Ling 151/Psych 156A: Acquisition of Language II

Lecture 22

Structure III

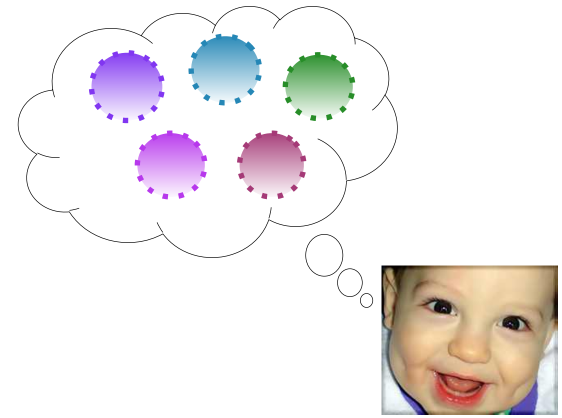
Announcements

Be working on HW8 and the structure review questions

Final review this Friday 3/16/18

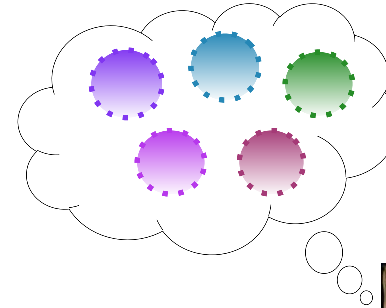
Consider taking more language science classes in the future!

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.

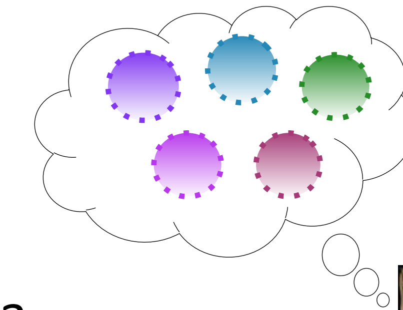
This can make it hard to figure out what parameter values might have produced the observable data - even if the child already knows what the parameters are.

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*

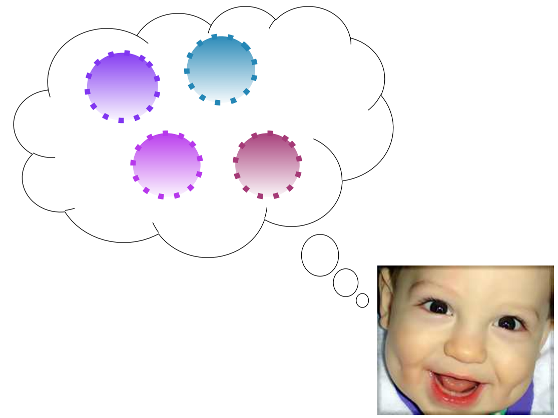
Kannada

Subject *Object* Verb Object

English

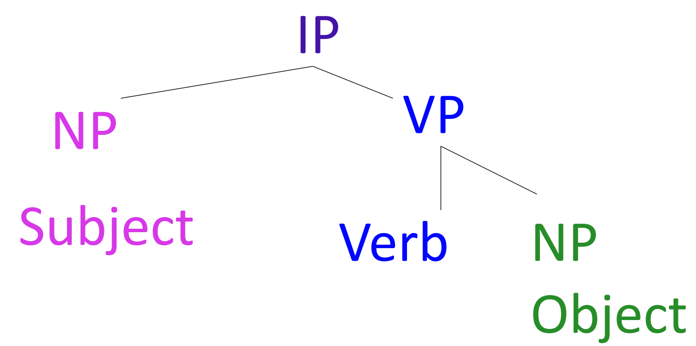
Subject Verb Object

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

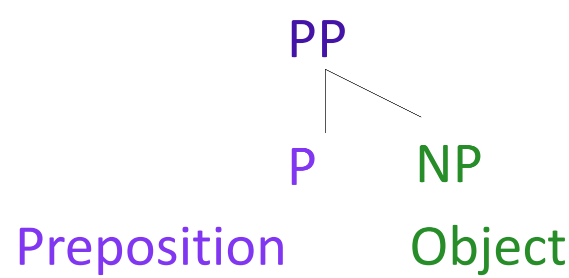


Basic word order:

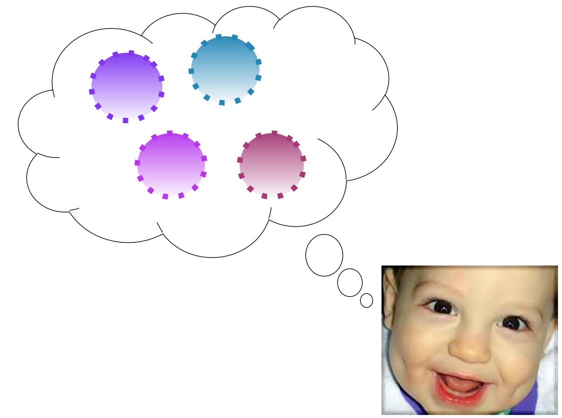
Subject Verb Object [SVO]

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]

IP

NP VP

Subject

NP Verb

Object

Postpositions:

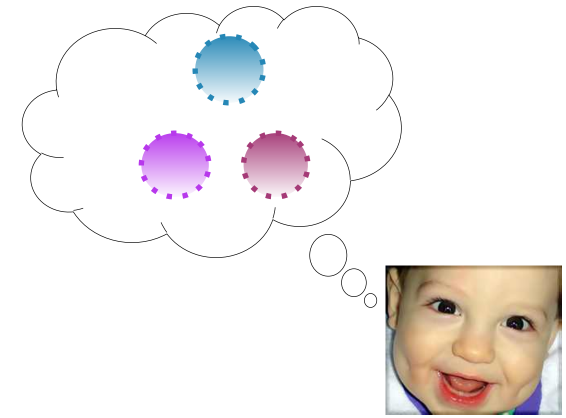
Noun Phrase Postposition

PP

NP P

Object Postposition

Interacting parameters



Example Parameter 1: Head-directionality

Edo/English: Head-first

Japanese/Navajo: Head-final

Example Parameter 2: Verb Second (V2)

German: +V2

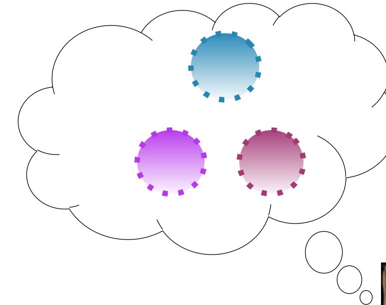
Verb moves to second phrasal position, some other phrase moves to the first position

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

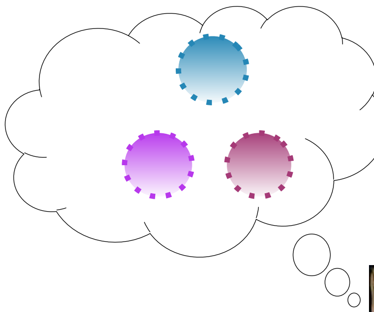
Verb moves to second phrasal position, some other phrase moves to the first position

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

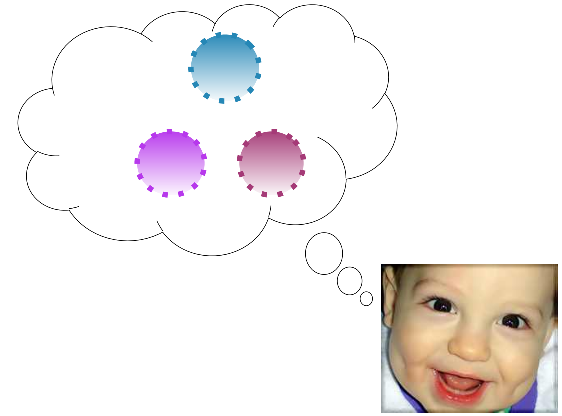
Verb moves to second phrasal position, some other phrase moves to the first position

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

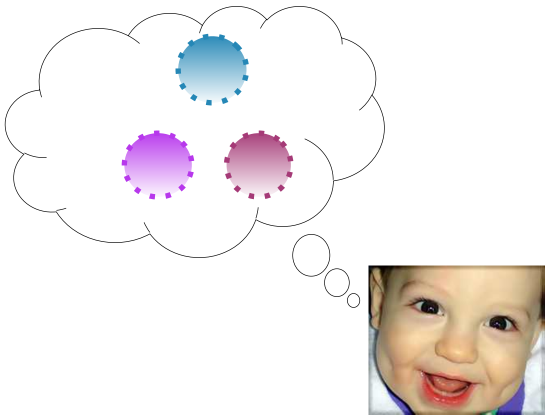
Verb moves to second phrasal position, some other phrase moves to the first position

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

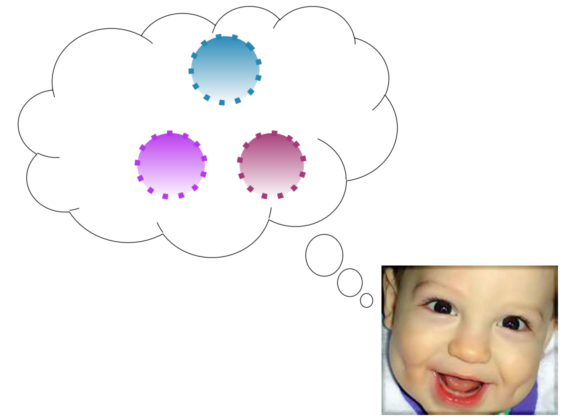
Verb doesn't move.

Sarah reads the book

Underlying form of the sentence

Observable form of the sentence

Interacting parameters



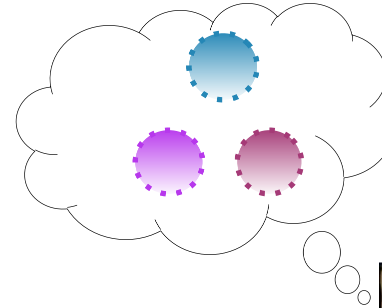
Head-directionality

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)

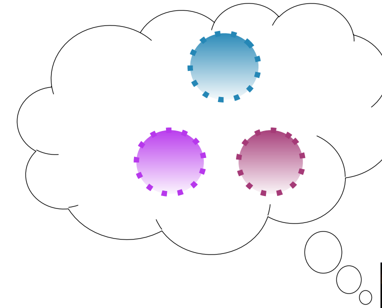
“I love kitties.”

Data point

Subject **Verb** **Object**

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

Interacting parameters



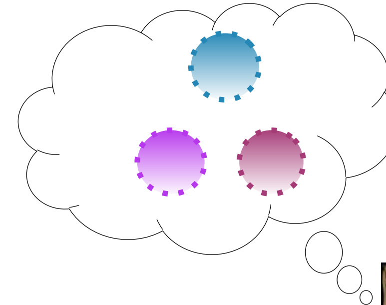
Head-directionality Verb Second (V2)

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

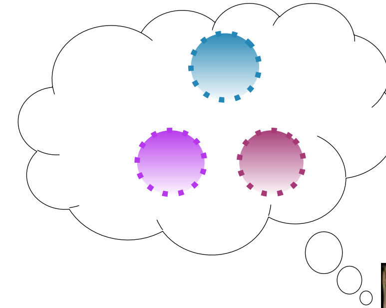
✔ +V2 predicts Verb moved to second position

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

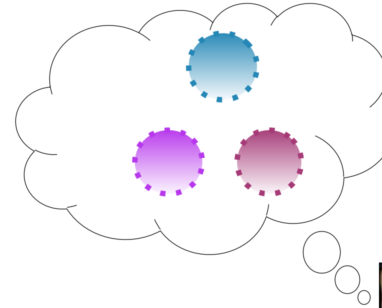
✔ +V2 predicts Verb moved to second position

✔ **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

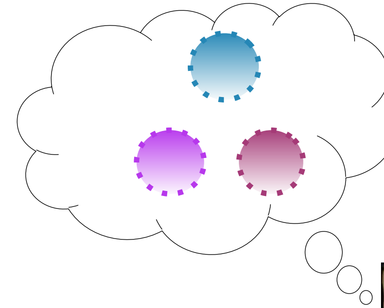
✔ -V2 predicts Verb doesn't move

✔ **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

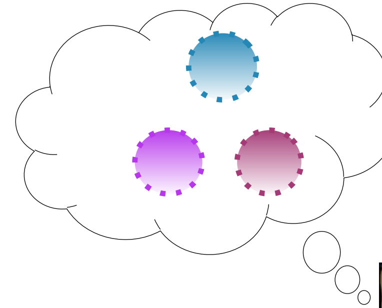
✓ -V2 predicts Verb doesn't move

✓ **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

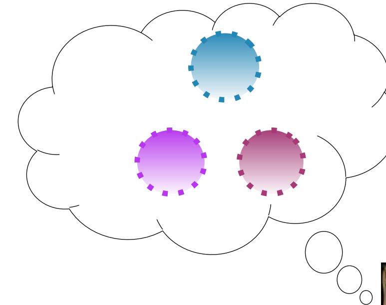
G2 Head-final +V2

G3 Head-first -V2

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

G4 Head-final -V2

Interacting parameters



"I love kitties."

Head-directionality — Verb Second (V2)

Subject **Verb** **Object**

G1 Head-first +V2

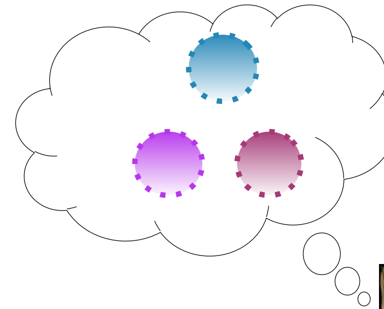
G2 Head-final +V2

G3 Head-first -V2

We don't know whether the true grammar is 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**

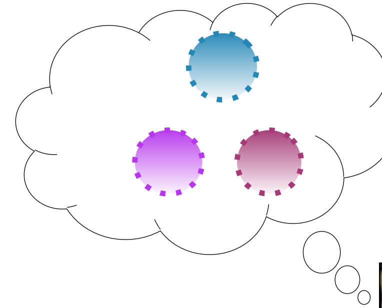
✔ **G1** Head-first +V2	✔ **G2** Head-final +V2	✔ **G3** Head-first -V2

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)

"I love kitties."

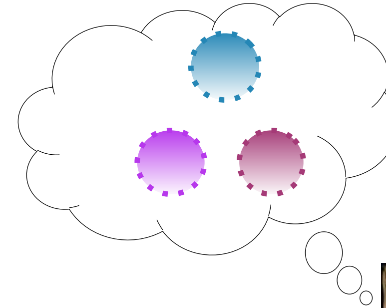
Subject Verb Object

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.

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

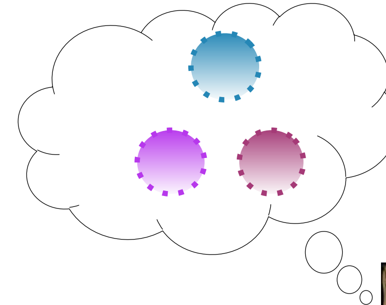
✔ **G3** Head-first -V2

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

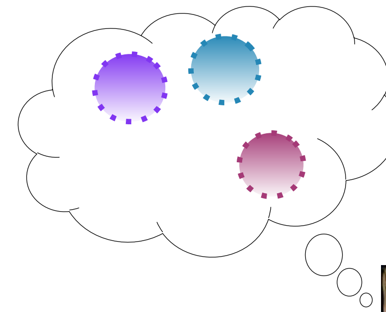
G2 Head-final +V2

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

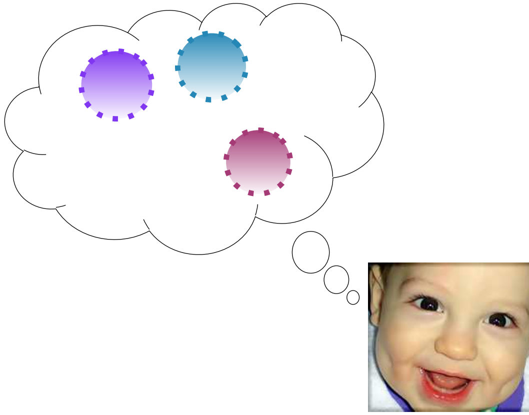
Allows Subject to be overt or dropped

✔ 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

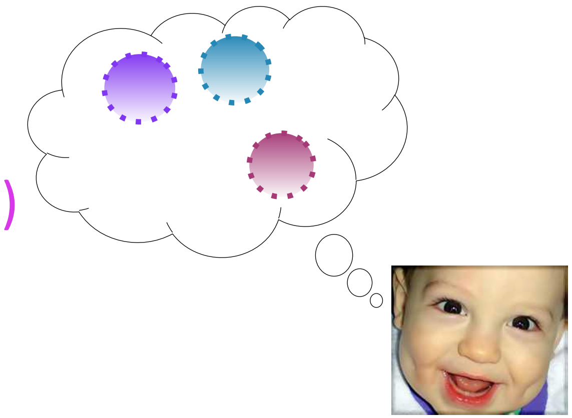
Subject must be overt

✔ They drink

✗ Drink

“They drink”

Interacting parameters

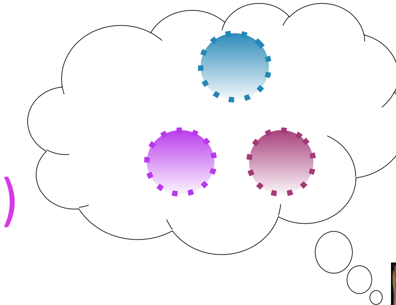


Head-directionality Subject drop (subj-drop)

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?

