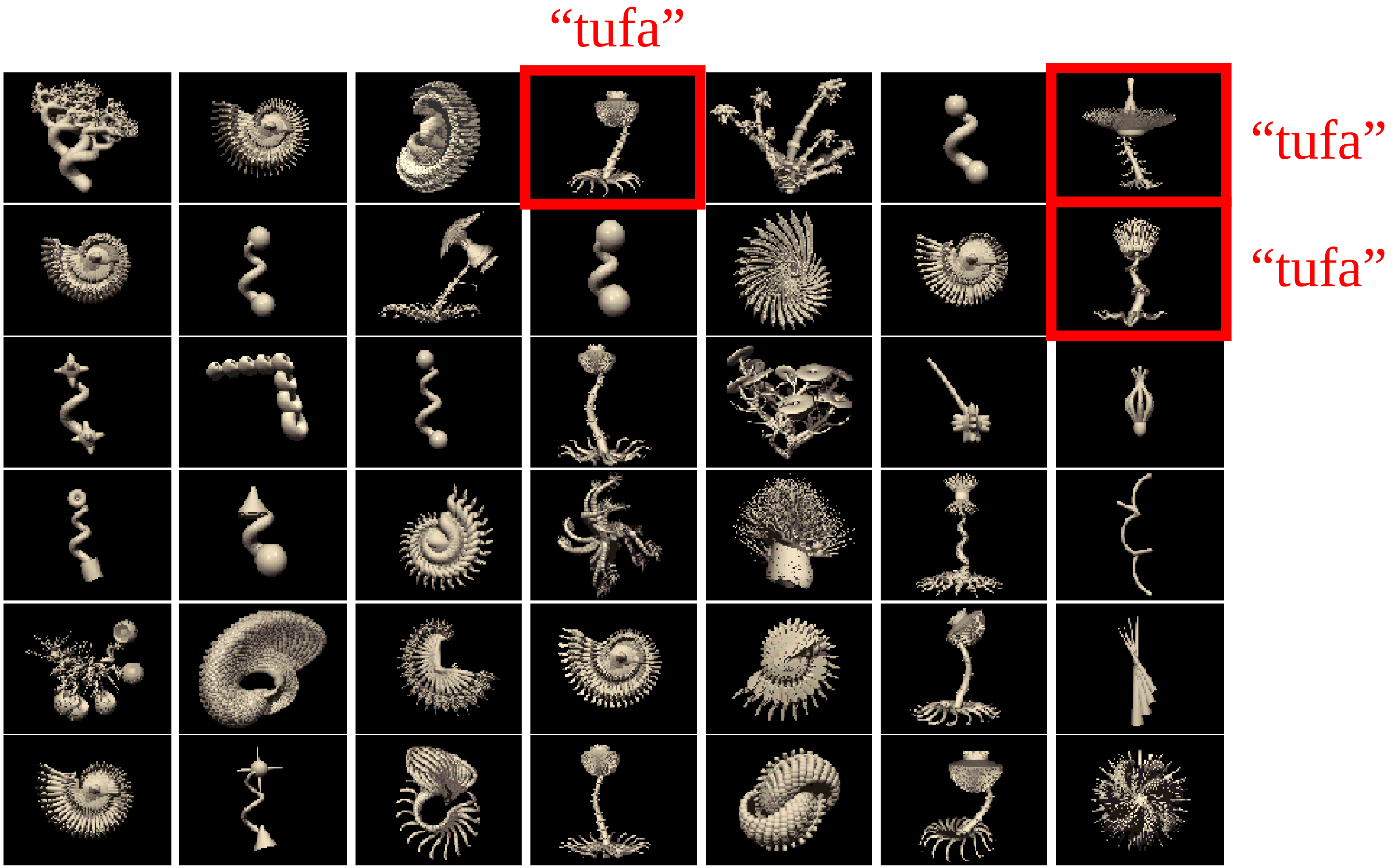
Outline for lectures

- Introduction
- Cognition as probabilistic inference
- Learning concepts from examples (continued)
- Learning and using intuitive theories (more structured systems of knowledge)

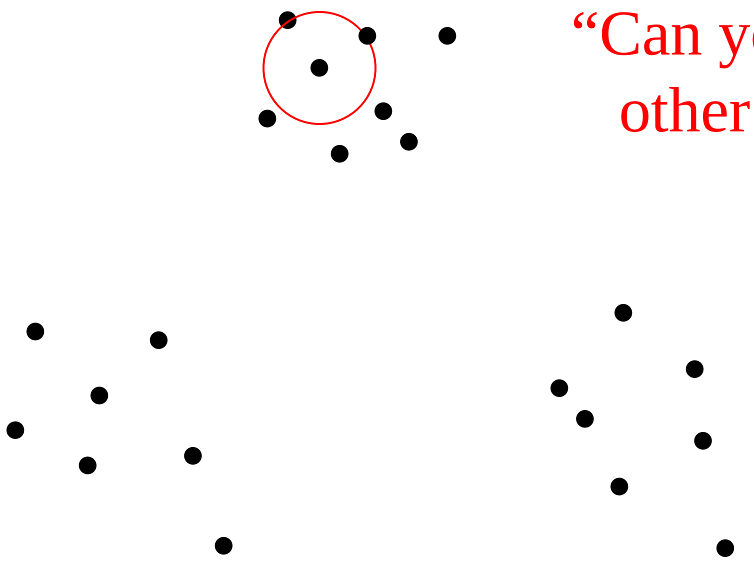


Learning from just one or a few examples, and mostly unlabeled examples ("semi-supervised learning").

Simple model of concept learning

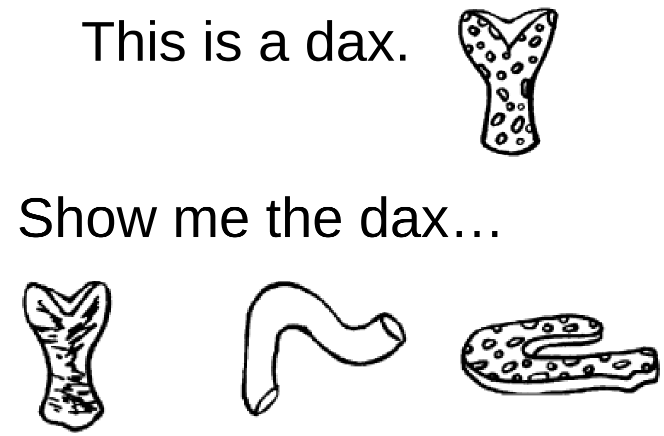
“This is a blicket.”

“Can you show me the other blickets?”

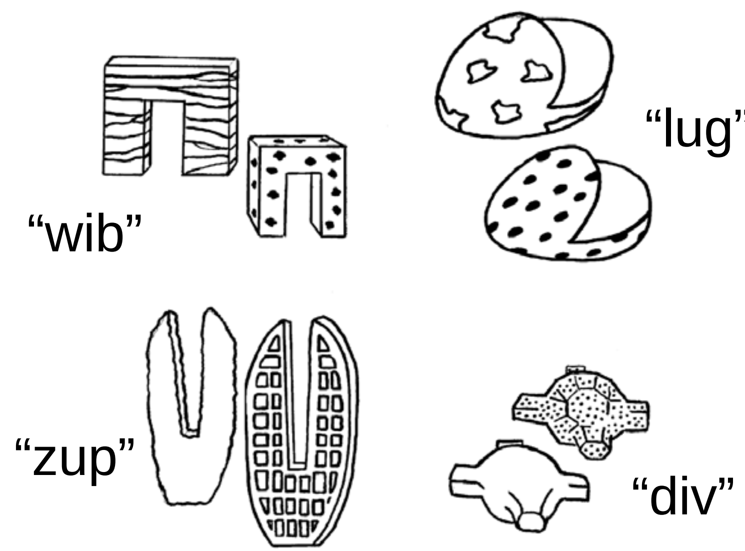


Learning to learn: what object features count for word learning?

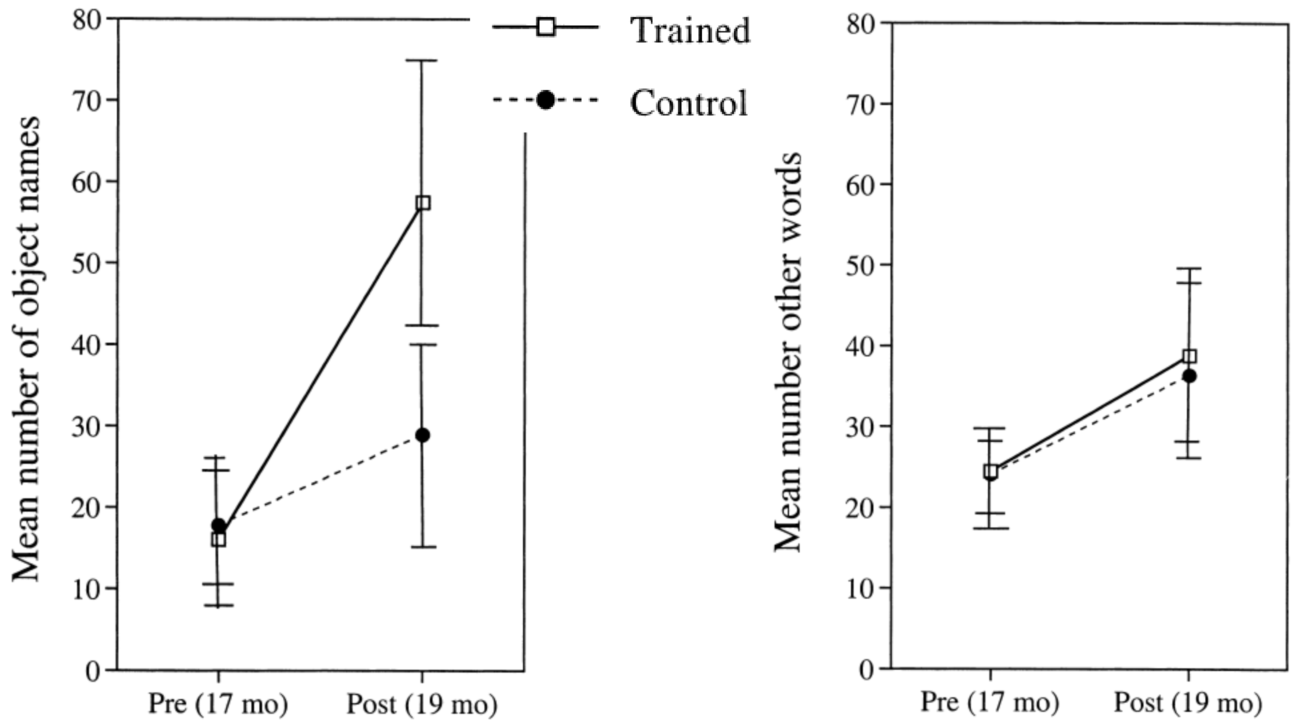
- 24-month-olds show the shape bias with simple novel objects. *20-month-olds do not.* (Landau, Smith, Jones 1988)



- Smith et al (2002) trained 17-month-olds on labels for 4 artificial categories:
- After 8 weeks of training (20 min/week), 19-month-olds show the shape bias.



Transfer to real-world vocabulary



The intuition: Learn that shape varies across categories but is relatively constant within nameable categories.

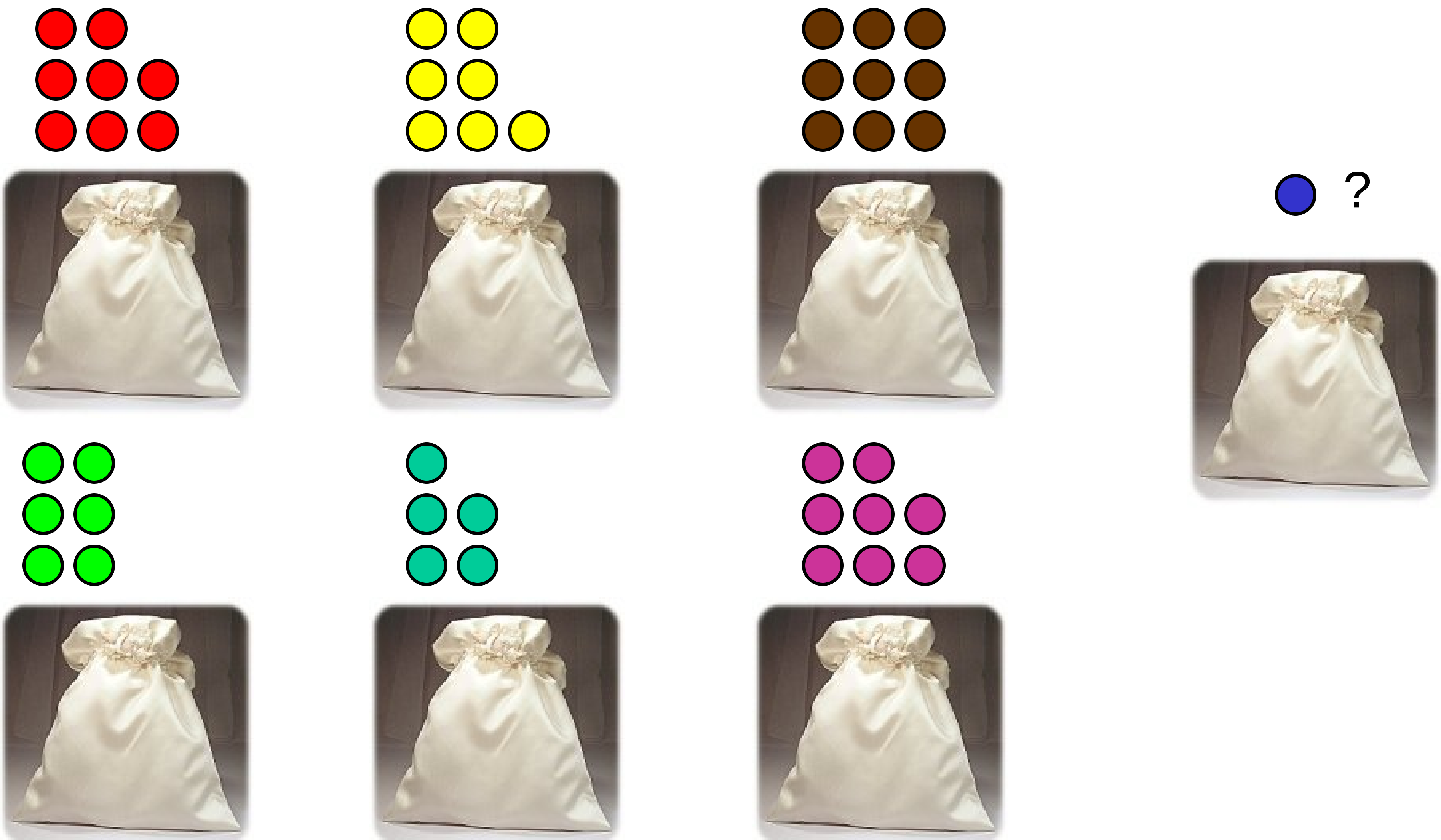
The puzzle: The shape bias is a powerful inductive constraint, yet can be learned from very little data.

Learning about feature variability

(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)

Learning about feature variability

(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)



A hierarchical Bayesian model

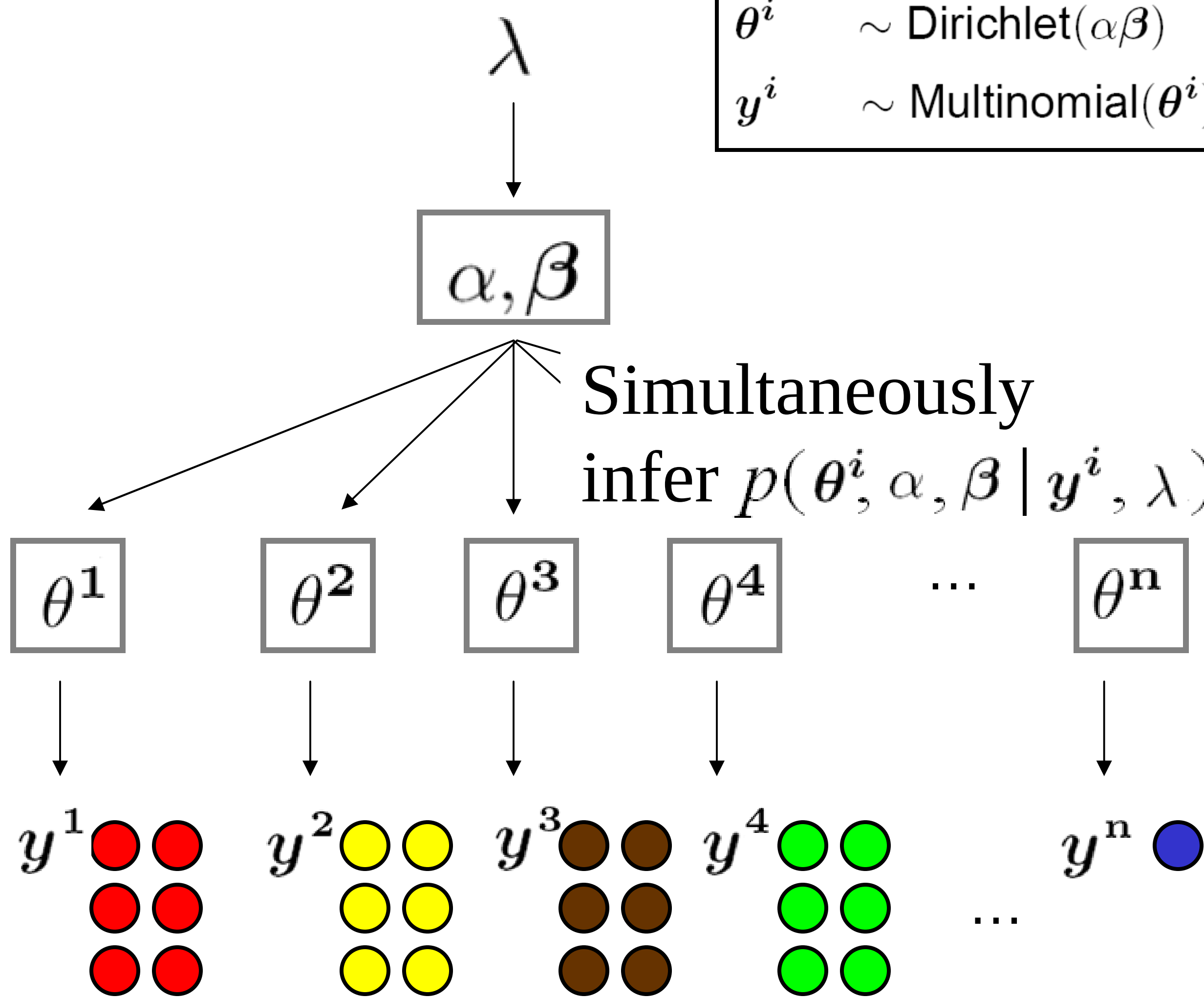
$$\begin{aligned}
\alpha &\sim \text{Exponential}(\lambda) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}^i &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\
\boldsymbol{y}^i &\sim \text{Multinomial}(\boldsymbol{\theta}^i)
\end{aligned}$$

