Computational Cognitive Science

Lecture 16: Compositionality

Guest Lecturer: Frank Mollica

Logical Hypotheses

Disjunctive Normal Form

RED

Conjunctive Normal Form

RED

(Sheperd et al., 2008)

Logical Hypotheses

Disjunctive Normal Form

$$(\mathrm{RED} \wedge \mathrm{SMALL}) \vee (\mathrm{GREEN} \wedge \mathrm{LARGE})$$

Conjunctive Normal Form

$$(\mathrm{RED} \vee \mathrm{LARGE}) \wedge (\mathrm{GREEN} \vee \mathrm{SMALL})$$

(Sheperd et al., 2008)

Logical Hypotheses

Disjunctive Normal Form

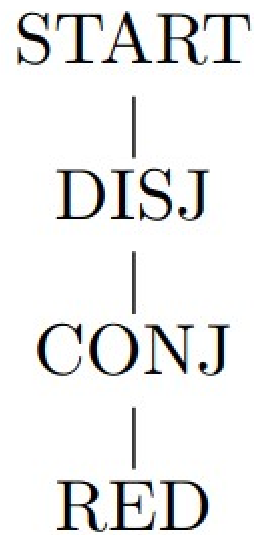
(GREEN ∧ TRIANGLE) ∨ (GREEN ∧ LARGE) ∨
(RED ∧ SQUARE ∧ SMALL)

Conjunctive Normal Form

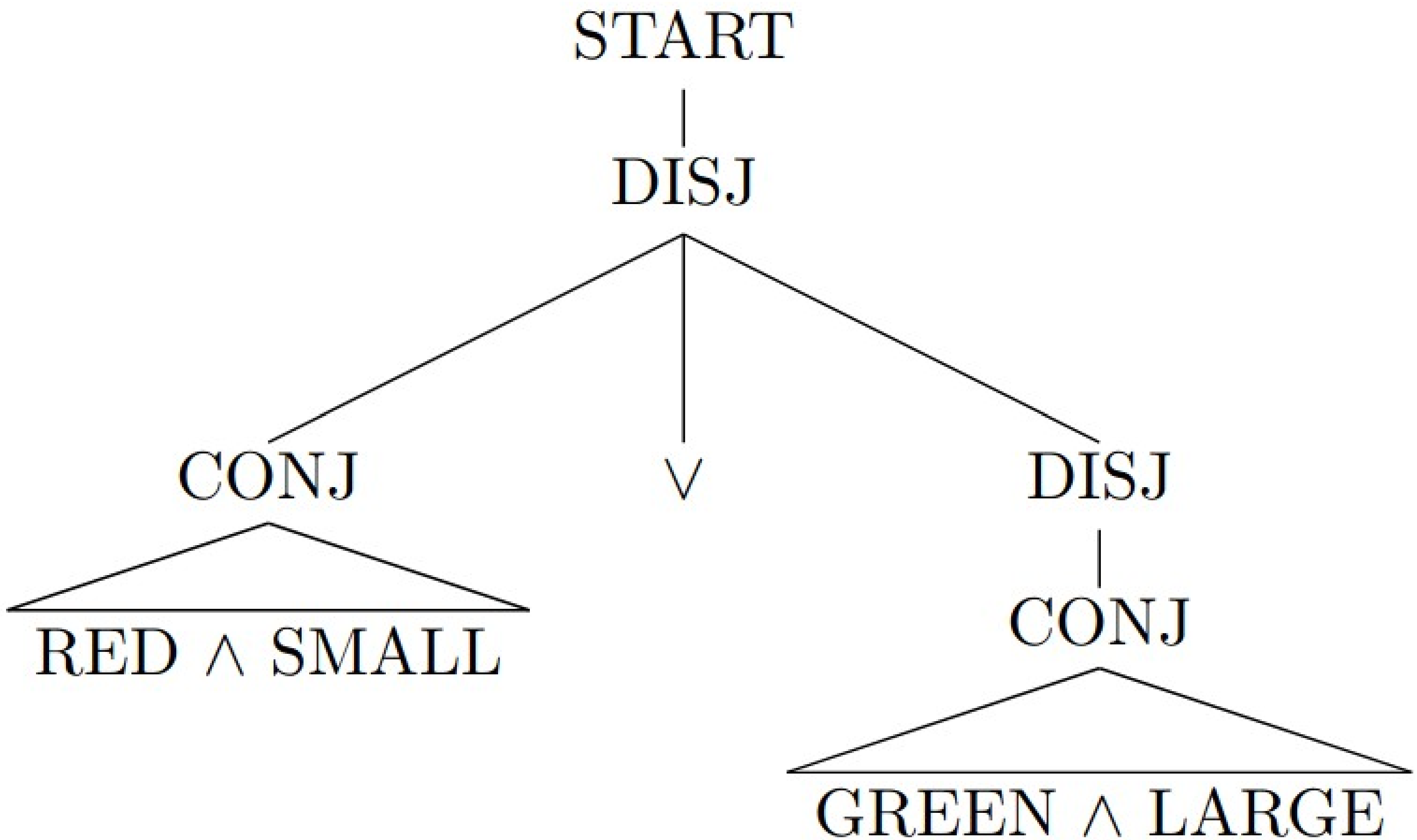
(GREEN ∨ SQUARE) ∧ (GREEN ∨ SMALL) ∧
(RED ∨ TRIANGLE ∨ LARGE)

(Sheperd et al., 2008)

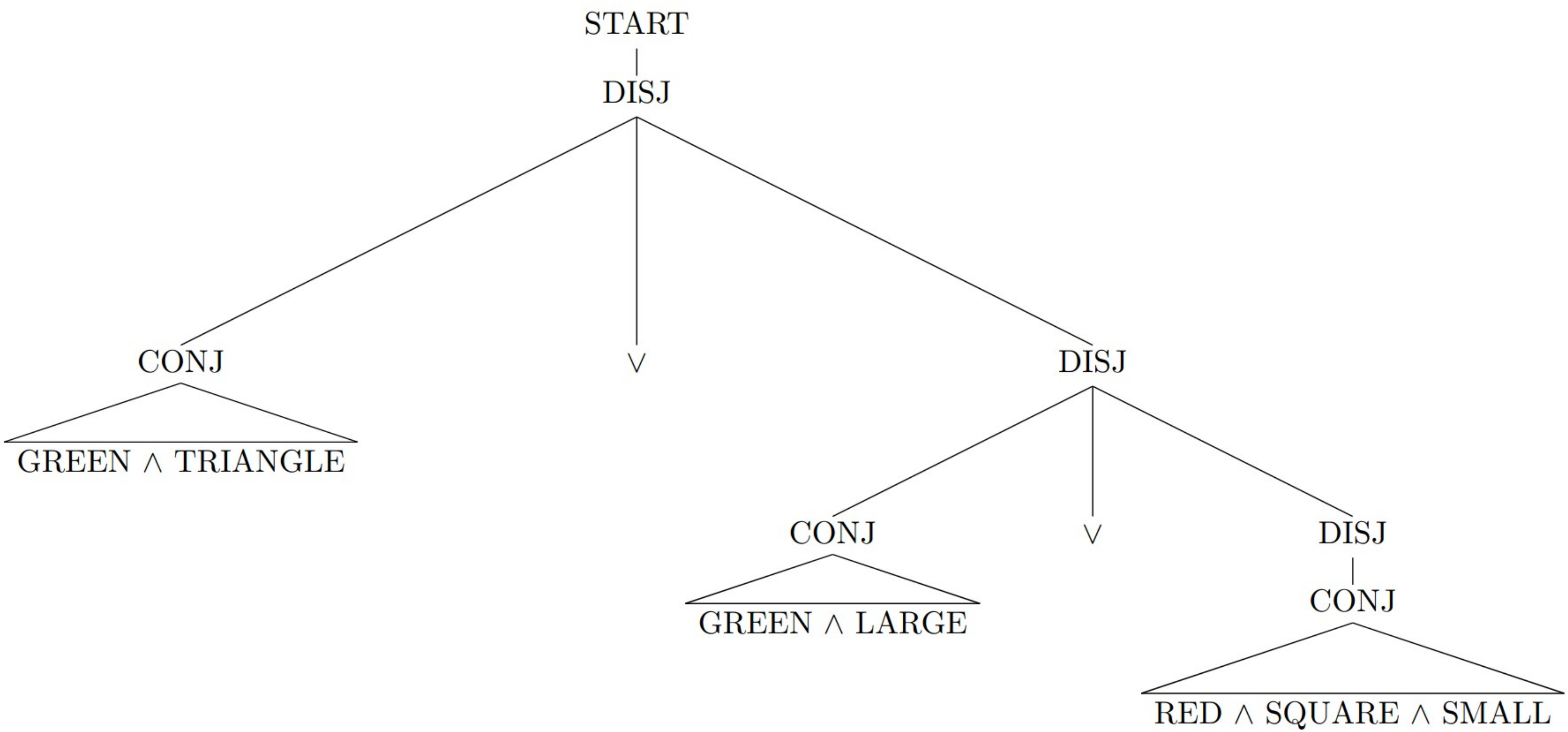
Compositionality



Compositionality



Compositionality



$$\text{START} \rightarrow \text{DISJ}$$
$$\text{DISJ} \rightarrow \text{CONJ}$$
$$\text{DISJ} \rightarrow \text{CONJ} \vee \text{DISJ}$$

Compositionality

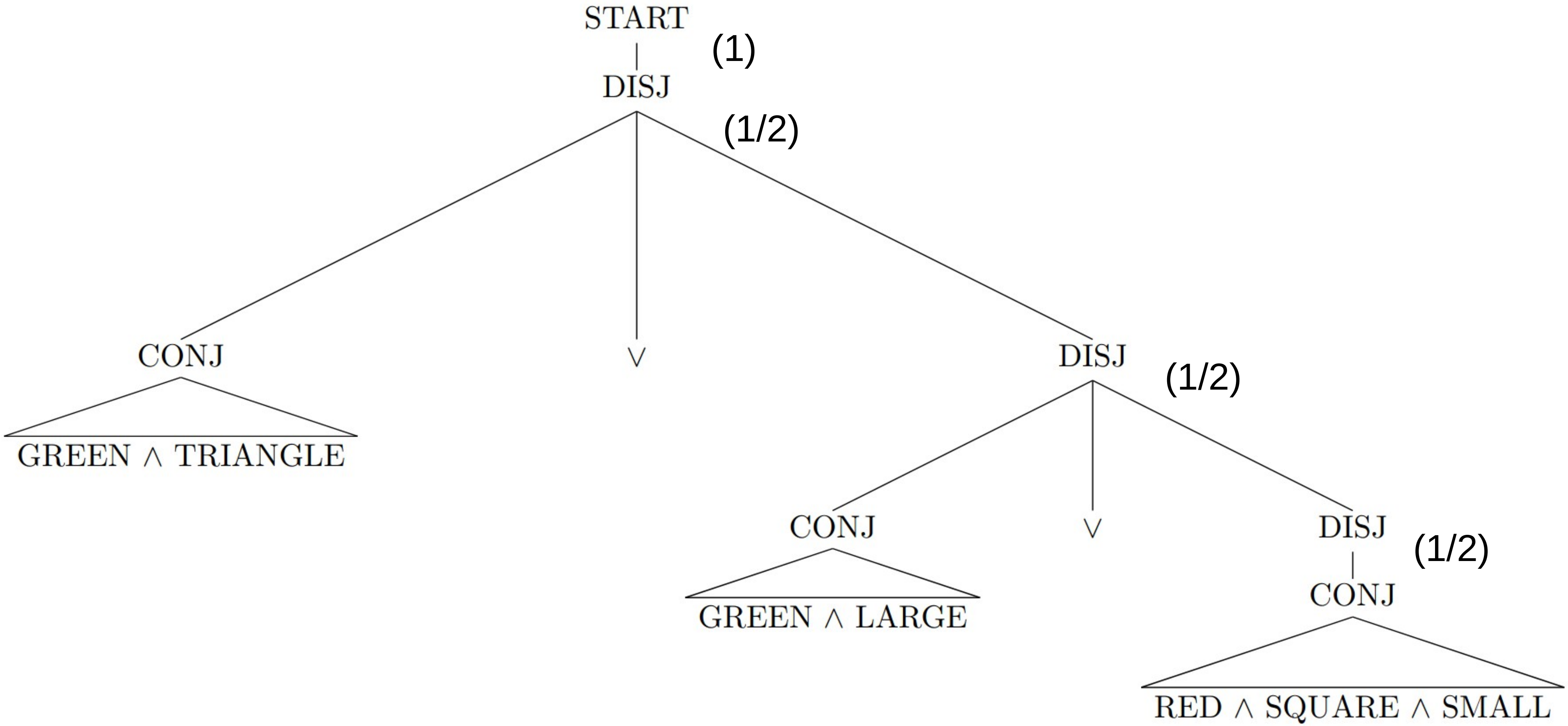
- Productivity:
  - The language generates all licit hypotheses even those unseen
- Systematicity:
  - The function/meaning of a complex hypothesis is determined by its structure and primitive components

(Fodor, 1975)