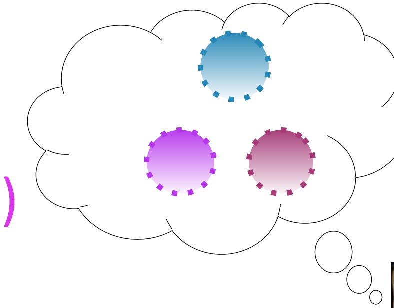
G1 Head-first
+subj-drop

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

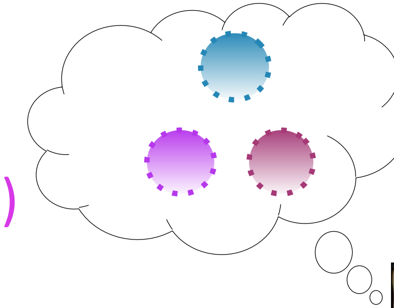
G1 Head-first
+subj-drop

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

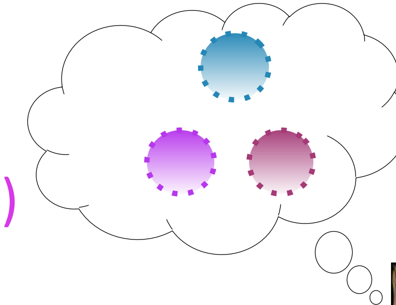
G2 Head-final +subj-drop

G1 Head-first +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 requires subject to be overt

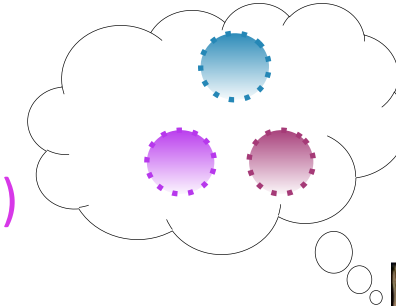
G3 Head-first -subj-drop

✓ **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

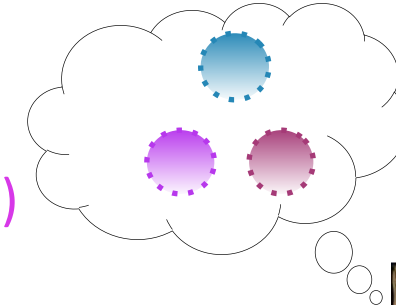
G4 Head-final -subj-drop

✔ **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

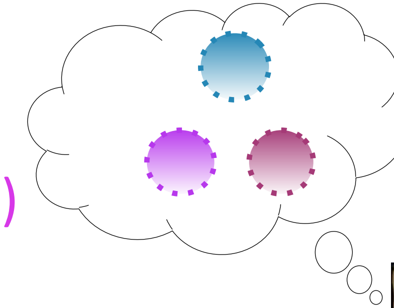
✔ **G4** Head-final -subj-drop

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

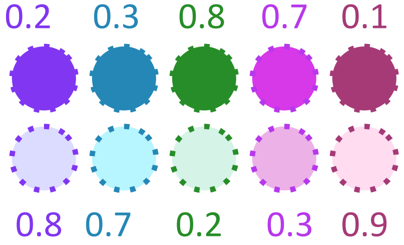
G4	Head-final	-subj-drop

So what can we do?

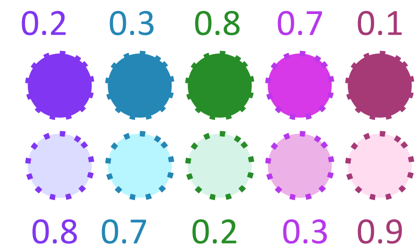
G1	Head-first	+subj-drop

G3	Head-first	-subj-drop

Learning with parameters



Learning with parameters

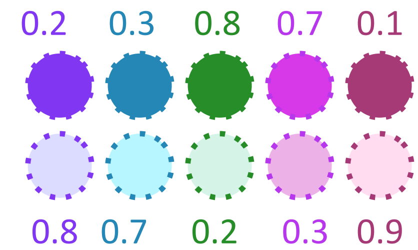


A language's grammar = combination of parameter values

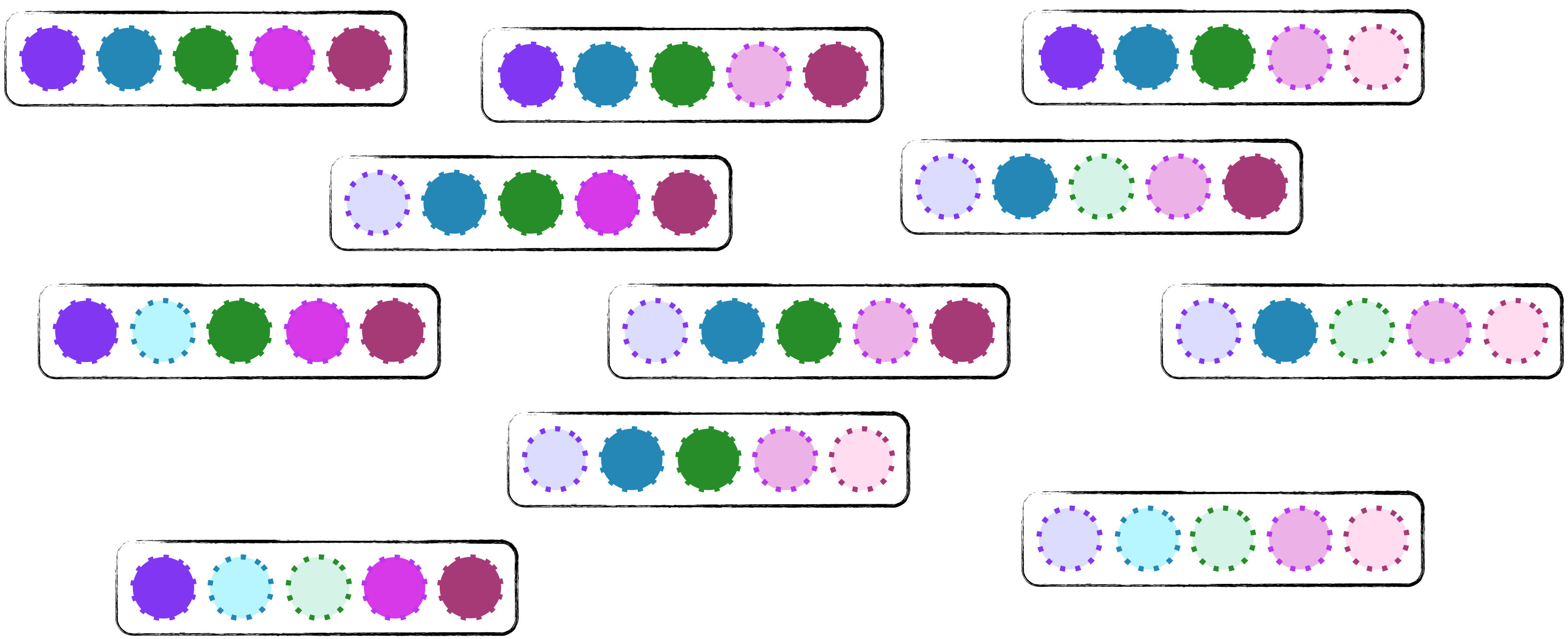
G2
Head-final
+subj-drop

G4
Head-final
-subj-drop

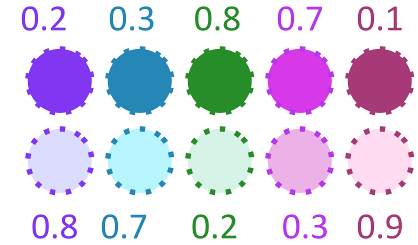
Learning with parameters

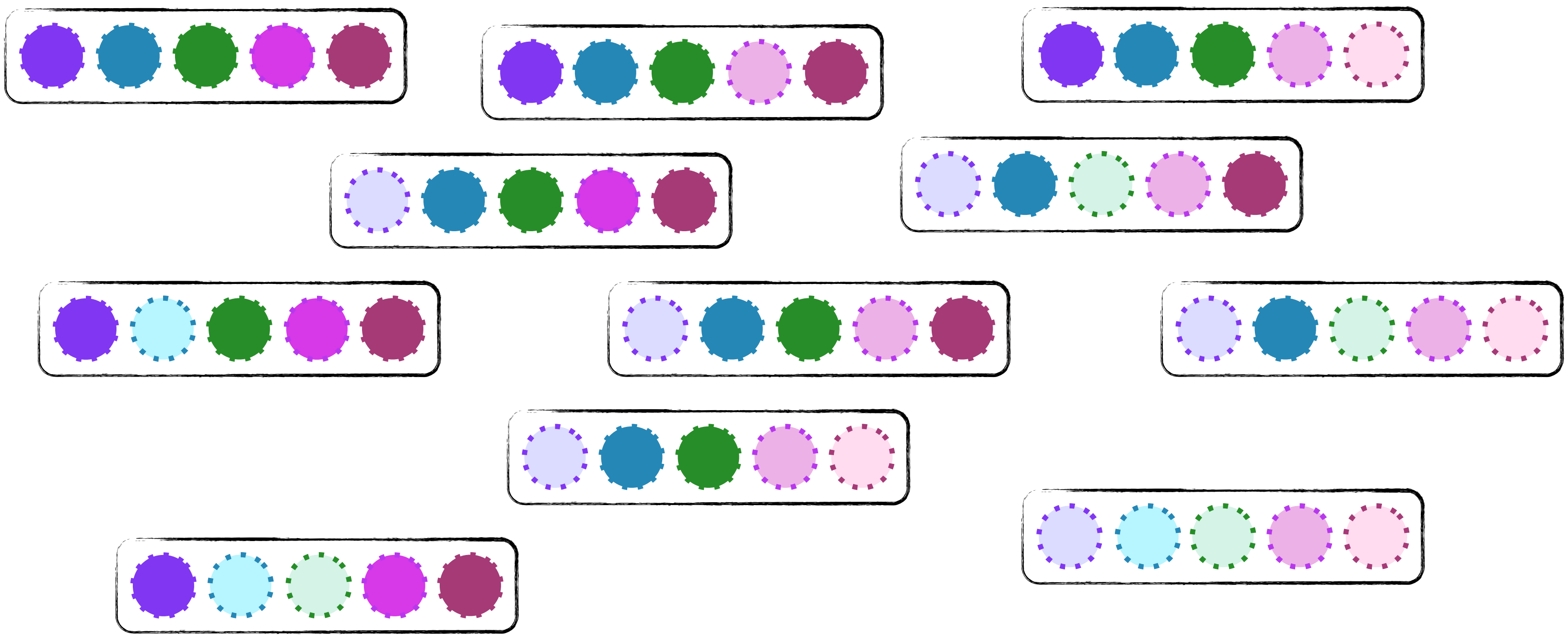


A language's grammar = combination of parameter values



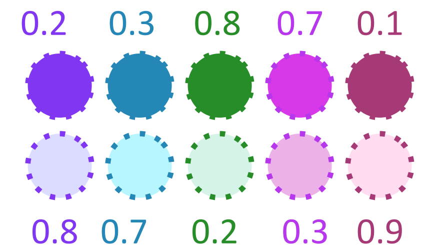
Learning with parameters





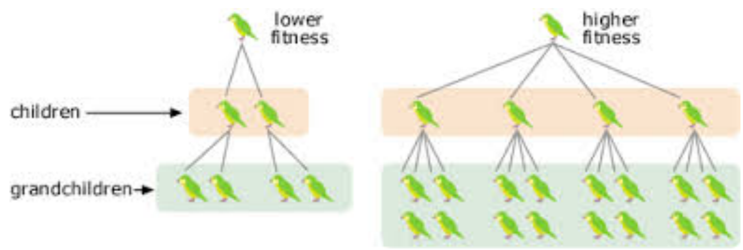
Variational learning (Yang 2002, 2004, 2012): use reinforcement learning to learn which value (for each parameter) that the native language uses for its grammar. This is a combination of using linguistic knowledge & statistical learning.

Learning with parameters



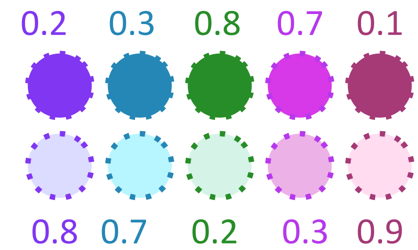
Variational learning

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?

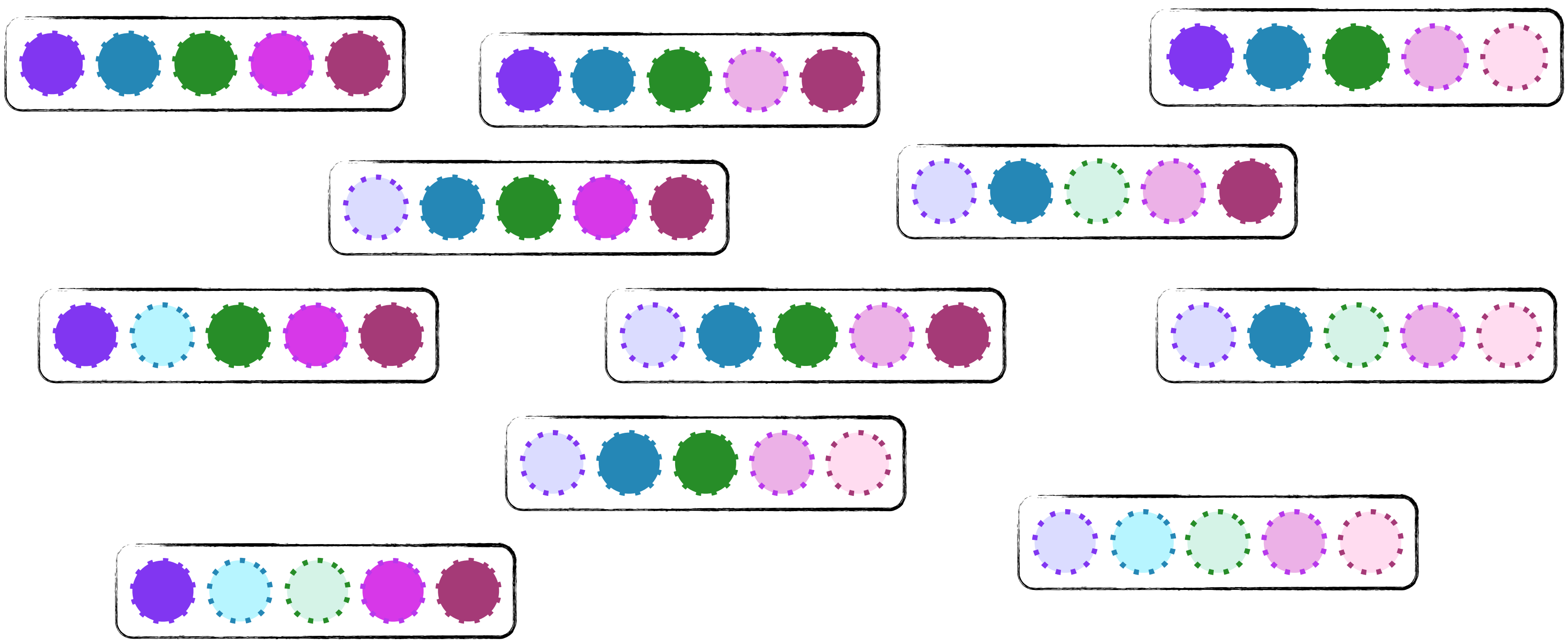
Learning with parameters



Variational learning

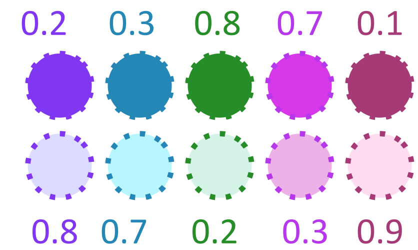
The fittest **individuals** survive while the others die out.

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



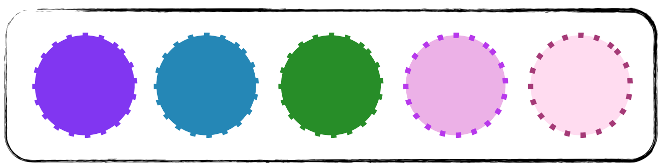
Learning with parameters

Variational learning

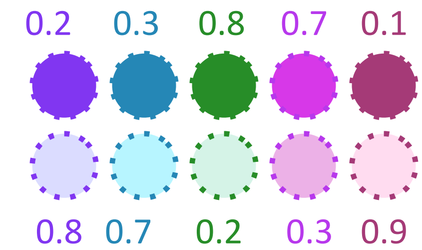


The **fittest** individuals survive while the others die out.