Level 3:
Prior expectations
on bags in general

Level 2:
Bags in general

Level 1:
Bag proportions

Data



A hierarchical Bayesian model

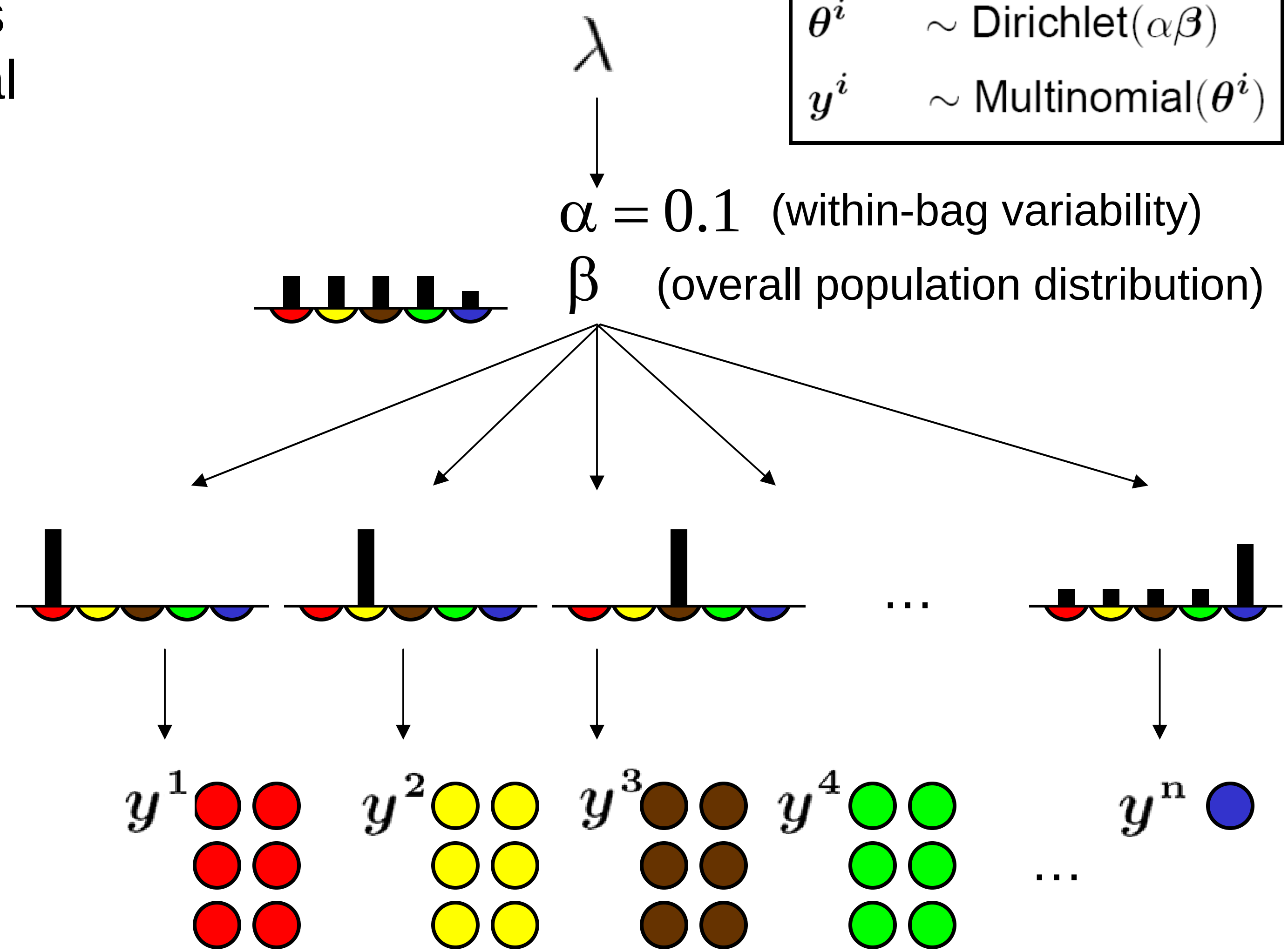
$$\alpha \sim \text{Exponential}(\lambda)$$
$$\boldsymbol{\beta} \sim \text{Dirichlet}(\mathbf{1})$$
$$\boldsymbol{\theta}^i \sim \text{Dirichlet}(\alpha\boldsymbol{\beta})$$
$$\boldsymbol{y}^i \sim \text{Multinomial}(\boldsymbol{\theta}^i)$$

Level 3:
Prior expectations
on bags in general

Level 2:
Bags in general

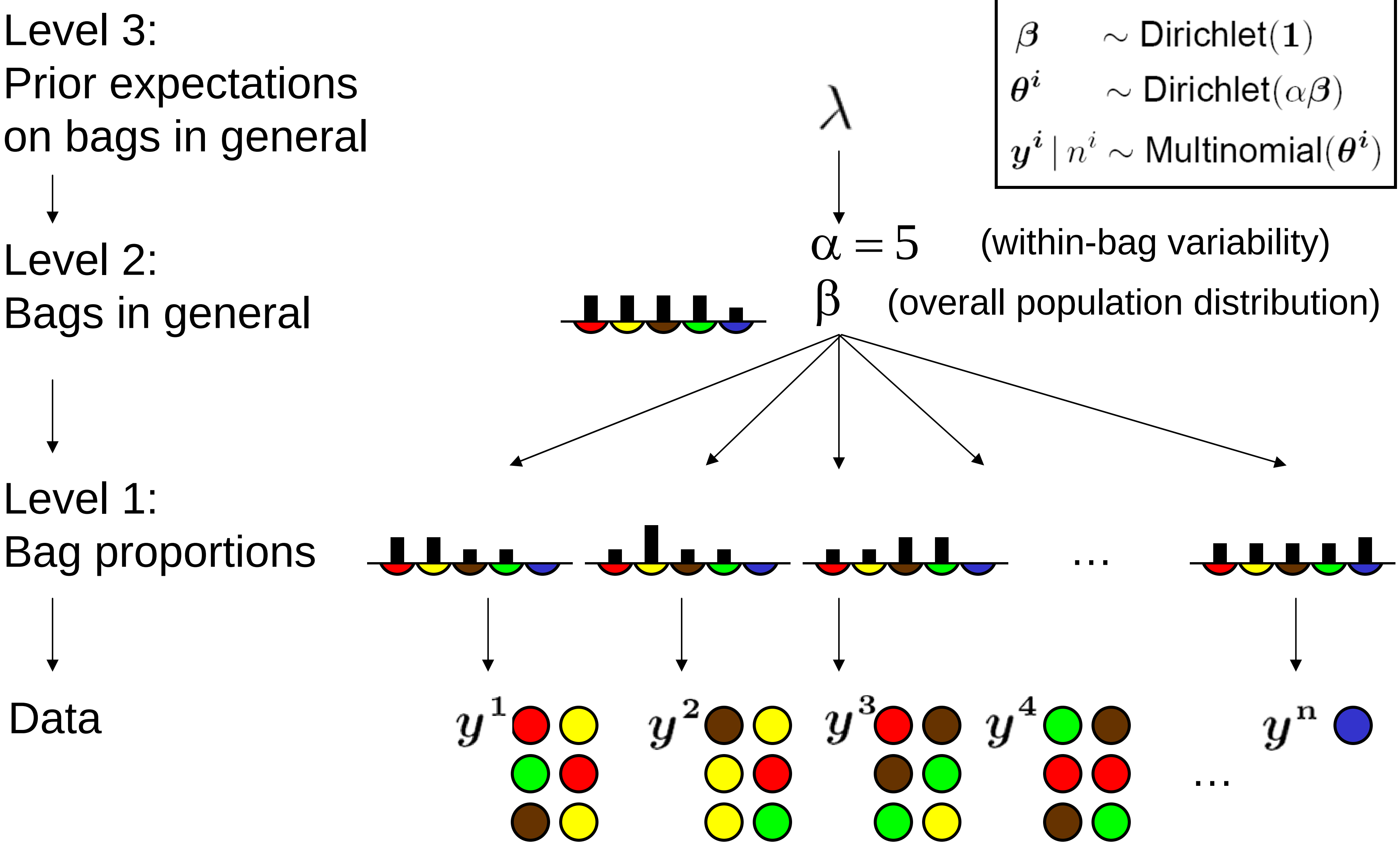
Level 1:
Bag proportions

Data



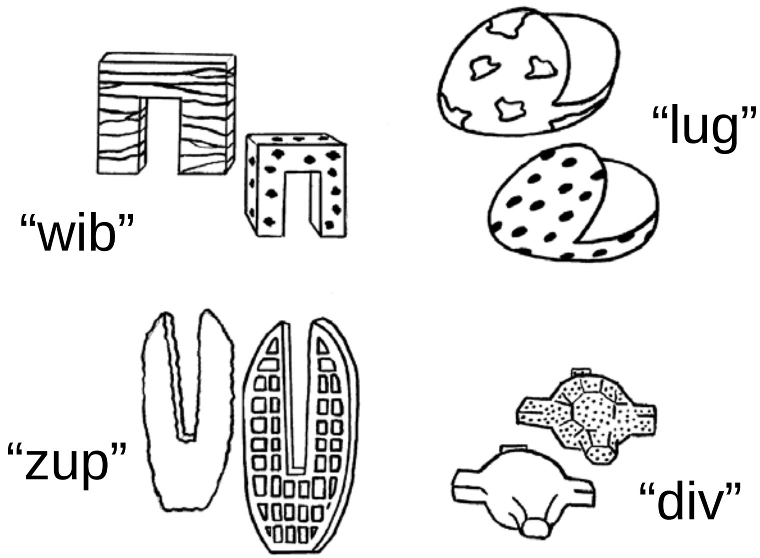
A hierarchical Bayesian model

$$
\begin{aligned}
\alpha &\sim \text{Exponential}(\lambda) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}^i &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\
\boldsymbol{y}^i \mid n^i &\sim \text{Multinomial}(\boldsymbol{\theta}^i)
\end{aligned}
$$



Learning the shape bias

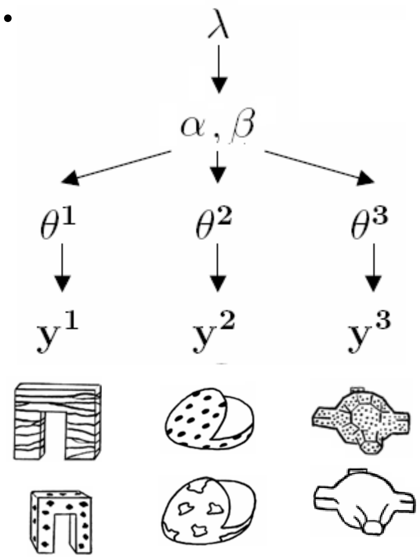
(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)



Training

Category	1 1	2 2	3 3	4 4
Shape	1 1	2 2	3 3	4 4
Texture	1 2	3 4	5 6	7 8
Color	1 2	3 4	5 6	7 8
Size	1 2	1 2	1 2	1 2

Assuming independent Dirichlet-multinomial models for each dimension ..

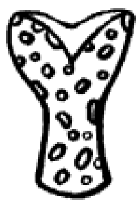


… we learn that:

- Shape varies across categories but not within categories.
- Texture, color, size vary across and within categories.

Second-order generalization test

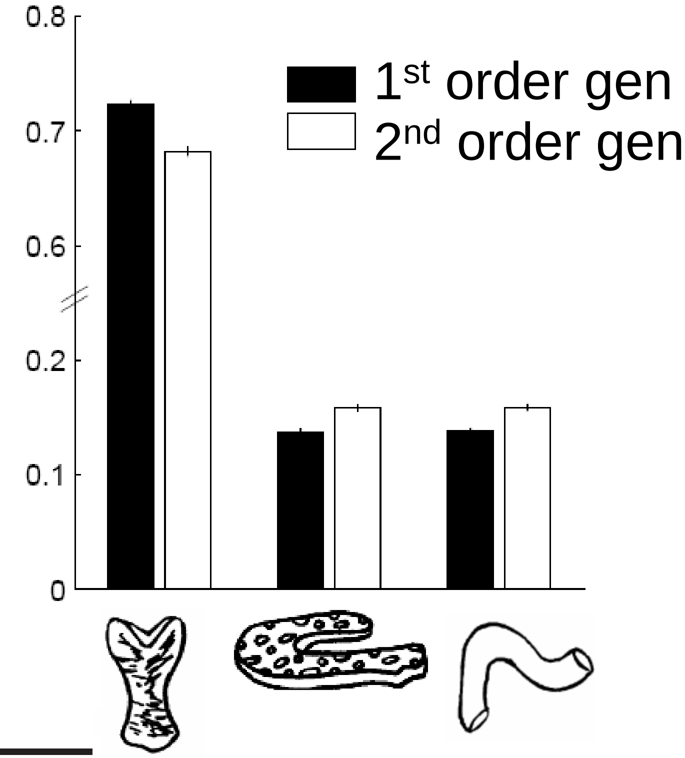
(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)



This is a dax.

Show me the dax…

Probability (normalized) that choice object belongs to the same category as the test exemplar



Training

Category	1 1	2 2	3 3	4 4
Shape	1 1	2 2	3 3	4 4
Texture	1 2	3 4	5 6	7 8
Color	1 2	3 4	5 6	7 8
Size	1 2	1 2	1 2	1 2

Test

5	?	?	?
5	5	6	6
9	10	9	10
9	10	10	9
1	1	1	1

"blessing of abstraction"

A more realistic model

$$
\begin{aligned}
\boldsymbol{z} &\sim \text{CRP}(\gamma) \\
\alpha &\sim \text{Exponential}(\lambda) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}^k &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\
\boldsymbol{y}^i &\sim \text{Multinomial}(\boldsymbol{\theta}^{z_i})
\end{aligned}
$$

Prior expectations on categories in general

λ

Categories in general

$\alpha, \boldsymbol{\beta}$

Individual categories

$\theta^{\mathbf{1}}$ $\theta^{\mathbf{2}}$ $\theta^{\mathbf{3}}$ $\theta^{\mathbf{4}}$ …?

Data

? ? ? ?

42571507 - 1
50614667 - ?
23648160 - ?
30746502 - 4
56315442 - ?
73046446 - ?
78640370 - 2
31242541 - ?
73616235 - ?
11577707 - ?
41502465 - ?
16616311 - ?
30252135 - ?
30746502 - ?
56643025 - ?
41670016 - ?

(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

A more realistic model

$$
\begin{aligned}
\boldsymbol{z} &\sim \text{CRP}(\gamma) \\
\alpha &\sim \text{Exponential}(\lambda) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}^k &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\
\boldsymbol{y}^i &\sim \text{Multinomial}(\boldsymbol{\theta}^{\boldsymbol{z}_i})
\end{aligned}
$$

Prior expectations on categories in general: λ

↓

Categories in general: $\alpha, \boldsymbol{\beta}$

↓

Individual categories: $\theta^{\mathbf{1}}$ $\theta^{\mathbf{2}}$ $\theta^{\mathbf{3}}$ $\theta^{\mathbf{4}}$

↓

Data

y *z*	*y* *z*	*y* *z*	*y* *z*
42571507 - 1	23648160 - 2	56643025 - 3	30746502 - 4
11577707 - 1	73046446 - 2	50614667 - 3	31242541 - 4
41670016 - 1	78640370 - 2	56315442 - 3	30252135 - 4
41502465 - 1	73616235 - 2	16616311 - 3	60240453 - 4

(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

A more realistic model

$$
\begin{aligned}
\boldsymbol{z} &\sim \text{CRP}(\gamma) \\
\alpha &\sim \text{Exponential}(\lambda) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}^k &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\
\boldsymbol{y}^i &\sim \text{Multinomial}(\boldsymbol{\theta}^{\boldsymbol{z}_i})
\end{aligned}
$$

Prior expectations on categories in general

λ

Categories in general

$\alpha, \boldsymbol{\beta}$

Individual categories

$\theta^{\mathbf{1}}$ $\theta^{\mathbf{2}}$ $\theta^{\mathbf{3}}$ $\theta^{\mathbf{4}}$ θ^{5}

Data

y *z*	*y* *z*	*y* *z*	*y* *z*	*y* *z*
42571507 - 1	23648160 - 2	56643025 - 3	30746502 - 4	88998899 - 5
11577707 - 1	73046446 - 2	50614667 - 3	31242541 - 4	
41670016 - 1	78640370 - 2	56315442 - 3	30252135 - 4	
41502465 - 1	73616235 - 2	16616311 - 3	60240453 - 4	

2nd order generalization:
88994271 - 5? or 42718899 - 5?

(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

A more realistic model

$$\begin{aligned} \mathbf{z} &\sim \text{CRP}(\gamma) \\ \alpha &\sim \text{Exponential}(\lambda) \\ \boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\ \boldsymbol{\theta}^k &\sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) \\ \boldsymbol{y}^i &\sim \text{Multinomial}(\boldsymbol{\theta}^{\boldsymbol{z}_i}) \end{aligned}$$

Prior expectations c
categories in gener

Learning the base distribution of a DP mixture

λ

Categories in general

$\boxed{\alpha, \boldsymbol{\beta}}$

Individual categories

$\theta^{\mathbf{1}}$ $\theta^{\mathbf{2}}$ $\theta^{\mathbf{3}}$ $\theta^{\mathbf{4}}$ θ^5

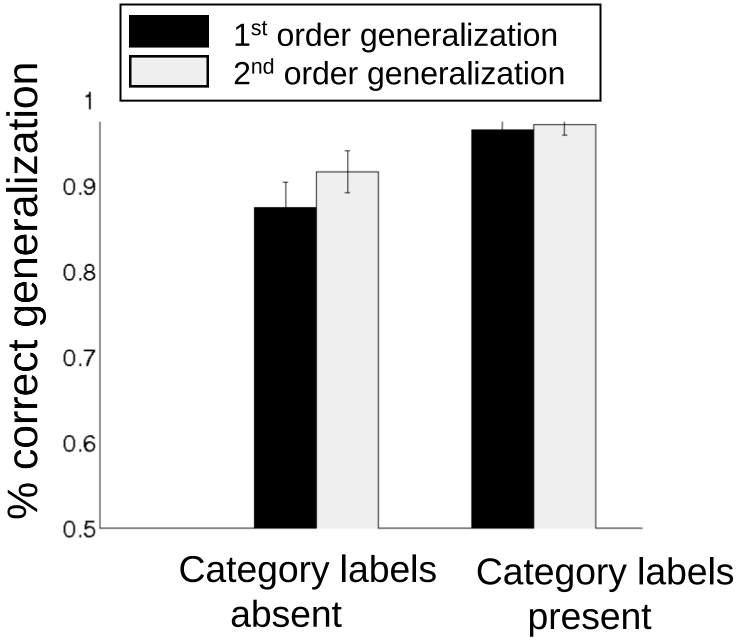
Data

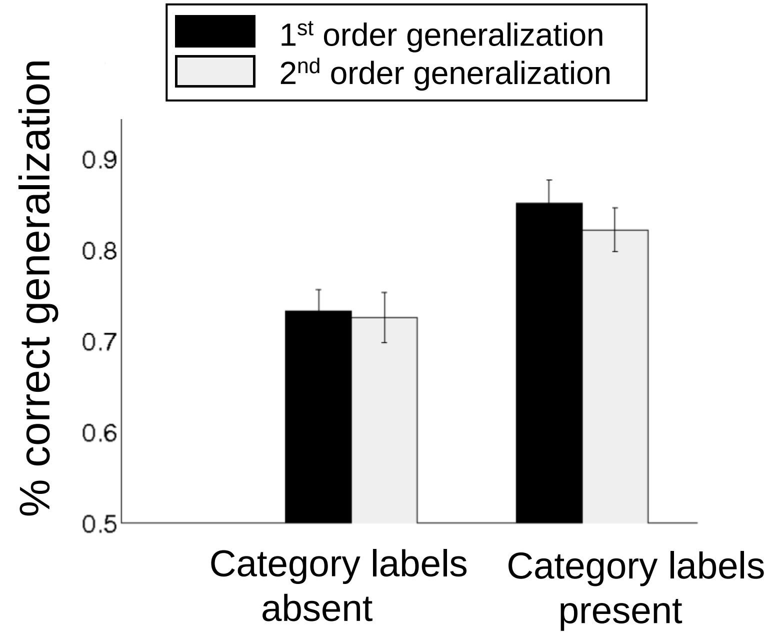
y	*z*	*y*	*z*	*y*	*z*	*y*	*z*	*y*	*z*
42571507	1	23648160	2	56643025	3	30746502	4	88998899	5
11577707	1	73046446	2	50614667	3	31242541	4		
41670016	1	78640370	2	56315442	3	30252135	4		
41502465	1	73616235	2	16616311	3	60240453	4		

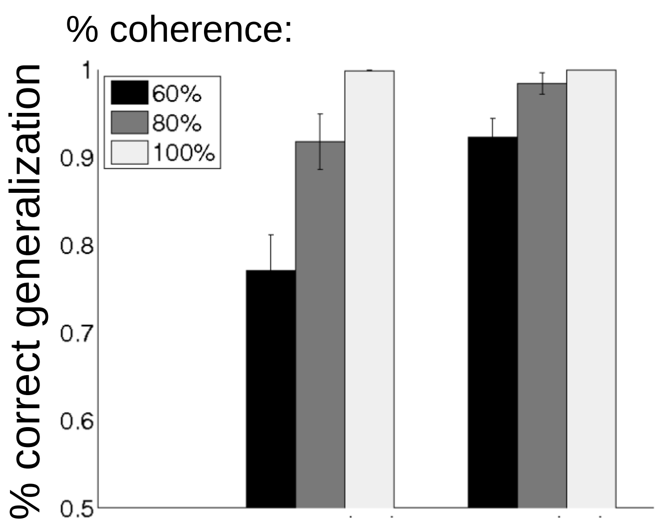
2nd order generalization:
88994271 - 5? or 42718899 - 5?