Rational Rules

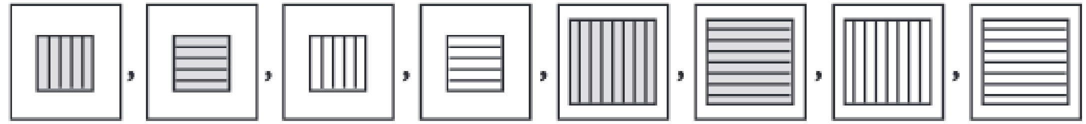
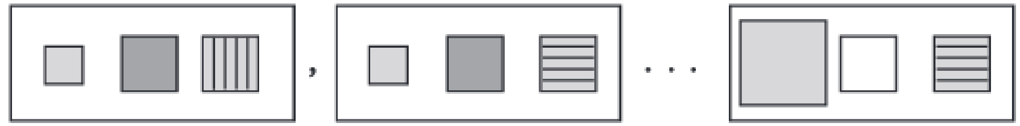
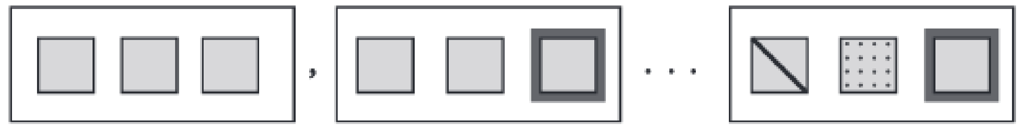
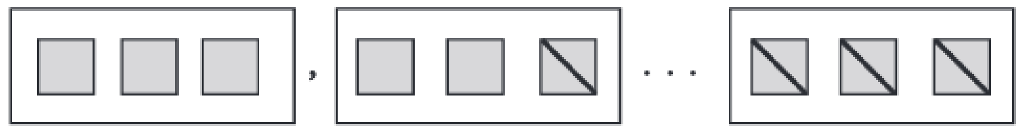
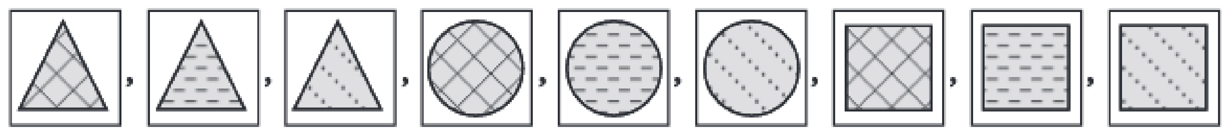
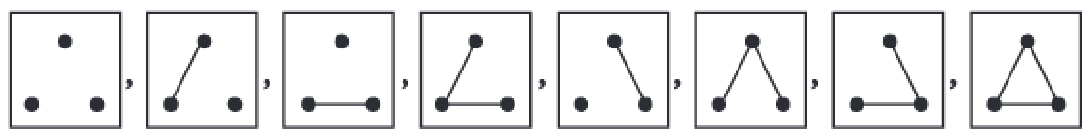
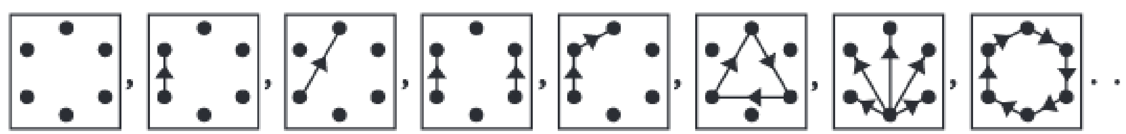
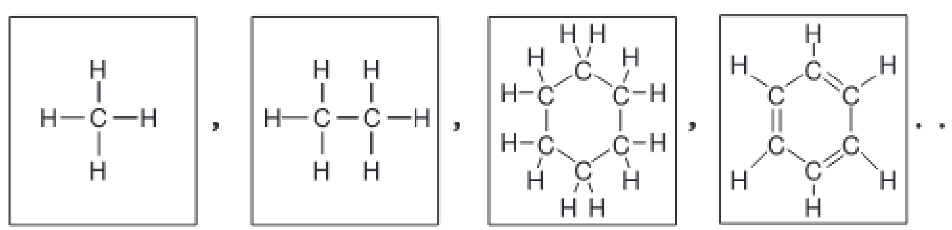
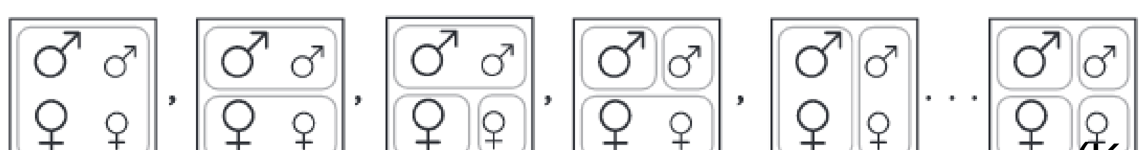
- Hypothesis Space:
  - All functions generated by the grammar
- Prior:
  - Augment the grammar with probabilities
  - Implicitly favors simple expressions

(Goodman et al., 2008; Piantadosi et al., 2016)

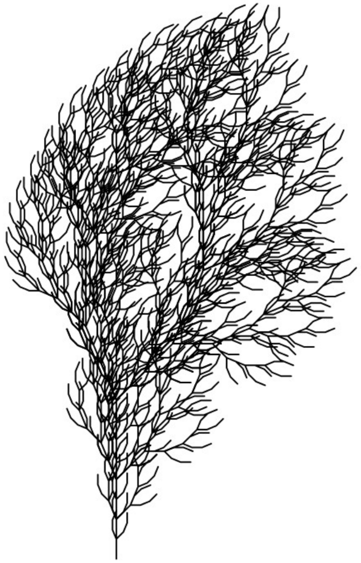


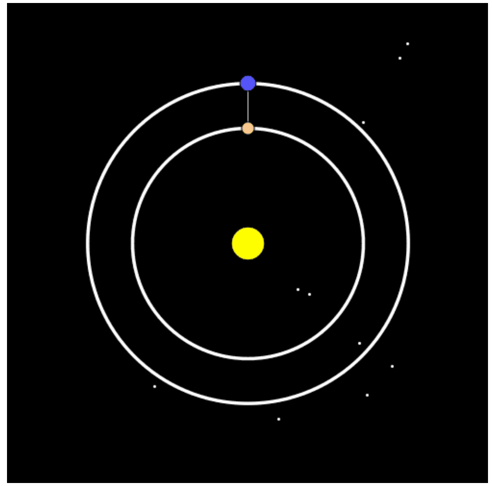
$$\text{START} \xrightarrow{(1)} \text{DISJ}$$
$$\text{DISJ} \xrightarrow{(1/2)} \text{CONJ}$$
$$\text{DISJ} \xrightarrow{(1/2)} \text{CONJ} \vee \text{DISJ}$$

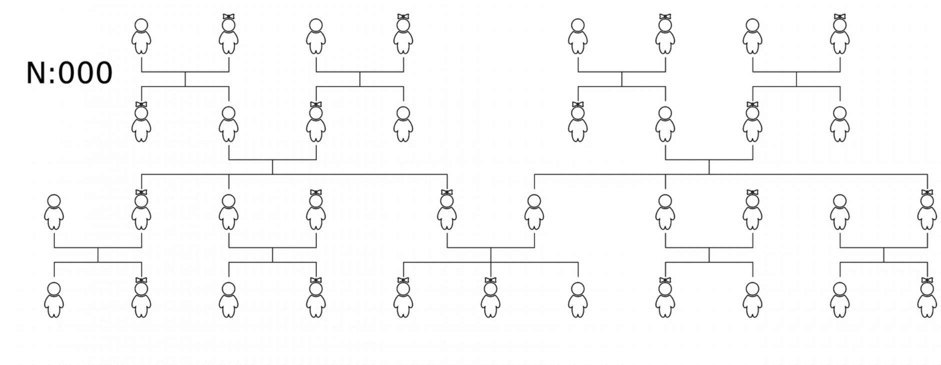
$$P(h) = (1)(1/2)(1/2)(1/2)$$
$$= (1/8)$$

Item description	Domain
1. 1 object, 3 substitutive binary features	
2. 3 objects, 3 substitutive binary features	
3. 3 objects, 3 additive binary features	
4. 3 objects, 1 additive binary feature	
5. 1 object, 2 substitutive ternary features	
6. 3 objects, 1 undirected binary relation	
7. 6 objects, 1 directed binary relation	
8. Multiple objects, multiple features and relations	
9. 4 objects, multiple features and relations	

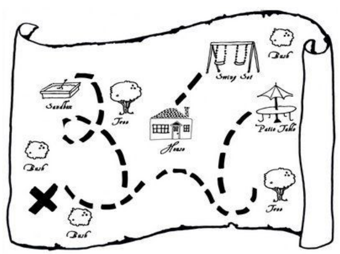
(Kemp, 2012)