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

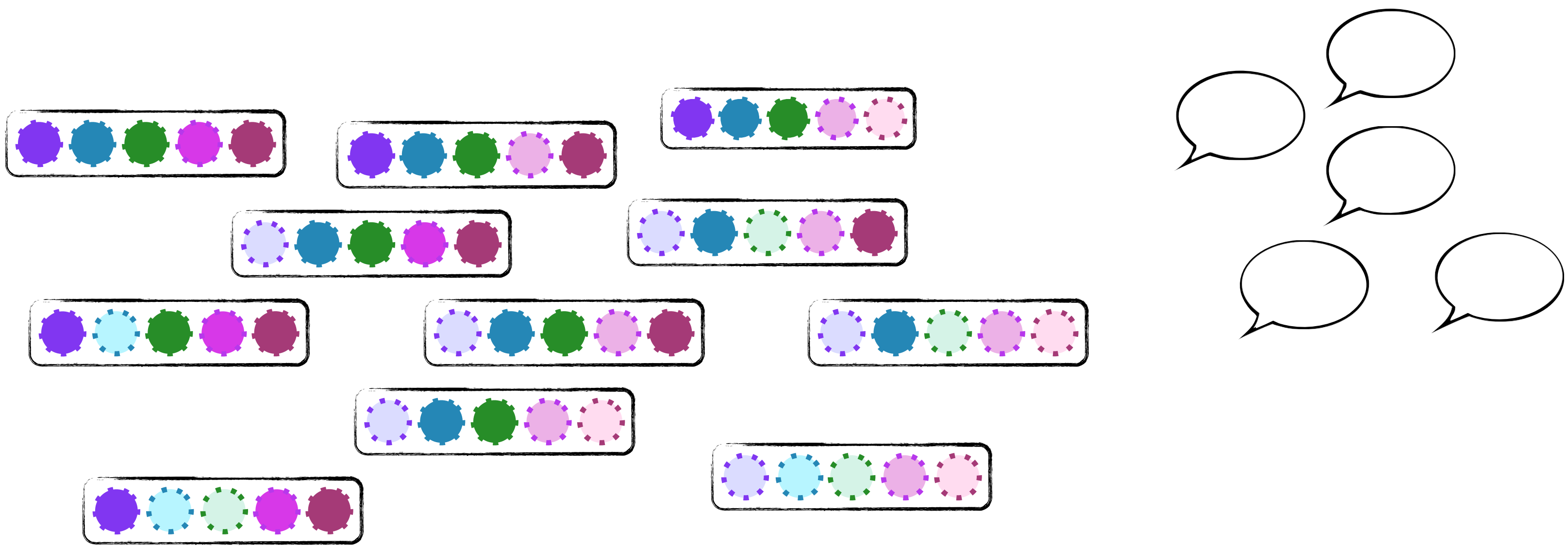


Learning with parameters



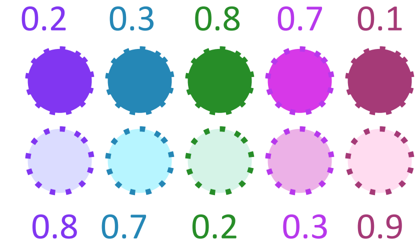
Variational learning

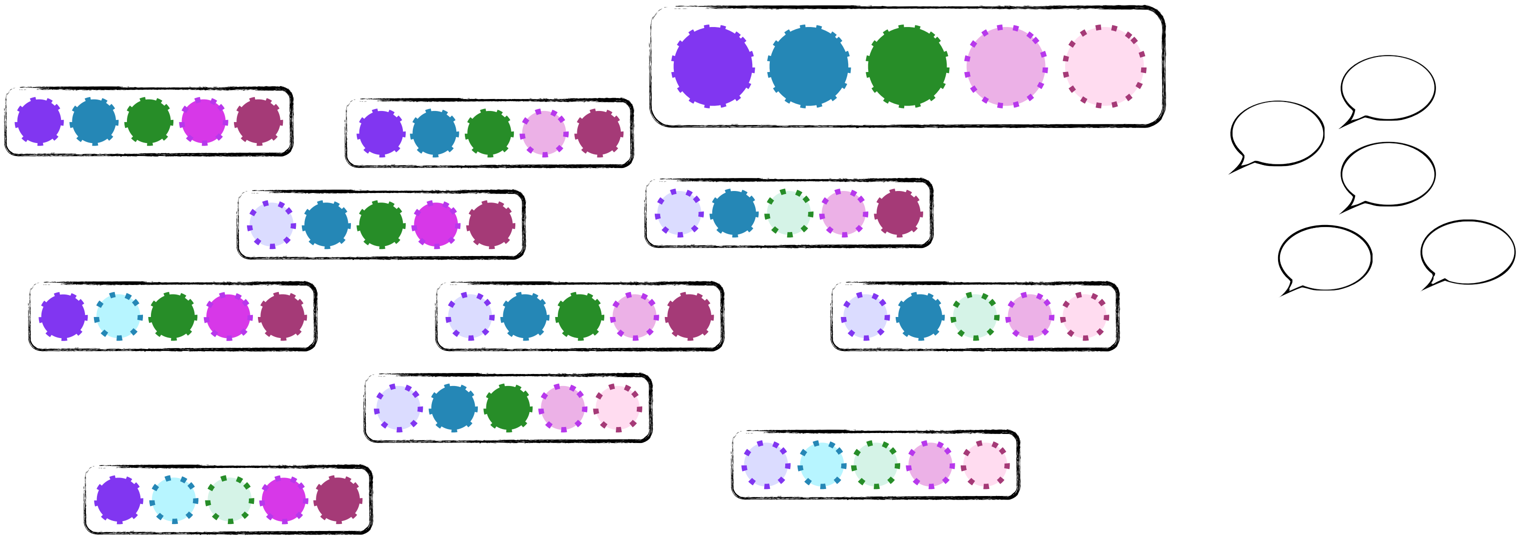
A child's mind consists of a population of grammars that are competing to analyze the data in the child's native language.



Learning with parameters

Variational learning

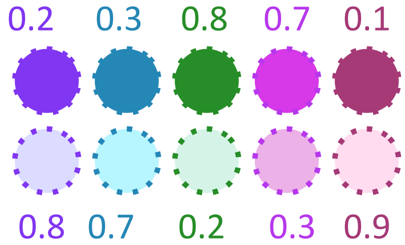


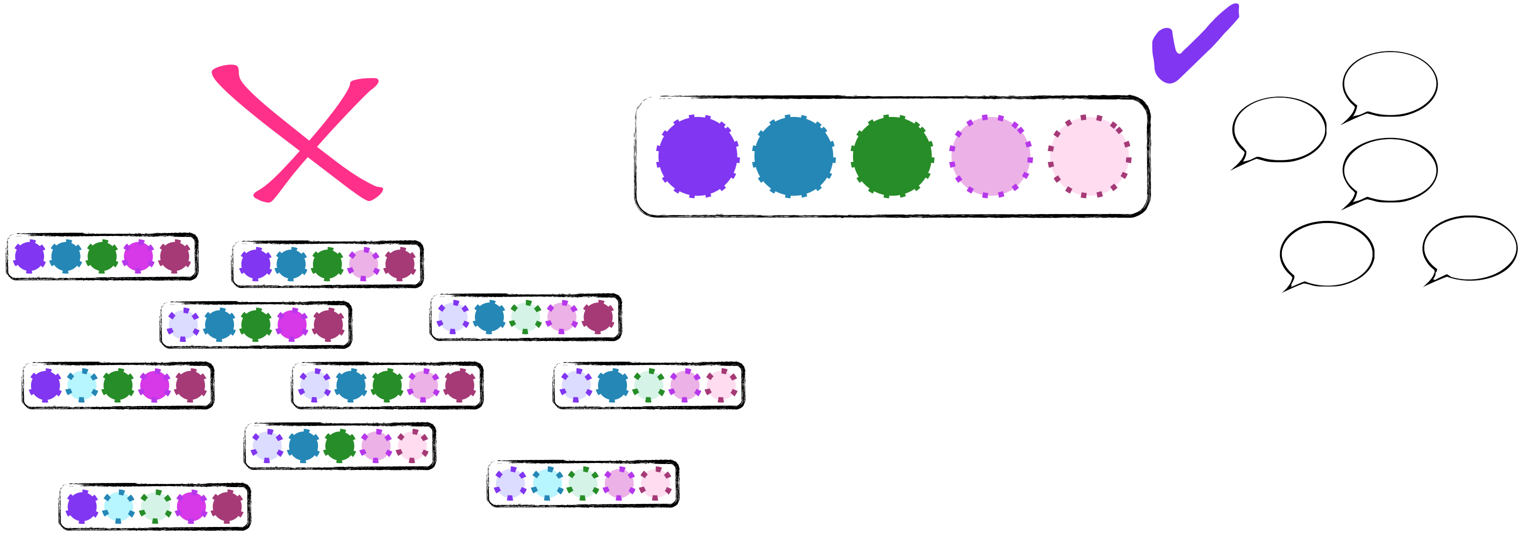


Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will “win”, once the child encounters enough native language data. This is because none of the other competing grammars can analyze all the data.

Learning with parameters

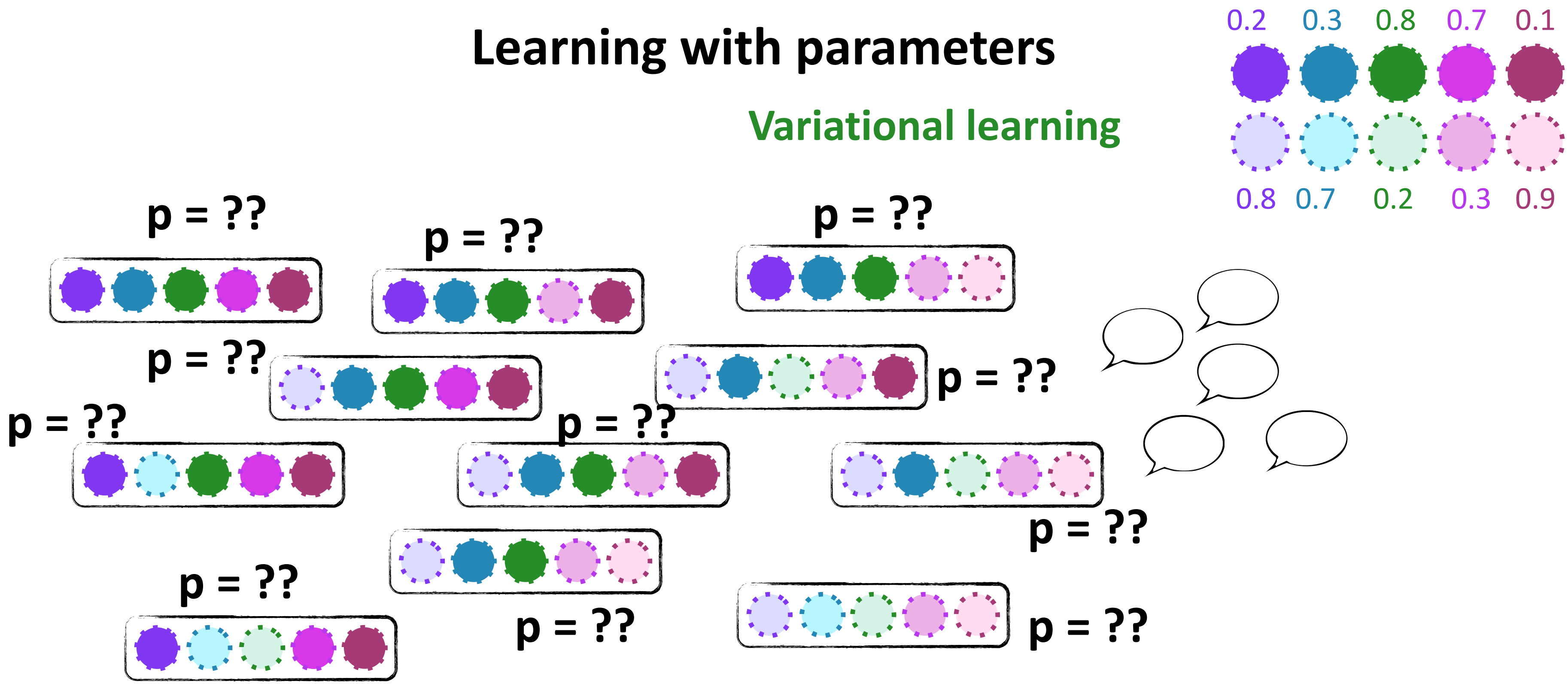
Variational learning





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

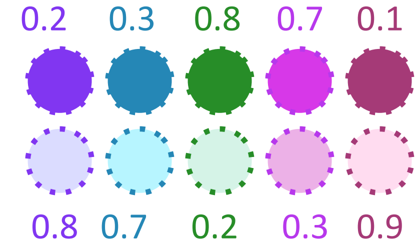
Learning with parameters

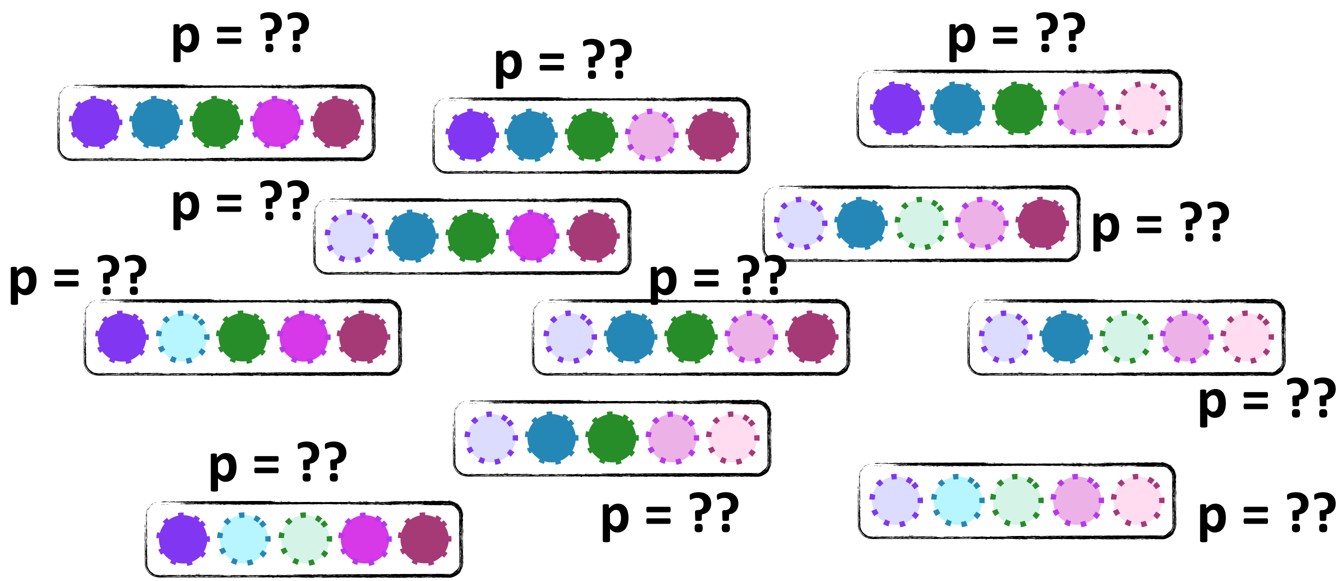


At any point in time, a grammar in the population will have a **probability** associated with it. This represents the child's belief that this grammar is the correct grammar for the native language.

Learning with parameters

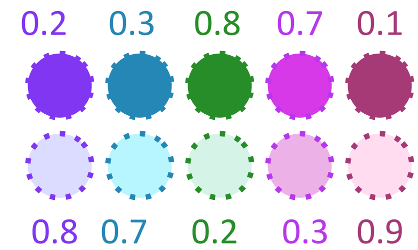
Variational learning



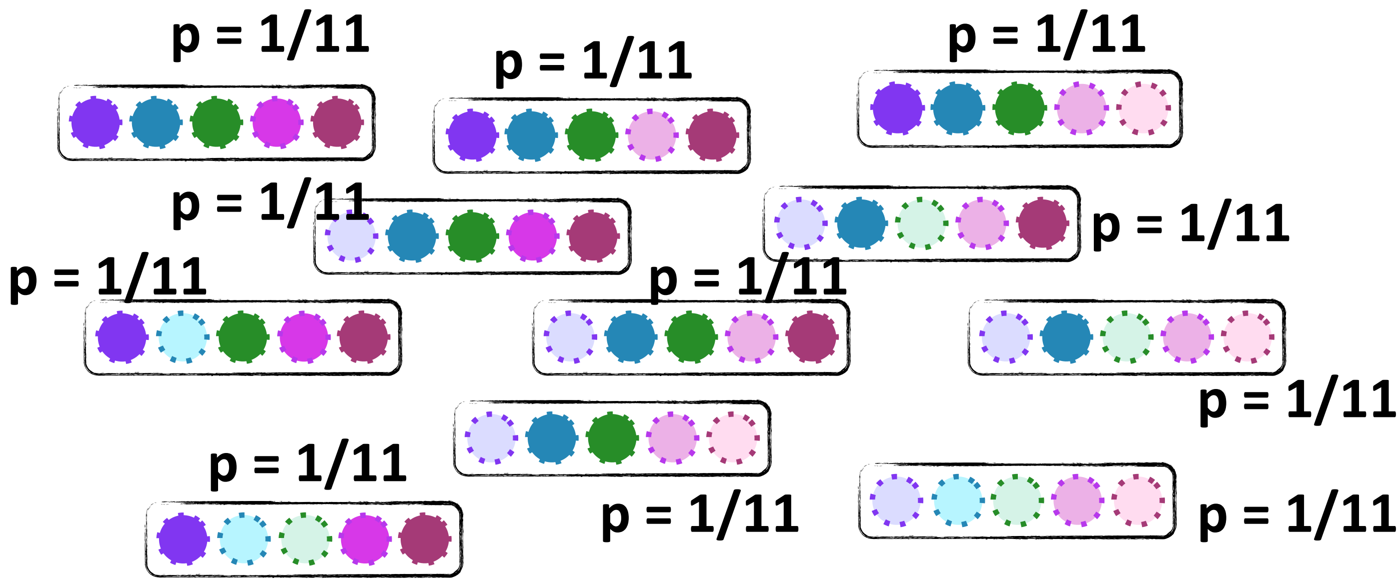


Before the child has encountered any native language data, all grammars are **equally likely**. So, initially all grammars have the same probability, which is 1 divided the number of grammars available.