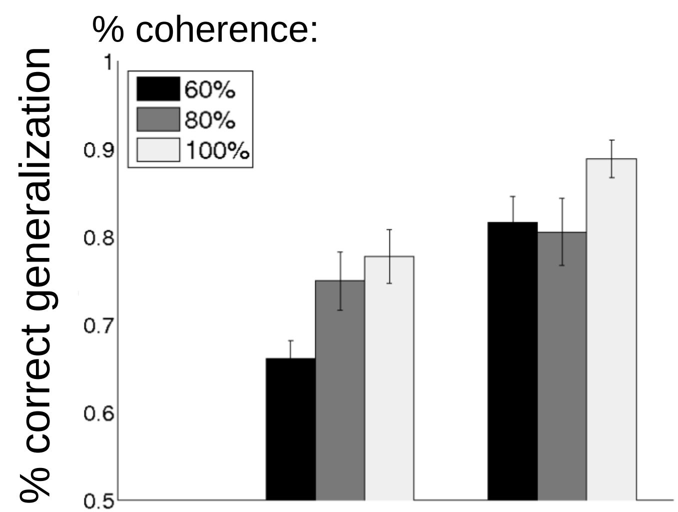
(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

Model vs. People



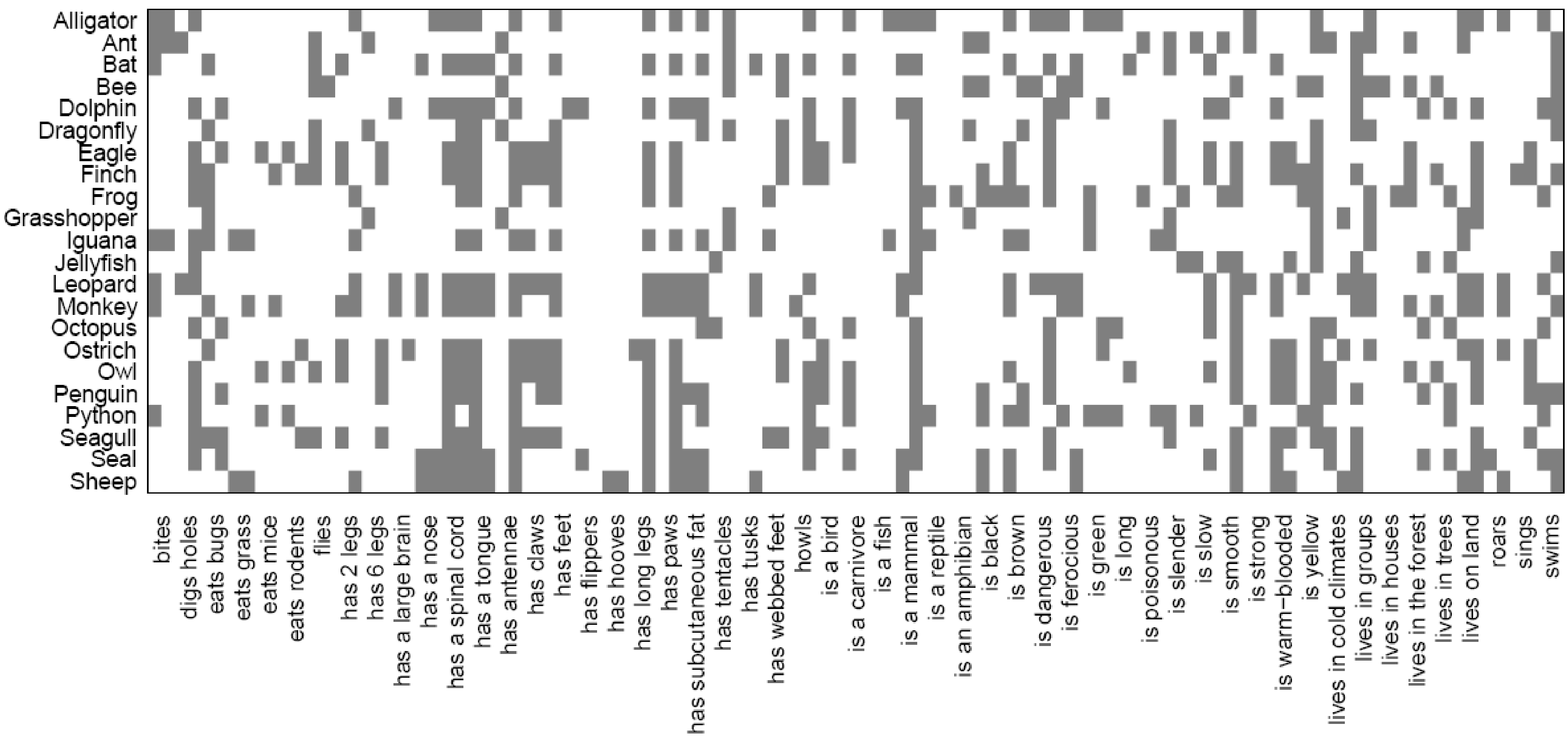






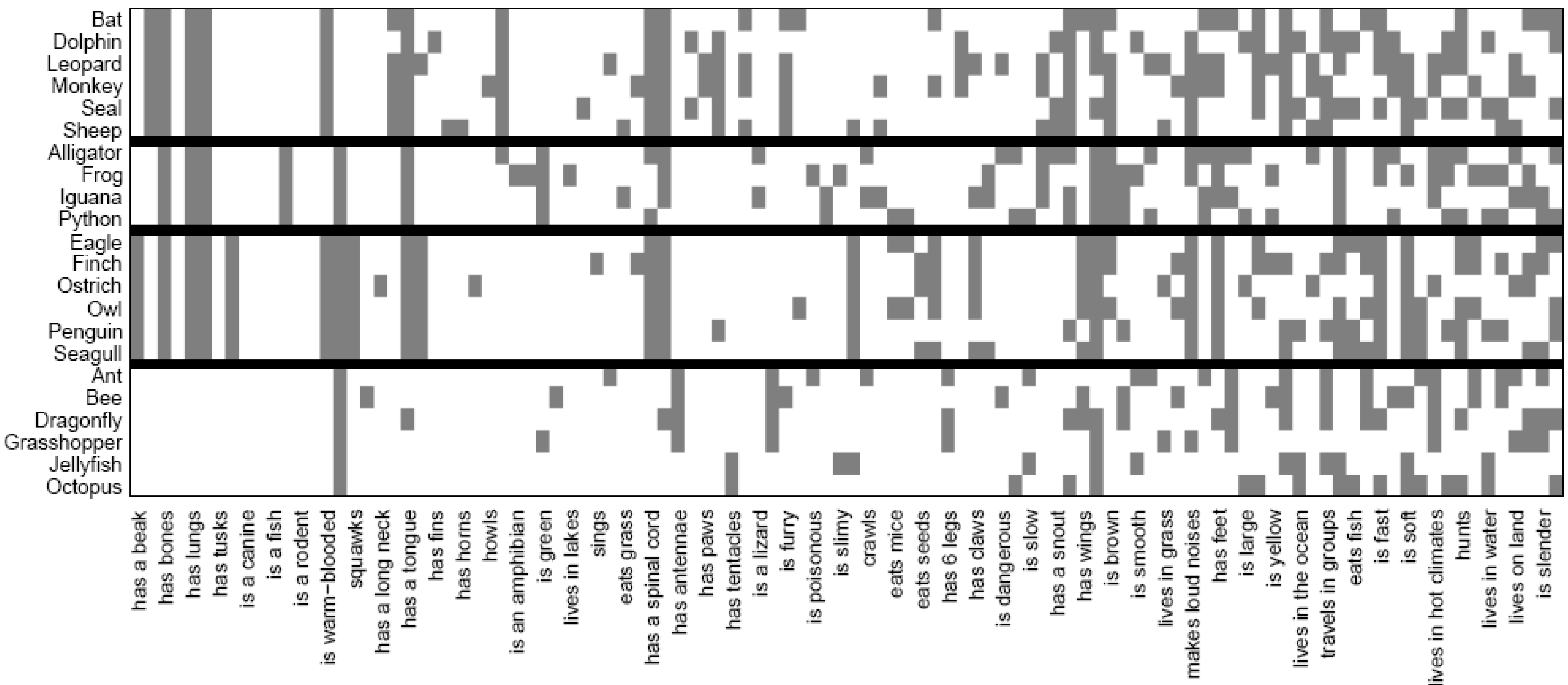
(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

Towards more natural concepts



Towards more natural concepts

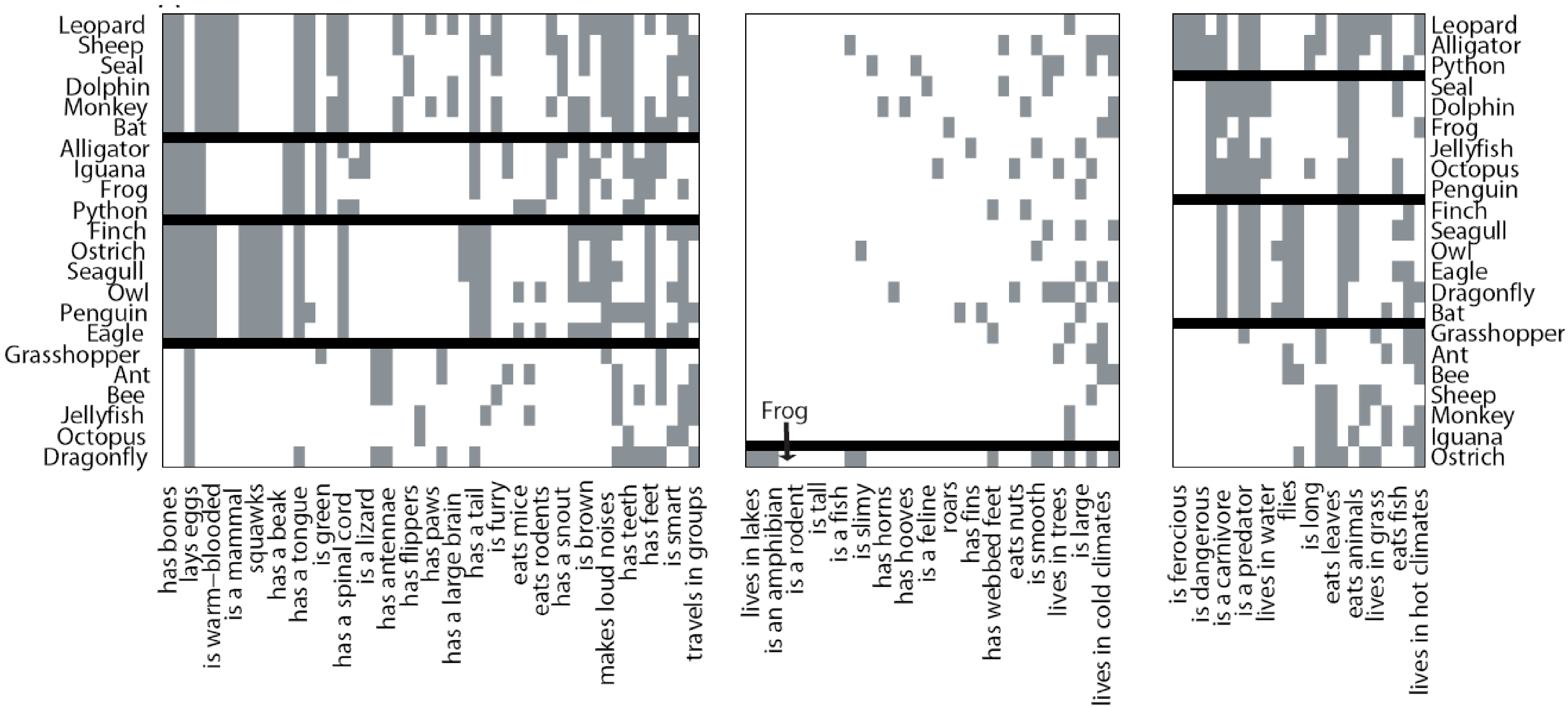
CRP mixture:



How many different ways to structure a domain?

(Shafto, Kemp, Mansingka, Tenenbaum, 2006; submitted)

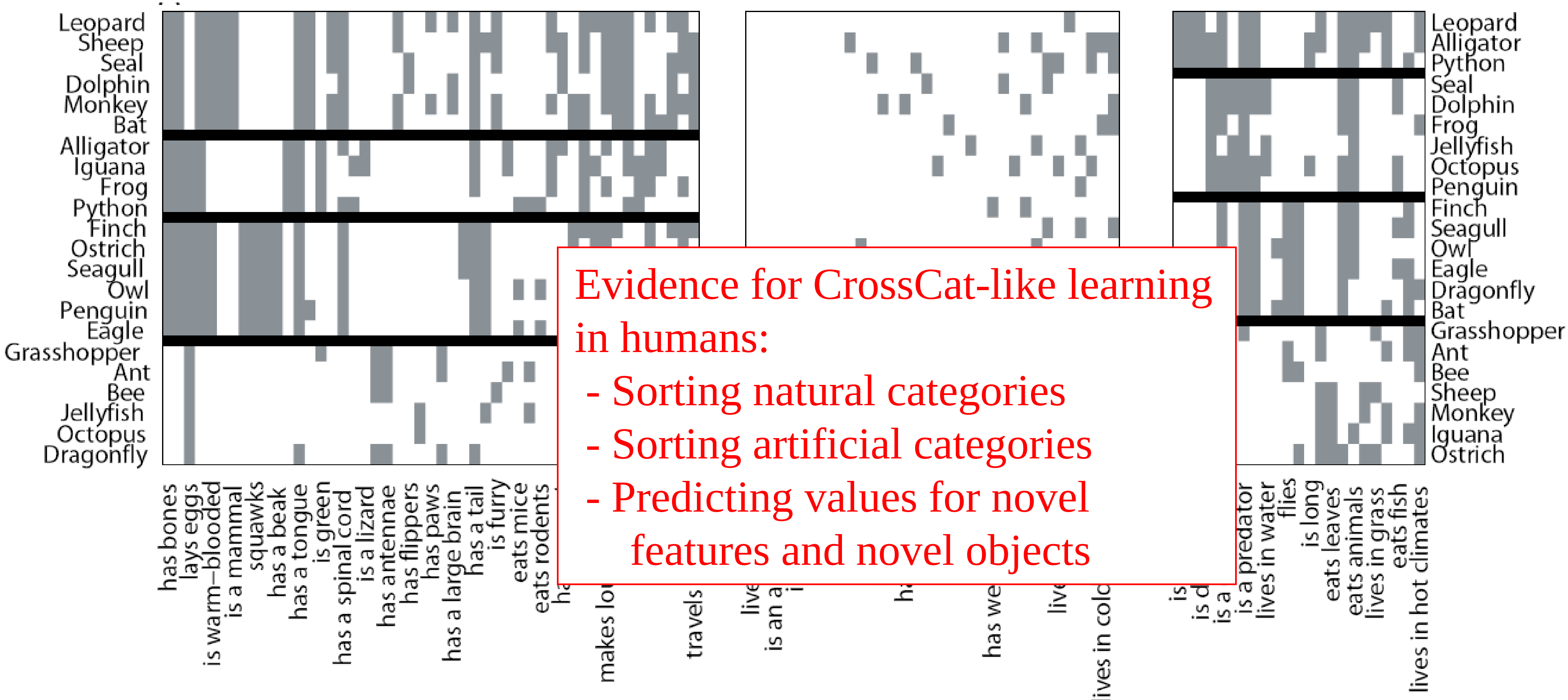
"CrossCat": nonparametric clustering over features, with a different clustering of objects for each feature-cluster.



How many different ways to structure a domain?

(Shafto, Kemp, Mansingka, Tenenbaum, 2006; submitted)

"CrossCat": nonparametric clustering over features, with a different clustering of objects for each feature-cluster.

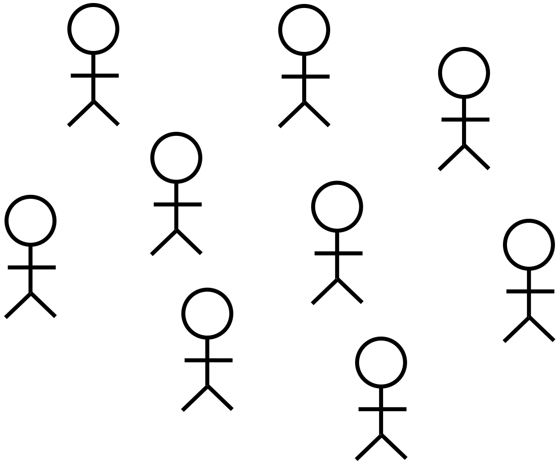


Evidence for CrossCat-like learning in humans:
- Sorting natural categories
- Sorting artificial categories
- Predicting values for novel features and novel objects

Learning relational concepts

CONCEPTS

Professors
Graduate students
Undergraduates

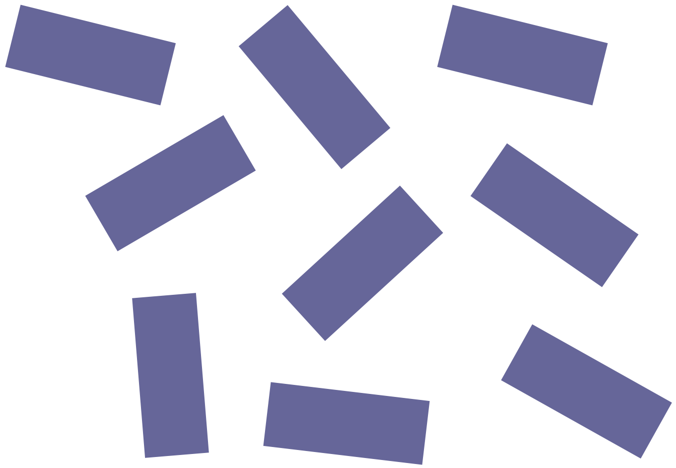


RELATIONSHIPS

Professors give advice to Grad students and Undergrads.
Grad students give advice to Undergrads.
Undergrads give advice to no one.