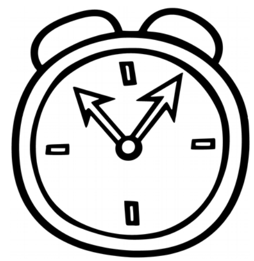


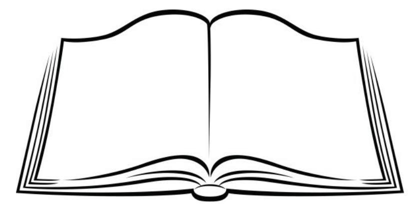


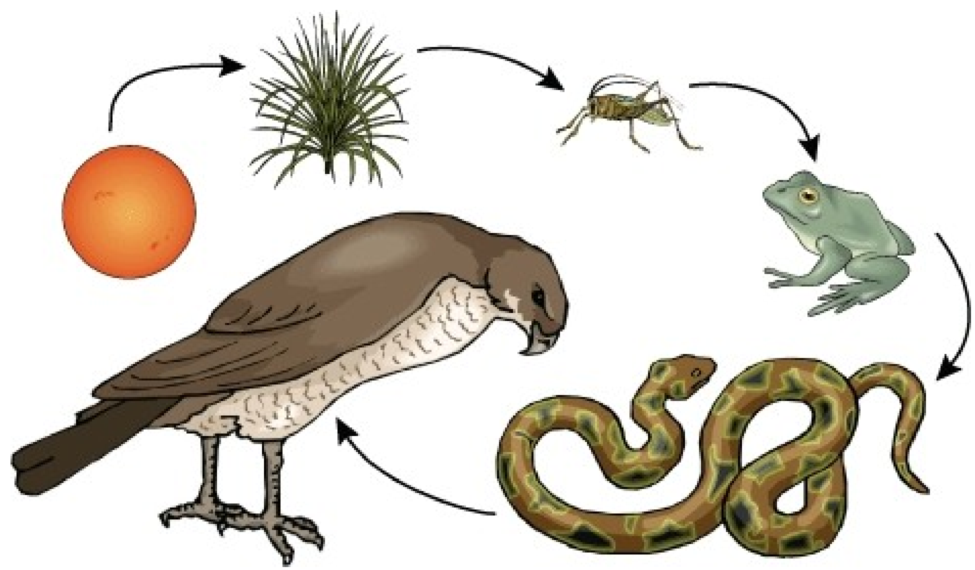


Language

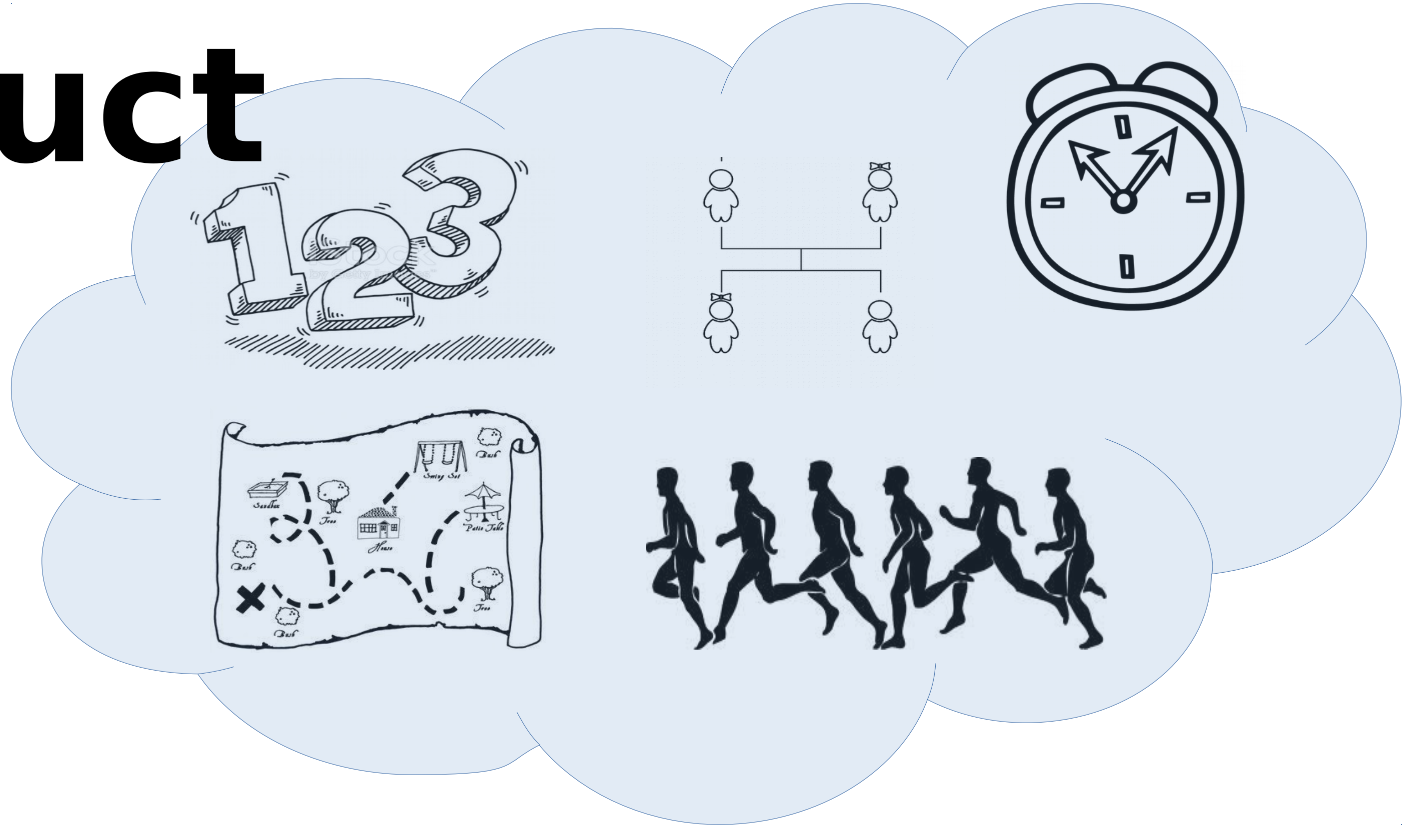


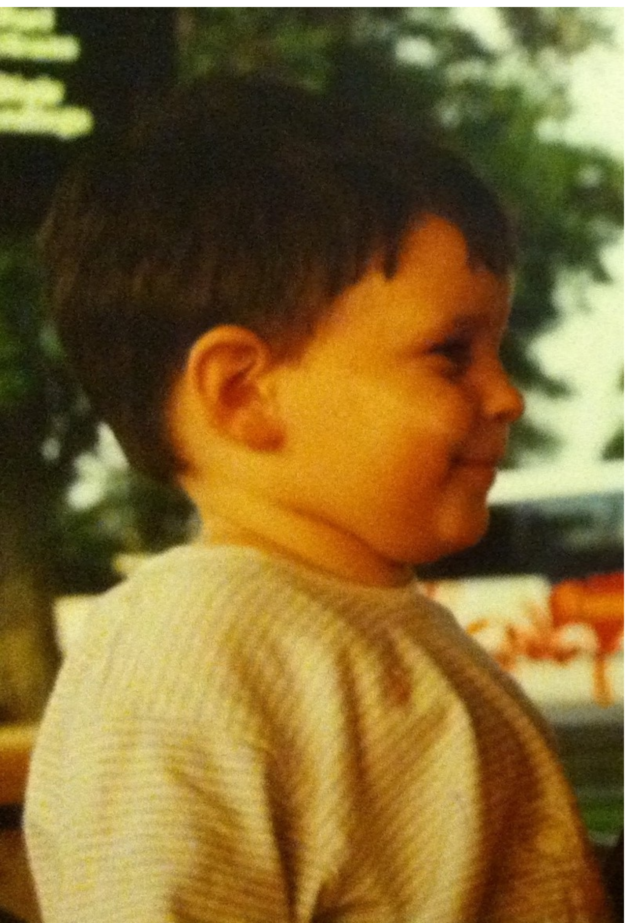






Construct

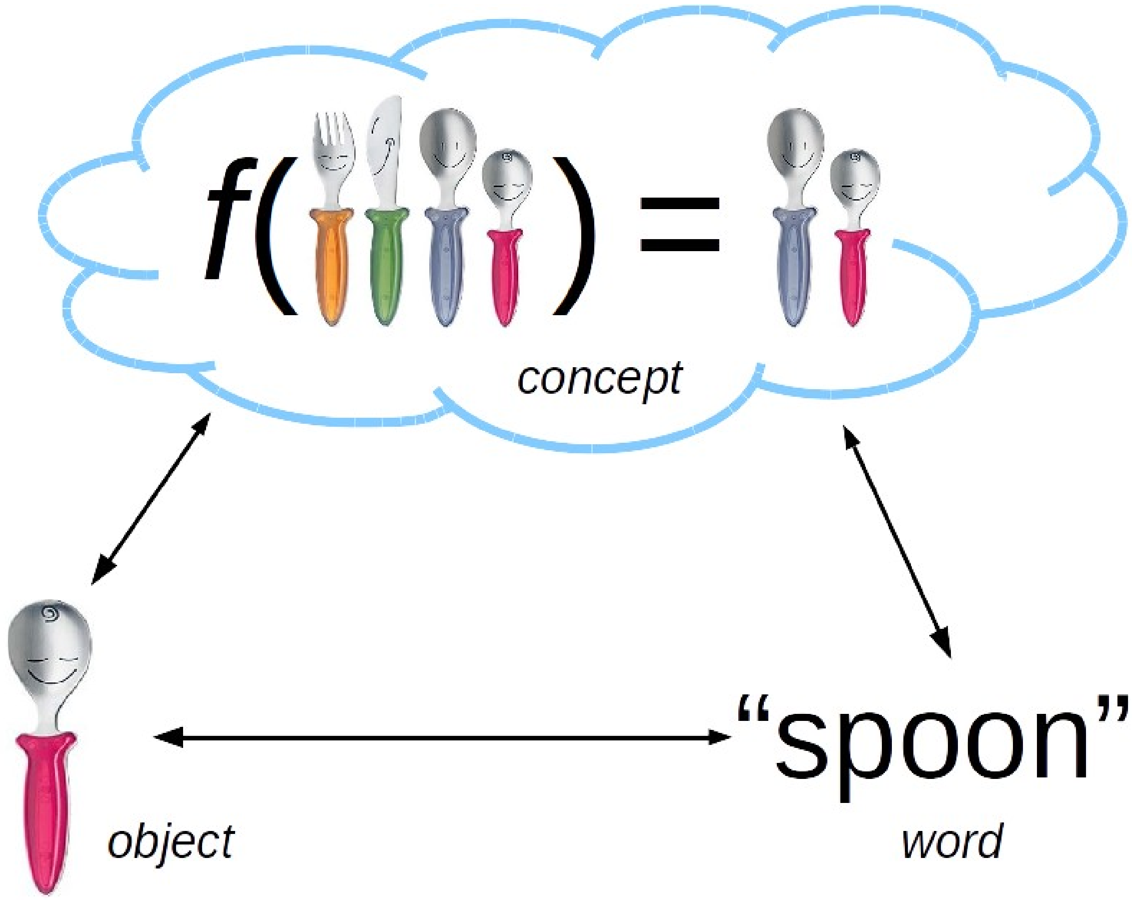




“Three koalas”

Our Approach

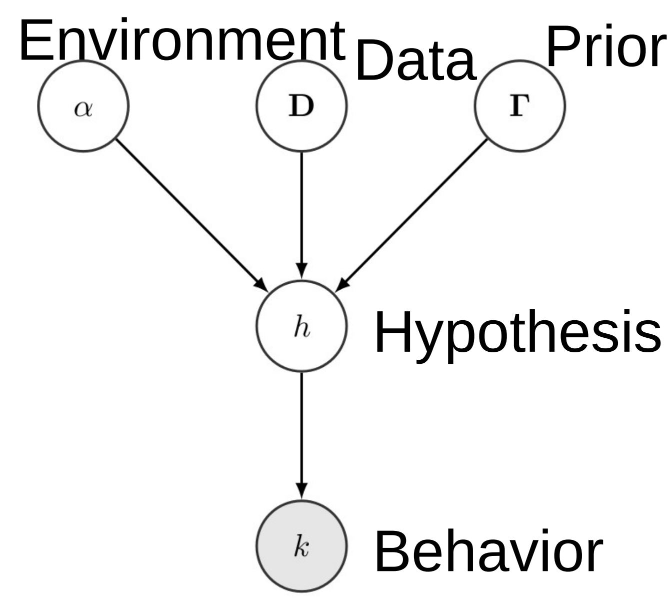
- Formalize word learning as logical **program** induction.



Peirce (1868)

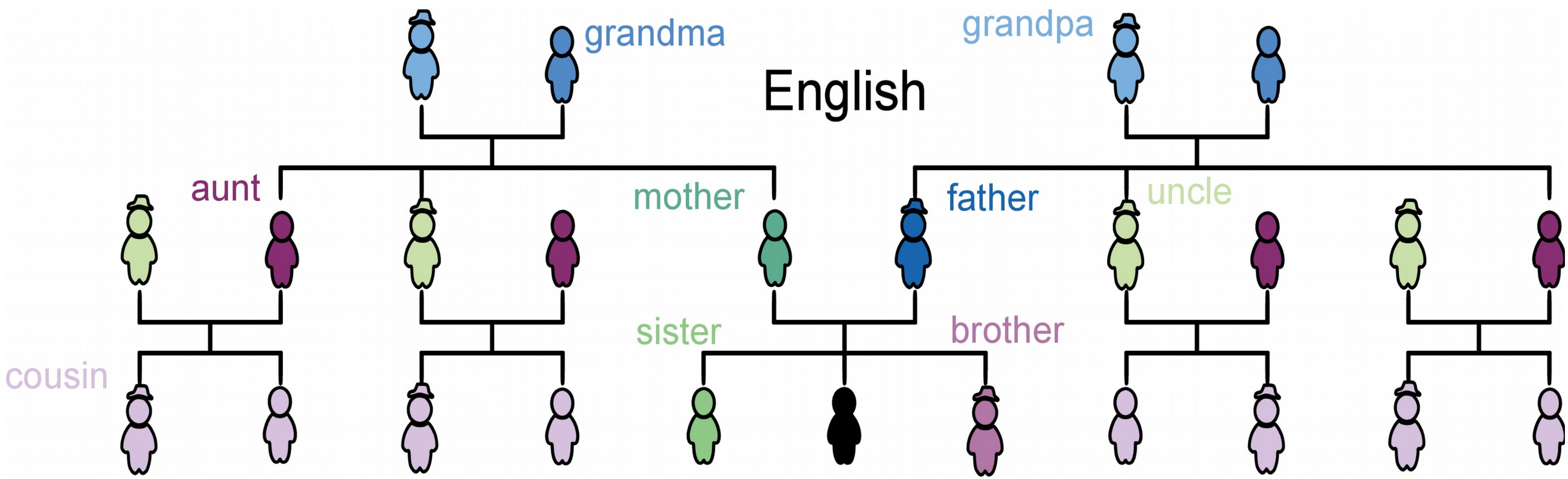
Ideal Learner Model

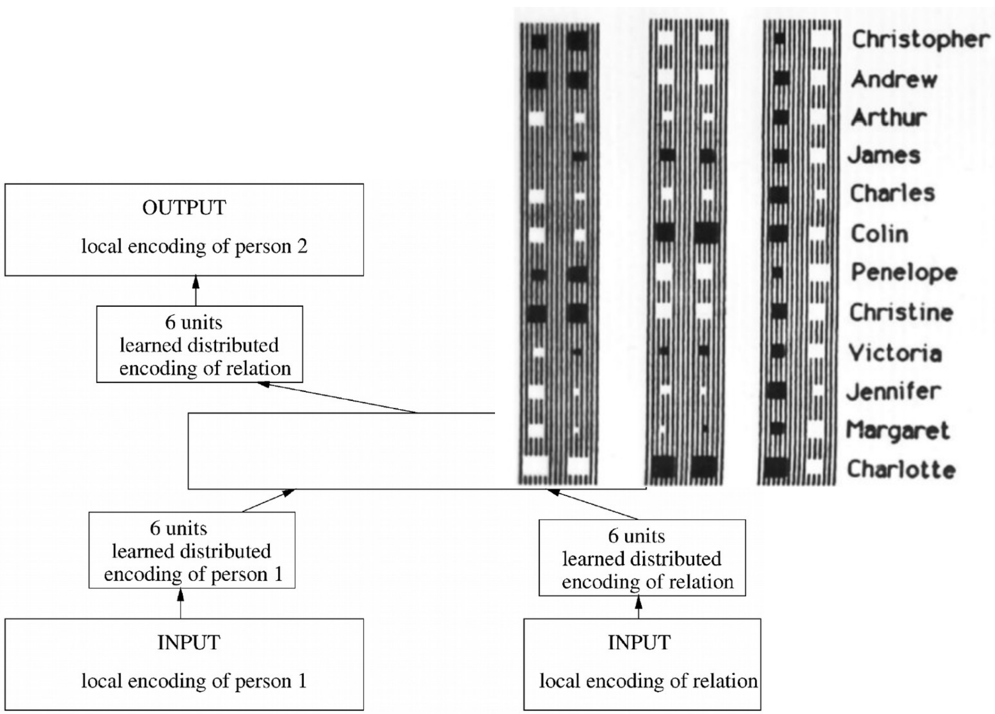
- Specify a **Hypothesis Space** of concepts
- Specify a **Prior** over hypotheses
- Specify a **Likelihood** function
- Specify the environment



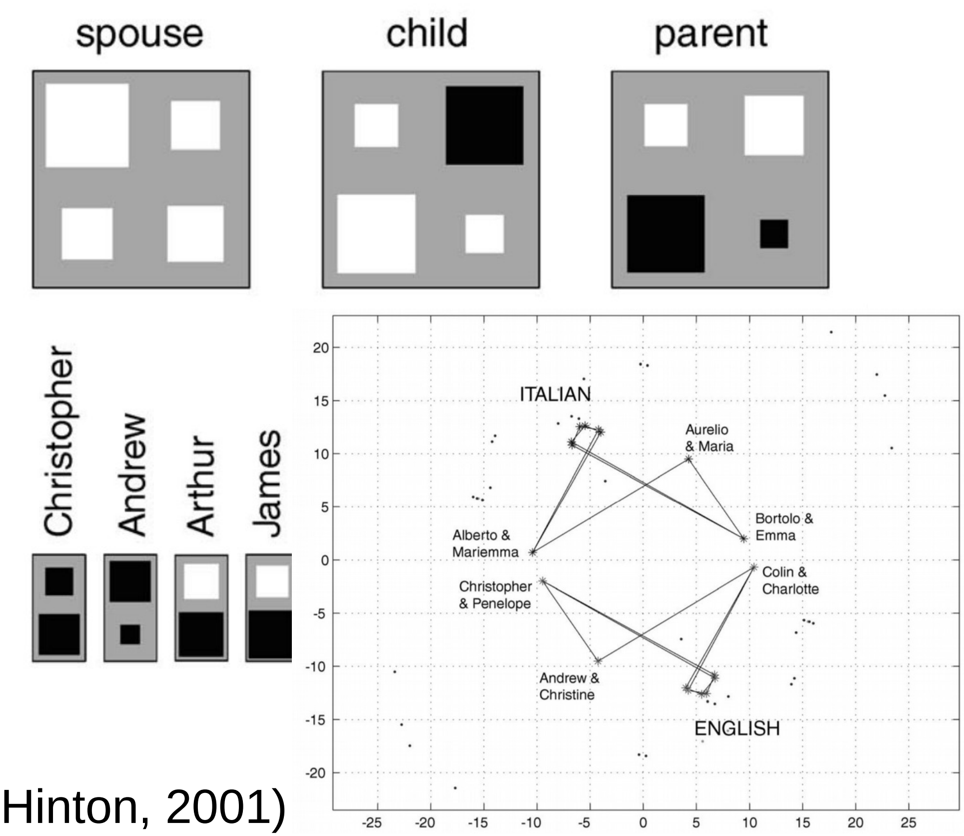
Ideal Learner Model

- Specify a **Hypothesis Space** of concepts
- Specify a **Prior** over hypotheses
- Specify a **Likelihood** function
- Specify the environment
- In a Bayesian learning model, learning corresponds to the movement of probability mass over a hypothesis space.