Learning with parameters

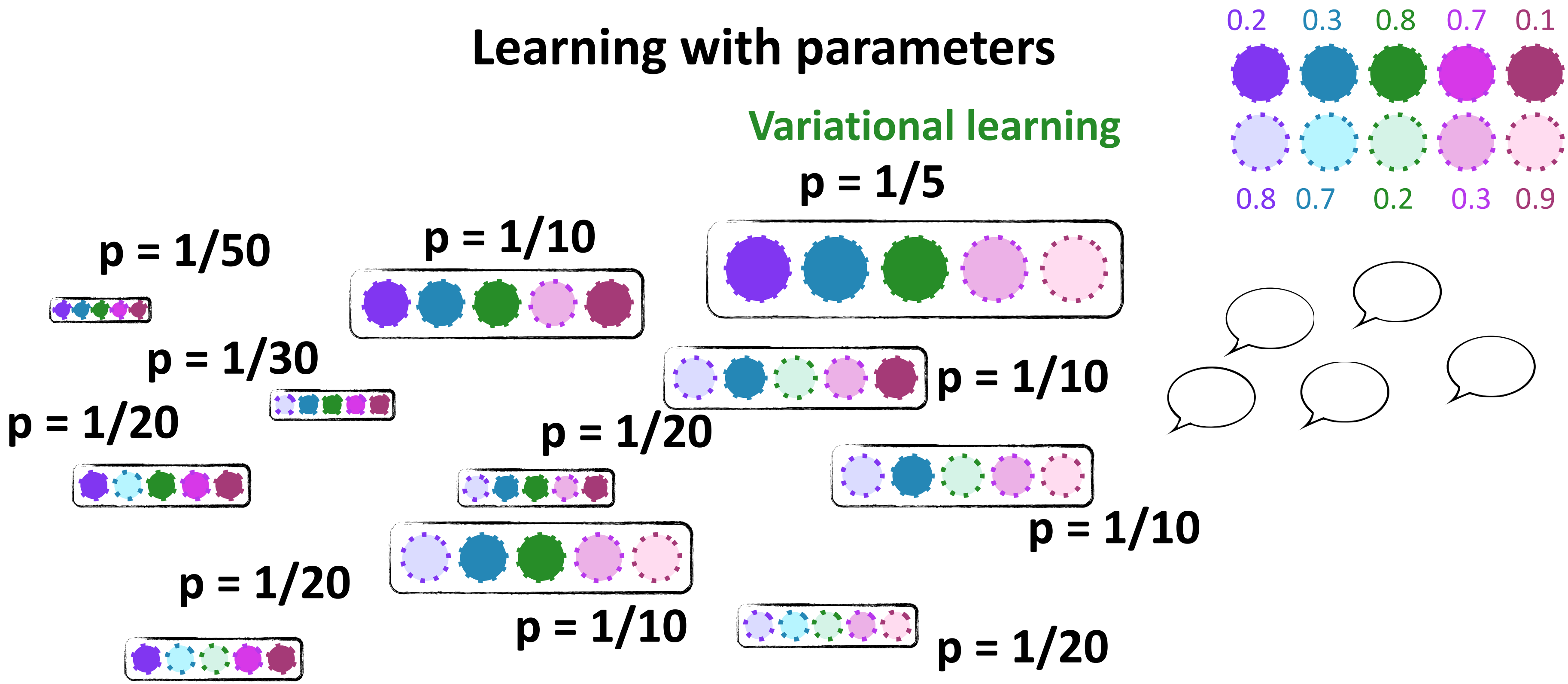


Variational learning



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

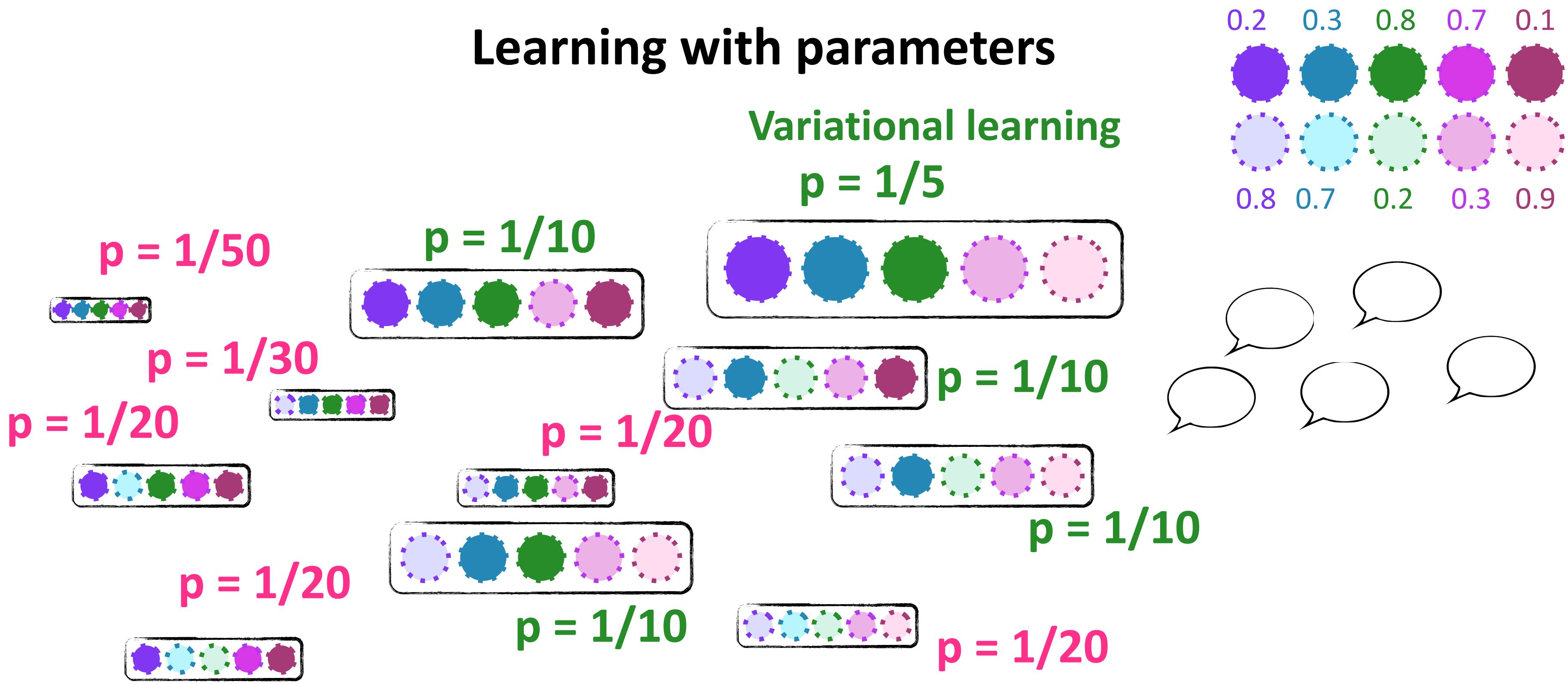
Learning with parameters



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.

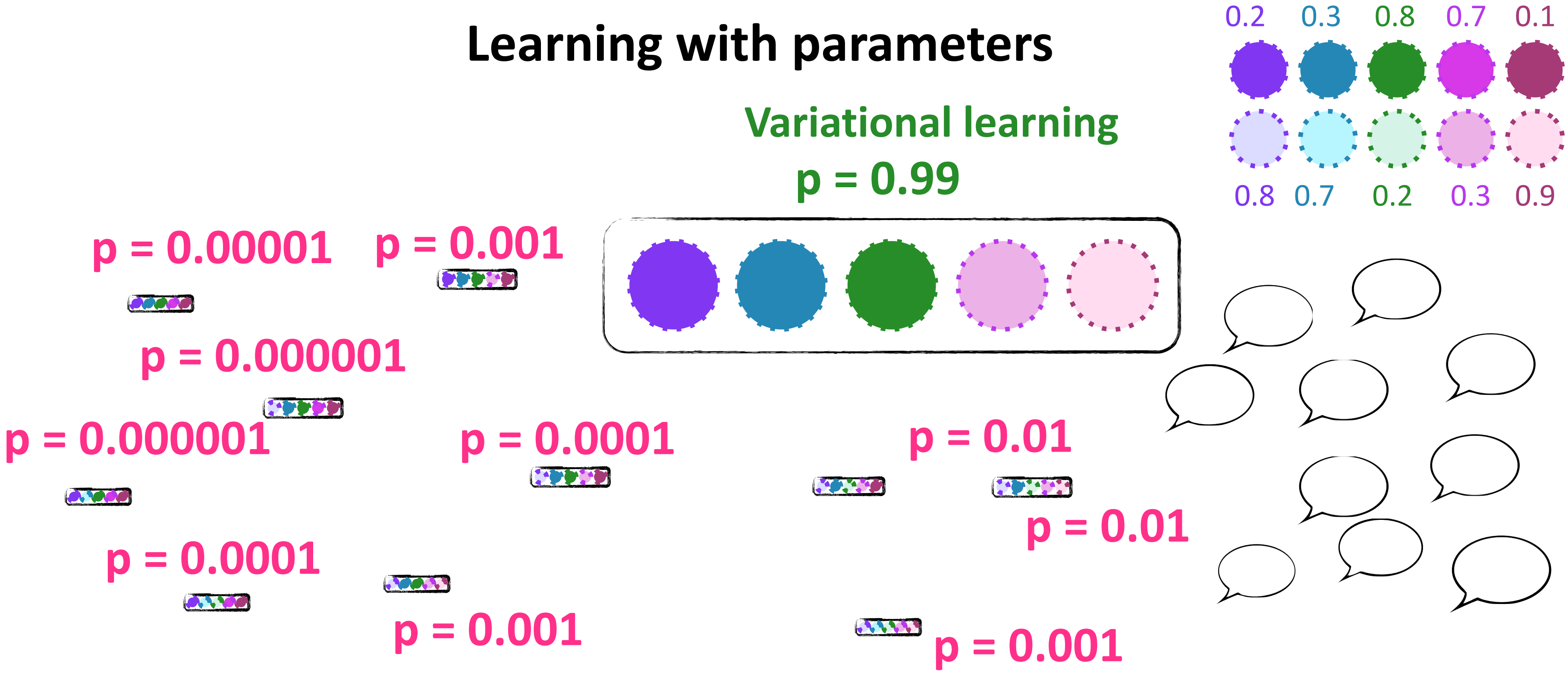
Other grammars will be less fit because they cannot account for some of the data encountered.

Learning with parameters



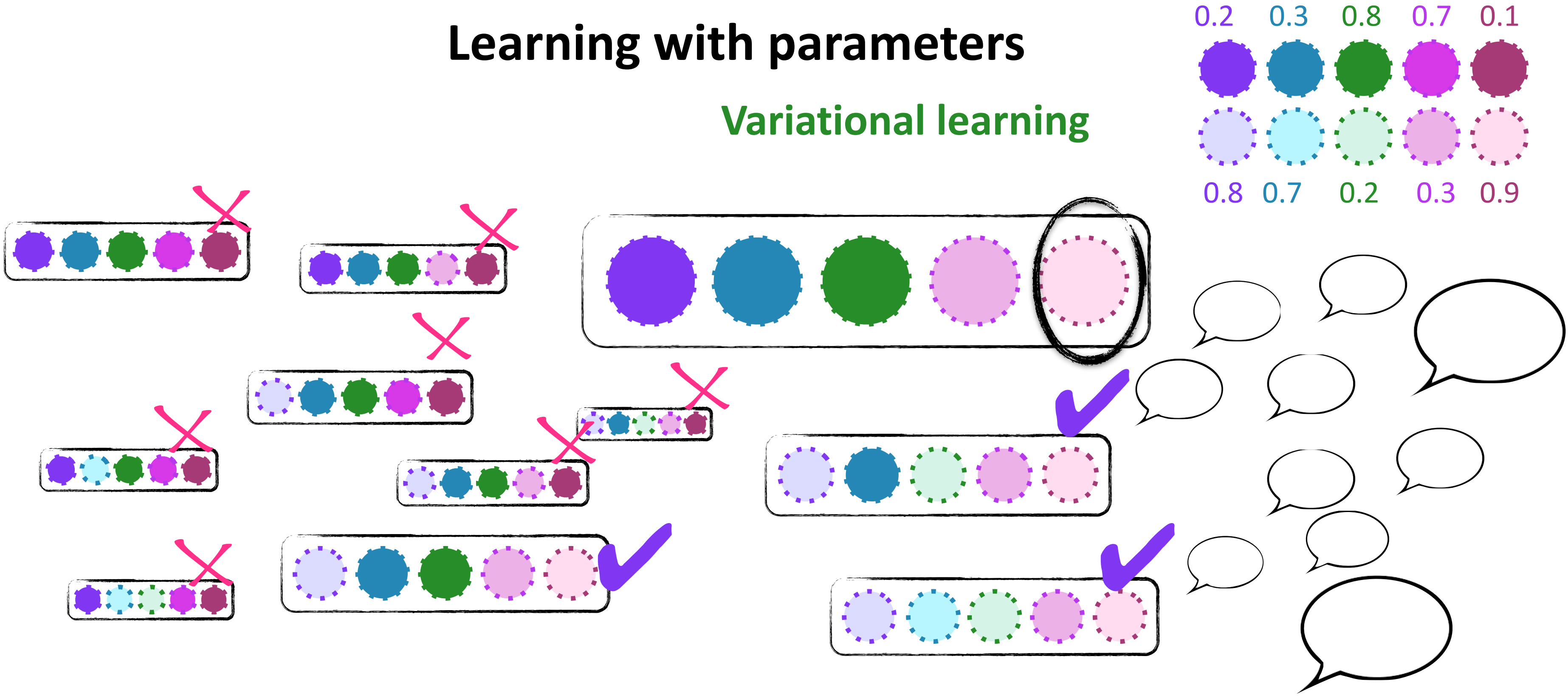
Grammars that are more compatible with the native language data intake will have their **probabilities increased** while grammars that are less compatible will have their **probabilities decreased** over time.