Learning relational concepts

CONCEPTS

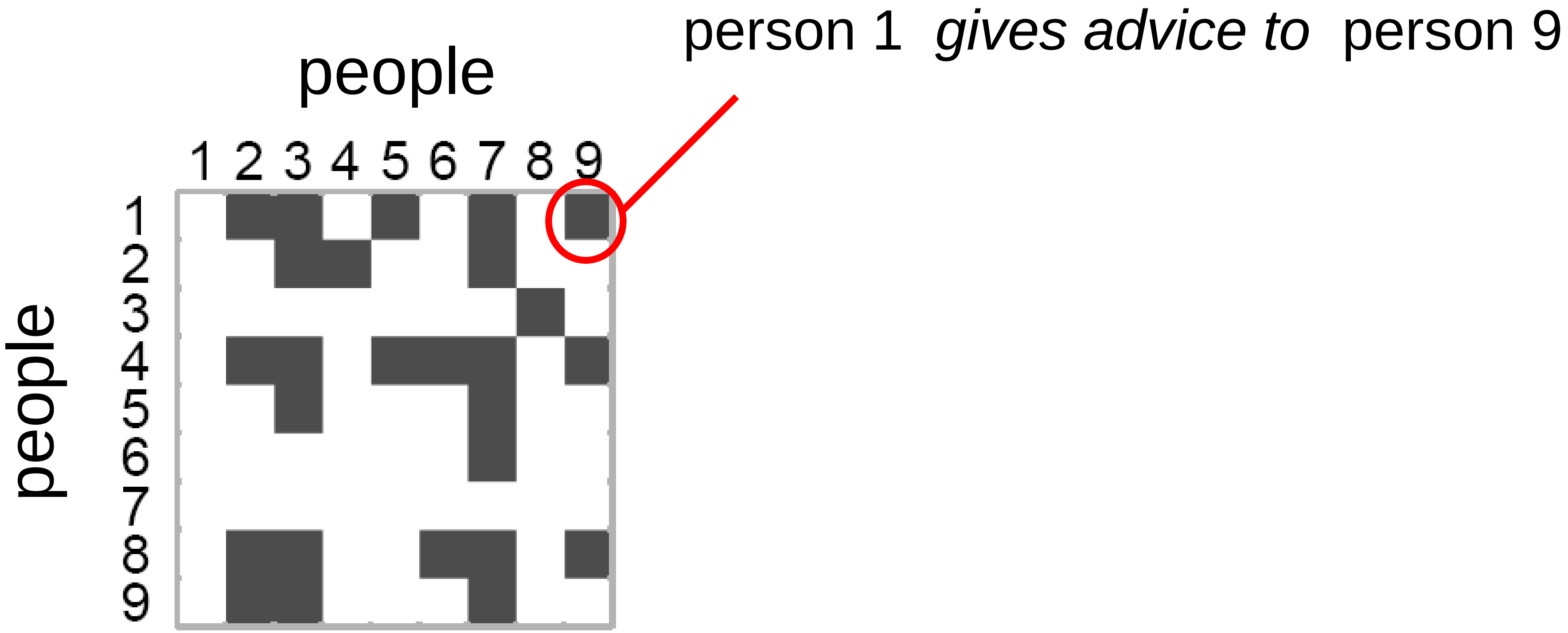


Magnets
Magnetic objects
Non-magnetic objects

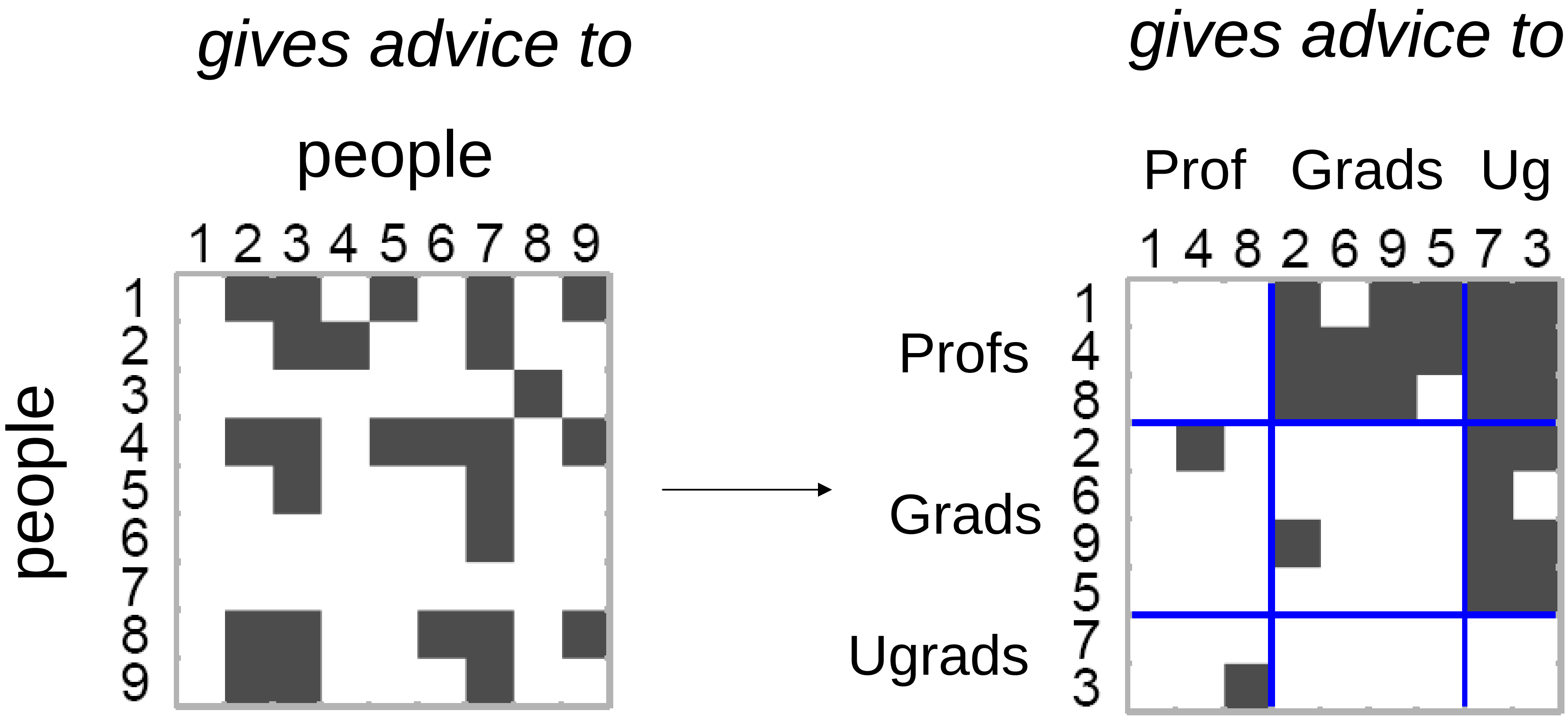
RELATIONSHIPS

Magnets interact with each other.
Magnets and Magnetic objects interact.
Magnetic objects do not interact with each other.
Non-magnetic objects do not interact with anything.

Learning relational concepts



Learning relational concepts



Infinite Relational Model (IRM)

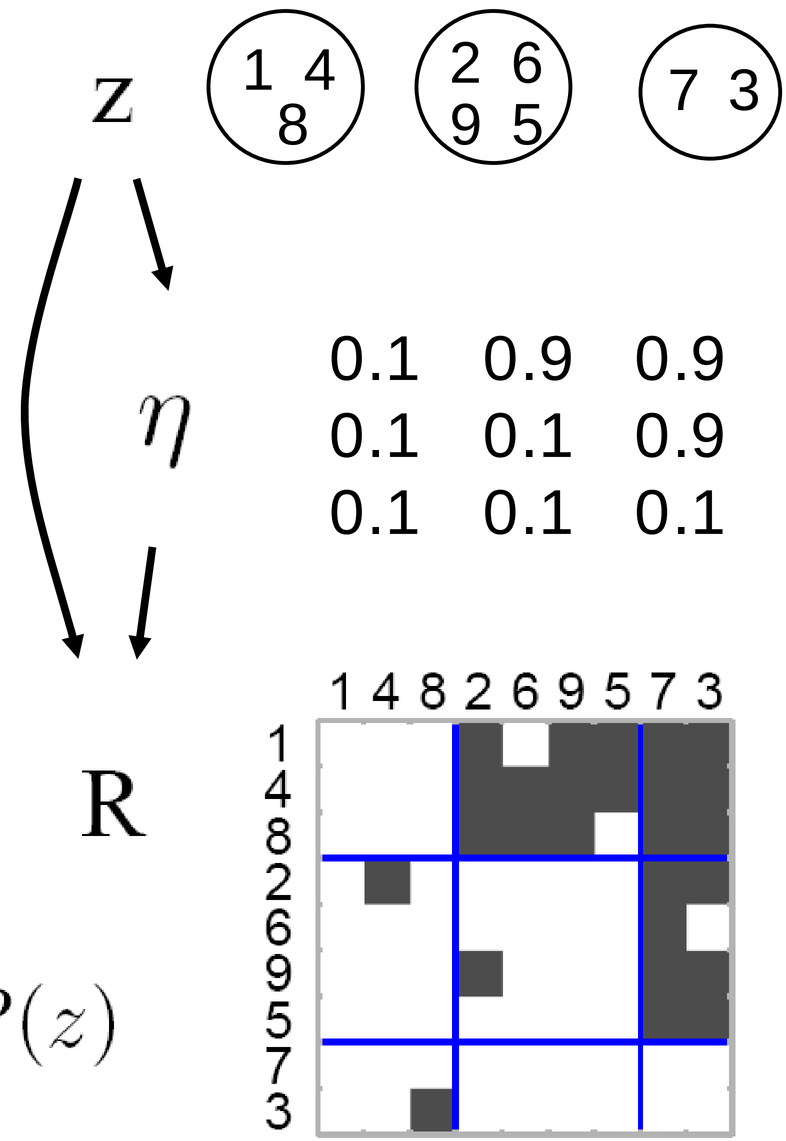
(Kemp, Griffiths, Tenenbaum, Yamada, & Ueda, 2006)

$$z \mid \gamma \quad \sim \mathrm{CRP}(\gamma)$$

$$\eta_{ab} \mid \alpha, \beta \sim \mathrm{Beta}(\alpha, \beta)$$

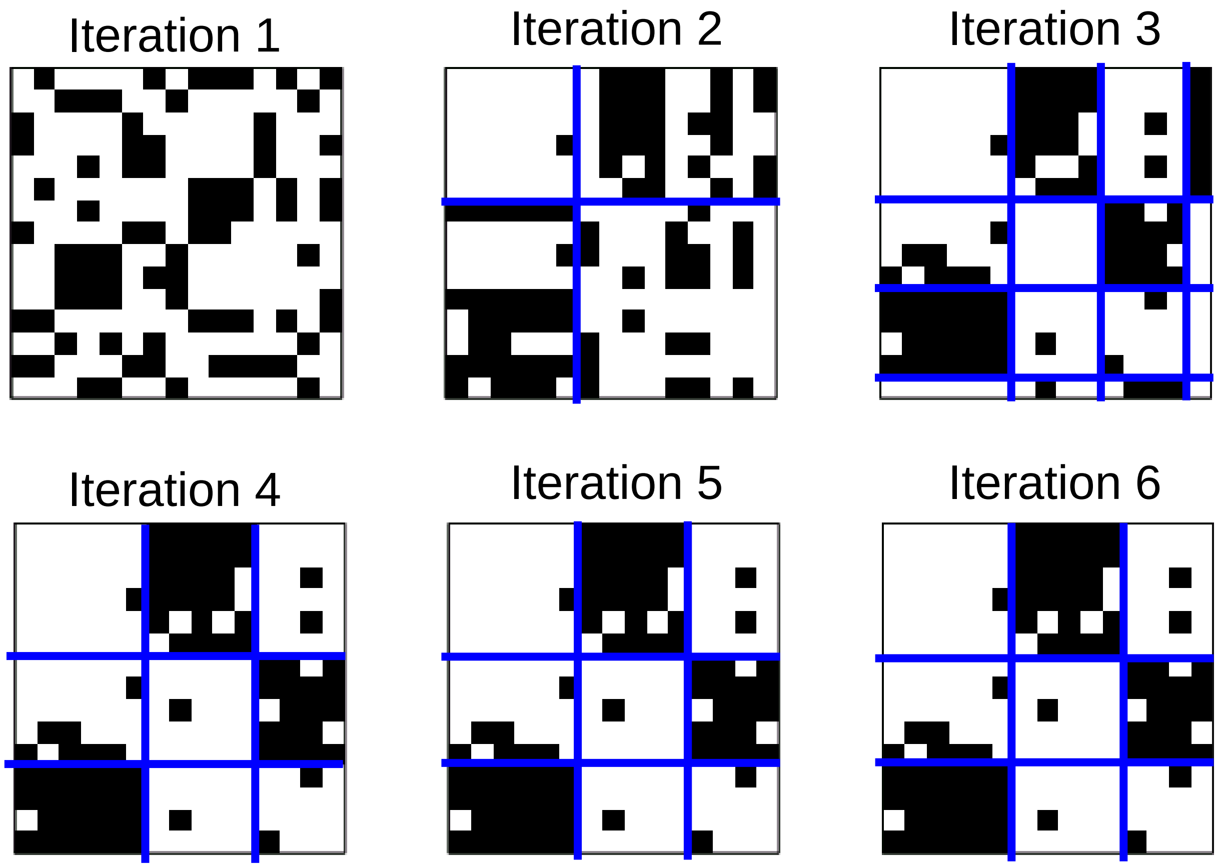
$$R_{ij} \mid z, \eta \sim \mathrm{Bernoulli}(\eta_{z_i z_j})$$

$$p(z, \eta | R) \propto P(R | z, \eta)\, p(\eta | z)\, P(z)$$



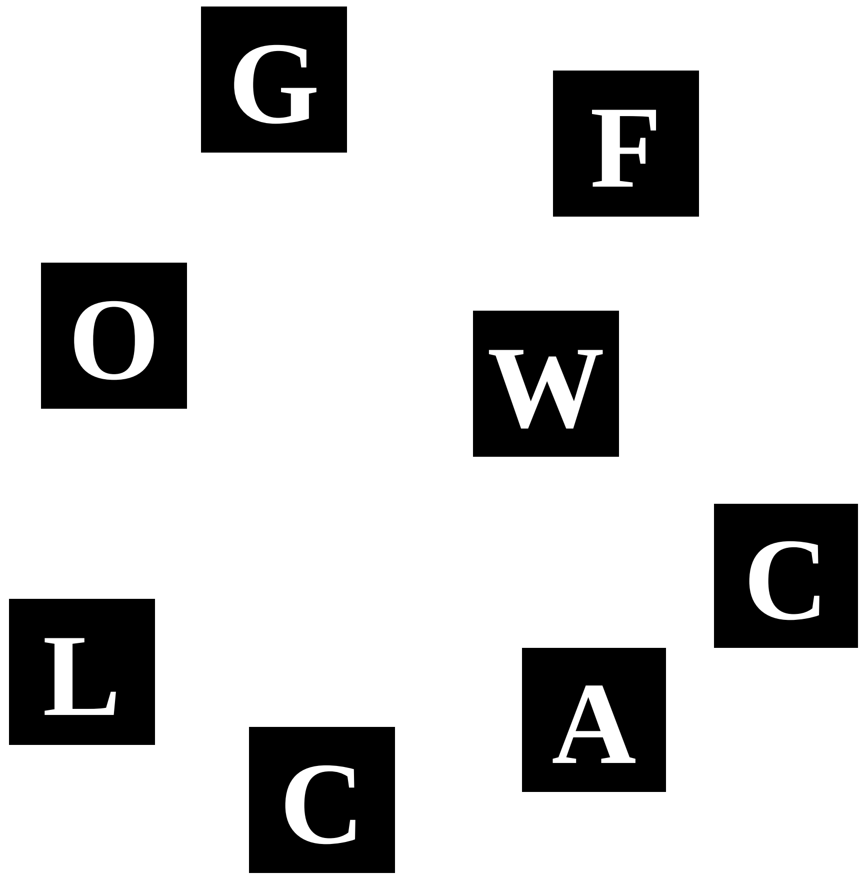
Learning algorithm

- Continuous parameters (weights/probabilities) integrated out analytically.
- Gibbs sampling + split-merge moves:



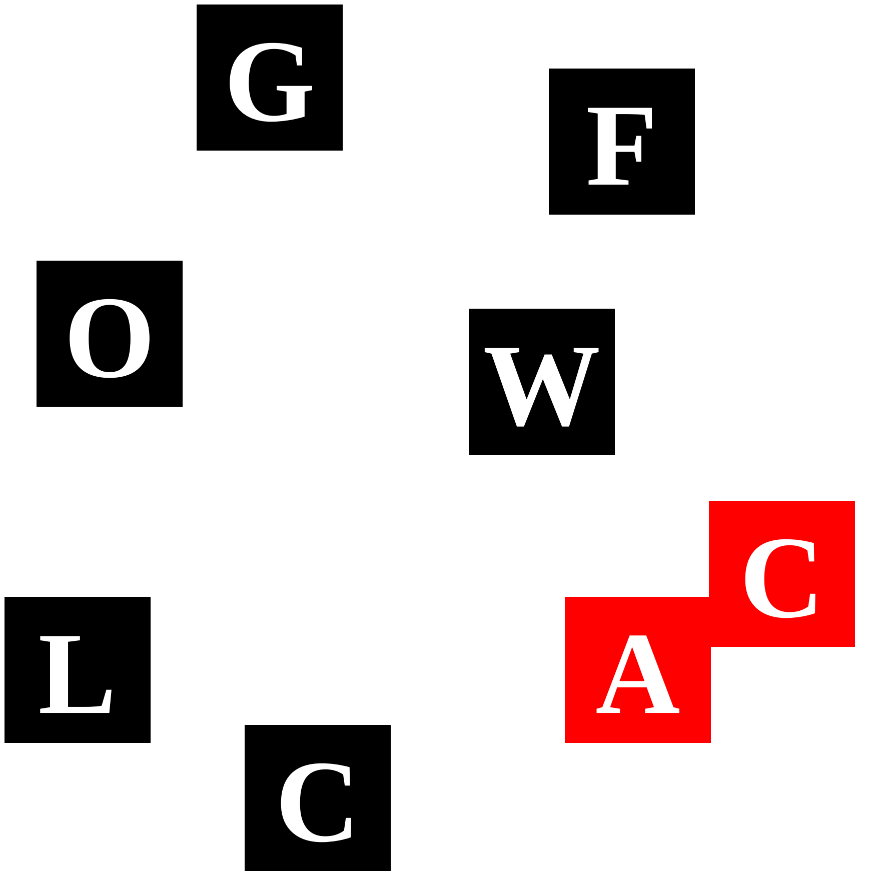
The causal blocks world

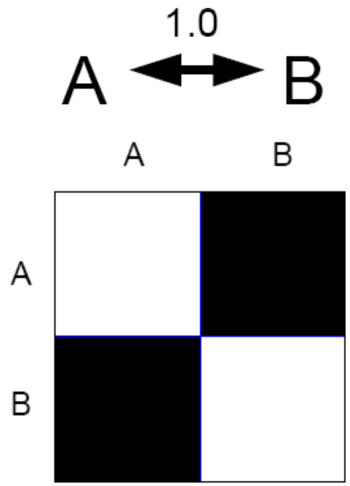
(Tenenbaum and Niyogi, 2003)

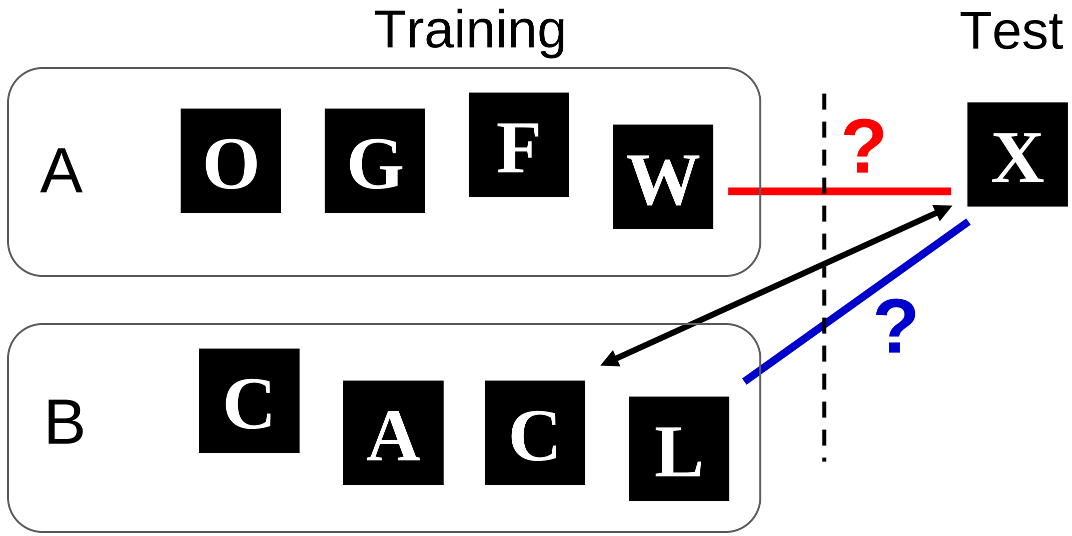


The causal blocks world

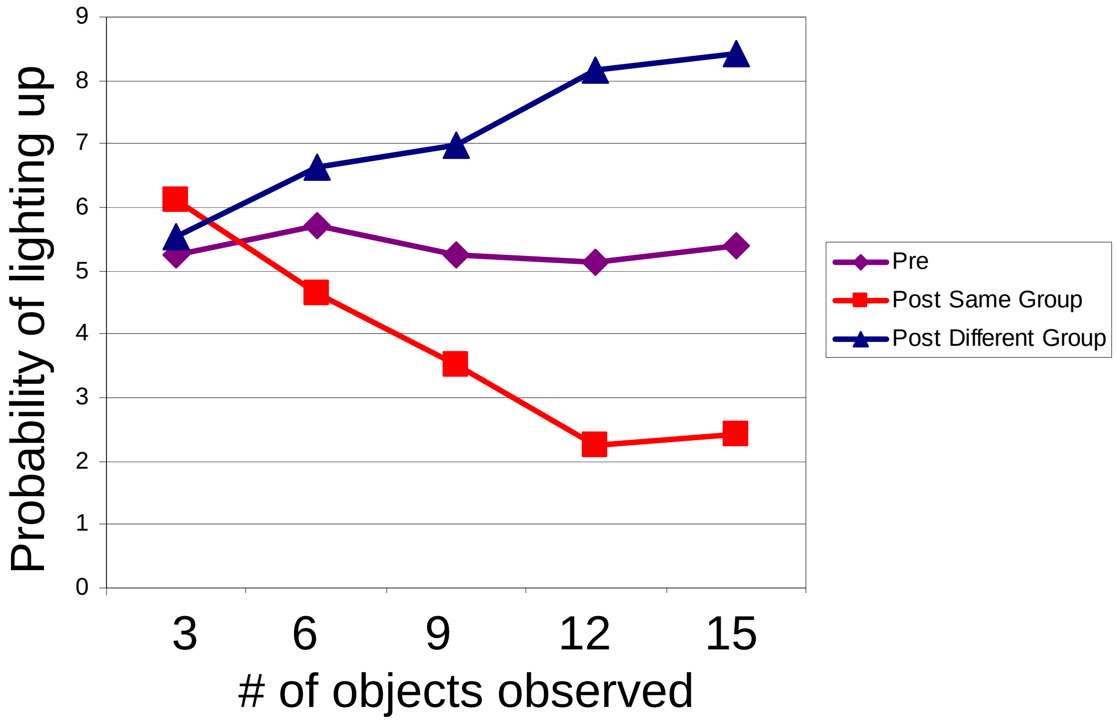
(Tenenbaum and Niyogi, 2003)