(Hinton, 1986)



(Paccanaro & Hinton, 2001)

Hypothesis Space

Tree Moving	Set Operations	Gender	Age	Inputs
Child Parent Spouse	Union Intersection Difference Complement	Female Male SameGender	SameGeneration ParentGeneration GparentGeneration	All Speaker (X) Individual

Hypothesis Space

Tree Moving	Set Operations	Gender	Age	Inputs
Child Parent Spouse	Union Intersection Difference Complement	Female Male SameGender	SameGeneration ParentGeneration GparentGeneration	All Speaker (X) Individual

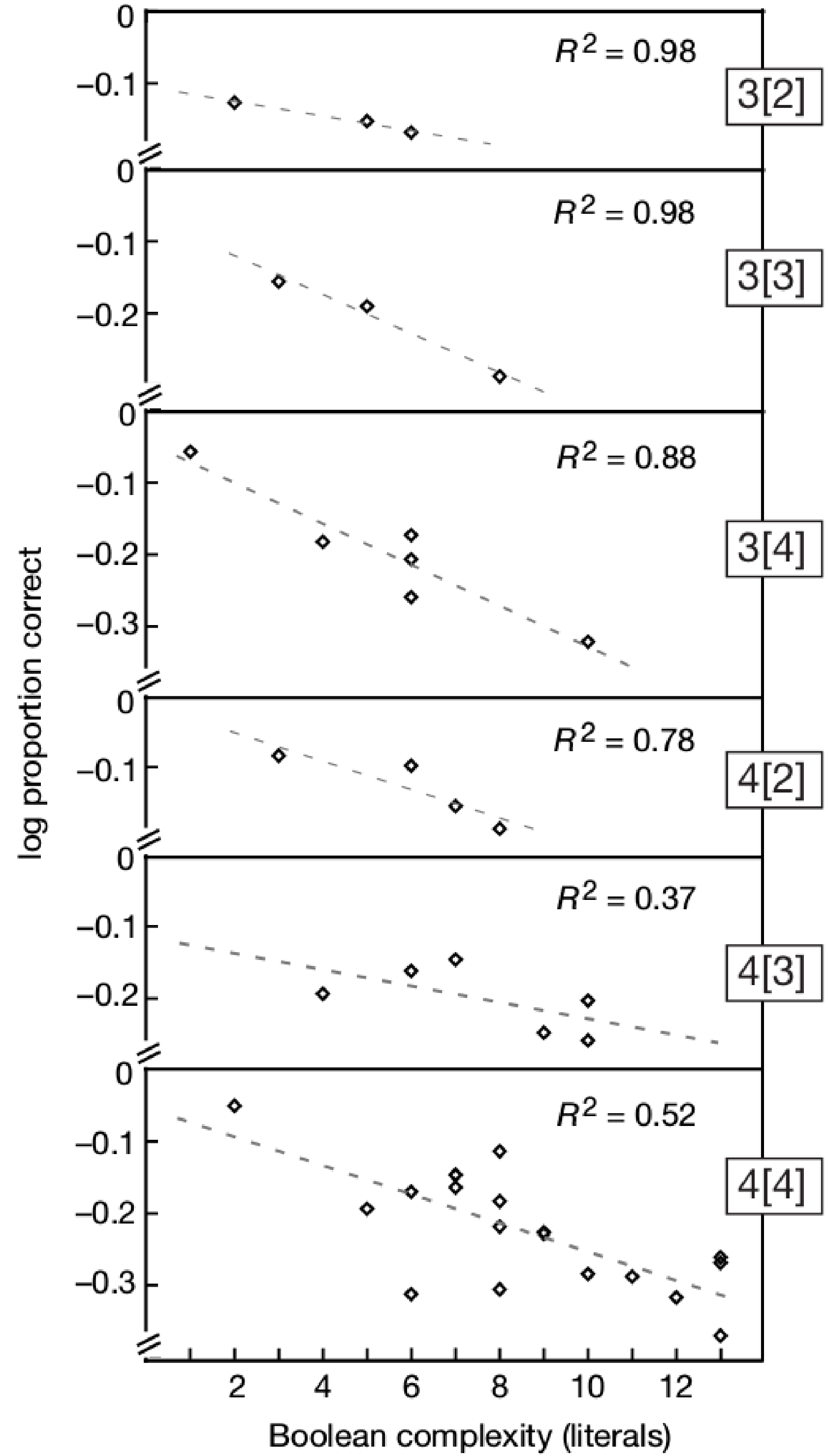
$\mathrm{SET} \xrightarrow{1} \mathrm{union(SET,SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{parent(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{generation0(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{male(SET)}$

$\mathrm{SET} \xrightarrow{1} \mathrm{intersection(SET,SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{child(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{generation1(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{female(SET)}$

$\mathrm{SET} \xrightarrow{1} \mathrm{difference(SET,SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{lateral(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{generation2(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{sameGender(SET)}$

$\mathrm{SET} \xrightarrow{1} \mathrm{complement(SET)}$ $\mathrm{SET} \xrightarrow{1} \mathrm{coreside(SET)}$ $\mathrm{SET} \xrightarrow{\frac{1}{37}} \mathrm{concreteReferent}$ $\mathrm{SET} \xrightarrow{1} \mathrm{all}$ $\mathrm{SET} \xrightarrow{10} \mathrm{X}$

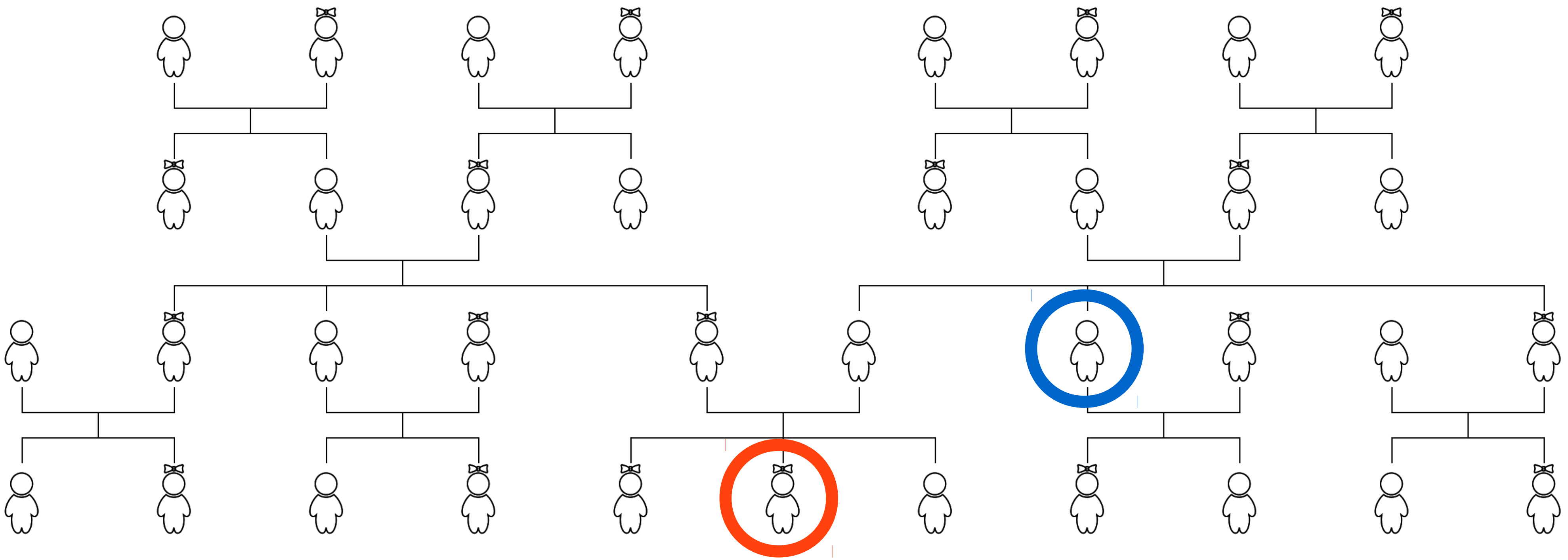
All
Tito
difference(generation0(X), sa
male(child(parent(par
female(child(parent(pa
intersection(lateral(child(parent(parent(X)))),
male(parent(X
ifference(male(generation0(X)), child(male(c
male(child(parent(female(
difference(generation0(X), c
male(difference(generation1
male(child(parent(female(difference(genera
female(parent(X
female(parent(pare
female(parent(male(pare
difference(female(generation1(X)), c
female(difference(generation
ifference(male(generation0(X)), child(female
rence(male(generation0(Tito)), child(female(
male(difference(child(parent(male(pa
Mary
X



(Feldman, 2000)

Where does data come from?

- Context:

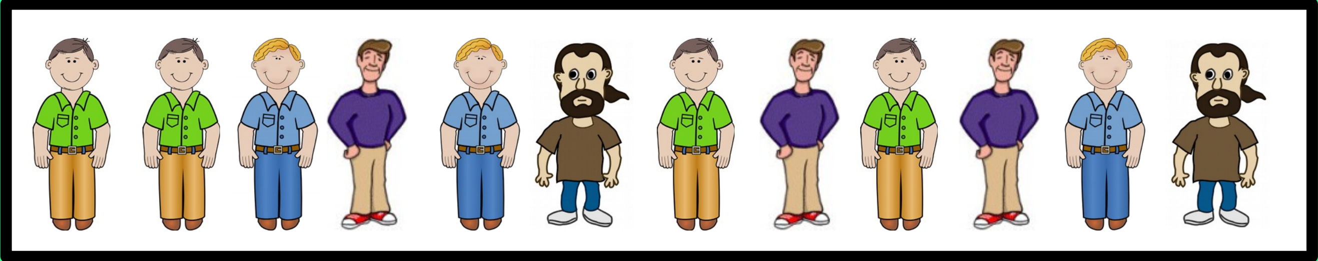


- Data Point:
  - Context
  - Word *uncle*
  - Speaker
  - Referent

How do we fit to the data?

- Size Principle Likelihood (e.g., Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007)

Data Distribution:



$$P(d|h) = \alpha \boxed{\frac{\delta_{d \in h}}{|h|}} + (1-\alpha) \boxed{\frac{1}{|O|}}$$

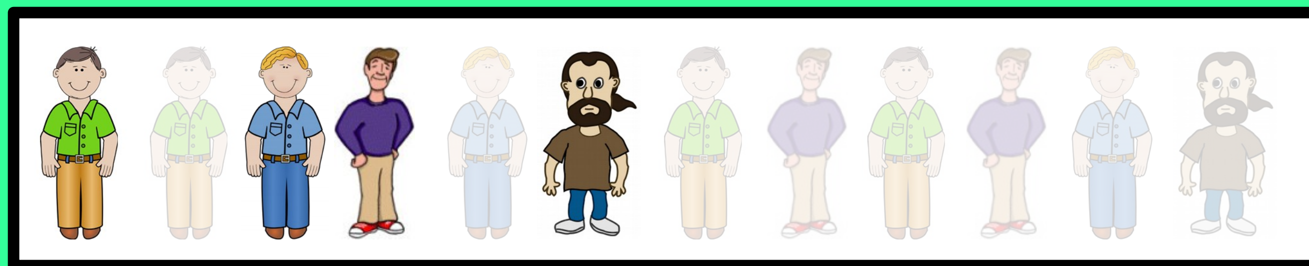
Sampling from the hypothesized concept.

Sampling from everything in the world.

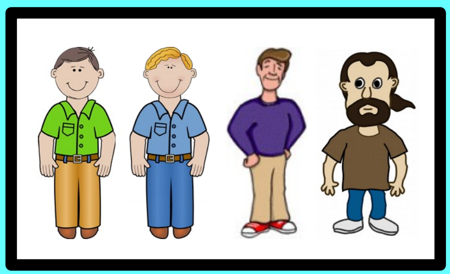
How do we fit to the data?

- Size Principle Likelihood (e.g., Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007)

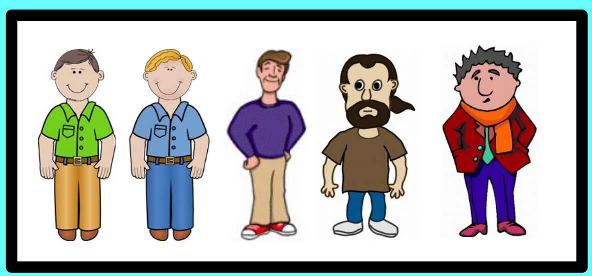
Data Distribution:



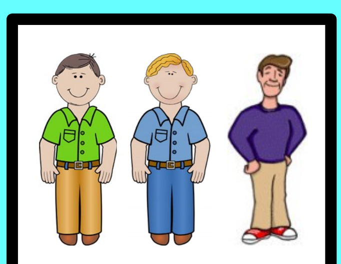
Hypothesis A:



Hypothesis B:



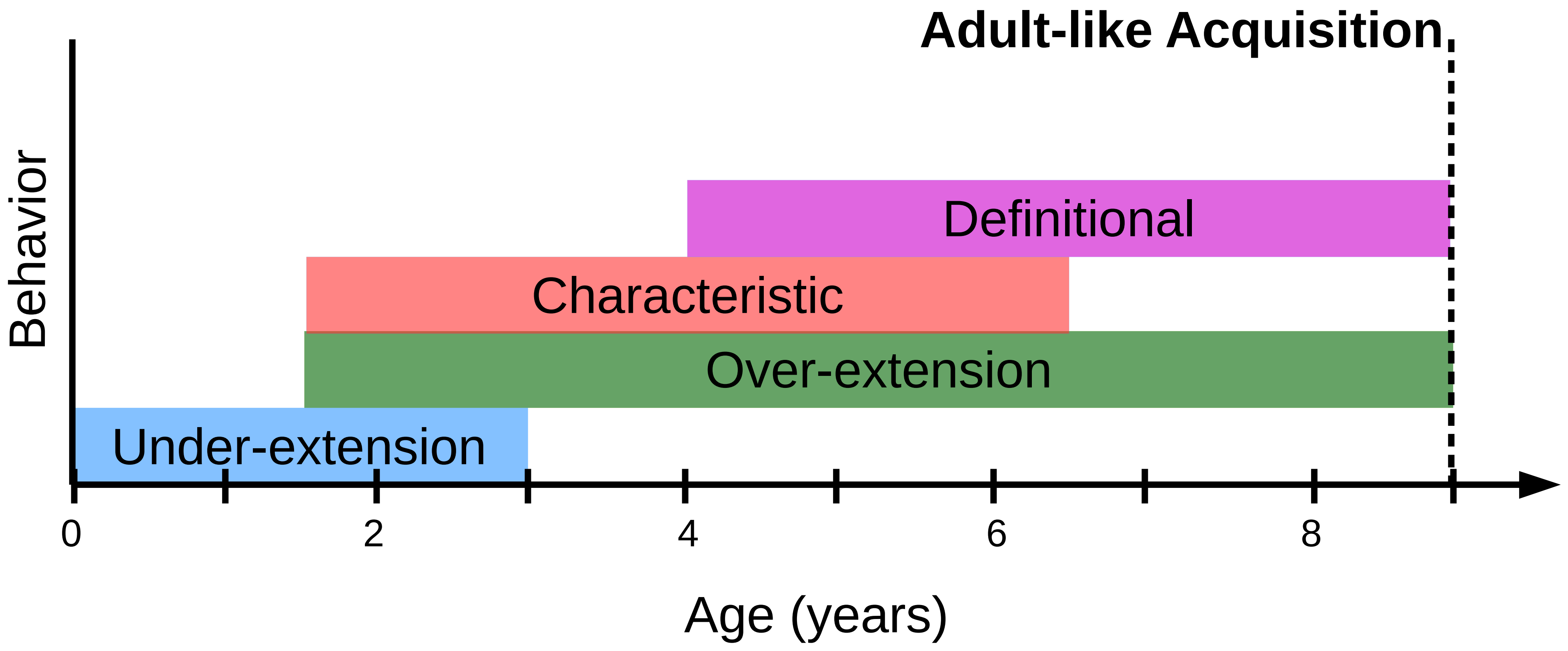
Hypothesis C:



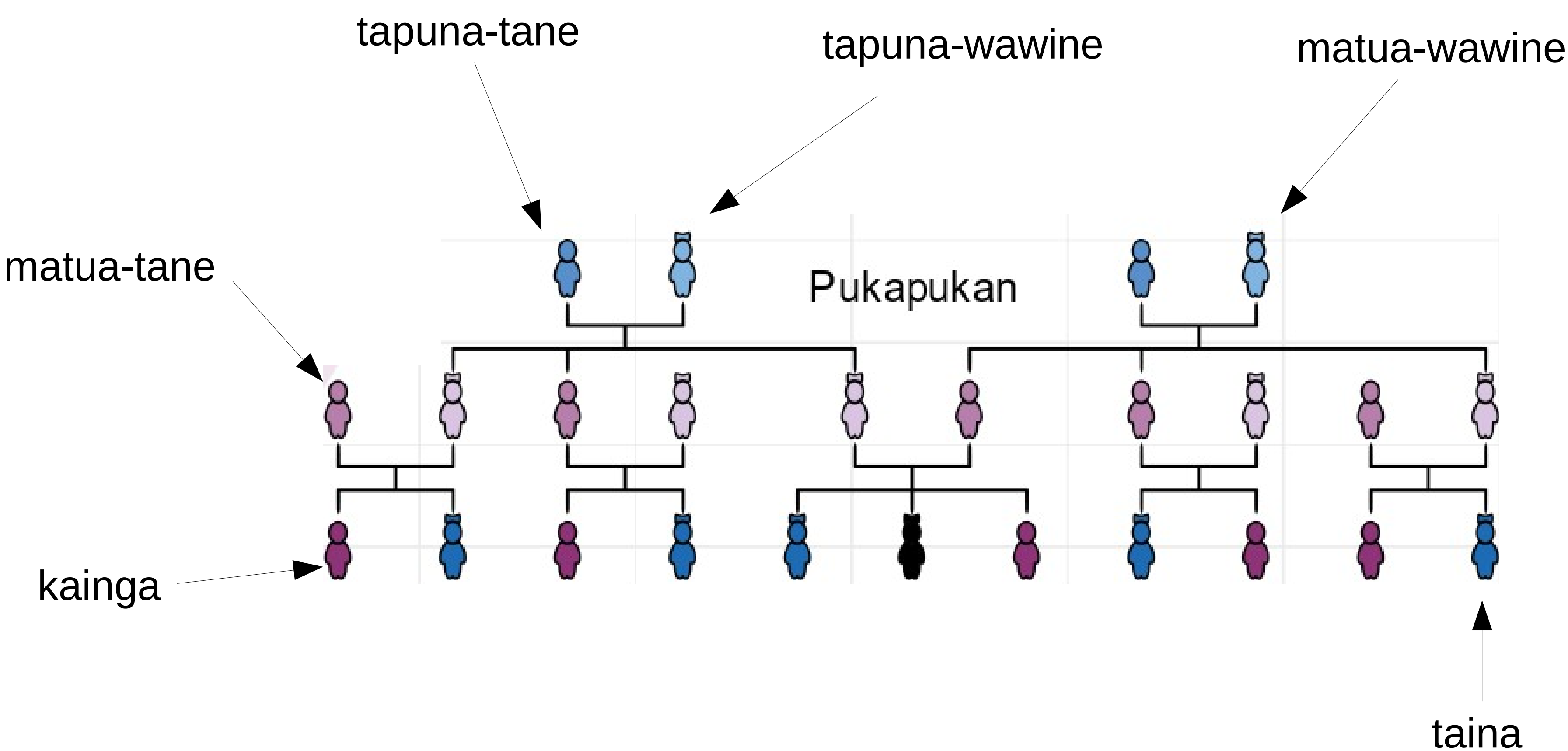
Ideal Learner Model

- Specify a **Hypothesis Space** of concepts
- Specify a **Prior** over hypotheses
- Specify a **Likelihood** function
- Specify the environment
- In a Bayesian learning model, learning corresponds to the movement of probability mass over a hypothesis space.

Kinship Acquisition Phenomena

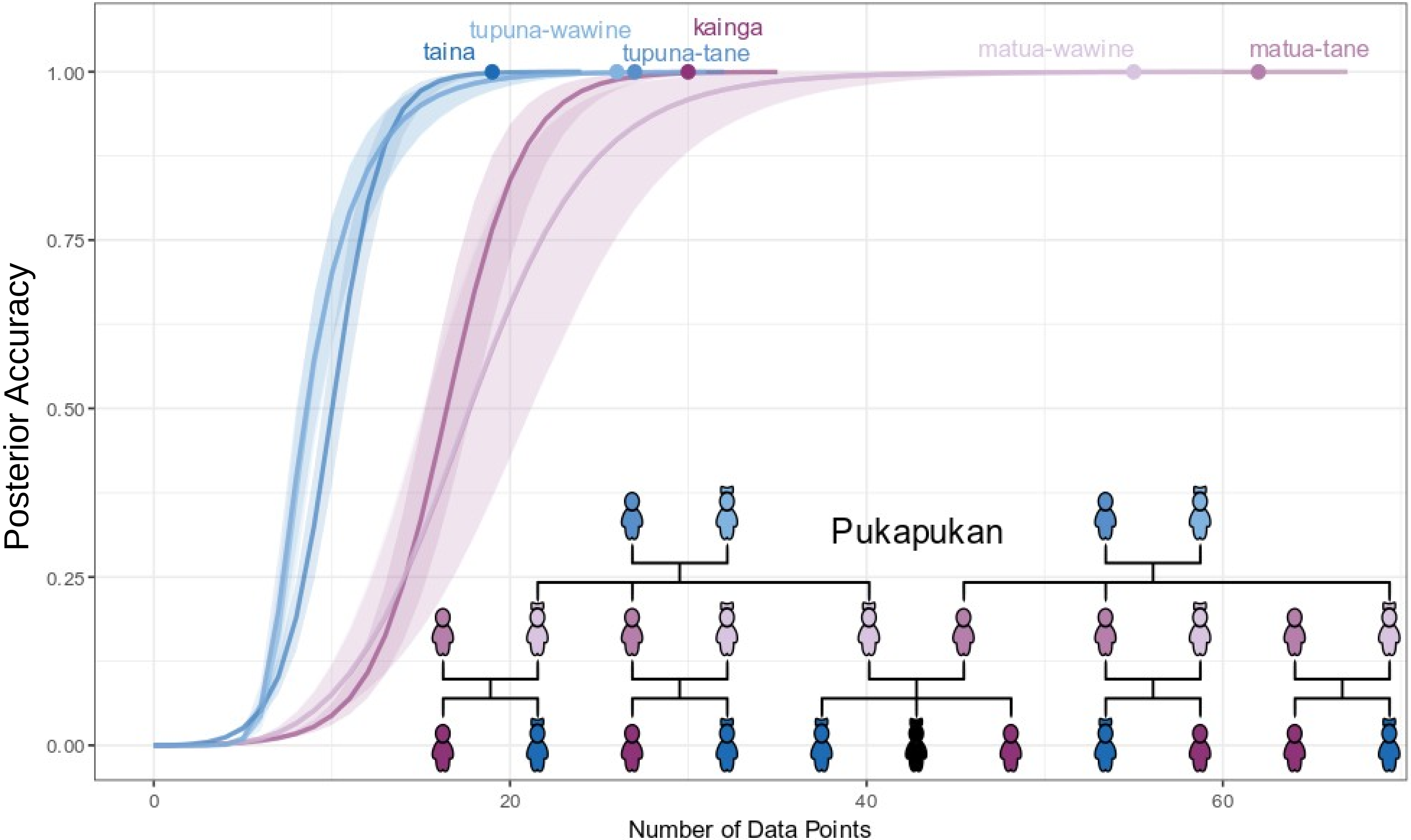


Kids learn **their** kinship system

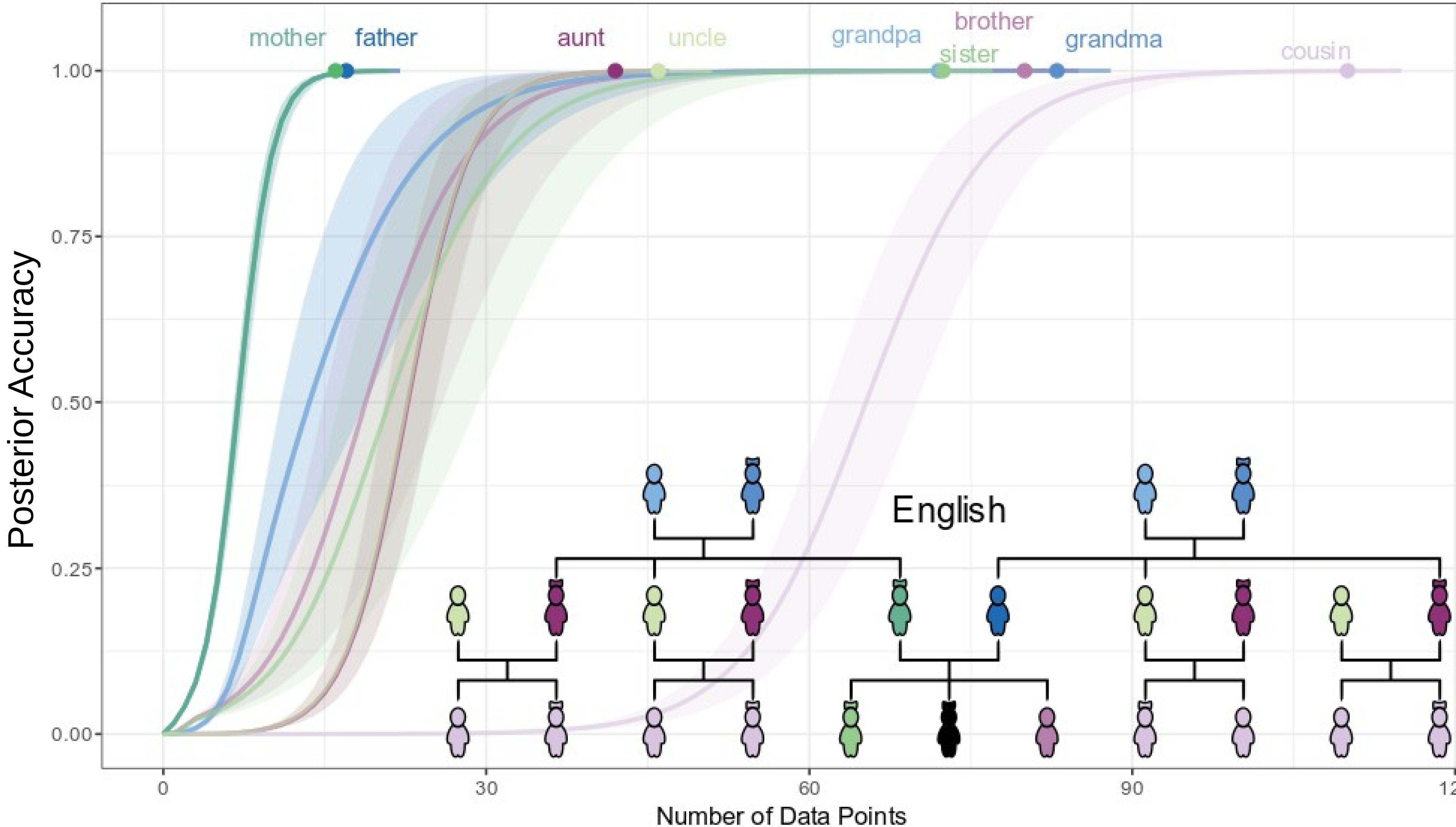


Marck (1996)

