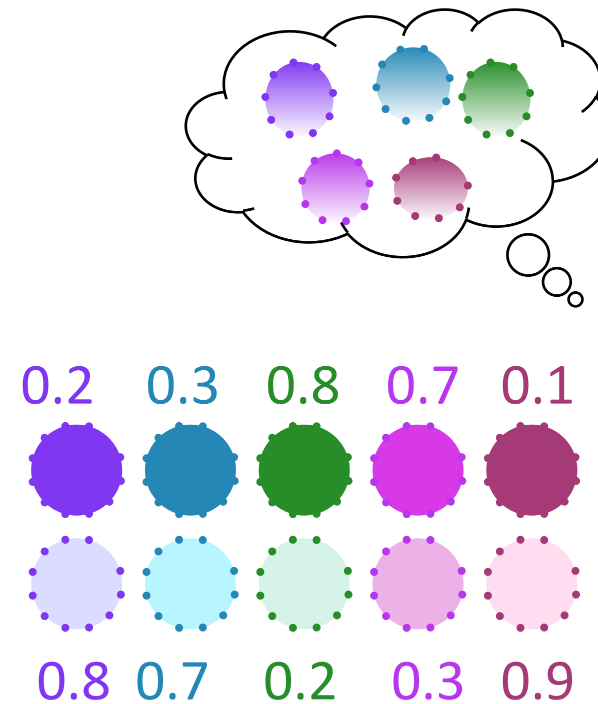
The more unambiguous data there are for one value over another (its advantage), the earlier it seems to be learned.

Thank you!

Computation of
Language
Laboratory
UC Irvine

Lisa S. Pearl
Associate Professor
Department of Linguistics
Department of Cognitive Sciences
SSPB 2219, SBSG 2314
University of California, Irvine
lpearl@uci.edu

This work was supported in part by NSF grants BCS-0843896 and BCS-1347028.




another one



Who does… is pretty?