Learning with parameters



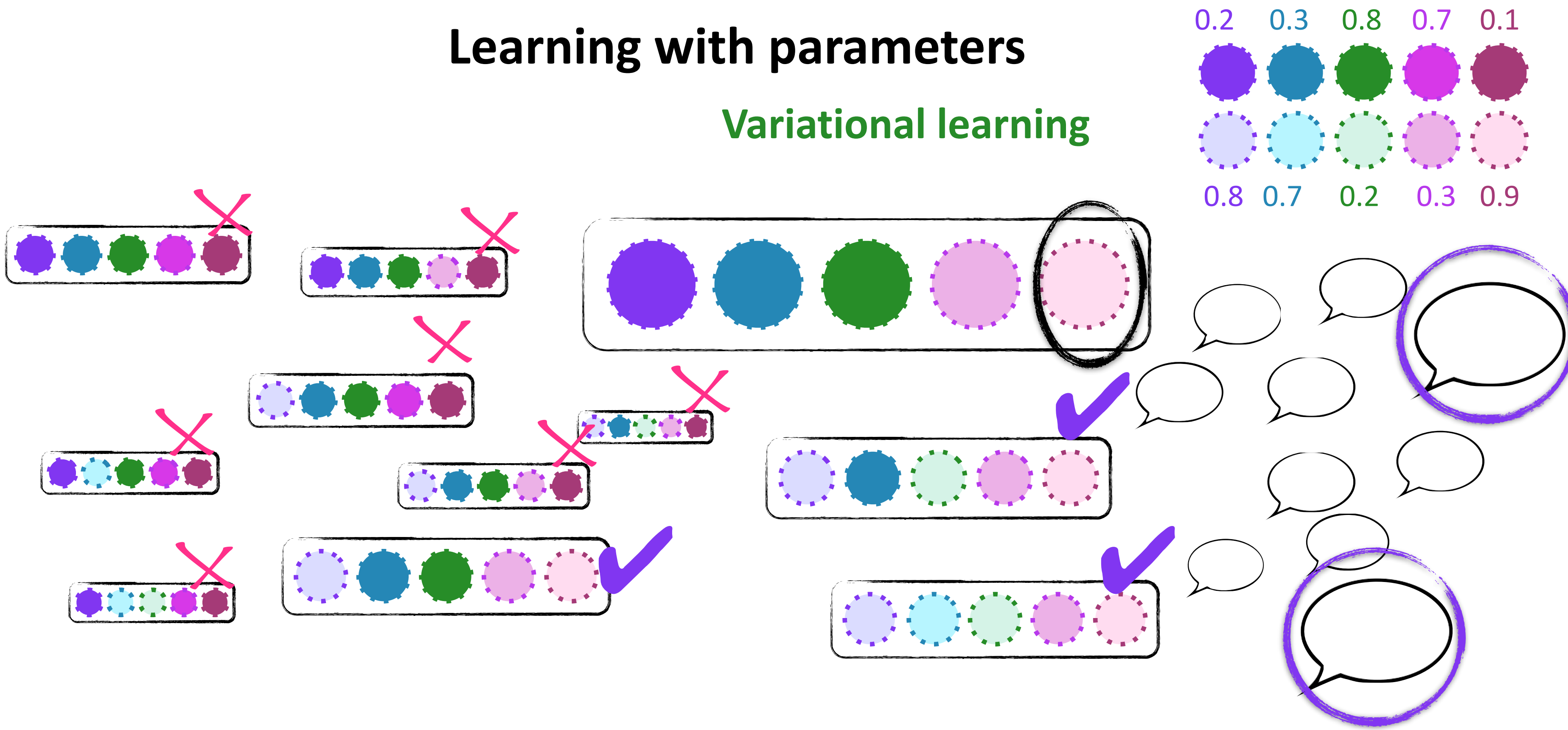
After the child has encountered enough data from the native language, the native language grammar should have a probability near 1.0 while the other grammars have a probability near 0.0.

Learning with parameters



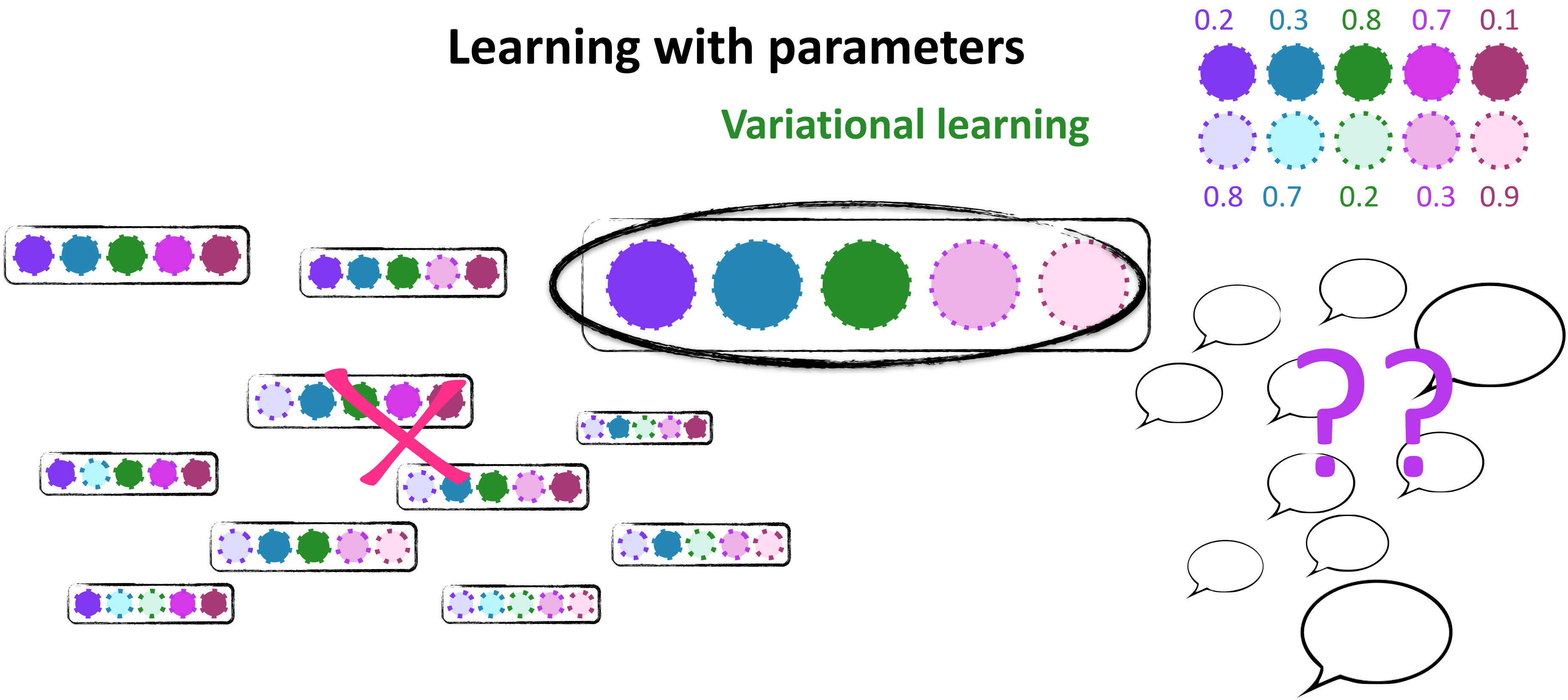
The power of **unambiguous data**:
Unambiguous data from the native language can only be analyzed by grammars that use the **native language's parameter value**.

Learning with parameters



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.

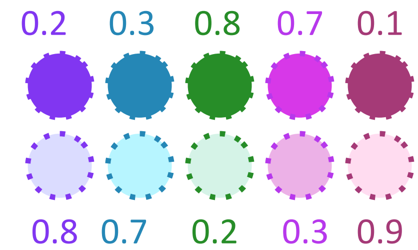
Learning with parameters



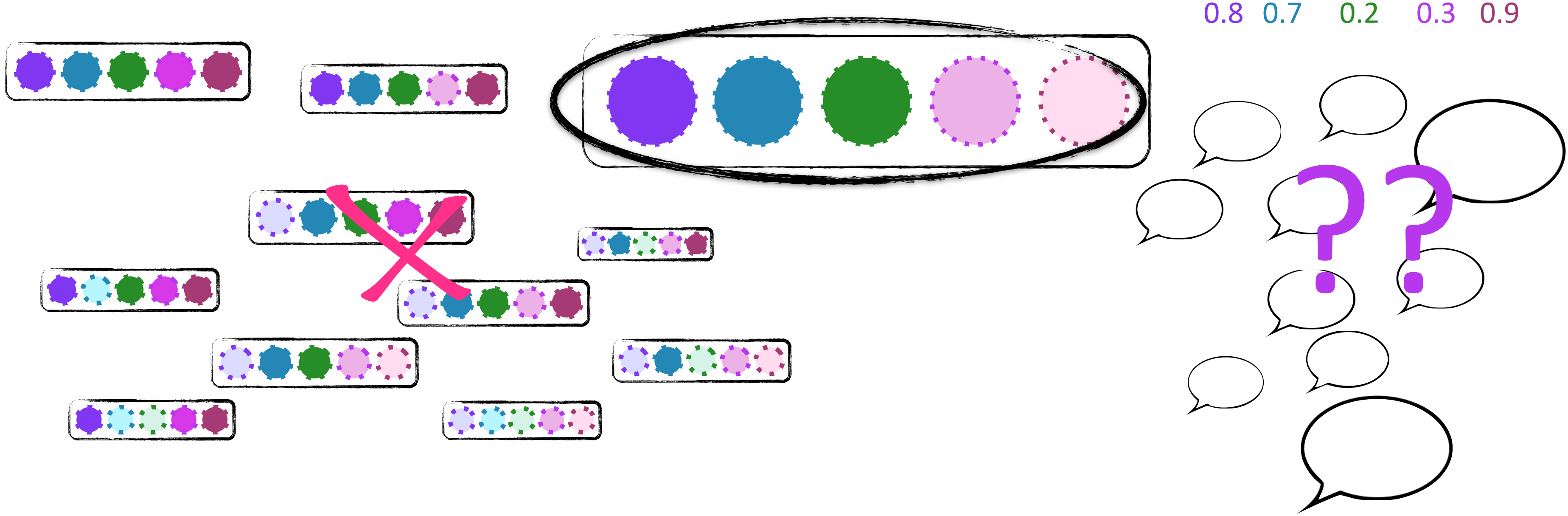
Problem: Do unambiguous data exist for entire grammars?

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

Learning with parameters

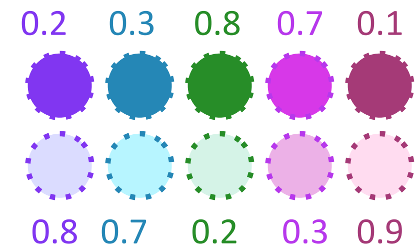


Variational learning

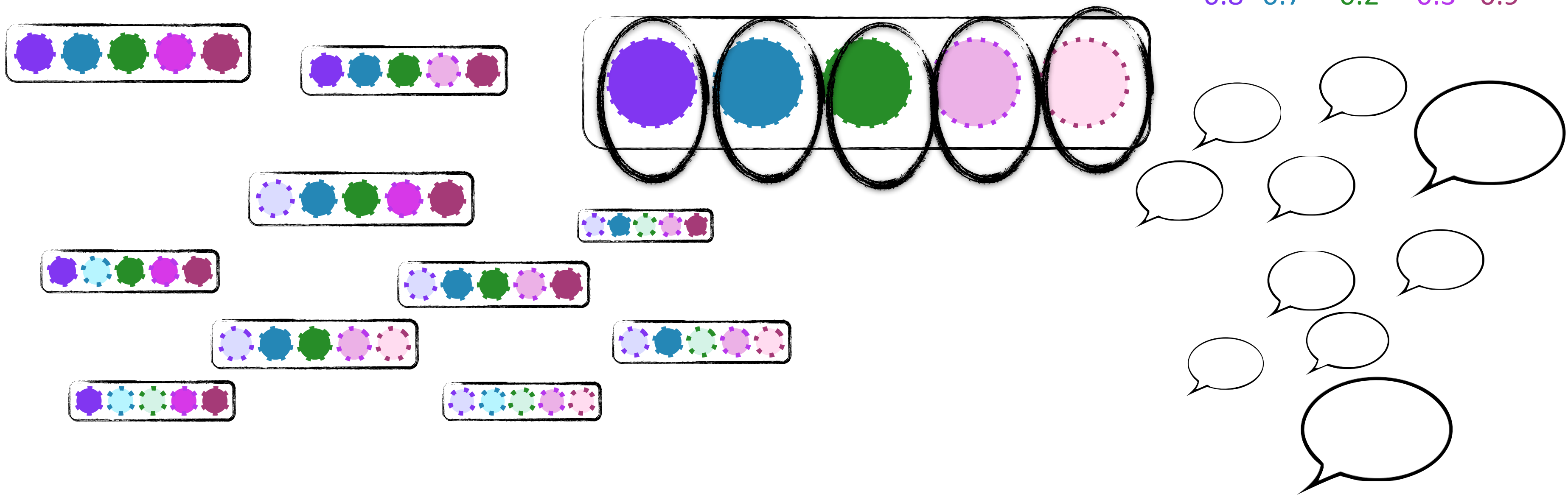


This seems unlikely for real language data because linguistic parameters connect with different types of patterns, which may have nothing to do with each other, or parameters may interact with each other.

Learning with parameters

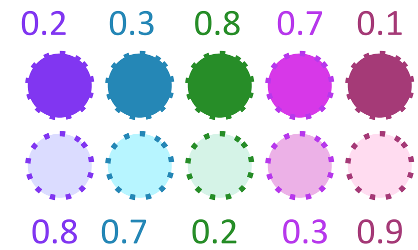


Variational learning

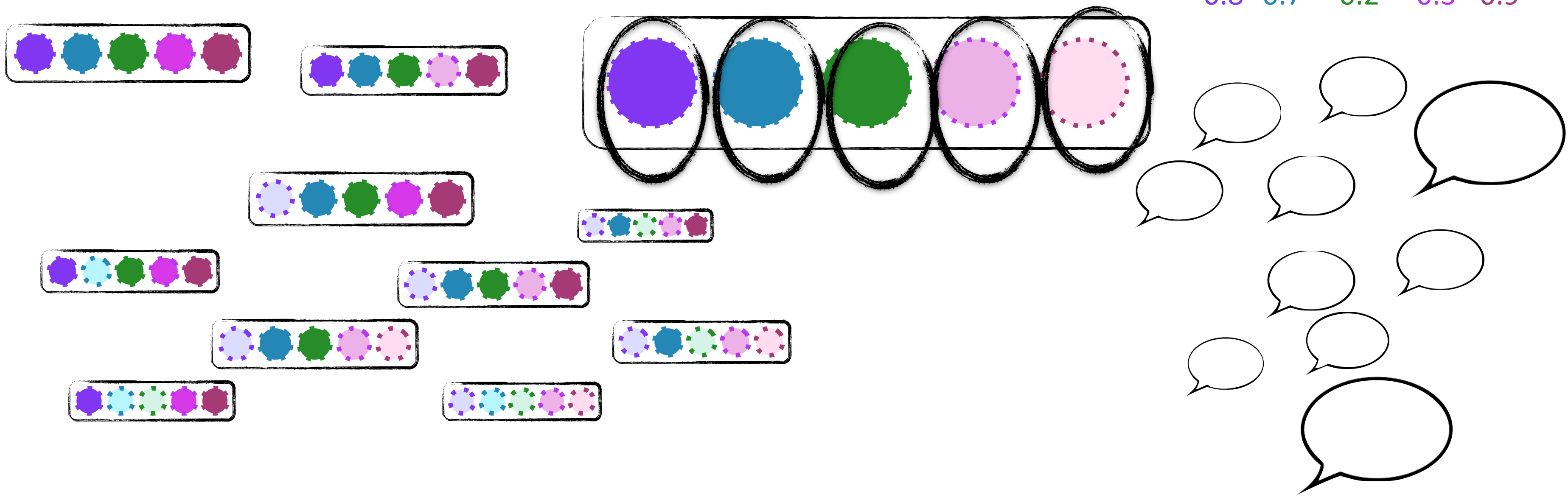


Key: Parameters are separable components of grammars

Learning with parameters



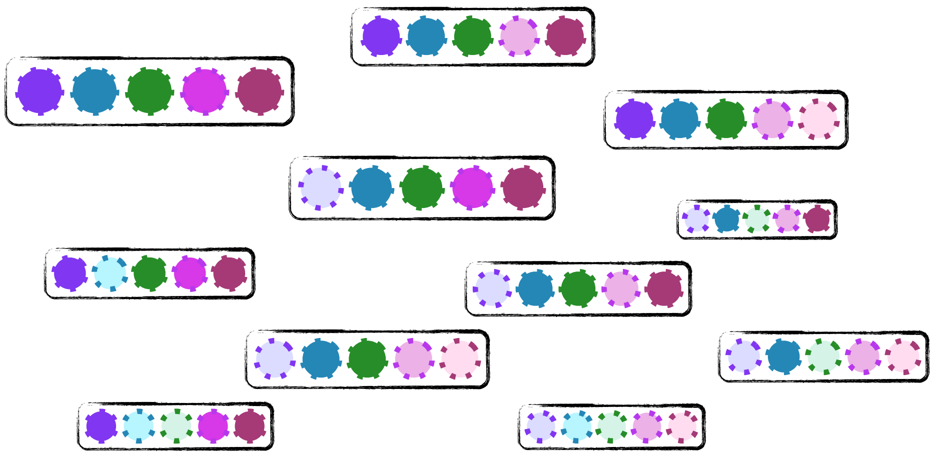
Variational learning

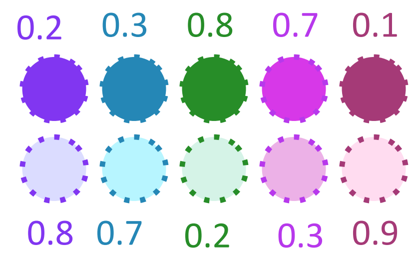


A variational learner can take advantage of the fact that grammars are really sets of parameter values.