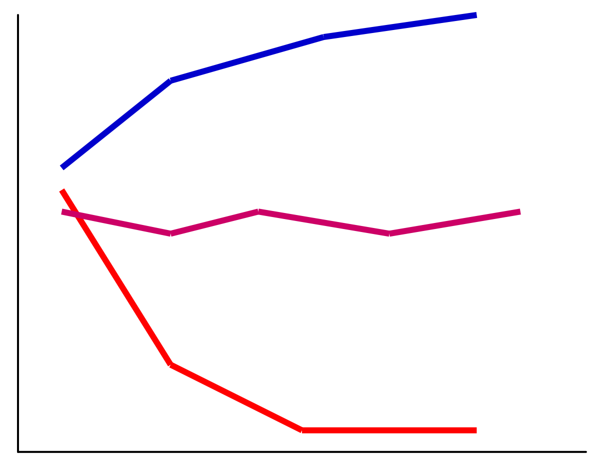






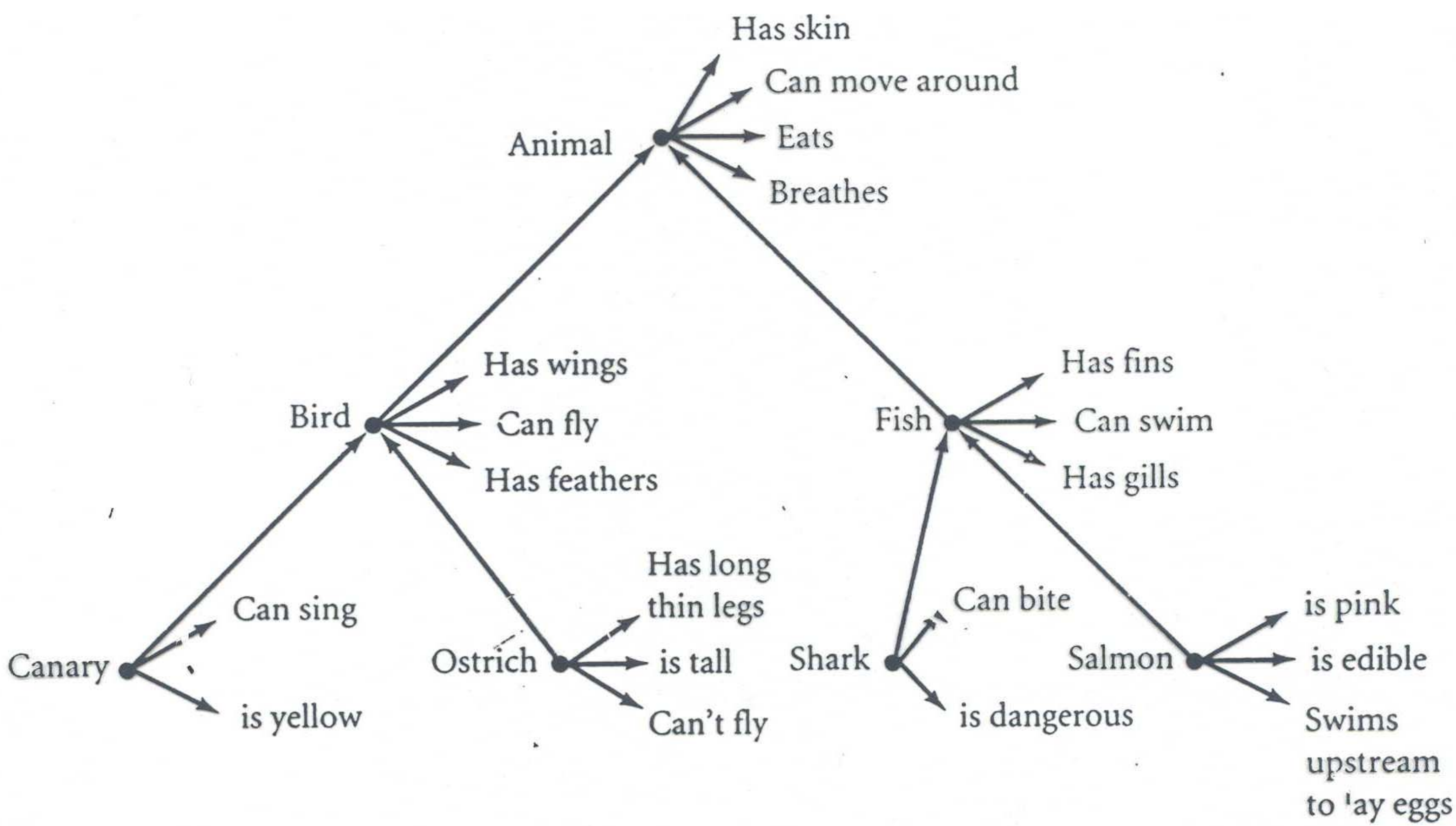
Learning curves





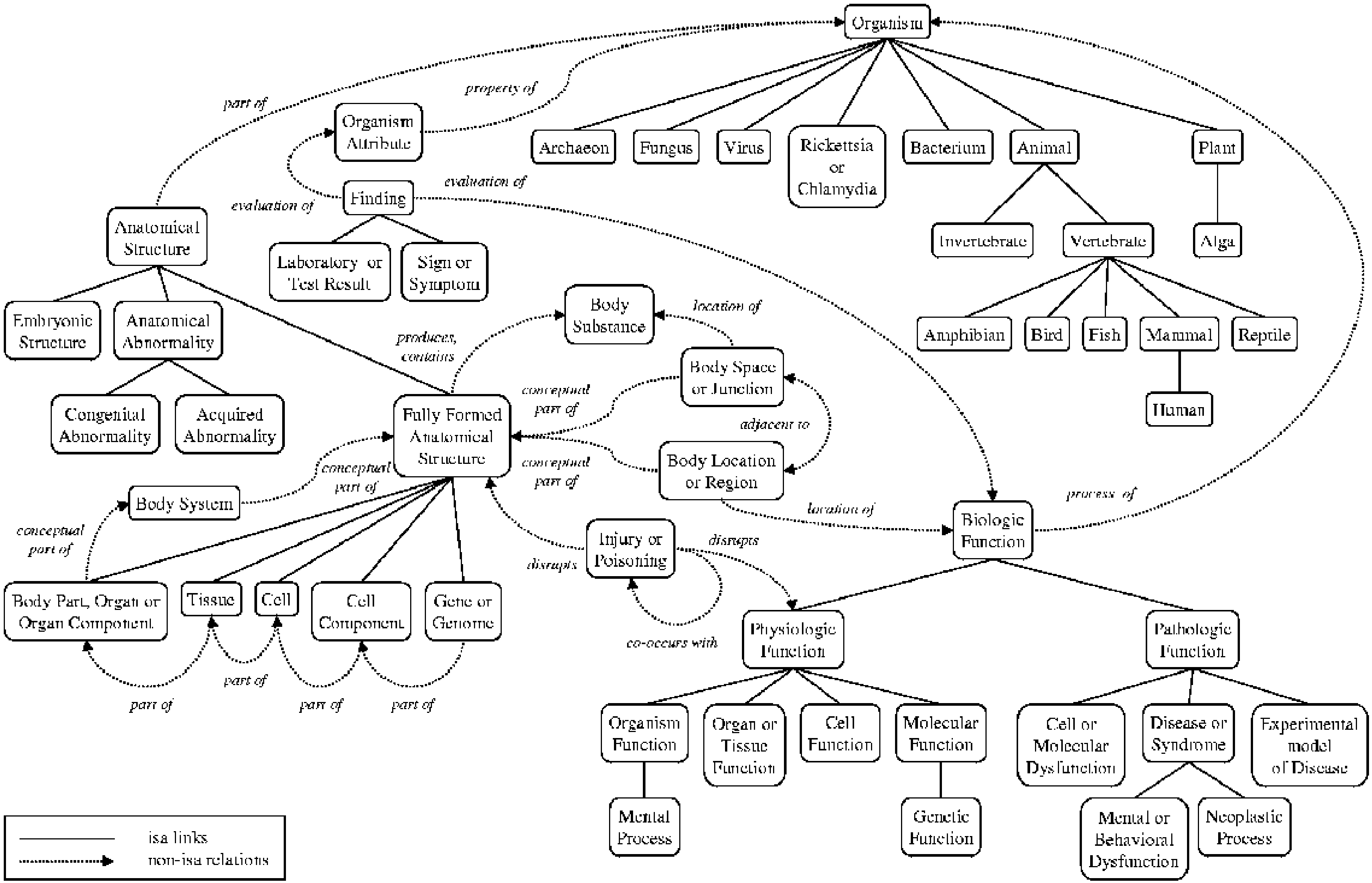
Model predictions

Constructing semantic networks

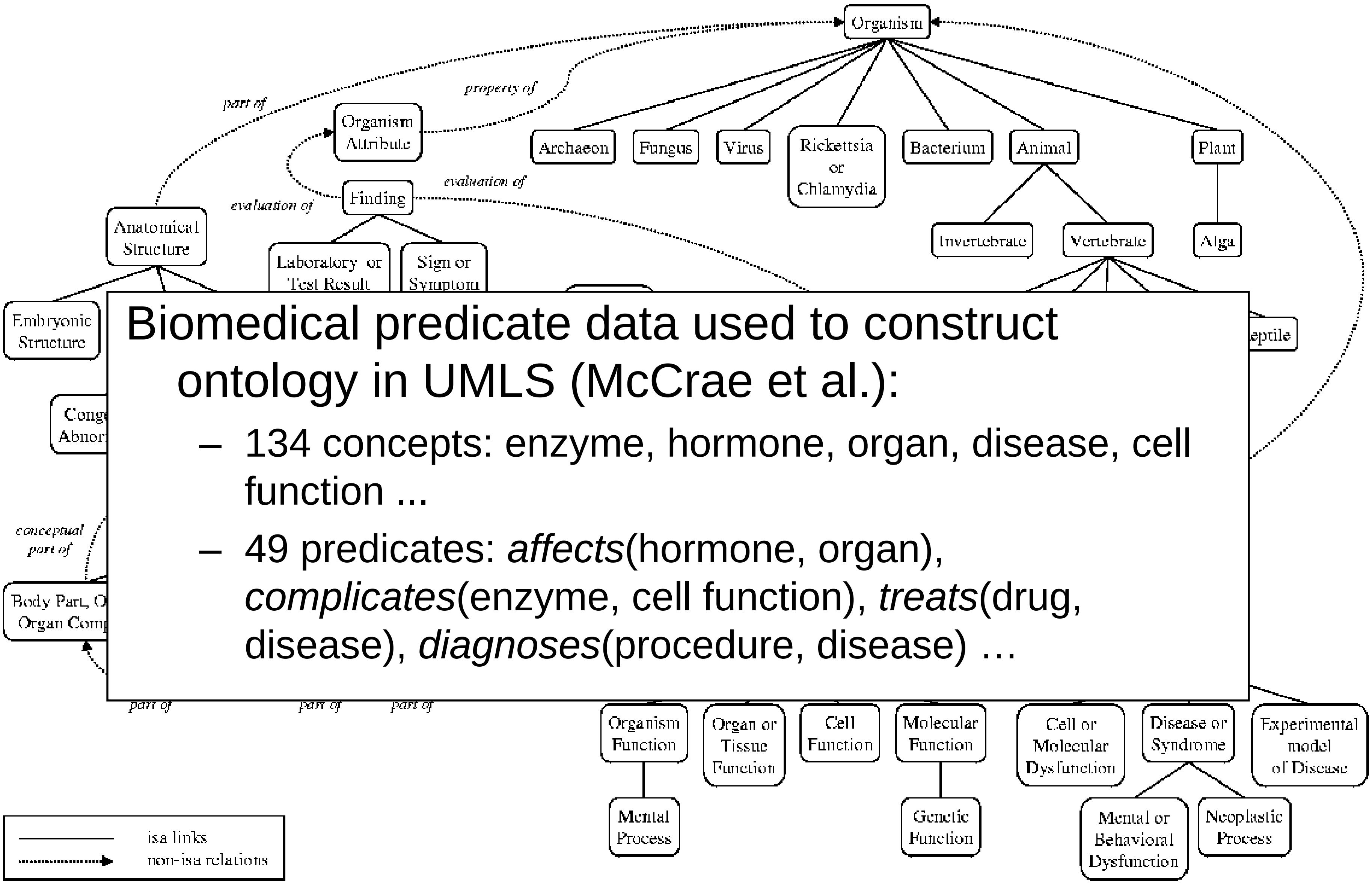


(Collins & Quillian, 1969)

Upper level medical ontology



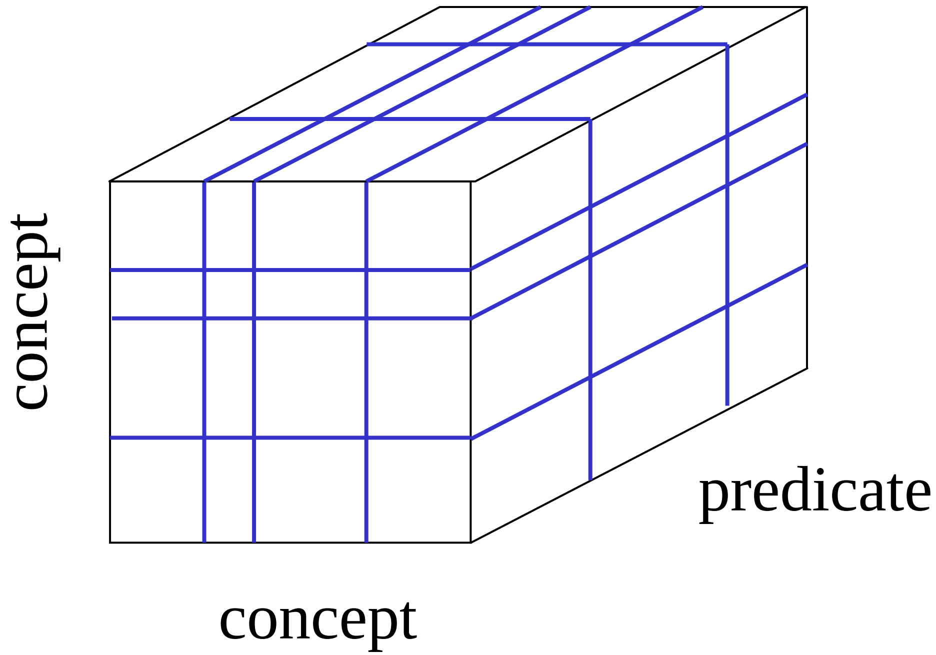
Upper level medical ontology



Biomedical predicate data used to construct ontology in UMLS (McCrae et al.):

- 134 concepts: enzyme, hormone, organ, disease, cell function ...
- 49 predicates: *affects*(hormone, organ), *complicates*(enzyme, cell function), *treats*(drug, disease), *diagnoses*(procedure, disease) …

Learning semantic networks with IRM



Biomedical predicate data from UMLS (McCrae et al.):

- 134 concepts: enzyme, hormone, organ, disease, cell function ...
- 49 predicates: *affects*(hormone, organ), *complicates*(enzyme, cell function), *treats*(drug, disease), *diagnoses*(procedure, disease) …

Learning semantic networks with IRM

a) Concept clusters

1.Organisms	2.Chemicals	3.Biological functions	4.Bio-active substances	5.Diseases
Alga	Amino Acid	Biological function	Antibiotic	Cell dysfunction
Amphibian	Carbohydrate	Cell function	Enzyme	Disease
Animal	Chemical	Genetic function	Poisonous substance	Mental dysfunction
Archaeon	Eicosanoid	Mental process	Hormone	Neoplastic process
Bacterium	Isotope	Molecular function	Pharmacologic substance	Pathologic function
Bird	Steroid	Physiological function	Vitamin	Expt. model of disease