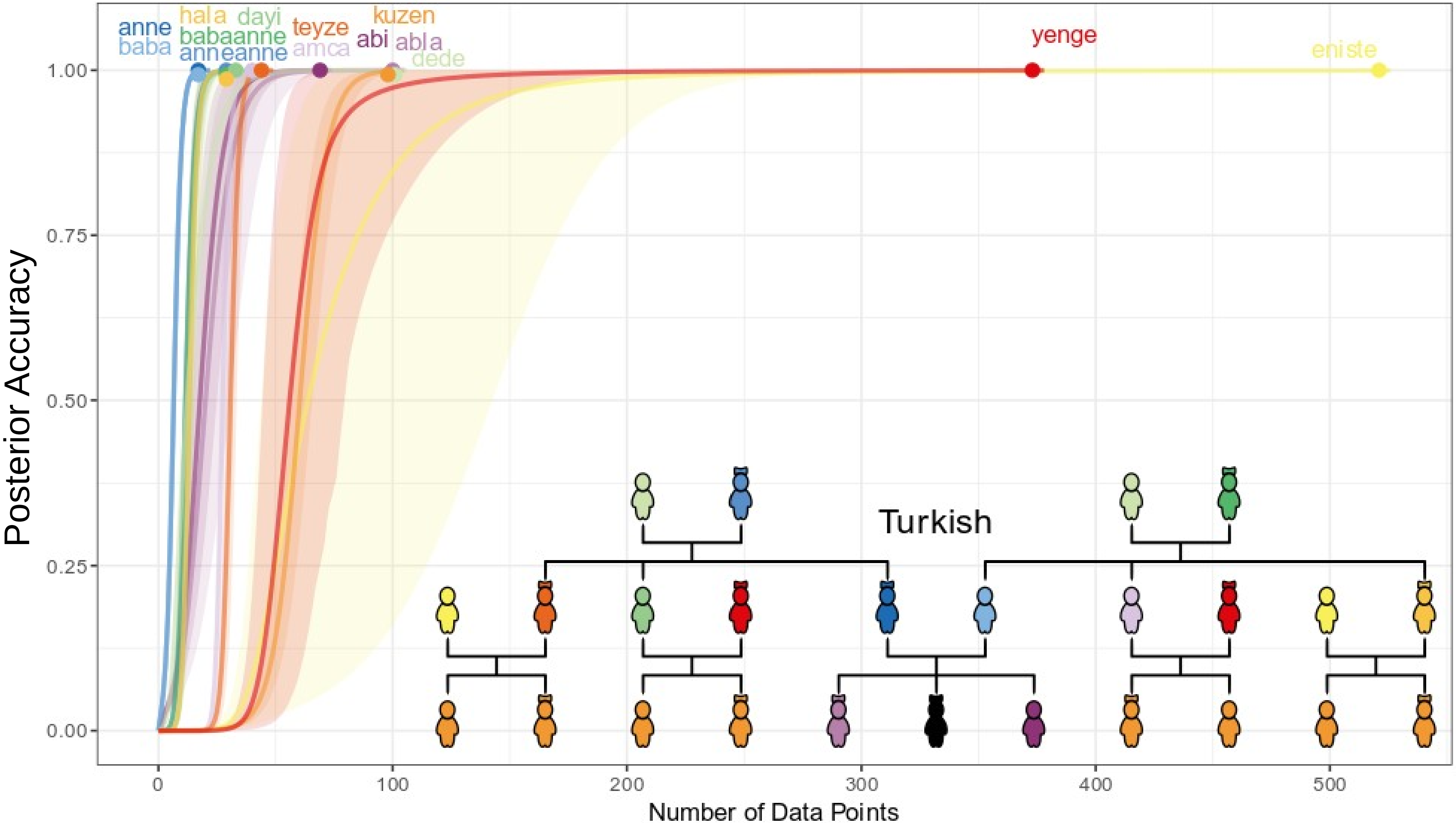
Kids learn **their** kinship system



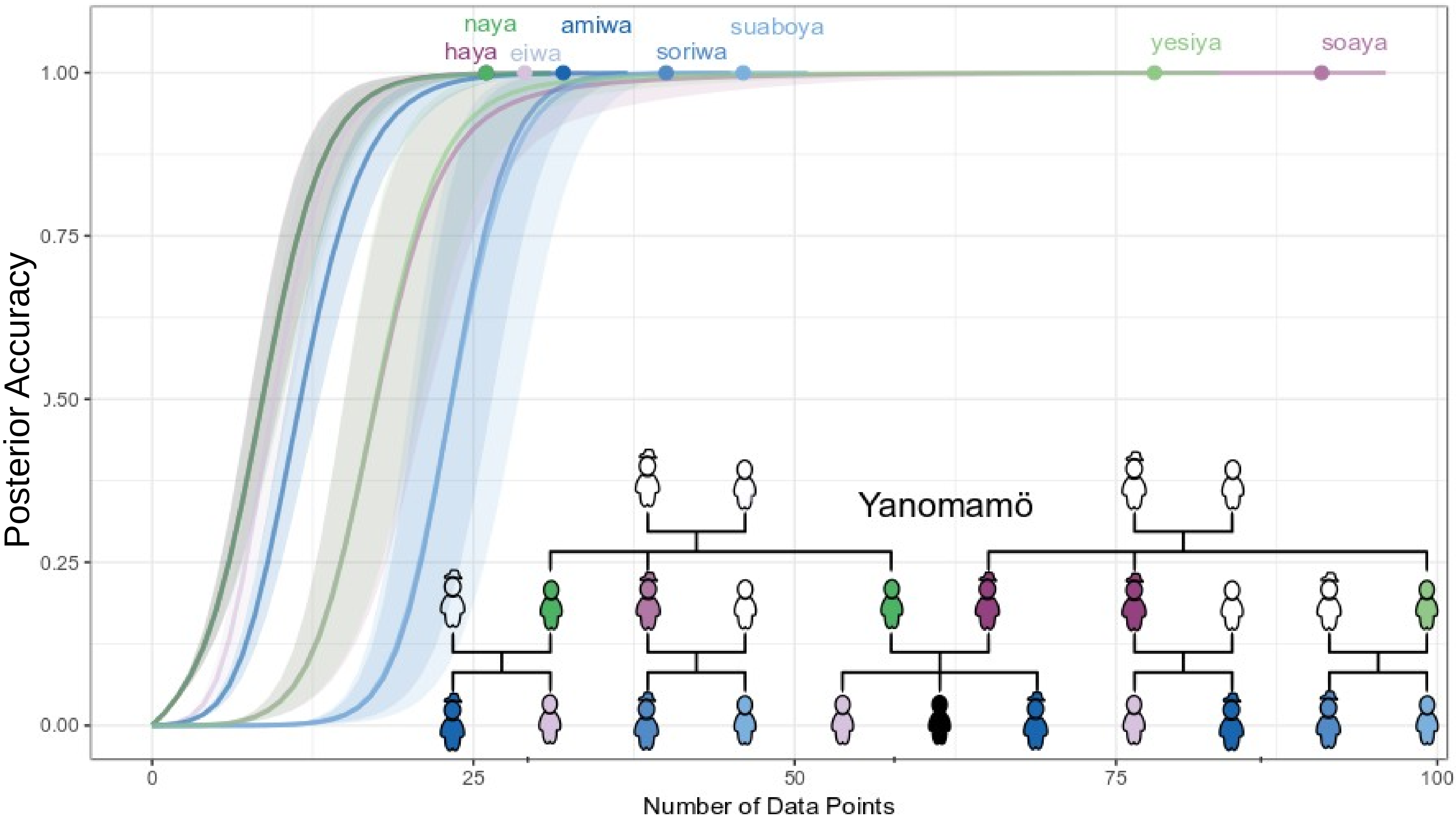
Kids learn **their** kinship system



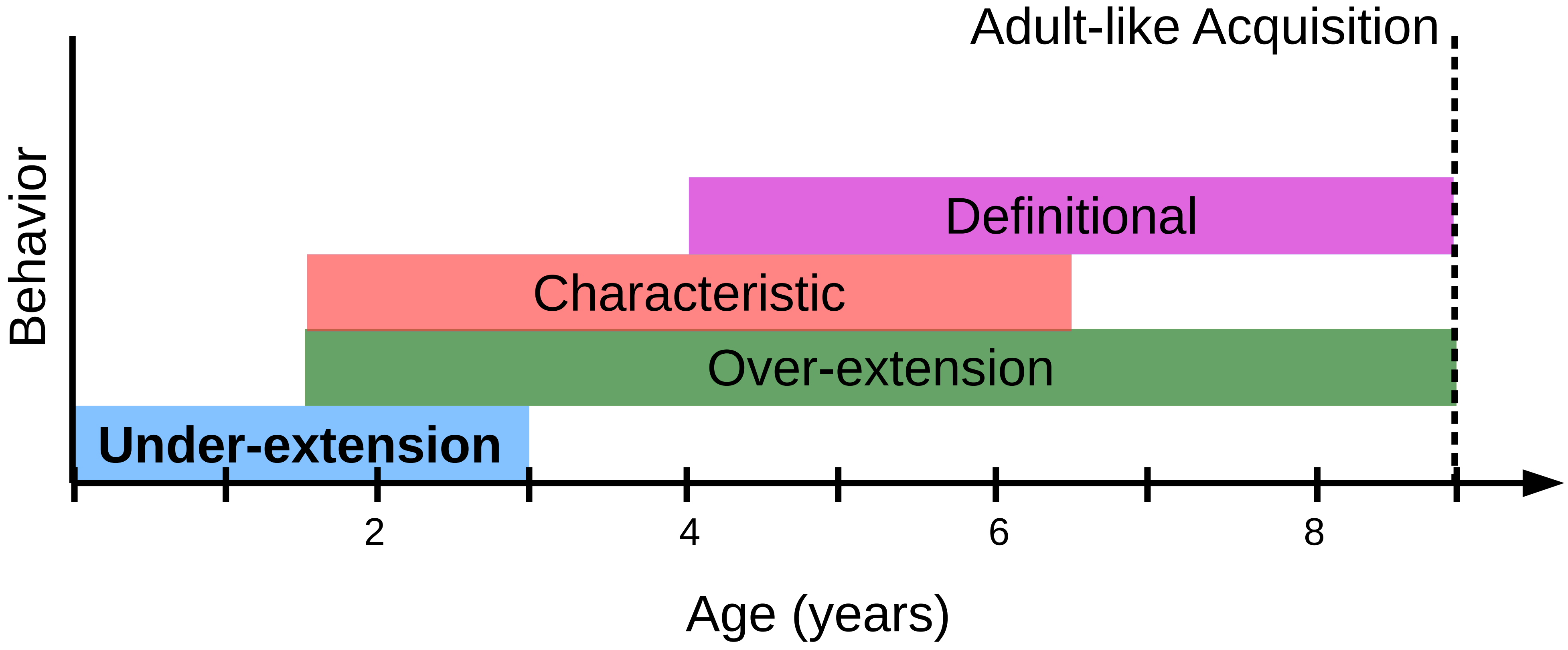
Kids learn **their** kinship system



Kids learn **their** kinship system



Young kids prefer concrete referents



Young kids prefer concrete referents

I: What is an uncle?

S: Uncle Anthony

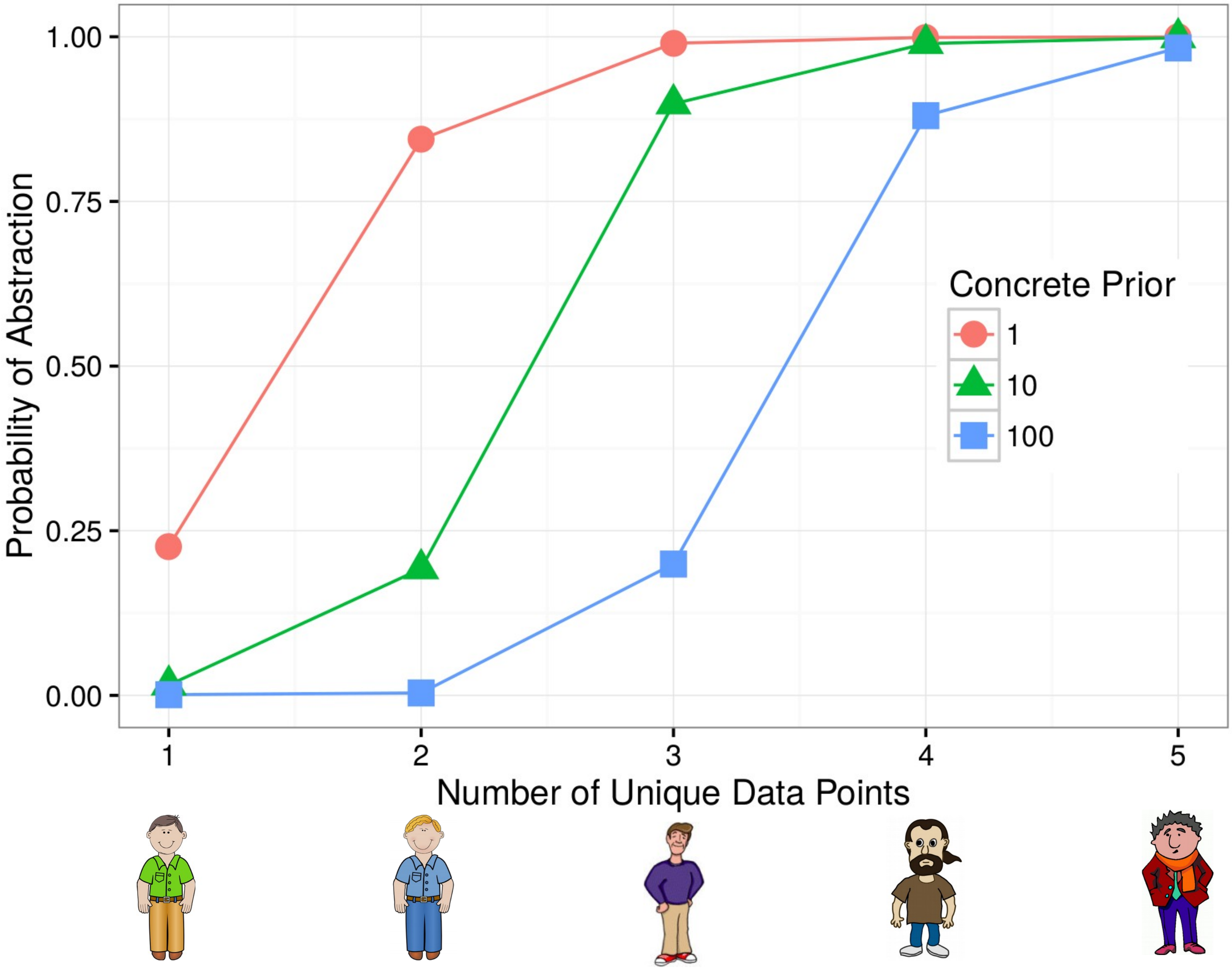
I: Tell me everything you know about an uncle.

S: Uncle Henry

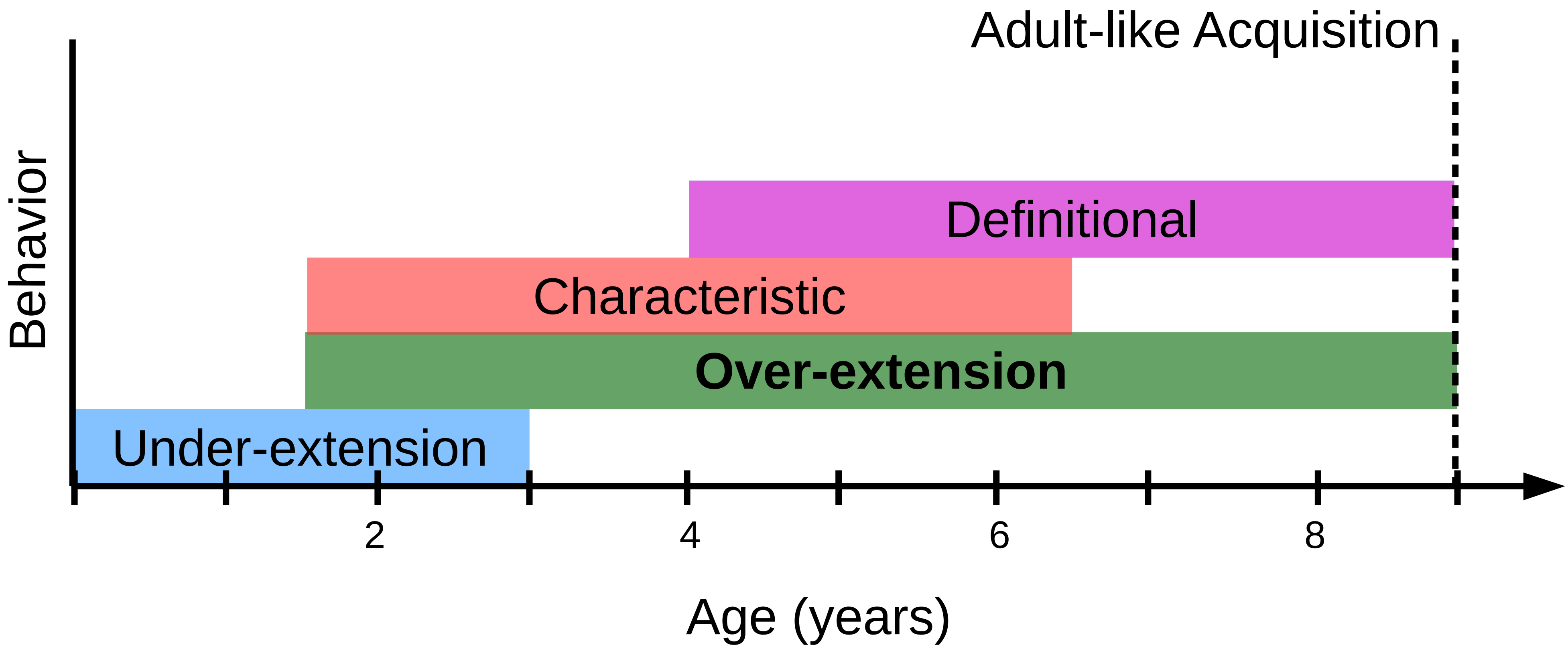
3;0 YO

(Benson & Anglin, 1987)

Young kids prefer concrete referents



Older kids over-generalize



Older kids over-generalize

I: Tell me everything you know about an uncle.