Learning with parameters

Variational learning

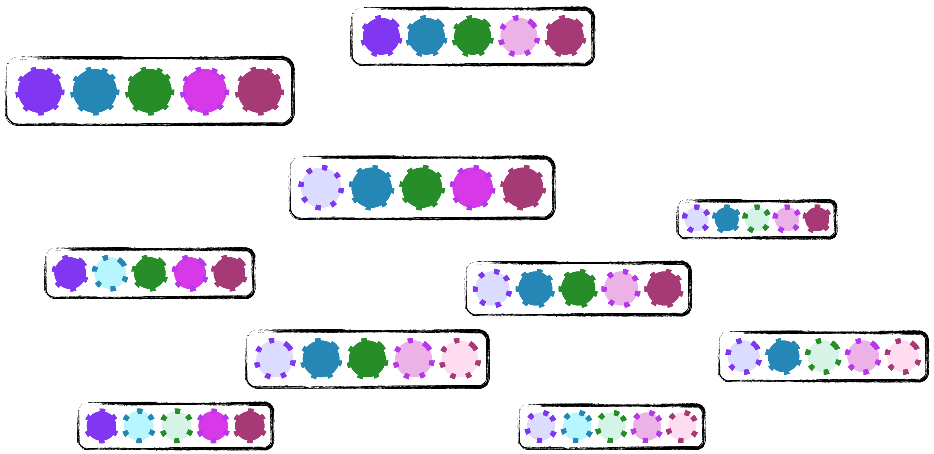


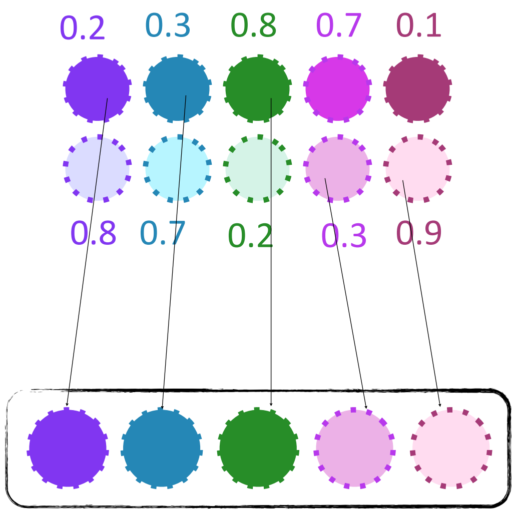


Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning



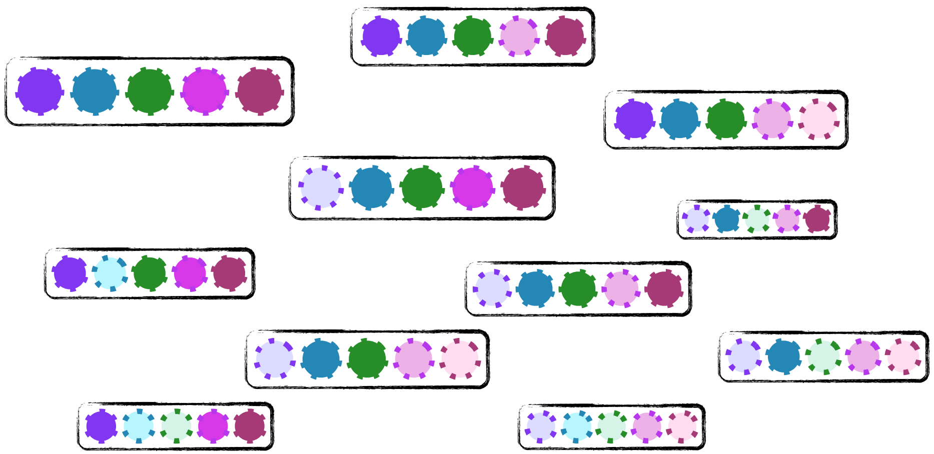


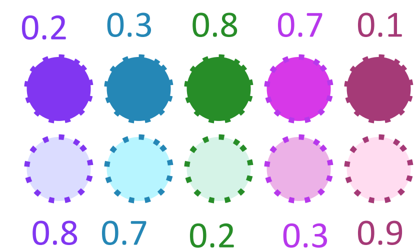
p = .2*.3*.8*.3*.9

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

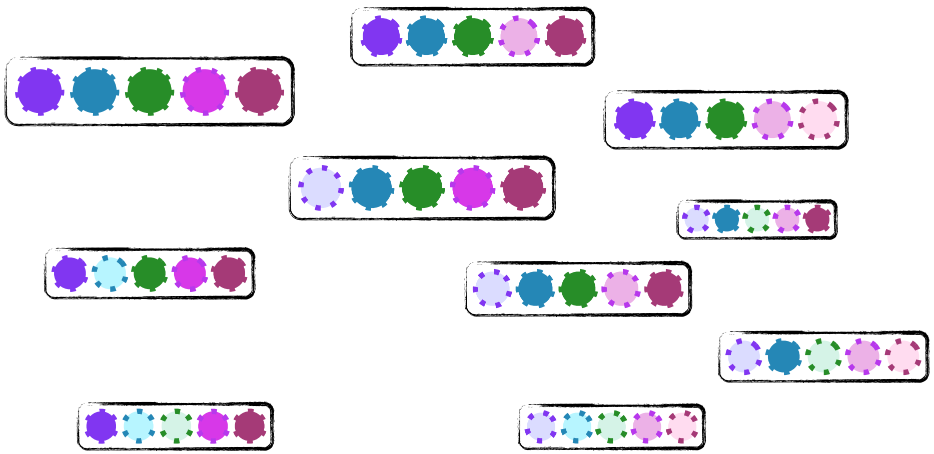


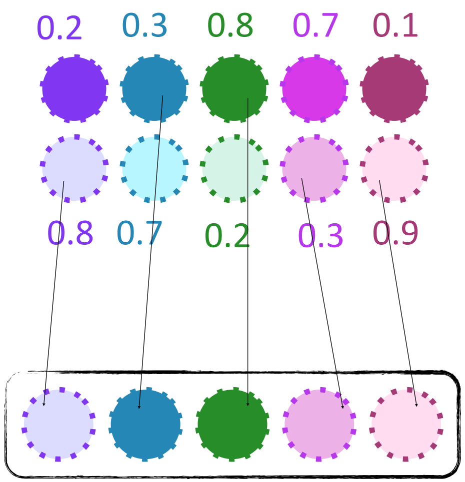


Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning



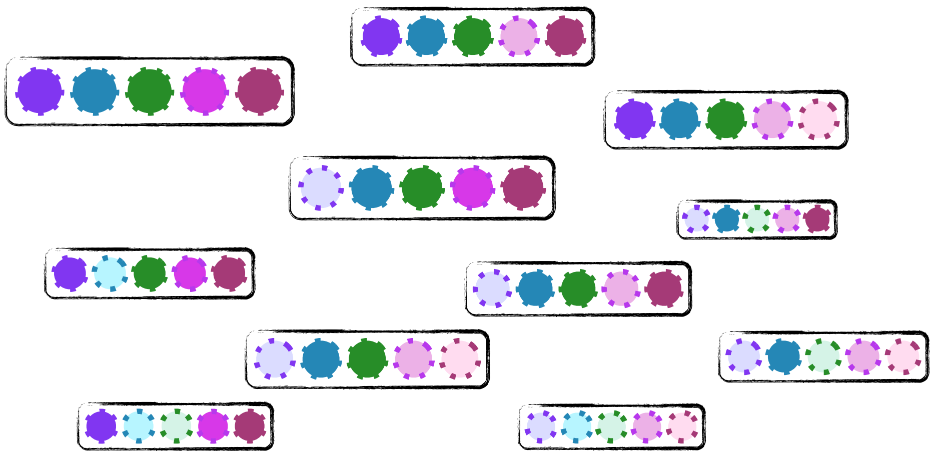


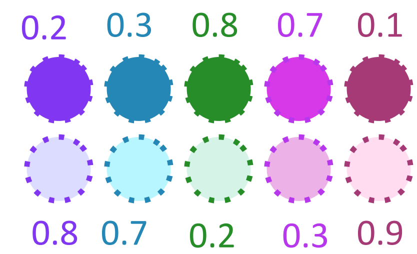
p = .8*.3*.8*.3*.9

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

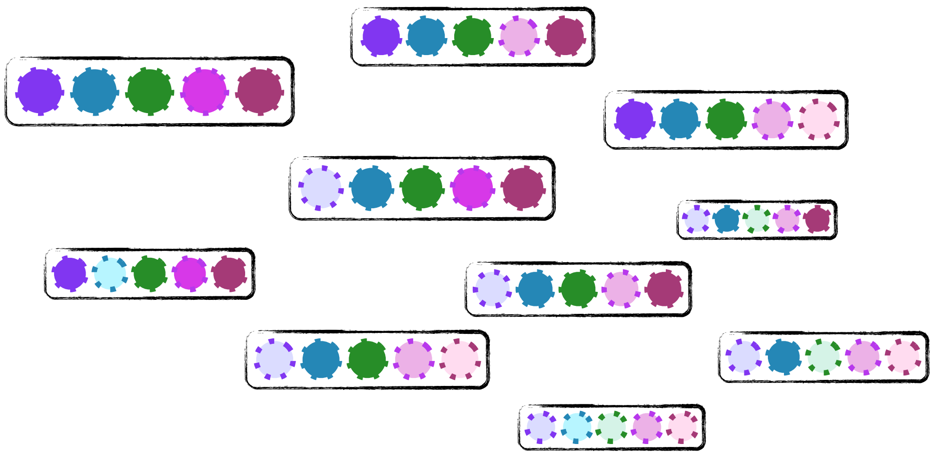


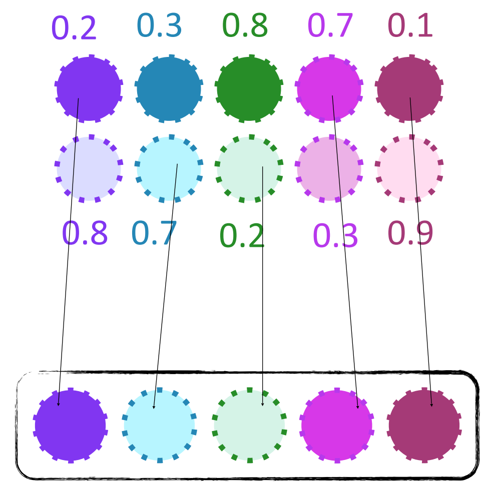


Parameter values can be probabilistically accessed, depending on the level of belief (probability) the learner currently has in each one.

Learning with parameters

Variational learning





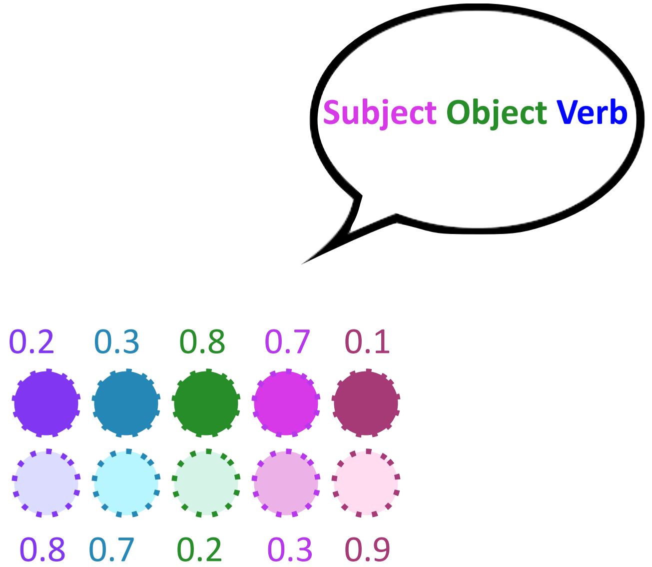
p = .2*.7*.2*.7*.1

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

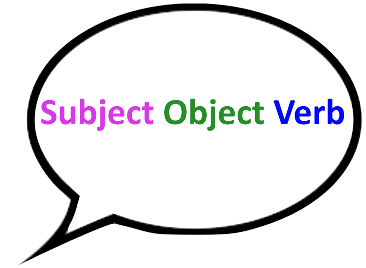
Variational learning

For each data point encountered in the input…



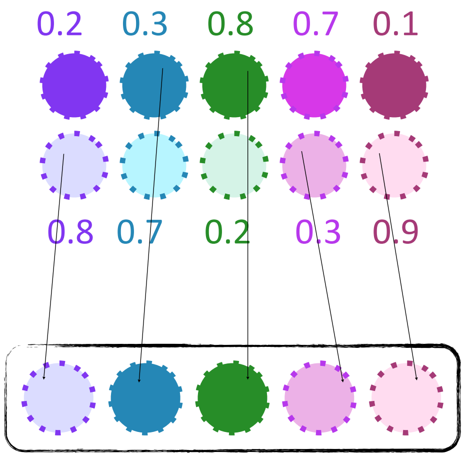
Learning with parameters
The learning algorithm

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.

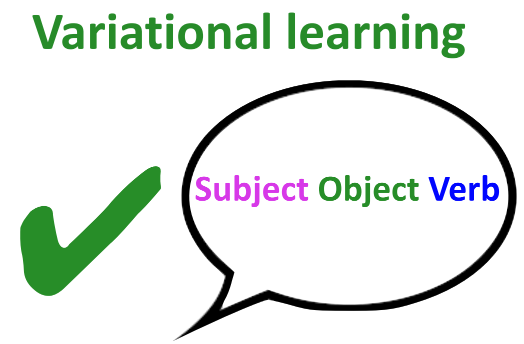
Learning with parameters
The learning algorithm

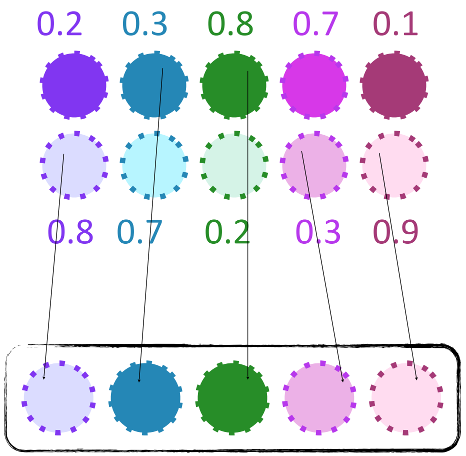
For each data point encountered in the input…

(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).

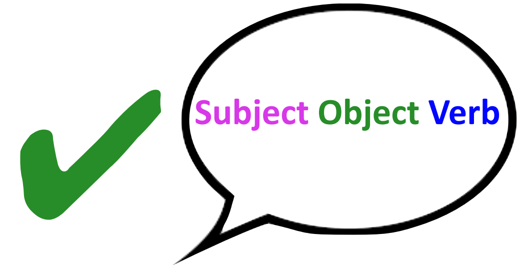




p = .8*.3*.8*.3*.9

Learning with parameters
The learning algorithm

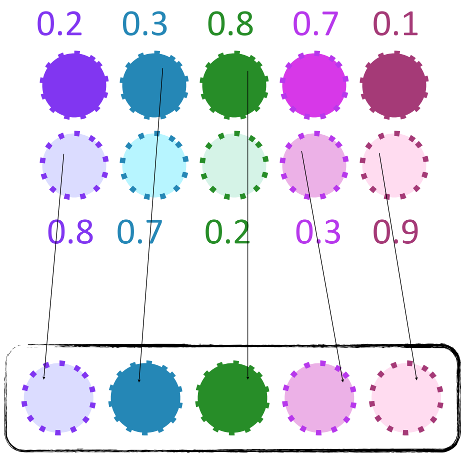
Variational learning



For each data point encountered in the input…

(1) Choose a grammar.

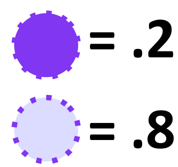
(2) Try to analyze the data point with this grammar.



(3) Update parameter value probabilities.

1st parameter

Actual update equation for reward:



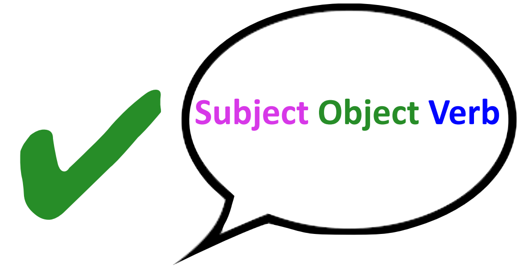
p = .8*.3*.8*.3*.9

p_v = previous value of successful parameter value

p_o = previous value of opposing parameter value

Learning with parameters
The learning algorithm

Variational learning

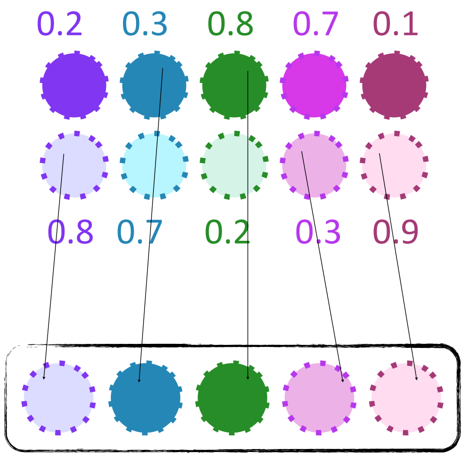


For each data point encountered in the input…

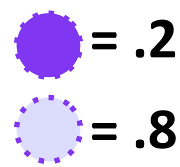
(1) Choose a grammar.

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

(3) Update parameter value probabilities.