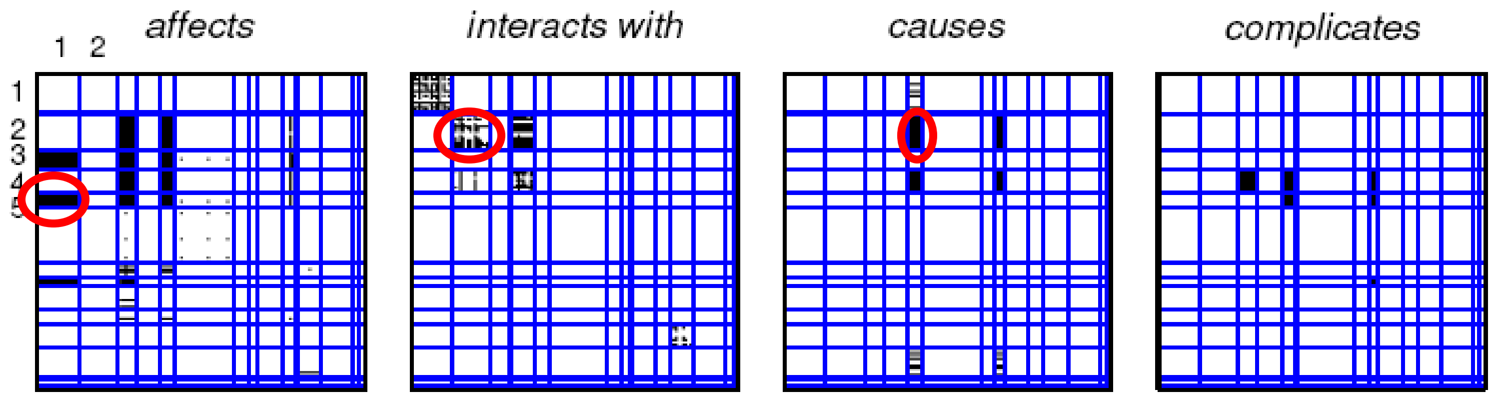
Predicate clusters

affects

analyzes
assesses effect of
measures

diagnoses
indicates
prevents
treats

carries out
exhibits
performs

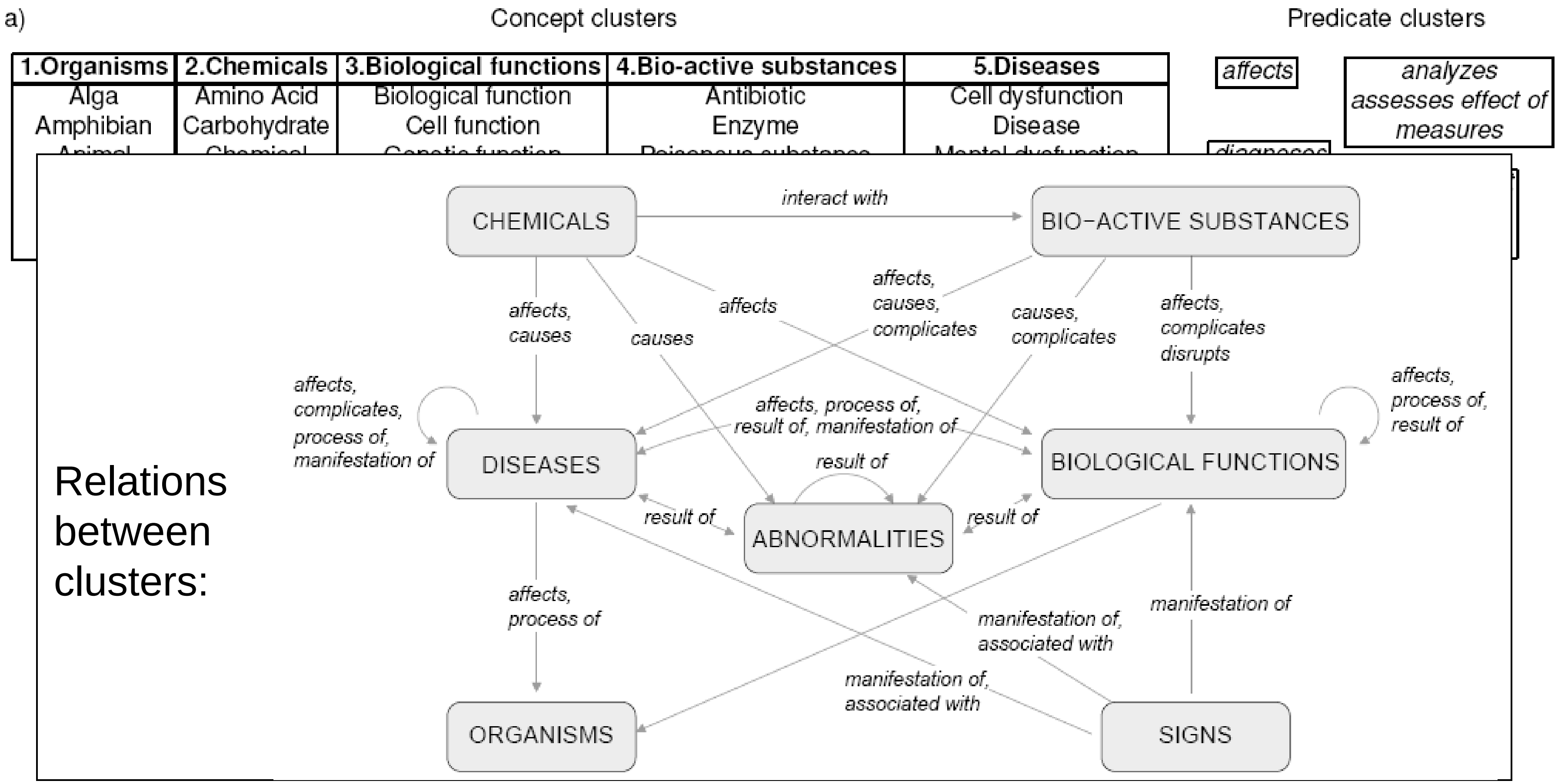


e.g., Diseases *affect* Organisms

Chemicals *interact with* Chemicals

Chemicals *cause* Diseases

Learning semantic networks with IRM

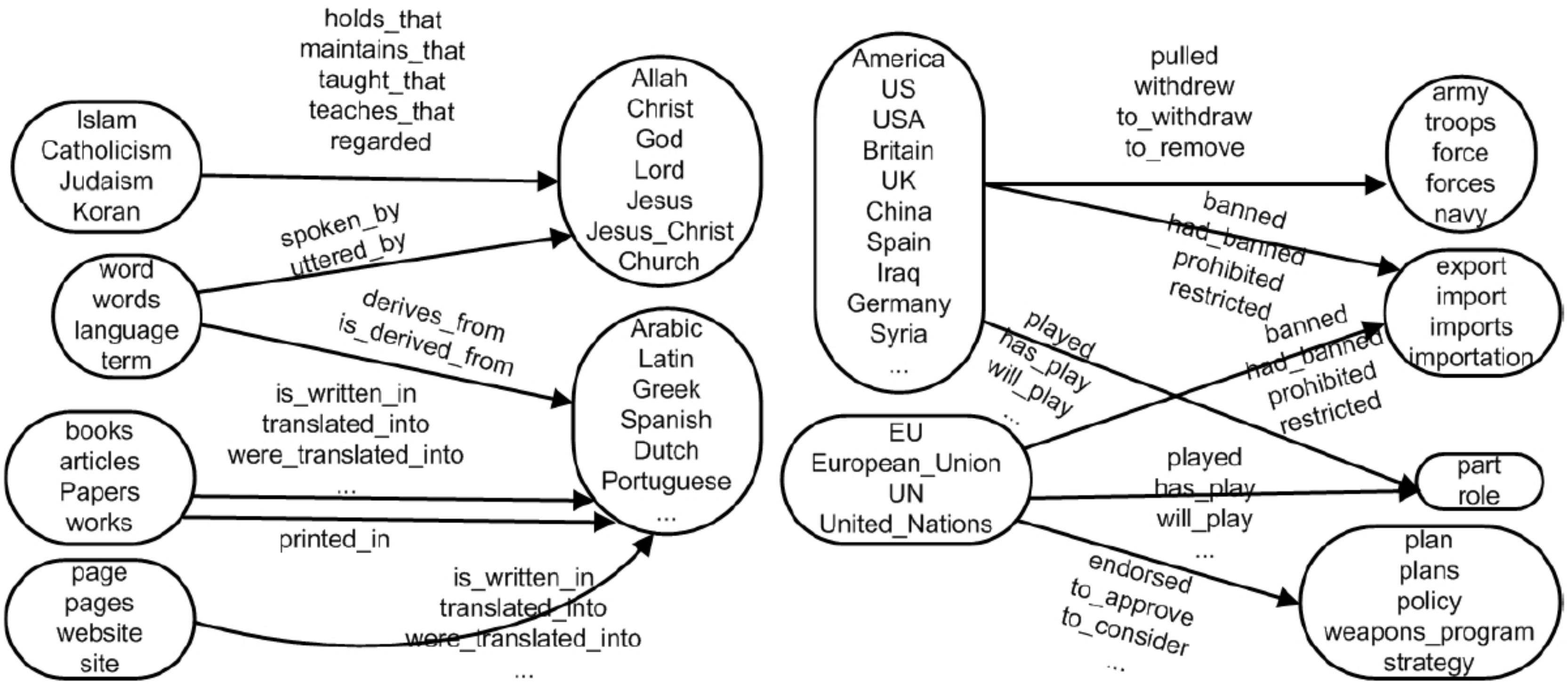


e.g., Diseases *affect* Organisms

Chemicals *interact with* Chemicals

Chemicals *cause* Diseases

Extracting semantic networks from text via relational clustering (Kok & Domingos 2008)



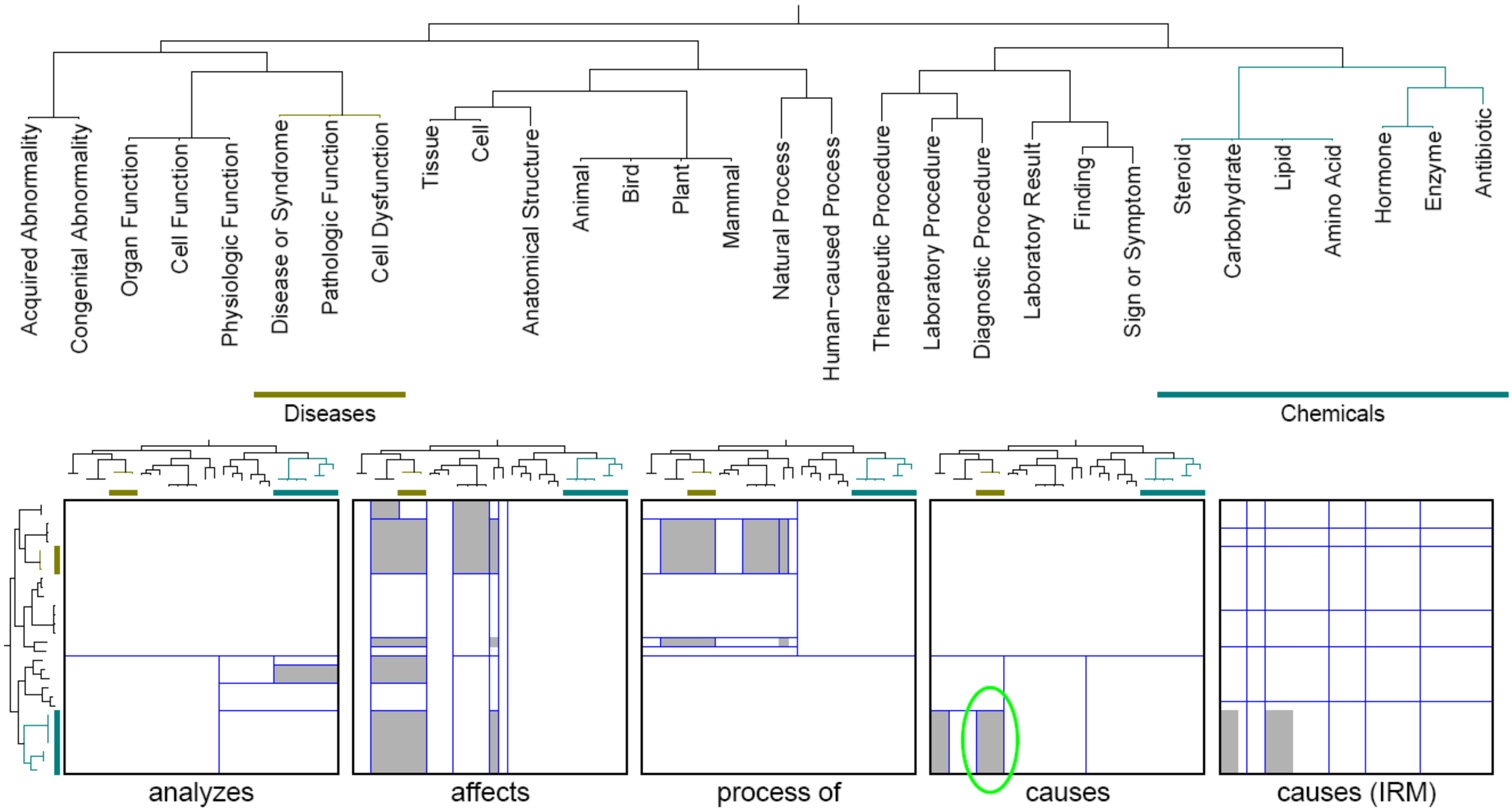
Tested several algorithms for relational clustering on TextRunner data:

~ 2 million triples of the form $R(x, y)$: e.g., upheld(Court, ruling), named_after(Jupiter, Roman_god).

~ 10,214 R symbols, 8942 x symbols, 7995 y symbols (each appears >25 times).

Annotated hierarchies model

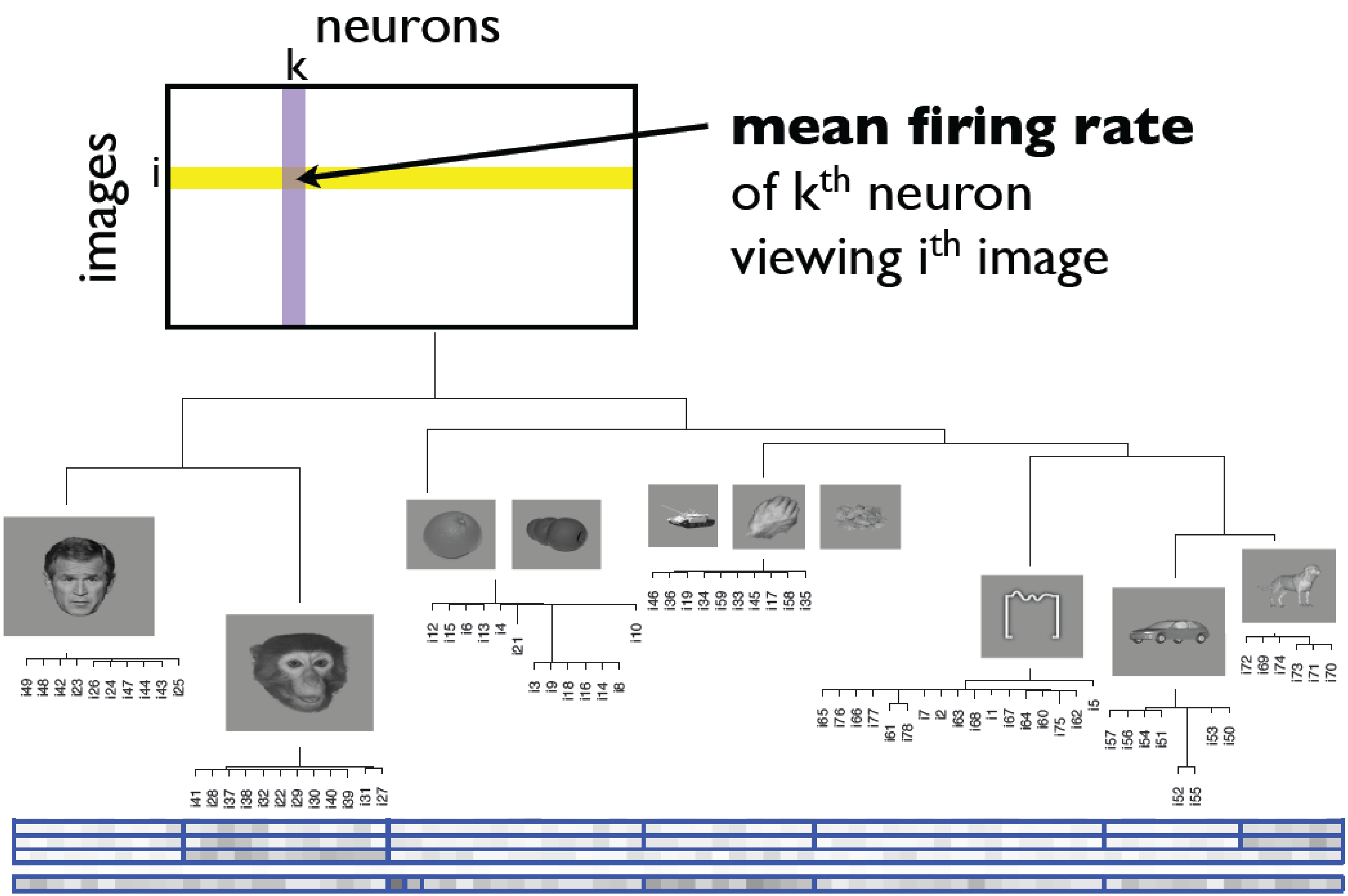
(Roy, Kemp, Mansinghka & Tenenbaum, 2007)



Annotated hierarchies model

170 neurons in Macaque Inferior Temporal (IT) cortex
78 grey-scale images

Hung, Kreiman, Poggio, and DiCarlo (Science 2005)



The Mondrian Process

(Roy & Teh, 2008; in prep)

We can also construct an exchangeable variant of the Annotated Hierarchies model (a hierarchical block model) by moving from the unit square to a product of random trees drawn from Kingman's coalescent prior (Kingman, 1982a). Let μ_d be Lebesgue measure.

$$T_d \sim \mathrm{KC}(\lambda), \forall d \in \{1, \ldots, D\} \qquad \textit{for each dimension, sample a tree} \tag{12}$$

$$M \mid T \sim \mathrm{MP}(2\alpha, T_1, \ldots, T_D) \qquad \textit{partition the cross product of trees} \tag{13}$$

$$\phi_S \mid M \sim \mathrm{Beta}(a_0, a_1),\ \forall S \in M. \qquad \textit{each block } S \textit{ gets a probability } \phi_S \tag{14}$$

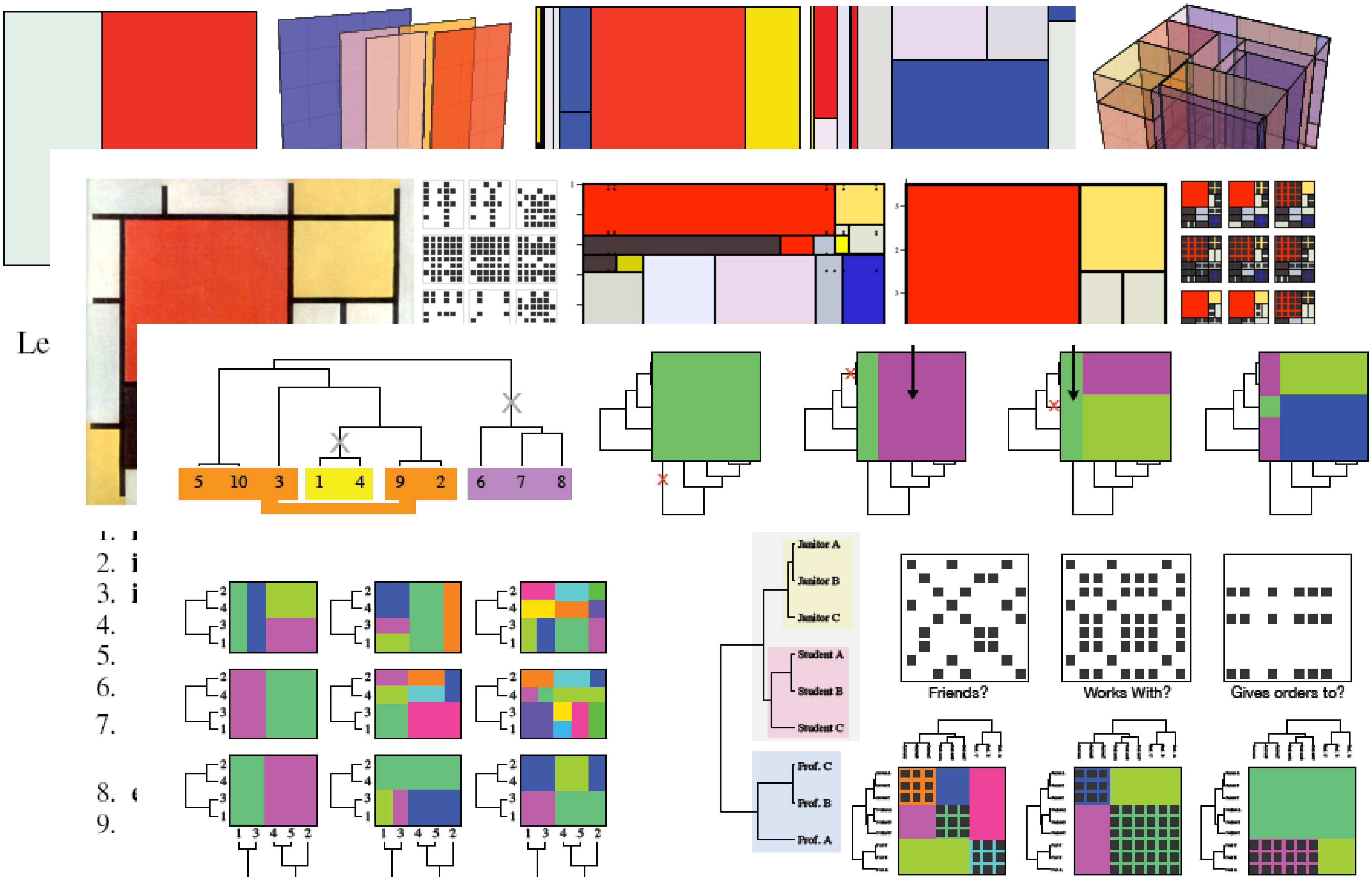
Let S_{ij} be the subset $S \in M$ where leaves (i, j) fall in S. Then

$$R_{ij} \mid \phi, M \sim \mathrm{Bernoulli}(\phi_{S_{ij}}), i, j \in \{1, \ldots, n\}. \qquad R_{ij} \textit{ is true w.p. } \phi_{S_{ij}} \tag{15}$$

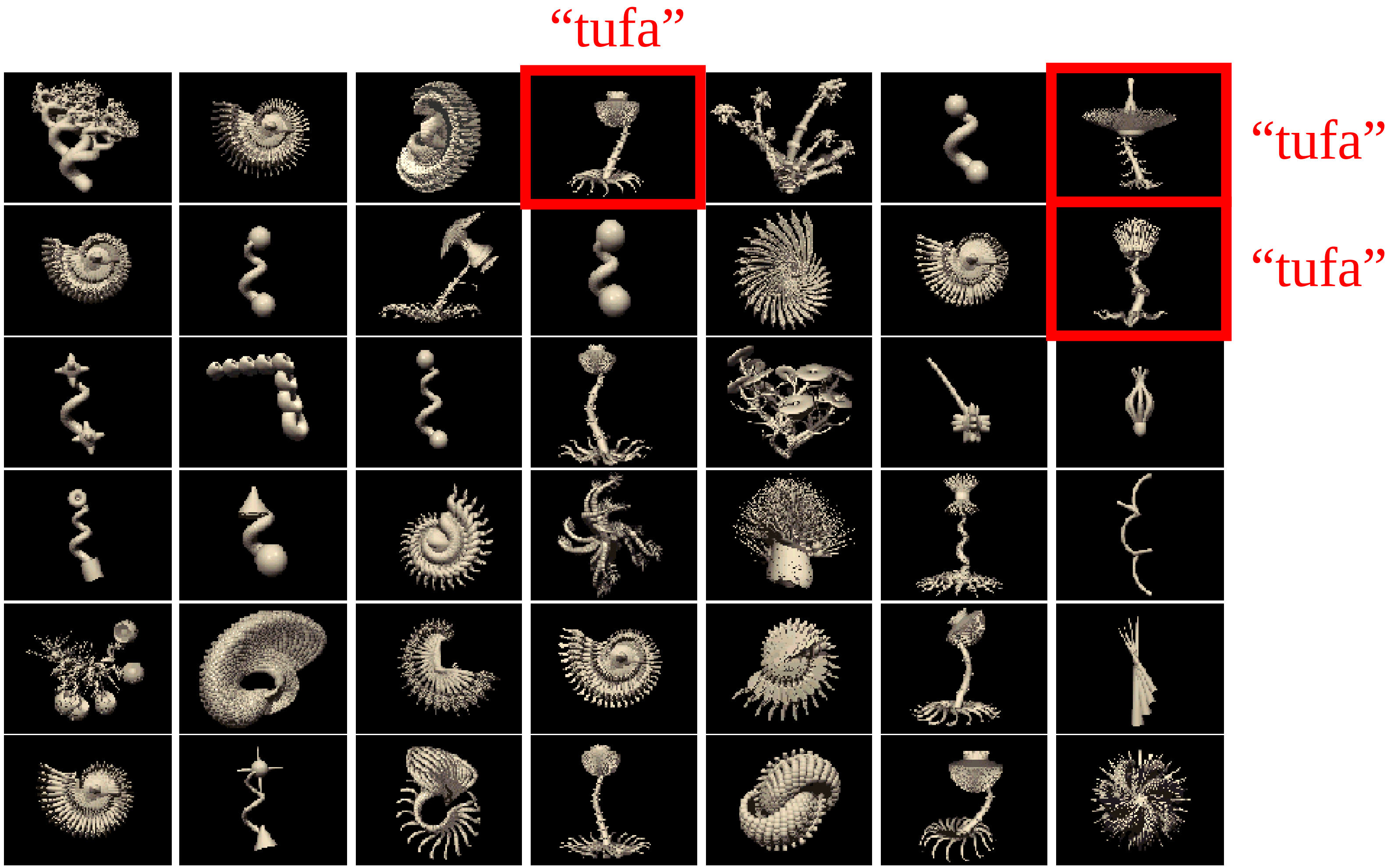
Algorithm 1 Conditional Mondrian $m \sim \mathrm{MP}(\lambda, \Theta_1, \ldots, \Theta_D \mid \rho)$ $\qquad \rho = \phi_d = \emptyset$ is unconditioned

1. **let** $\lambda' \leftarrow \lambda - E$ where $E \sim \mathrm{Exp}(\sum_{d=1}^{D} \mu_d(\Theta_d \setminus \Phi_d))$.
2. **if** ρ has no cuts **then** $\lambda'' \leftarrow 0$ **else** $\langle d', x', \lambda'', \rho_<, \rho_> \rangle \leftarrow \rho$.
3. **if** $\lambda' < \lambda''$ **then** *take root form of* ρ
4.   **if** ρ has no cut **then**
5.   **return** $m \leftarrow \Theta_1 \times \cdots \times \Theta_D$.
6.   **else** (d', x') *is the first cut in* m
7.   **return** $m \leftarrow \langle d', x', \lambda'', \mathrm{MP}(\lambda'', \Theta_1, \ldots, \Theta_{d'}^{<x'}, \ldots, \Theta_D \mid \rho_<),$
   $\mathrm{MP}(\lambda'', \Theta_1, \ldots, \Theta_{d'}^{>x'}, \ldots, \Theta_D \mid \rho_>)\rangle$.
8. **else** $\lambda'' < \lambda'$ *and there is a cut in* m *above* ρ
9.   draw a cut (d, x) outside ρ, i.e., $p(d) \propto \mu_d(\Theta_d \setminus \Phi_d)$, $x|d \sim \frac{\mu_d}{\mu_d(\Theta_d \setminus \Phi_d)}$
   without loss of generality suppose $\Phi_d \subset \Theta_d^{<x}$
10.   **return** $m \leftarrow \langle d, x, \lambda', \mathrm{MP}(\lambda', \Theta_1, \ldots, \Theta_d^{<x}, \ldots, \Theta_D \mid \rho),$
   $\mathrm{MP}(\lambda', \Theta_1, \ldots, \Theta_d^{>x}, \ldots, \Theta_D)\rangle$.