S: He’s a man.

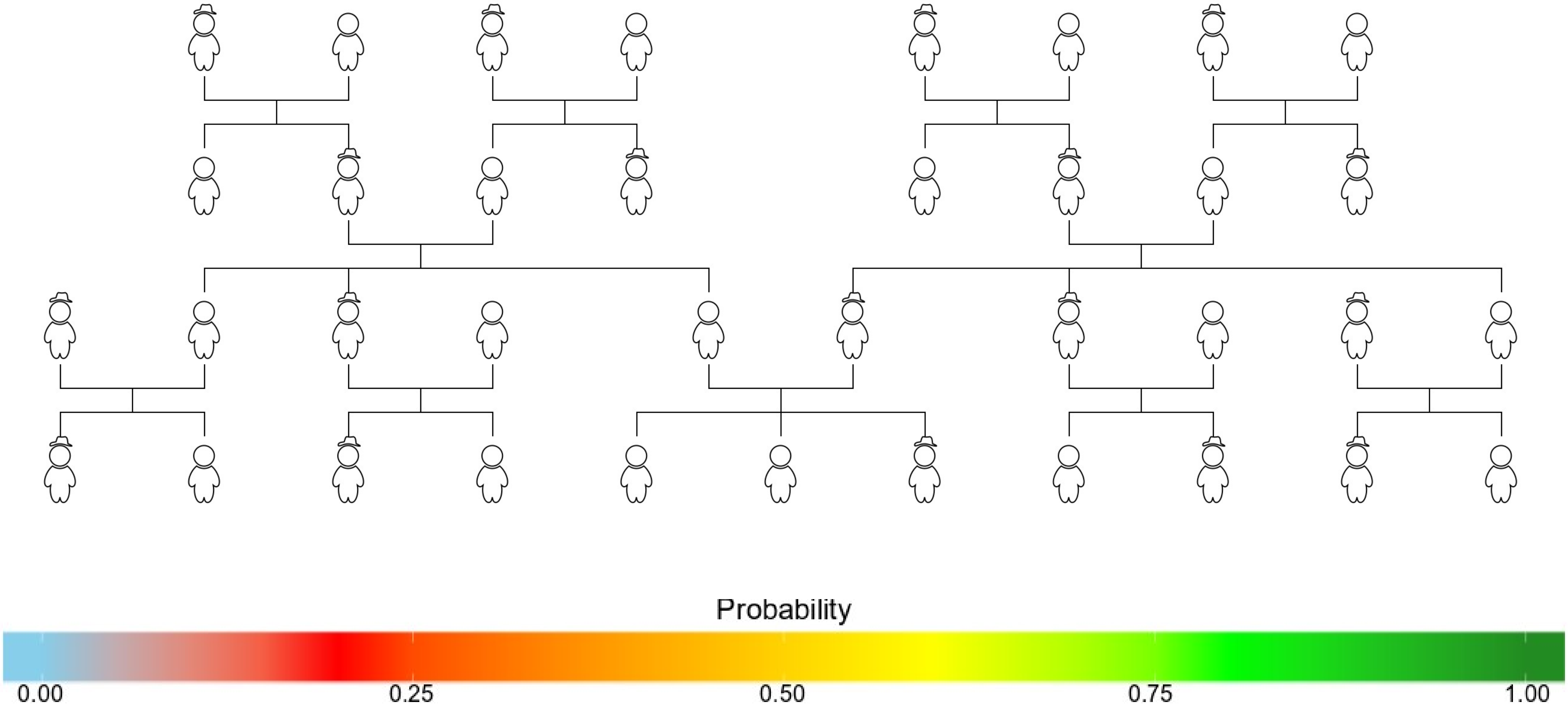
I: What kind of a thing is an uncle?

S: He’s a man.

5;4 YO

(Benson & Anglin, 1987)

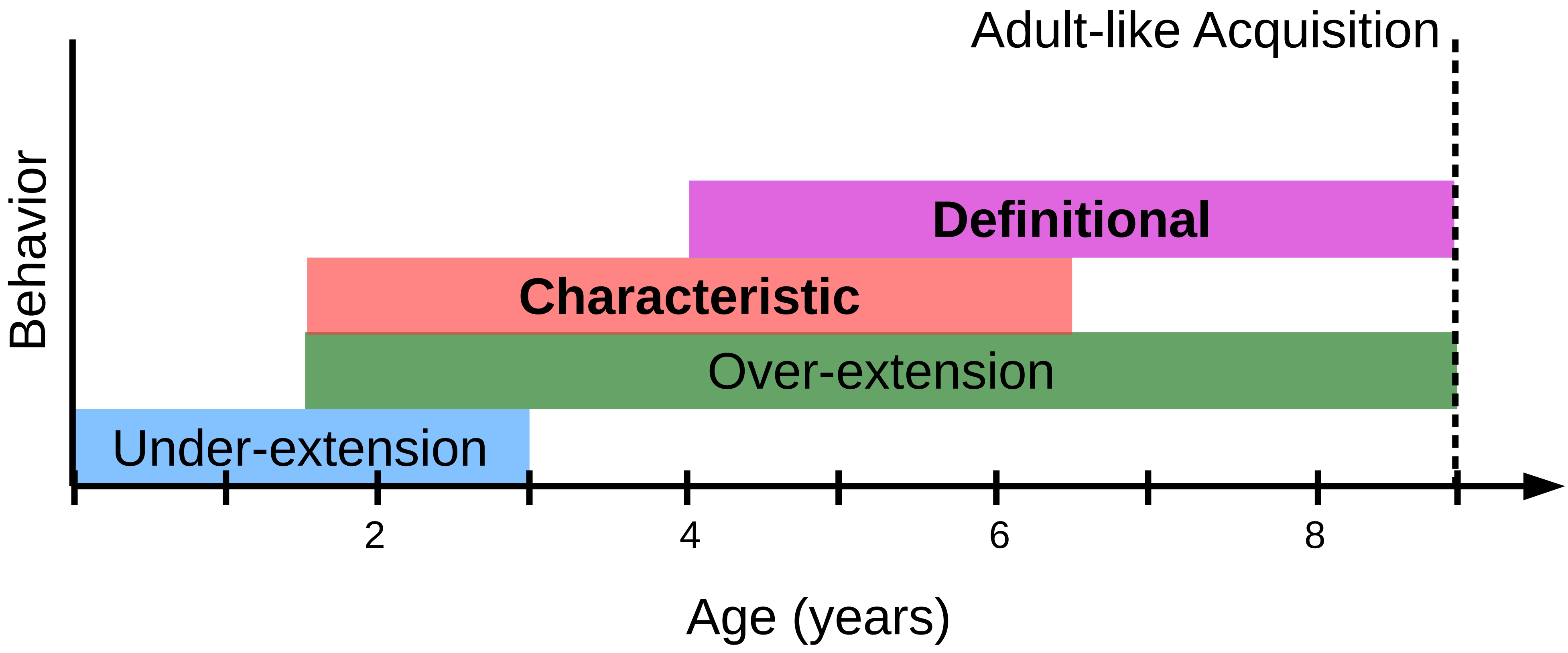
Older kids over-generalize



uncle

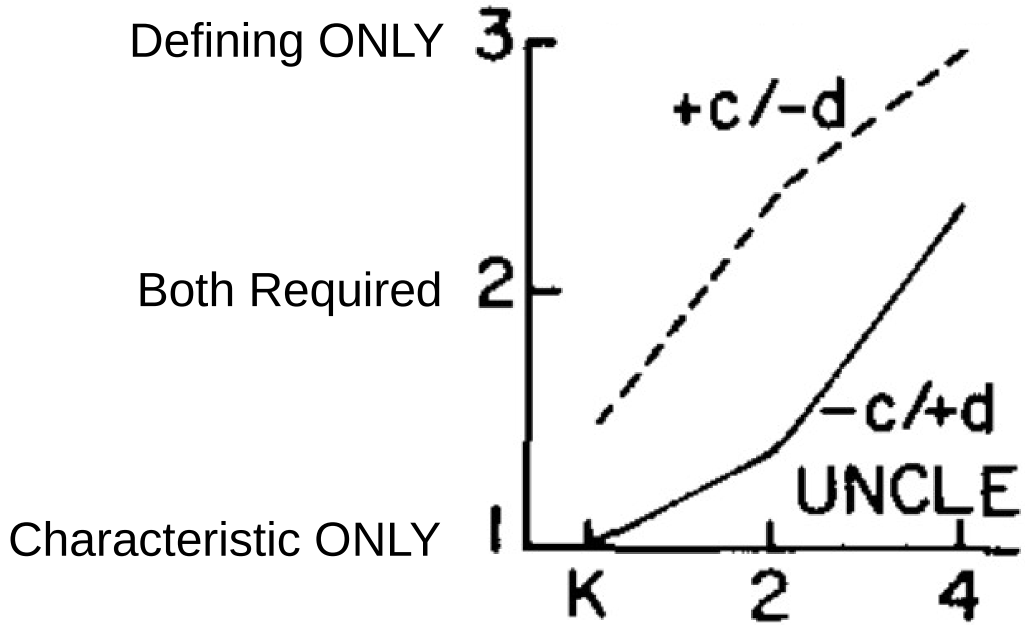
https://mollicaf.github.io/kinship.html

Generalization shifts from characteristic to defining features



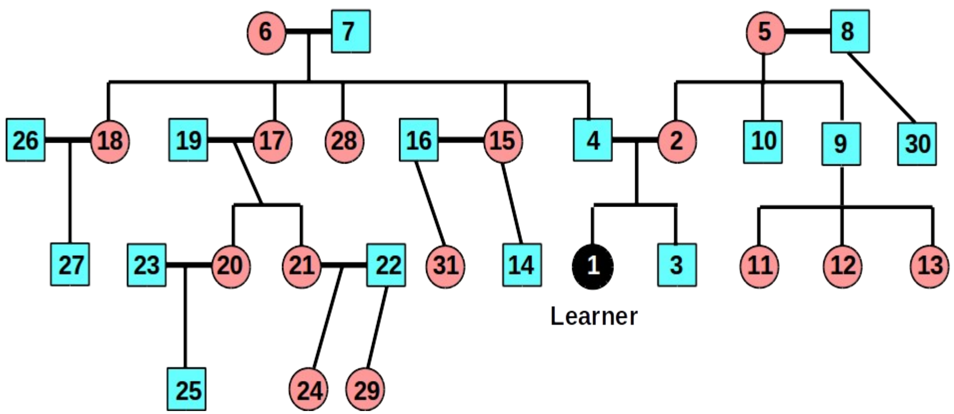
Generalization shifts from characteristic to defining features

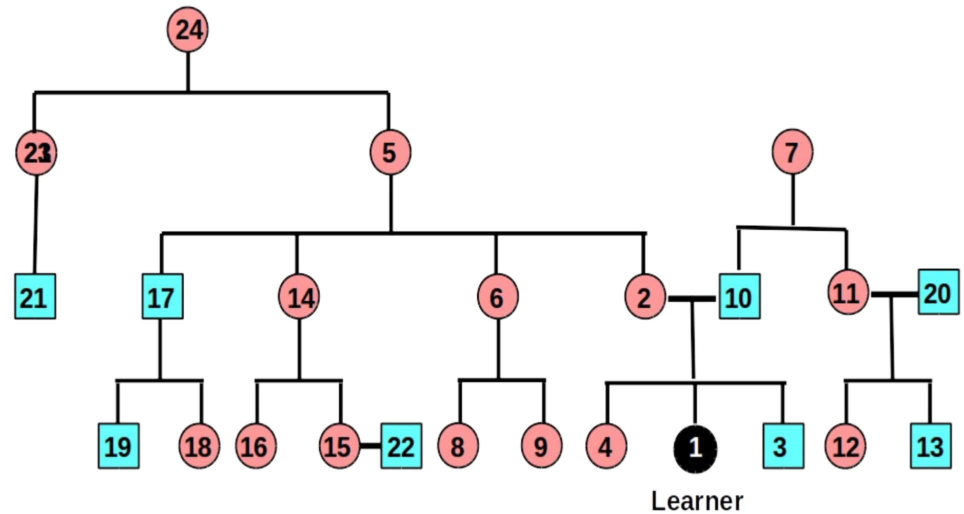
- This man your daddy's age loves you and your parents and loves to visit and bring presents, but he's not related to your parents at all. He's not your mommy or daddy's brother or sister or anything like that. Could that be an uncle?
- Suppose your mommy has all sorts of brothers, some very old and some very, very young. One of your mommy's brothers is so young he's only 2 years old. Could that be an uncle?