1st parameter



Actual update equation for reward:

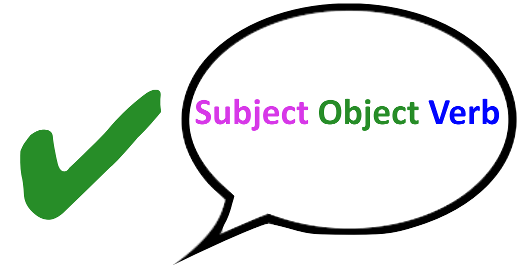
$p = .8 * .3 * .8 * .3 * .9$

$p_v = 0.8$

$p_o = 0.2$

Learning with parameters
The learning algorithm

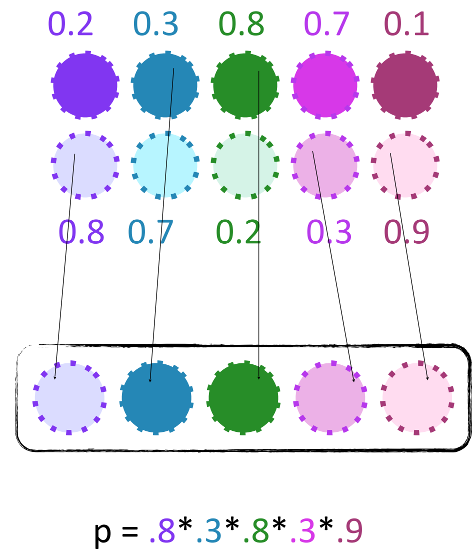
Variational learning



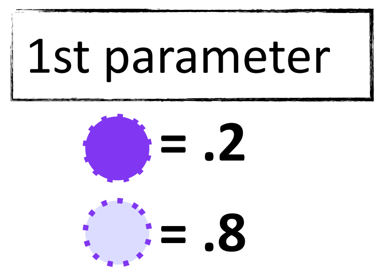
For each data point encountered in the input…

(1) Choose a grammar.

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



(3) Update parameter value probabilities.



Actual update equation for reward:

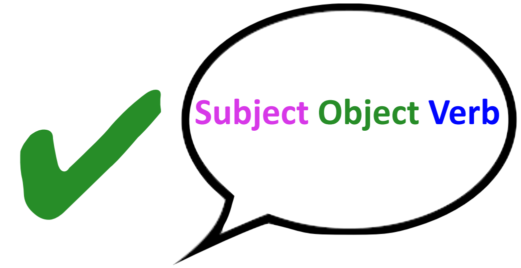
$p_v = 0.8$
$p_o = 0.2$

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

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

Learning with parameters
The learning algorithm

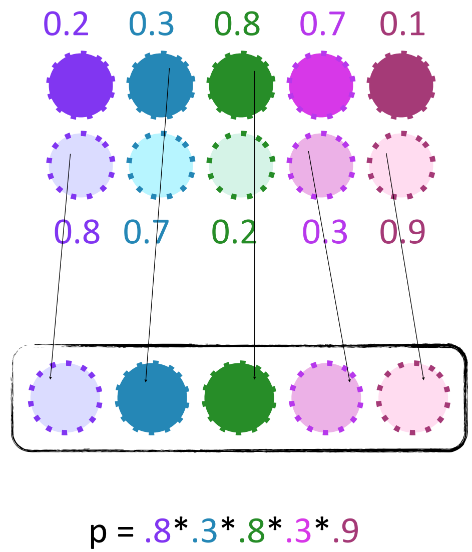
Variational learning



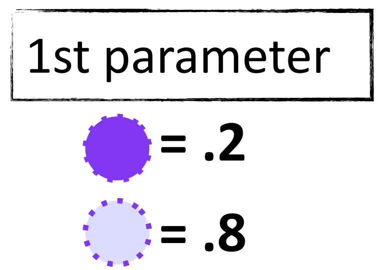
For each data point encountered in the input…

(1) Choose a grammar.

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



(3) Update parameter value probabilities.



Actual update equation for reward:

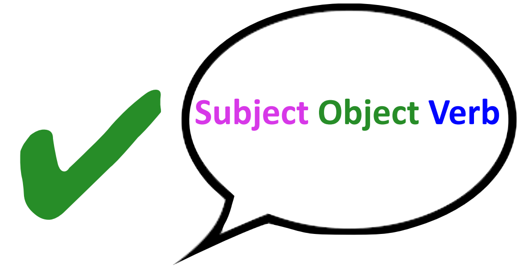
$p_v = 0.8$
$p_o = 0.2$

$p_{v\_updated} = 0.8 + 0.125(1- 0.8)$
$p_{o\_updated} = (1-0.125)0.2$

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

Learning with parameters
The learning algorithm

Variational learning

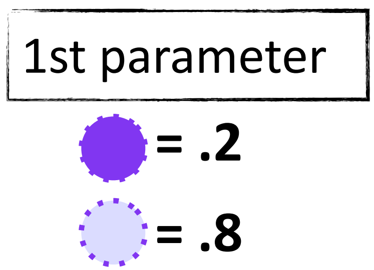


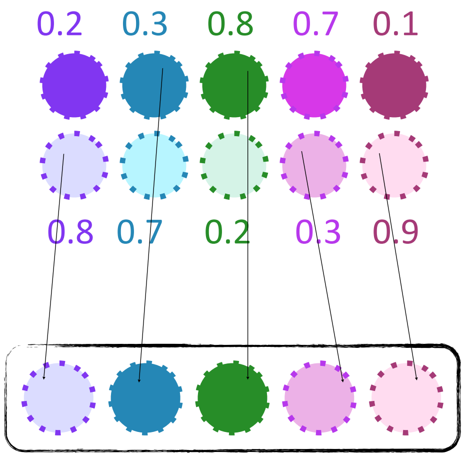
For each data point encountered in the input…

(1) Choose a grammar.

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

(3) Update parameter value probabilities.





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

Actual update equation for reward:

$p_v = 0.8$

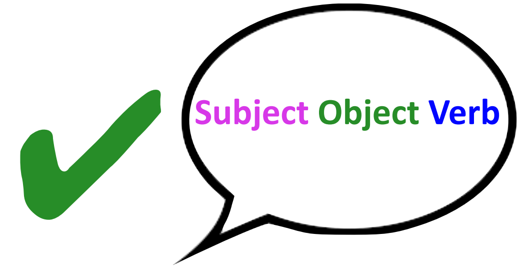
$p_o = 0.2$

$p_{v\_updated} = 0.825$

$p_{o\_updated} = 0.175$

Learning with parameters
The learning algorithm

Variational learning

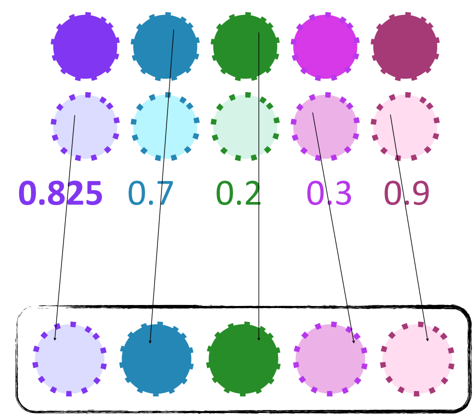


For each data point encountered in the input…

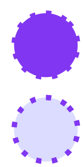
(1) Choose a grammar.

(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

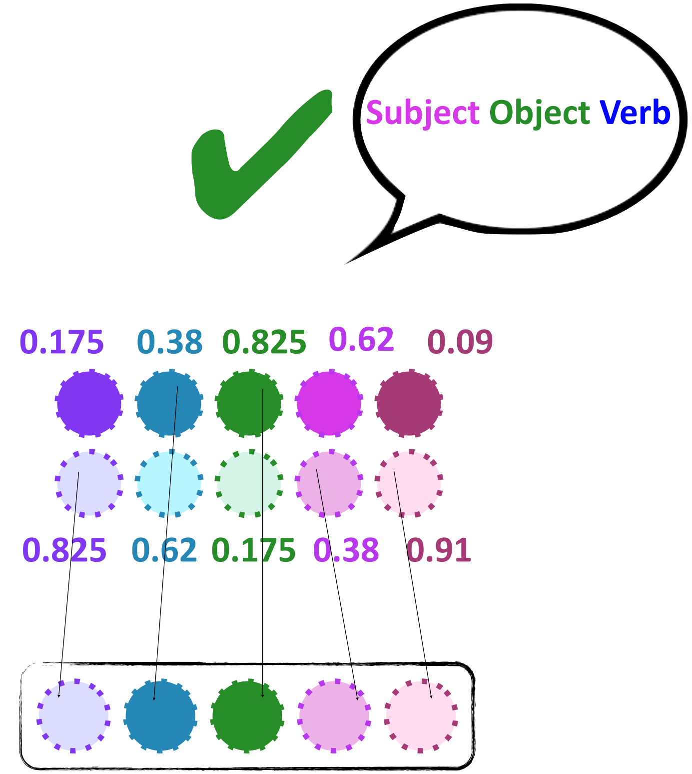
Variational learning

For each data point encountered in the input…

(1) Choose a grammar.

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

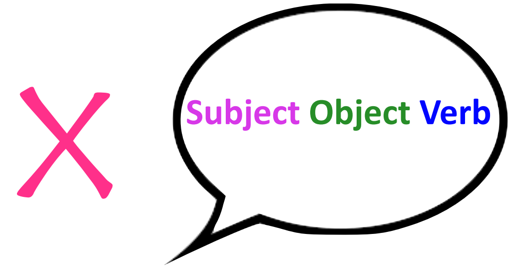
(3) Update parameter value probabilities.



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

Learning with parameters
The learning algorithm

Variational learning

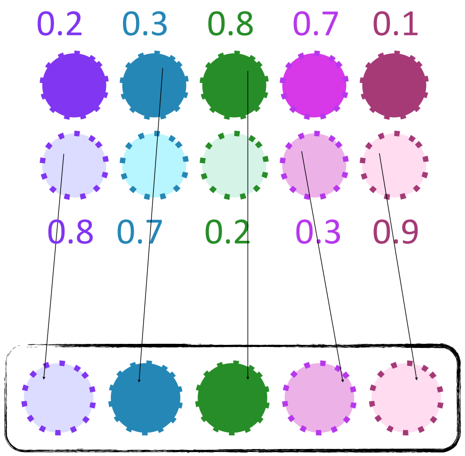


For each data point encountered in the input…

(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

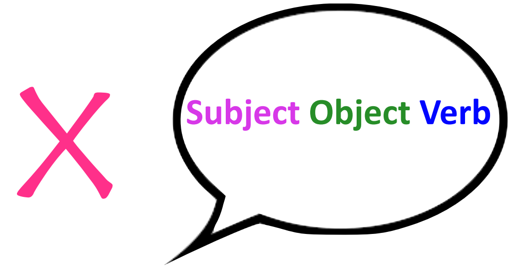
But what happens if the selected grammar can't account for the data point?

Then all the participating parameter values are punished.

Learning with parameters

The learning algorithm

Variational learning

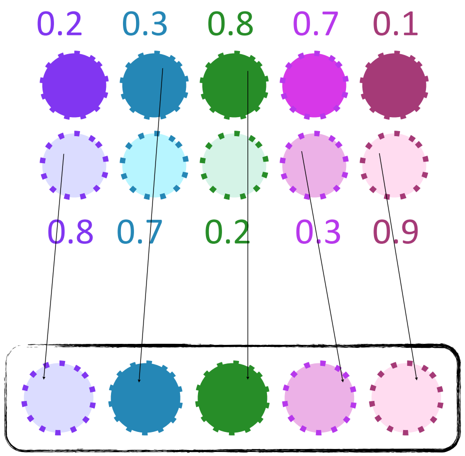


For each data point encountered in the input…

(1) Choose a grammar.

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

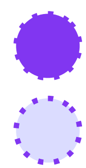
(3) Update parameter value probabilities.



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

1st parameter

Actual update equation for punishment:



= .2

= .8

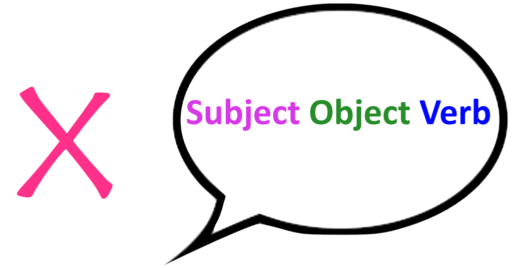
p_v = previous value of unsuccessful parameter value

p_o = previous value of opposing parameter value

Learning with parameters

The learning algorithm

Variational learning

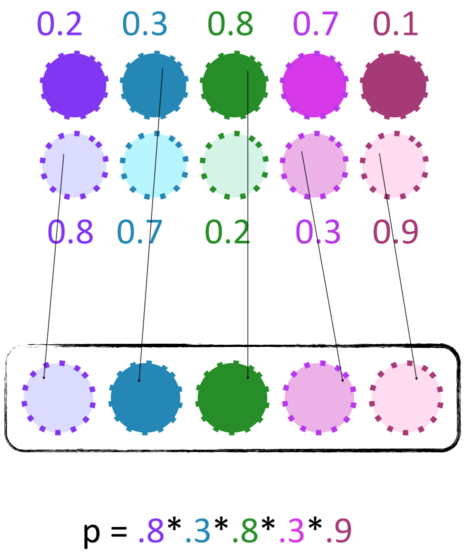


For each data point encountered in the input…

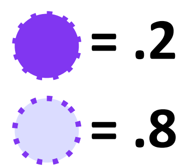
(1) Choose a grammar.

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

(3) Update parameter value probabilities.



1st parameter



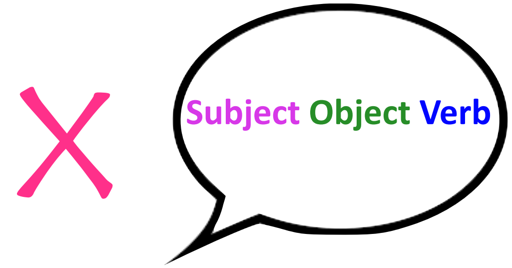
Actual update equation for punishment:

$p_v = 0.8$

$p_o = 0.2$

Learning with parameters
The learning algorithm

Variational learning

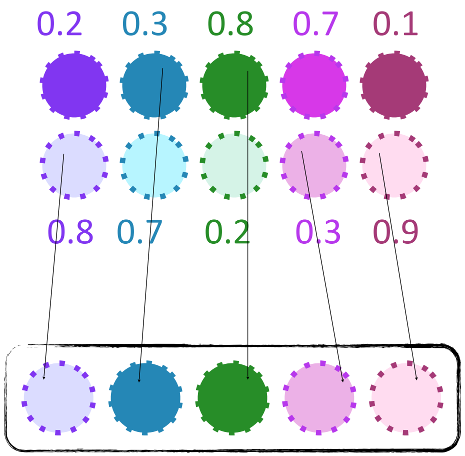


For each data point encountered in the input…

(1) Choose a grammar.

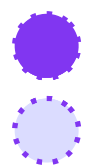
(2) Try to analyze the data point with this grammar.

(3) Update parameter value probabilities.



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

1st parameter


= .2
= .8

Actual update equation for punishment:

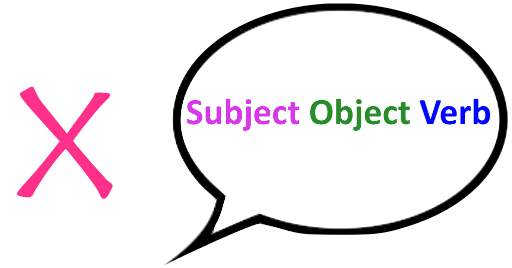
$p_v = 0.8$
$p_o = 0.2$

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

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

Learning with parameters
The learning algorithm

Variational learning

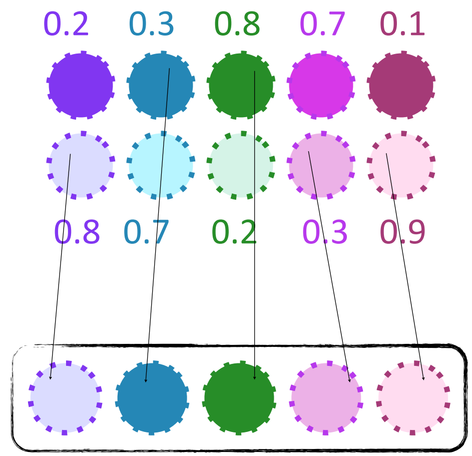


For each data point encountered in the input…

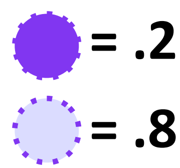
(1) Choose a grammar.

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

(3) Update parameter value probabilities.



1st parameter



Actual update equation for punishment:

$p_v = 0.8$

$p_o = 0.2$

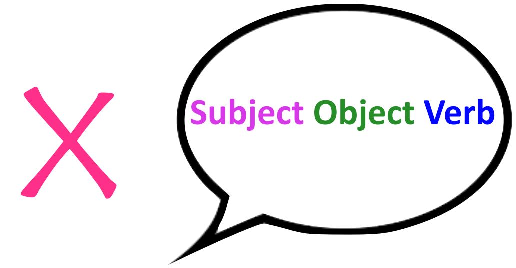
p = .8*.3*.8*.3*.9

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

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

Learning with parameters
The learning algorithm

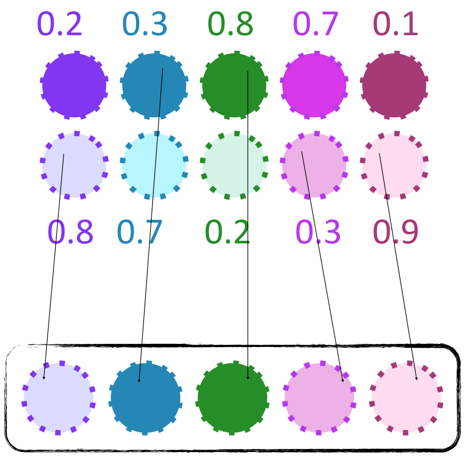
Variational learning



For each data point encountered in the input…

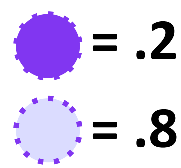
(1) Choose a grammar.