The Mondrian Process

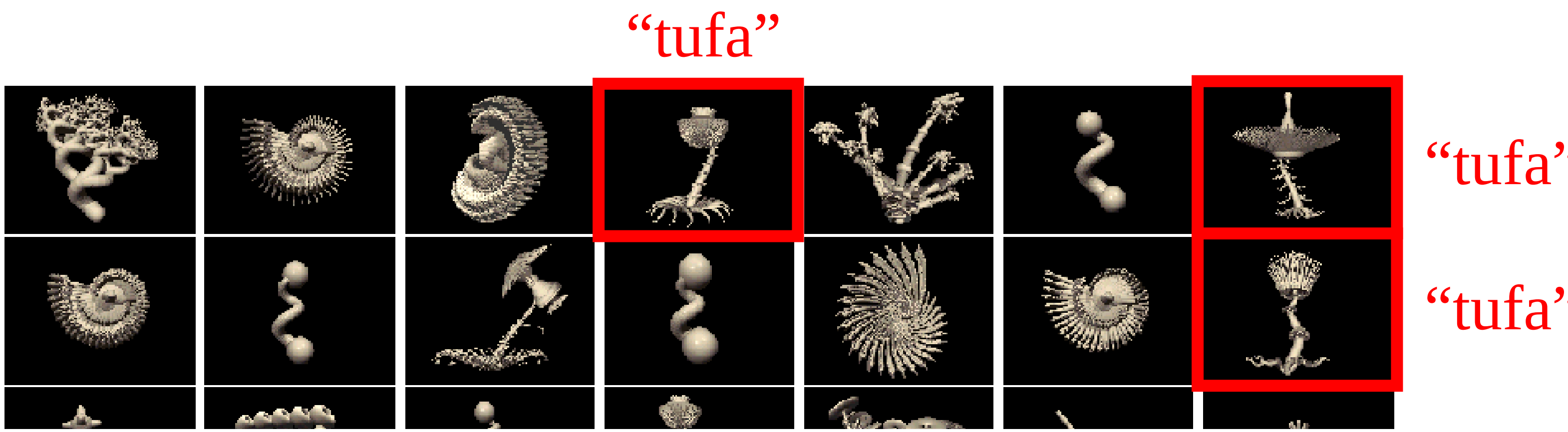


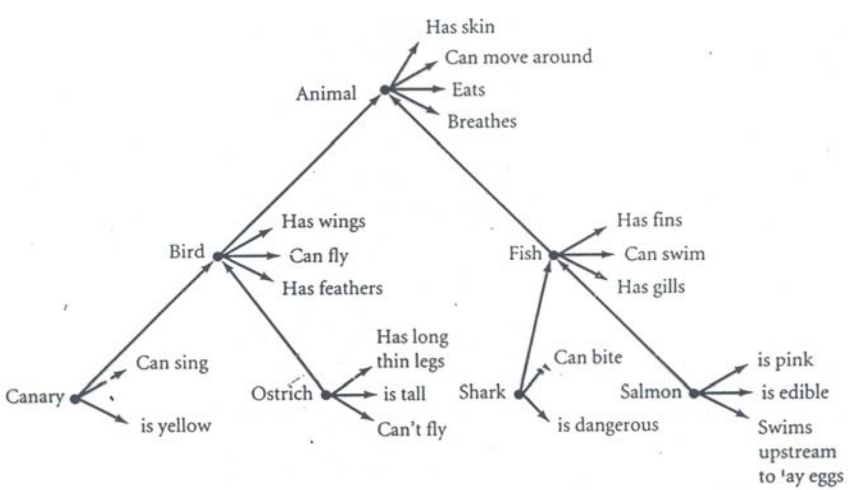
$$\mathrm{MP}(\lambda', \Theta_1, \ldots, \Theta_d^{>x}, \ldots, \Theta_D)\rangle.$$



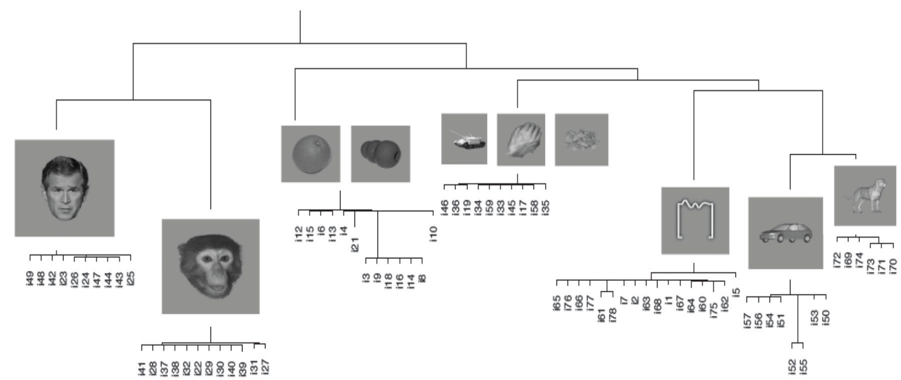
Learning from just one or a few examples, and mostly unlabeled examples ("semi-supervised learning").

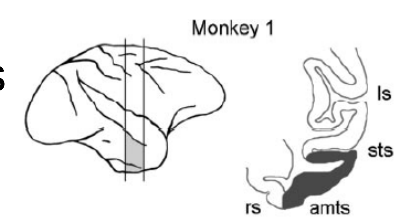
Learning words for objects





(Collins & Quillian, 1969)



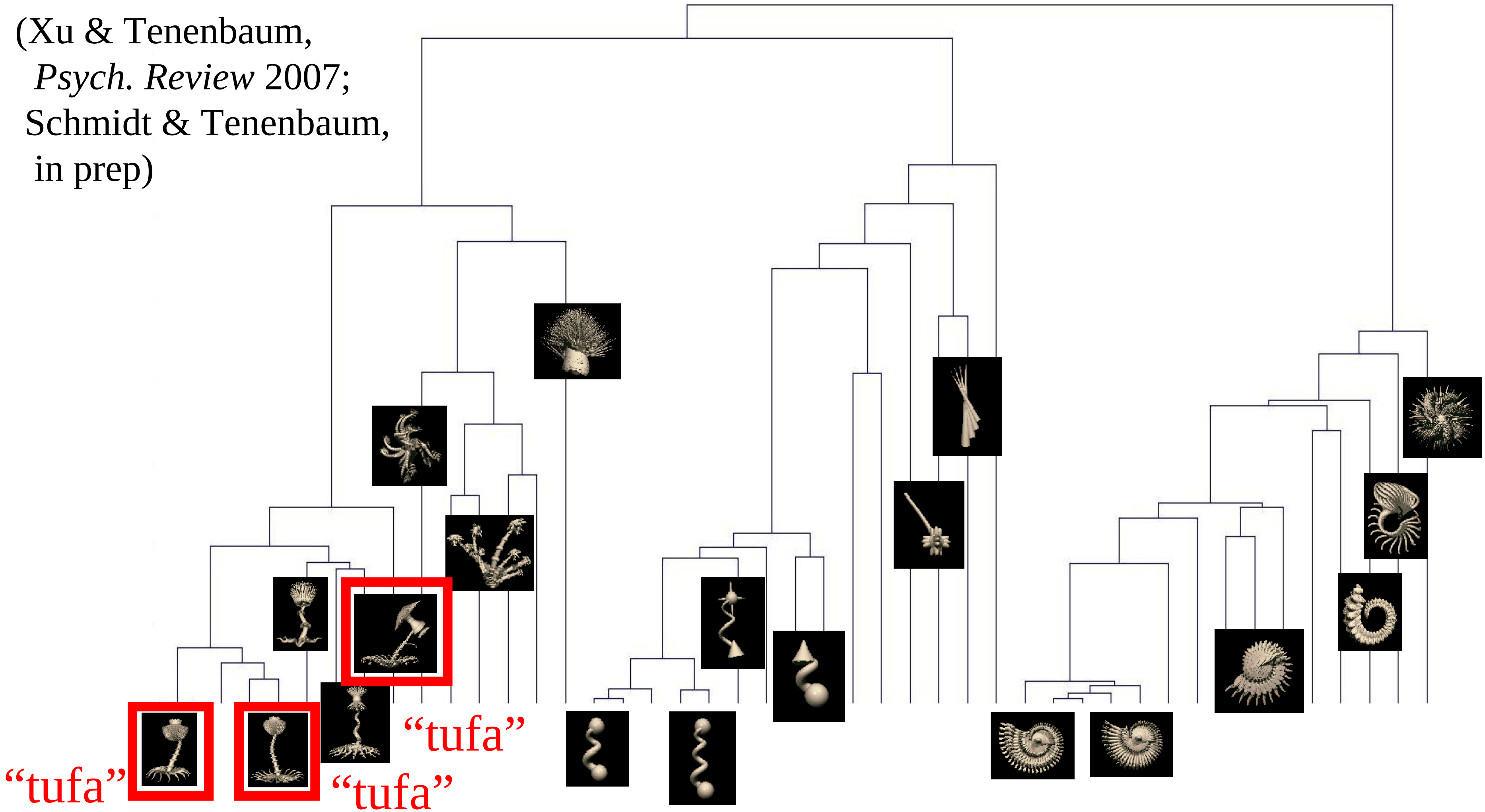


(IT population responses Hung et al., 2005; c.f. Kiani et al. 2007)

Learning words for objects

Bayesian inference over tree-structured hypothesis space:

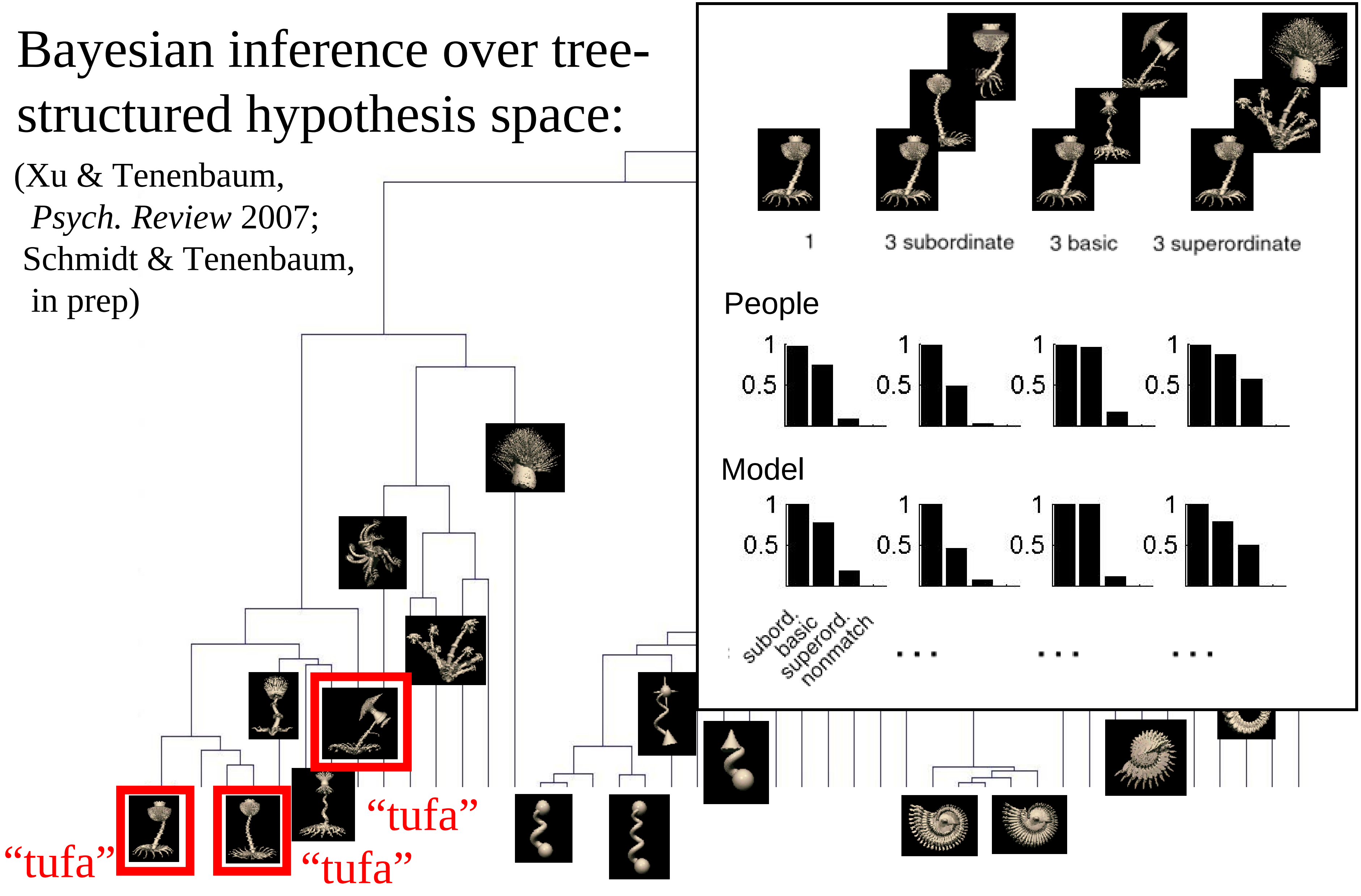
(Xu & Tenenbaum, *Psych. Review* 2007; Schmidt & Tenenbaum, in prep)



Learning words for objects

Bayesian inference over tree-structured hypothesis space:

(Xu & Tenenbaum, *Psych. Review* 2007; Schmidt & Tenenbaum, in prep)



Hierarchical Bayesian framework

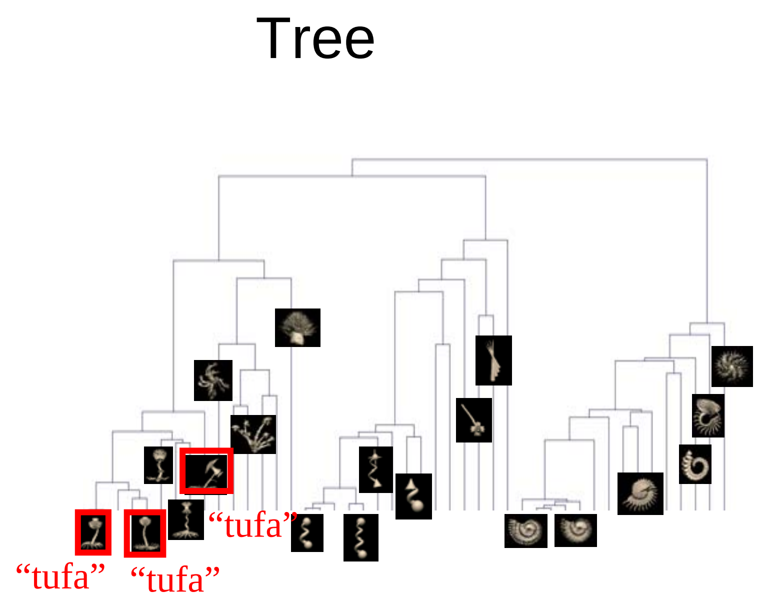
F: form

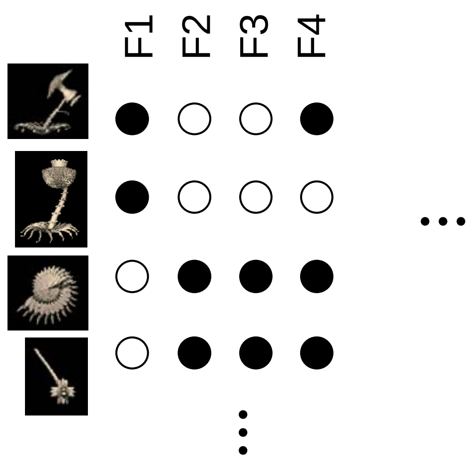
↓

S: structure

↓

D: data



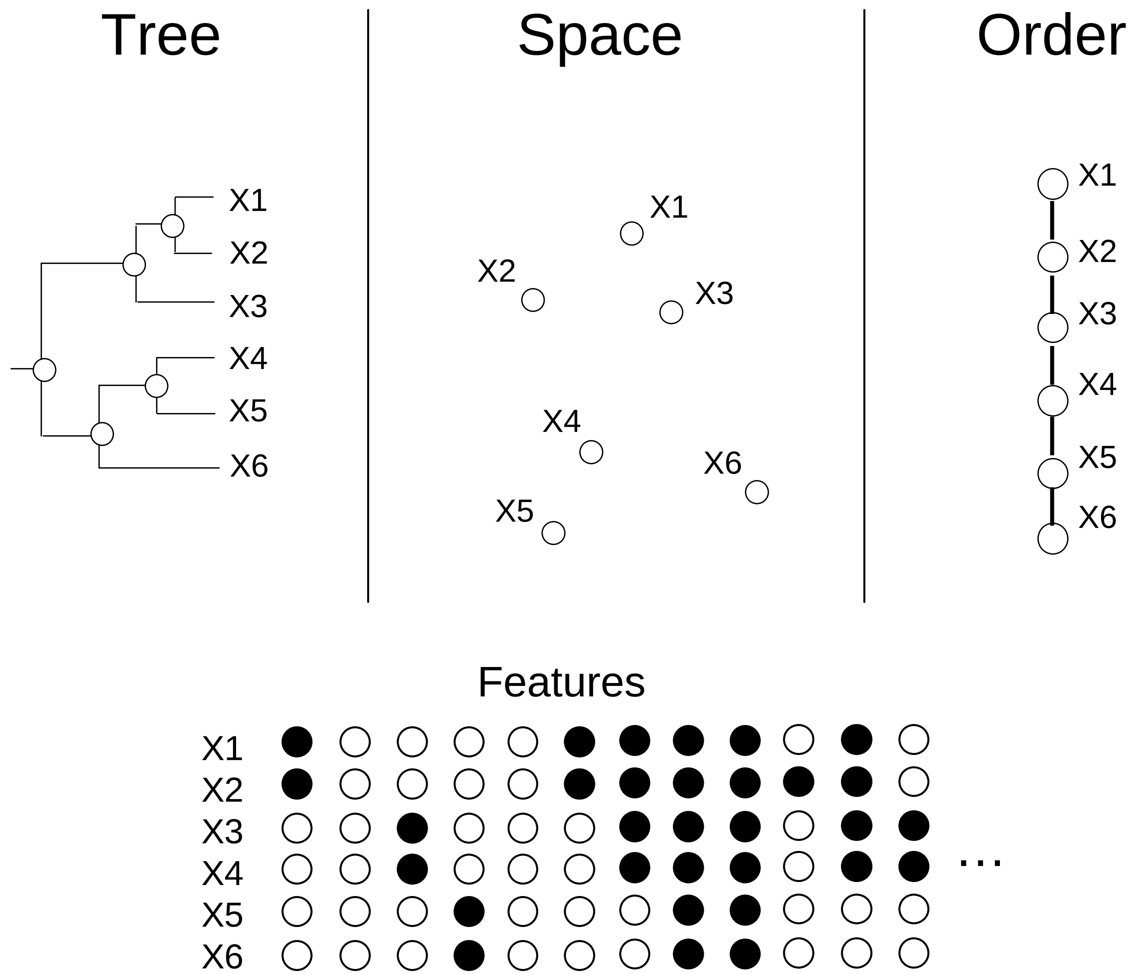


Learning to learn: what is the right form of structure for the domain?