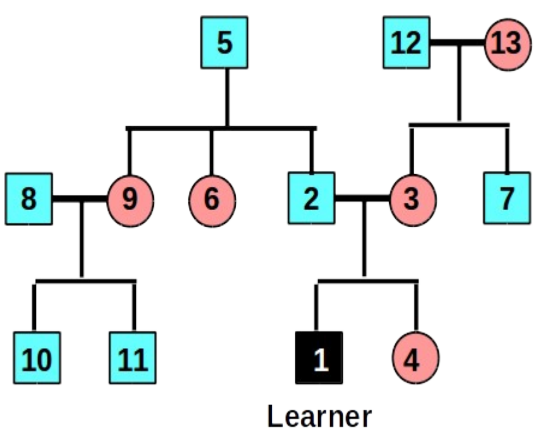


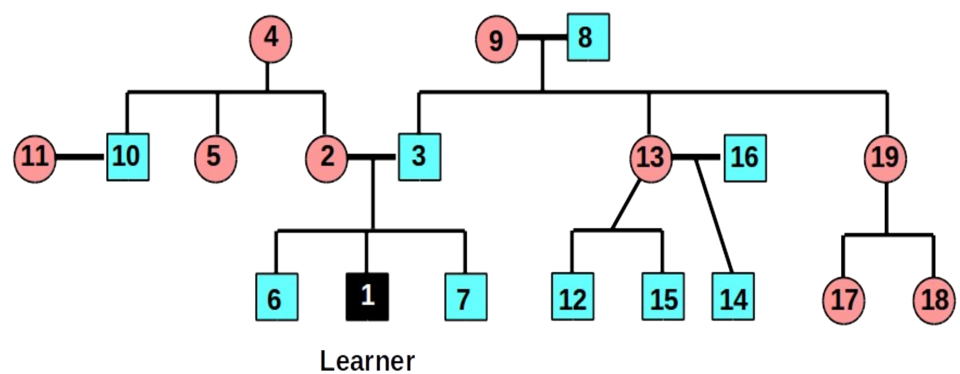
(Keil & Batterman, 1984)

Family Tree Data Collection (N=4)

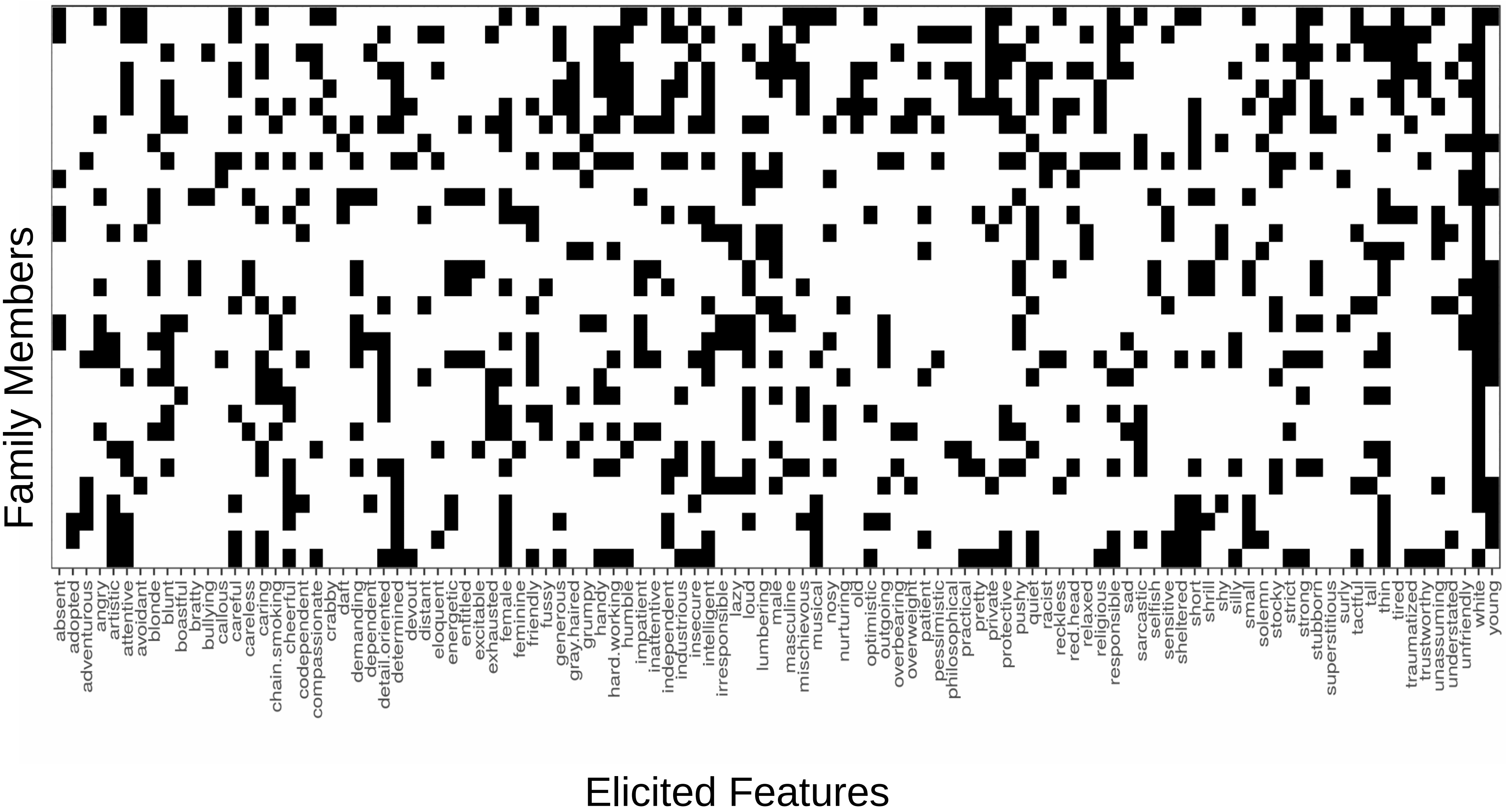








Feature Matrix



Hypothesis Space

Defining:

Tree Moving	Set Operations	Gender	Age	Inputs
Child Parent Spouse	Union Intersection Difference Complement	Female Male SameGender	SameGeneration ParentGeneration GparentGeneration	All Speaker Individual

Characteristic:

Set Operations	Inputs
Union Intersection Difference Complement	Feature

Hypothesis Space

Defining Examples:

1. All
2. Tito
3. Male(Parent(X))

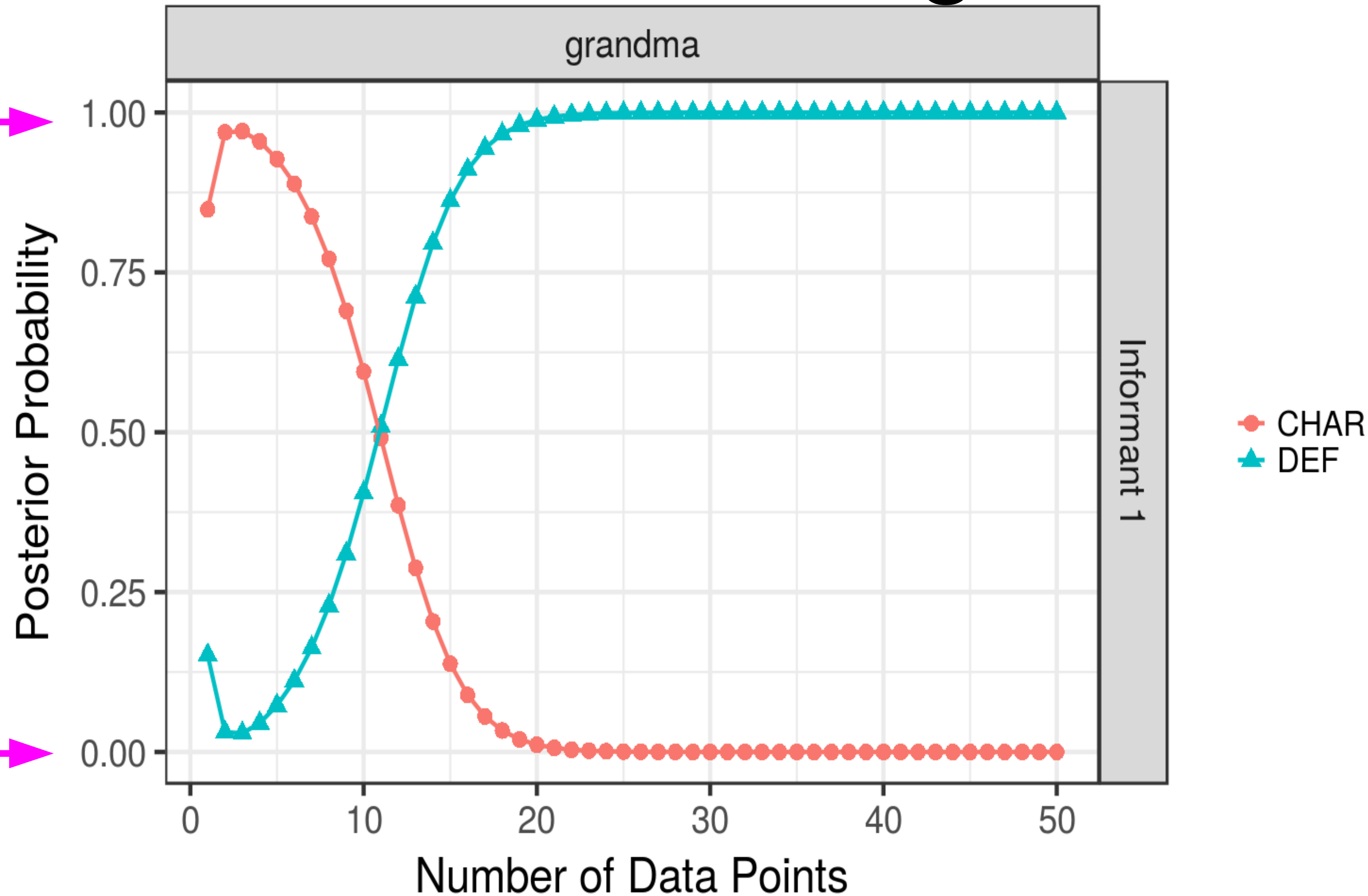
Characteristic Examples:

1. outgoing
2. union(outgoing, nosy)
3. difference(red-head, sarcastic)

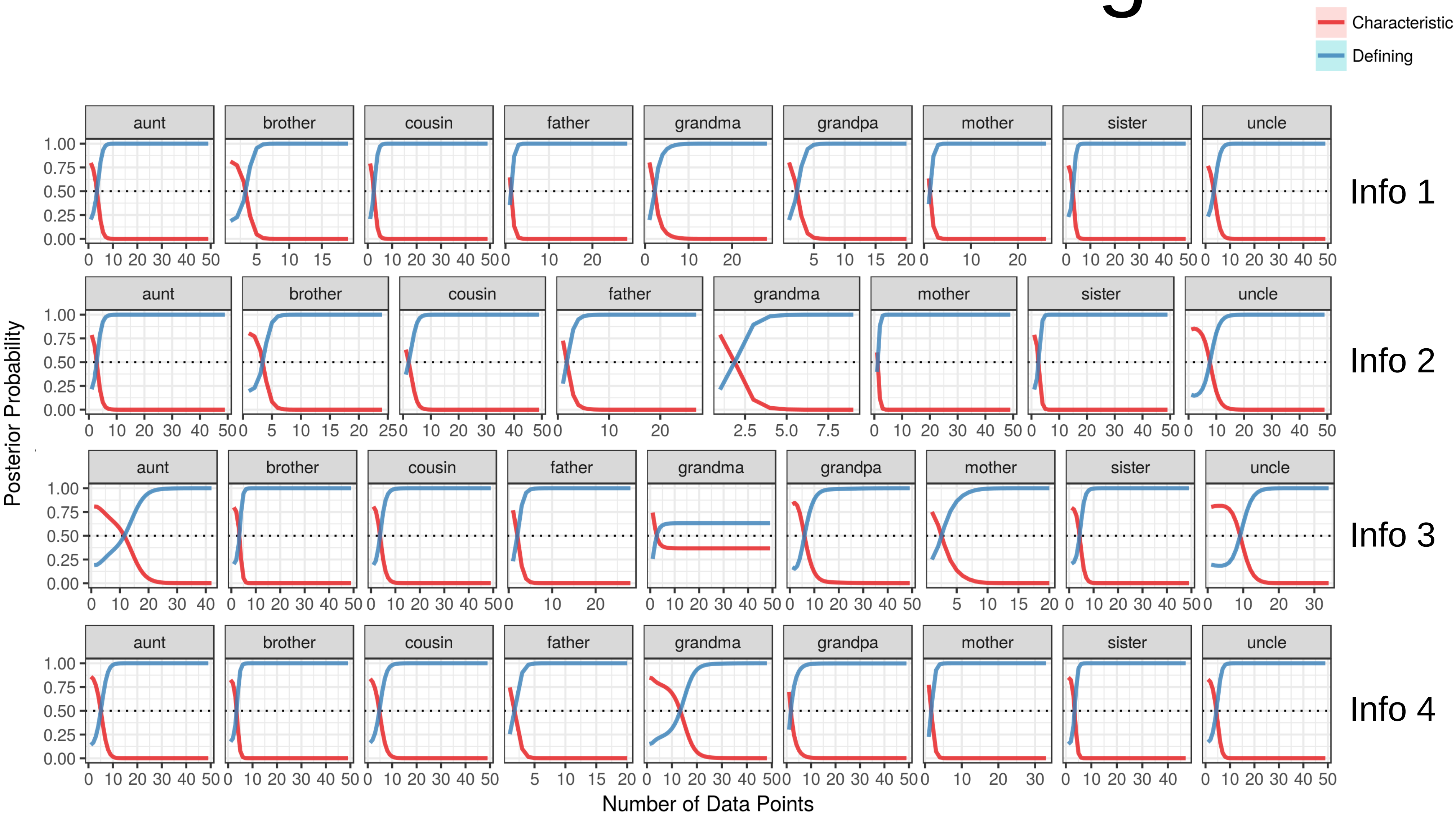
Characteristic-to-Defining Shift

Likely to use

Unlikely to use



Characteristic-to-Defining Shift



References

- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407(6804)*, 630-633.
- Fodor, J. A. (1975). The language of thought. Harvard University Press.
- Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science, 32(1)*, 108-154.
- Kemp, C. (2012). Exploring the conceptual universe. *Psychological review, 119(4)*, 685.
- Mollica, F., & Piantadosi, S. T. (in press). Logical word learning: The case of kinship. *Psychonomic Bulletin & Review*
- Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition, 123(2)*, 199-217.
- Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological review, 123(4)*, 392.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological monographs: General and applied, 75(13)*, 1.