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



(3) Update parameter value probabilities.

1st parameter



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

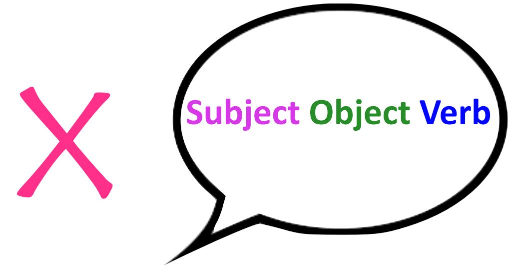
Actual update equation for punishment:

$p_v = 0.8$
$p_o = 0.2$

$p_{v\_updated} = 0.70$
$p_{o\_updated} = 0.30$

Learning with parameters
The learning algorithm

Variational learning

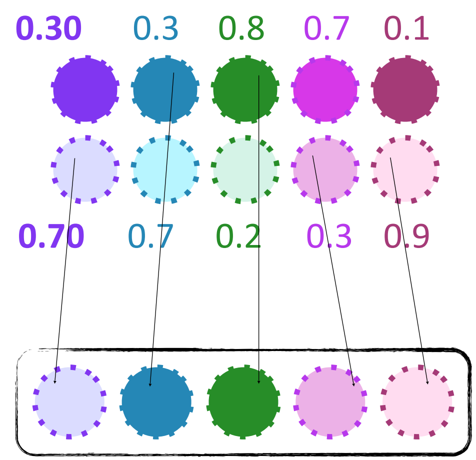


For each data point encountered in the input…

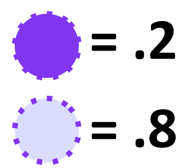
(1) Choose a grammar.

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

(3) Update parameter value probabilities.



1st parameter



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

Actual update equation for punishment:

$p_v = 0.8$

$p_o = 0.2$

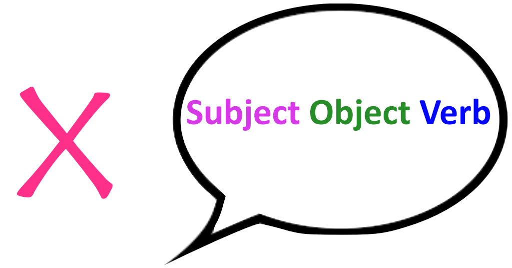
$p_{v\_updated} = 0.70$

$p_{o\_updated} = 0.30$

Do this for all the other parameters, too.

Learning with parameters
The learning algorithm

Variational learning

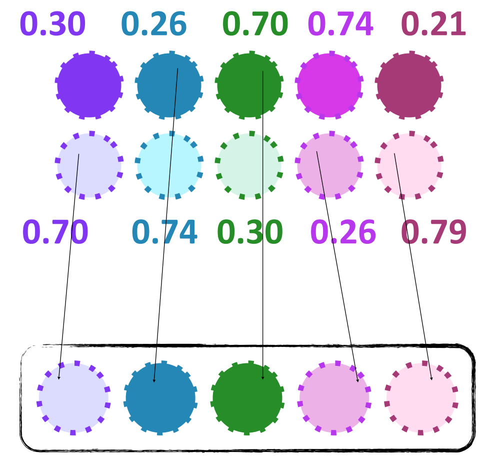


For each data point encountered in the input…

(1) Choose a grammar.

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

(3) Update parameter value probabilities.

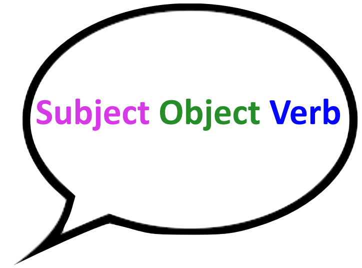


p = .8*.3*.8*.3*.9

Learning with parameters

The learning algorithm

Variational learning

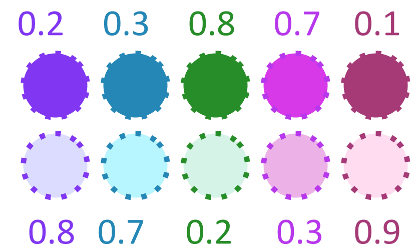


For each data point encountered in the input…

(1) Choose a grammar.

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

(3) Update parameter value probabilities.

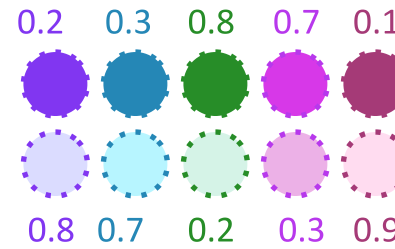


Problem ameliorated!
Unambiguous data are much more likely to exist for **individual parameter values** instead of entire grammars.

Learning with parameters

The learning algorithm

Variational learning



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

“...dass ich Kätzchen liebe.”

...that I Kitties love

Head-directionality Subject drop (subj-drop)

Subject Object Verb

G2 Head-final +subj-drop

G4 Head-final -subj-drop

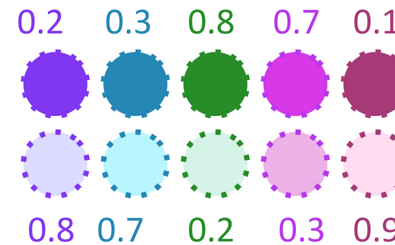
G1 Head-first +subj-drop

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

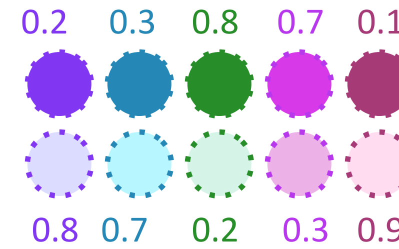
G1 Head-first +subj-drop

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

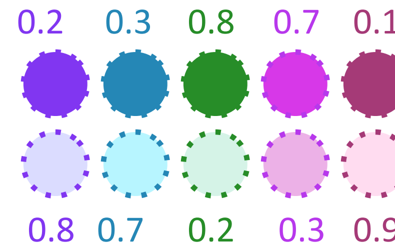
G2 Head-final +subj-drop

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

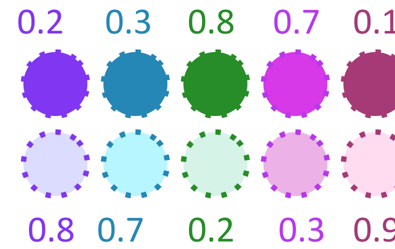
Subject Object Verb

✔ **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.

"...dass ich Kätzchen liebe."

...that I Kitties love

Head-directionality Subject drop (subj-drop)

Subject Object Verb

G1 Head-first +subj-drop

G3 Head-first -subj-drop

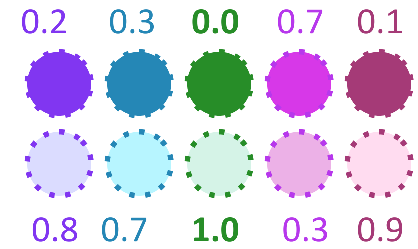
G2 Head-final +subj-drop

G4 Head-final -subj-drop

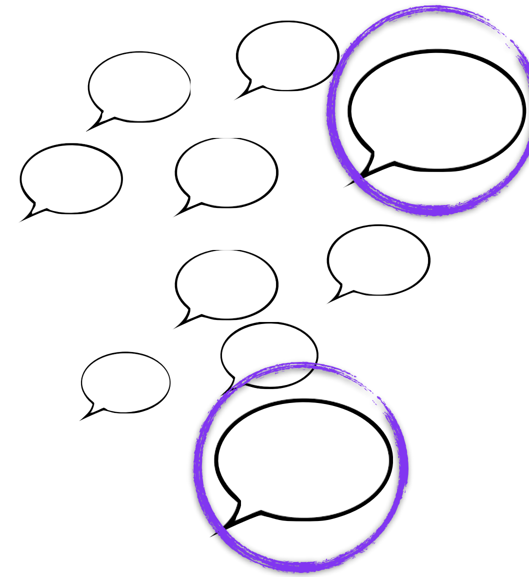
Learning with parameters

The learning algorithm

Variational learning



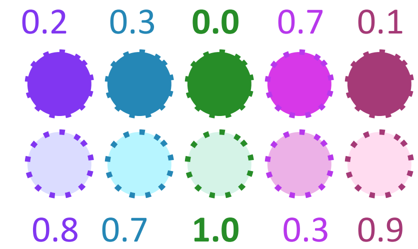
Implication: The more unambiguous data there are, the faster the native language's parameter value will "win" (reach a probability near 1.0). This means that the child will learn the associated structural pattern faster.



Learning with parameters

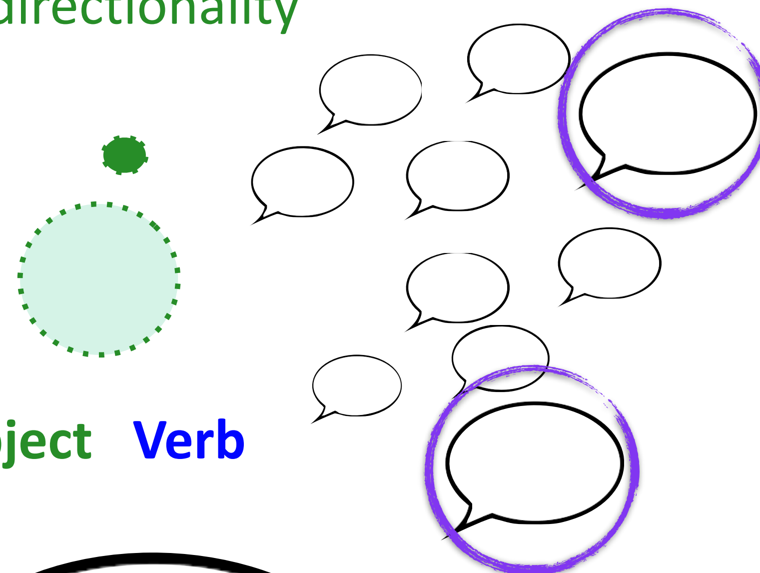
The learning algorithm

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**

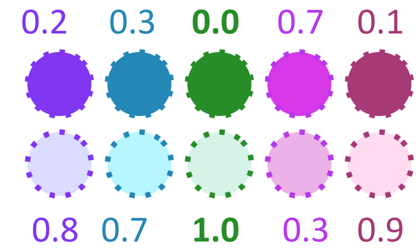
"...dass ich Kätzchen liebe."

...that I Kitties love

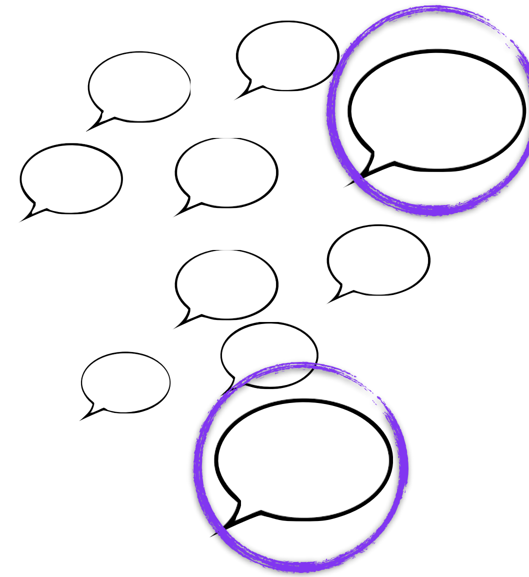
Learning with parameters

The learning algorithm

Variational learning



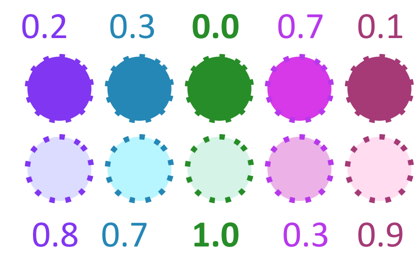
Is it true that the amount of unambiguous data the child encounters for a particular parameter strongly impacts when the child learns that structural property of the language?



Learning with parameters

The learning algorithm

Variational learning



Striking evidence that this is true

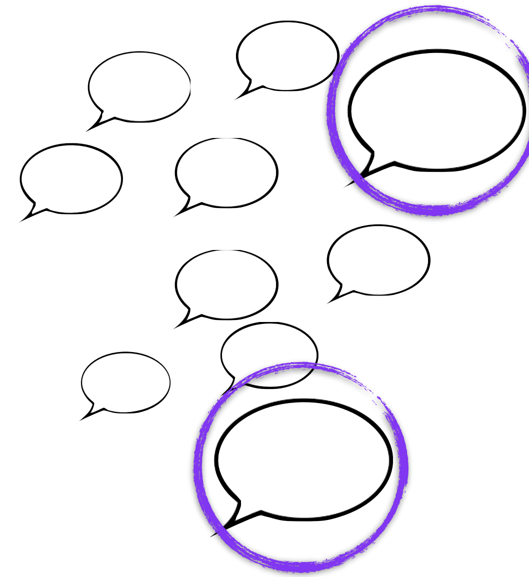
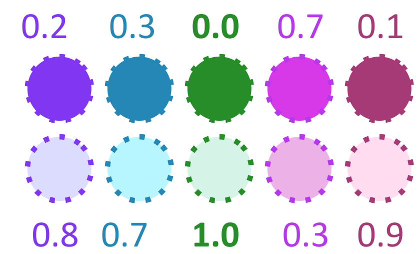


Table 1: The qualitative fit Yang discovered between the unambiguous data advantage (Adv) perceived by a VarLearner in its acquisitional intake and the observed age of acquisition (AoA) in children for six parameter values across different languages.

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

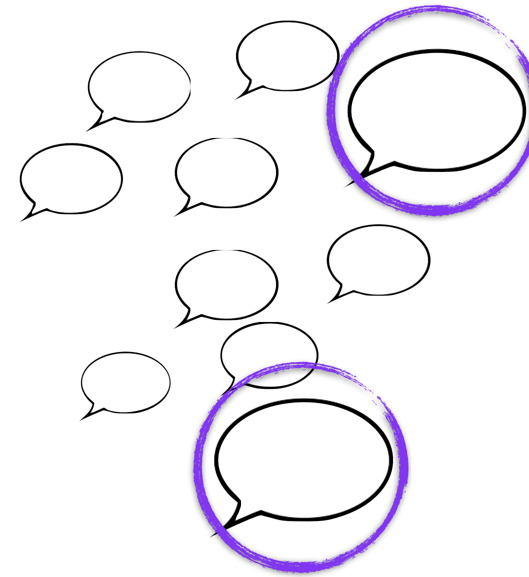
Learning with parameters
The learning algorithm
Variational learning



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

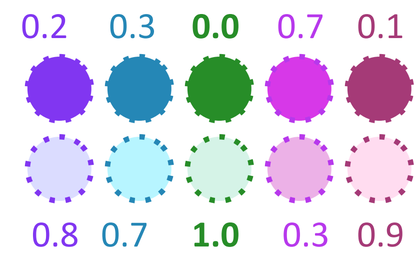


The more unambiguous data there are for one value over another (its advantage)…

Learning with parameters

The learning algorithm

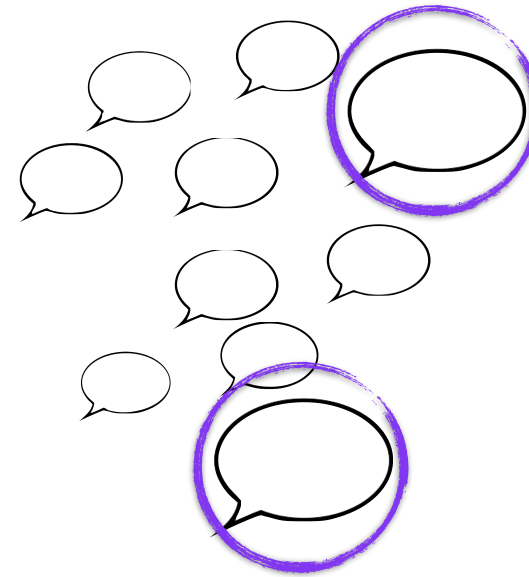
Variational learning



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.

Recap

Even with parameters, acquisition of linguistic structure can be hard because a child has to figure out which parameter values produce the observable data. This isn't always easy because parameters interact.

Variational learning leverages the fact that grammars can be divided into parameters, and a data point can be informative for one parameter but not others.

Recap

Predictions of variational learning:

Parameters set early: more unambiguous data available

Parameters set late: less unambiguous data available

These predictions seem to be born out by available data on when children learn certain structural patterns (parameter values) about their native language.

Questions?

You should be able to do up through 21 the structure review questions and up through 4 on HW8.