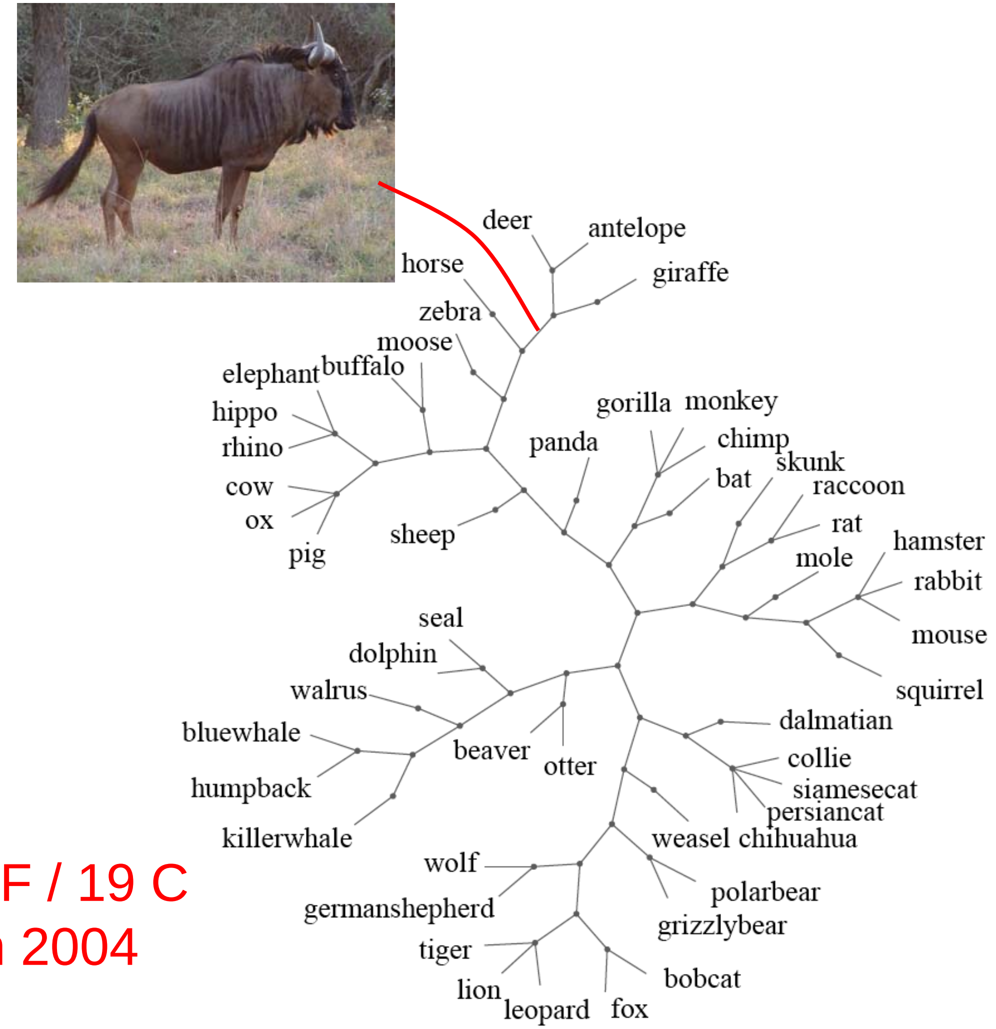
F: form

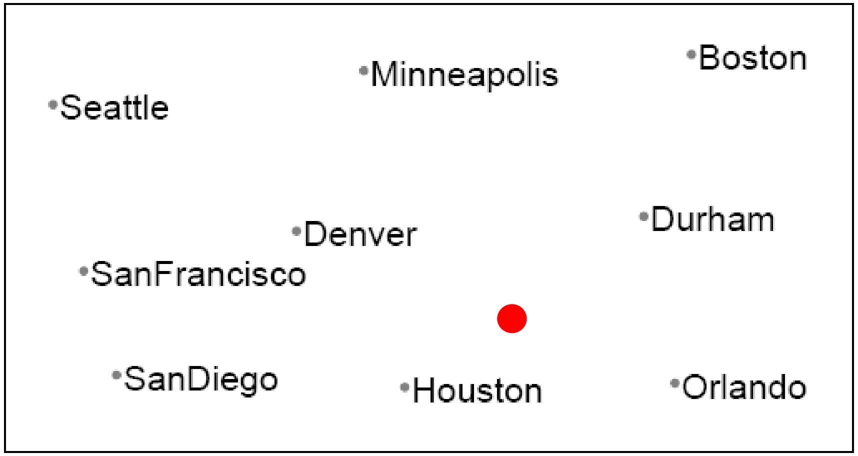
S: structure

D: data



The value of structural form knowledge: inductive constraints (bias)





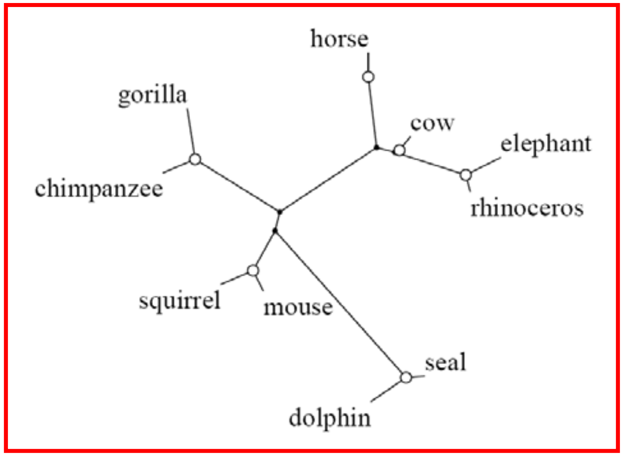
Mystery city …

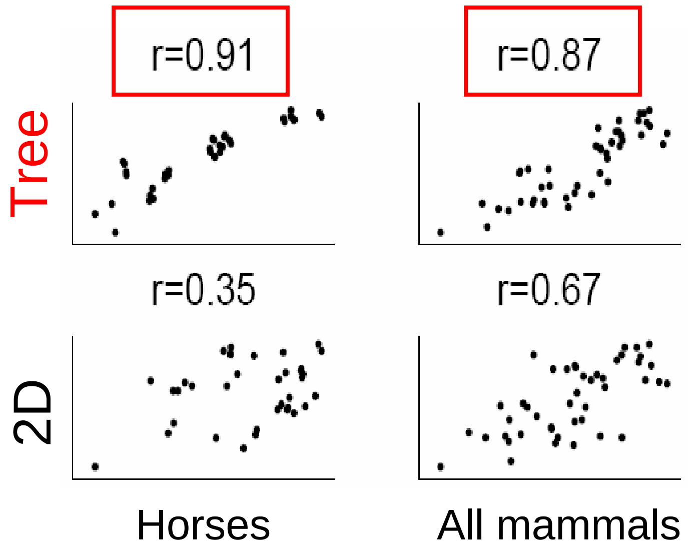
- average annual temperature: 66 F / 19 C
- voted 60% for George W Bush in 2004
- popular foods are fried and BBQ

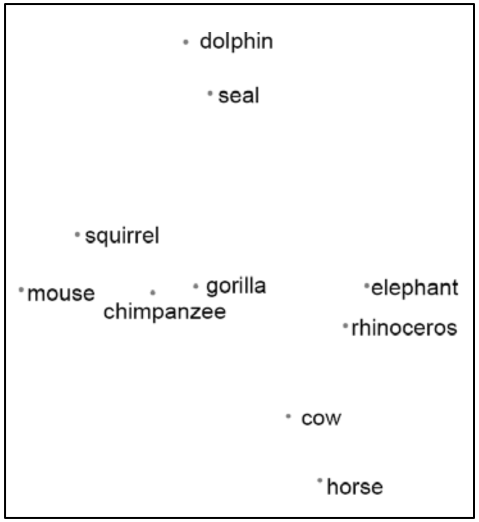
Property induction

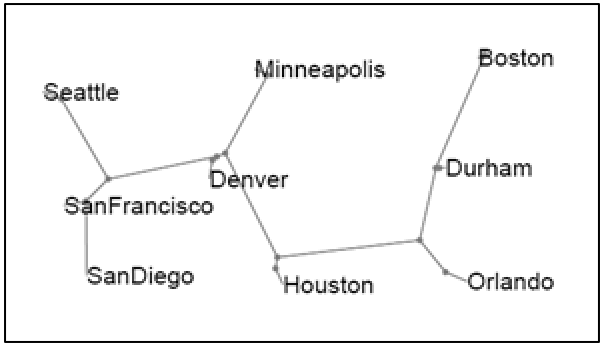
(Kemp & Tenenbaum, *Psych. Review* 2009; Shafto et al., *Cognition* 2008)

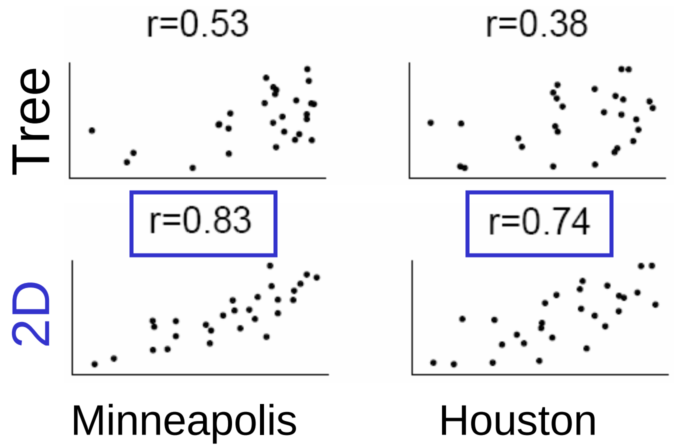
Given that $\{X_1, \ldots, X_n\}$ have property P, how likely is it that Y does?

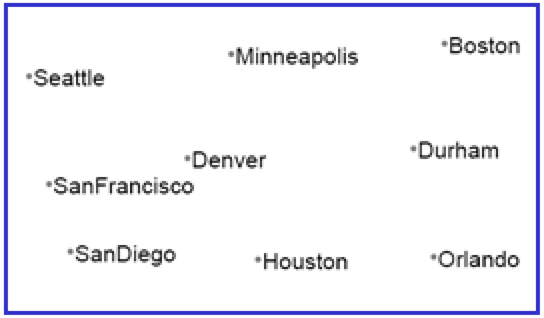








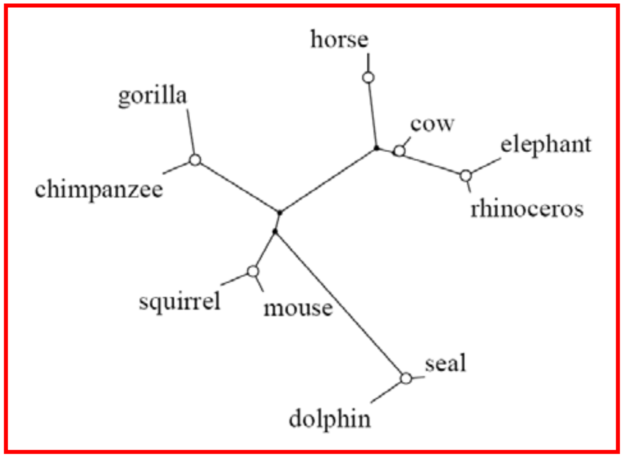


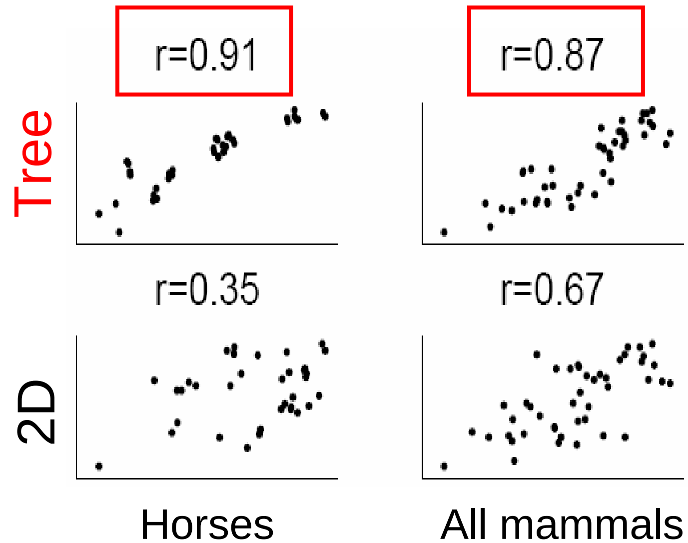


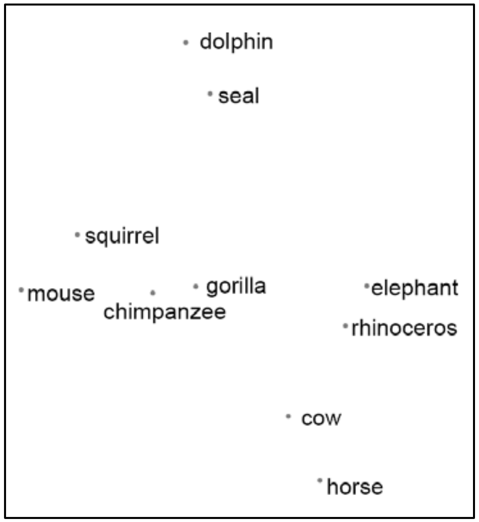
Property induction

(Kemp & Tenenbaum, *Psych. Review* 2009)

Given that $\{X_1, \ldots, X_n\}$ have property P, how likely is it that Y does?





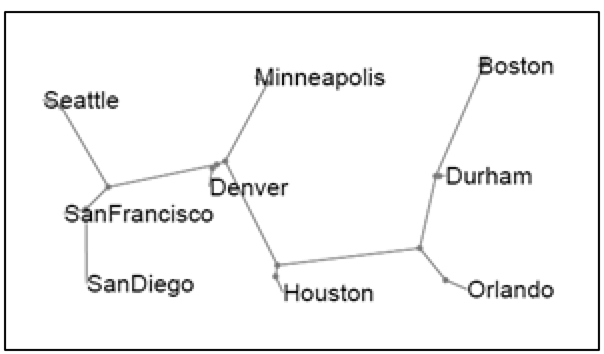


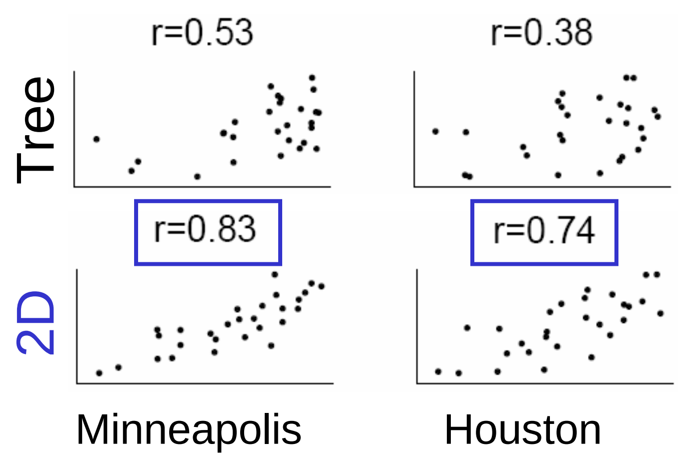
Property ~ $N\left(0, (\Delta + \frac{1}{\sigma^2} I)^{-1}\right)$

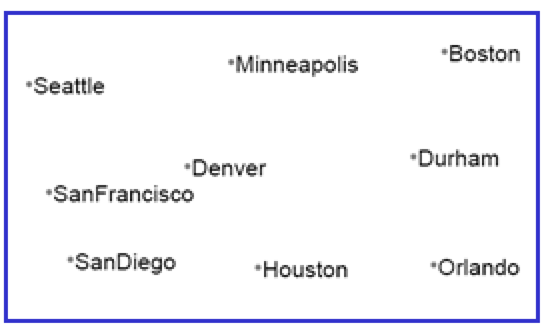
(Zhu, Lafferty & Ghahramani, 2003)

$N\left(0, \exp(-\frac{1}{\sigma} \| x_i - x_j \|)\right)$

(c.f. Lawrence, 2004)



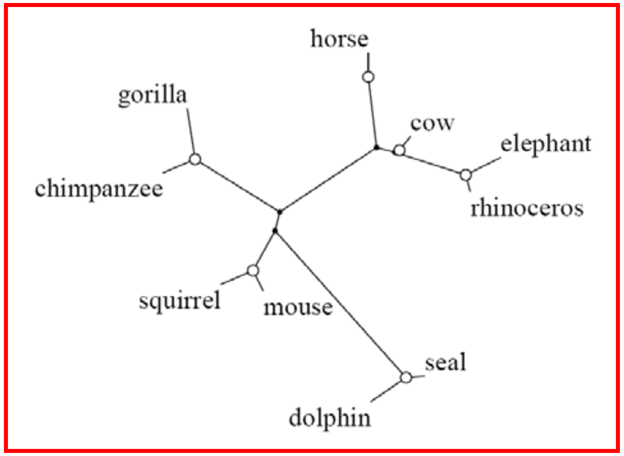


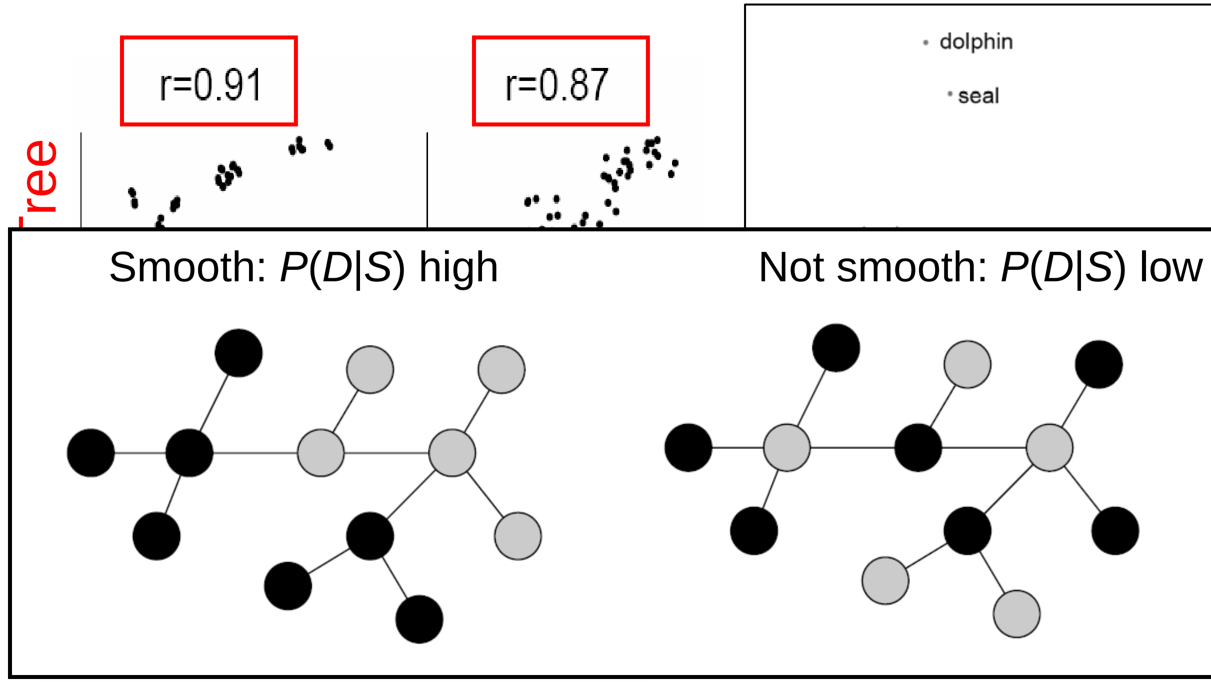


Property induction

(Kemp & Tenenbaum, *Psych. Review* 2009; Shafto et al., *Cognition* 2008)

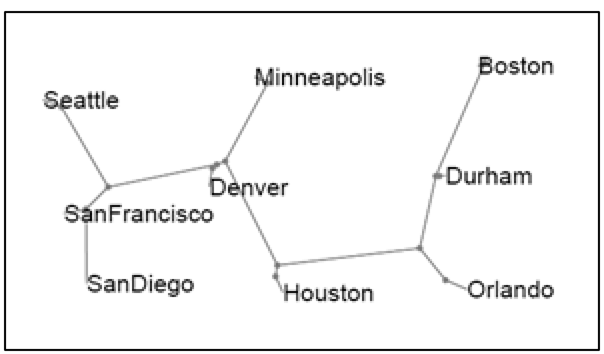
Given that $\{X_1, \ldots, X_n\}$ have property P, how likely is it that Y does?

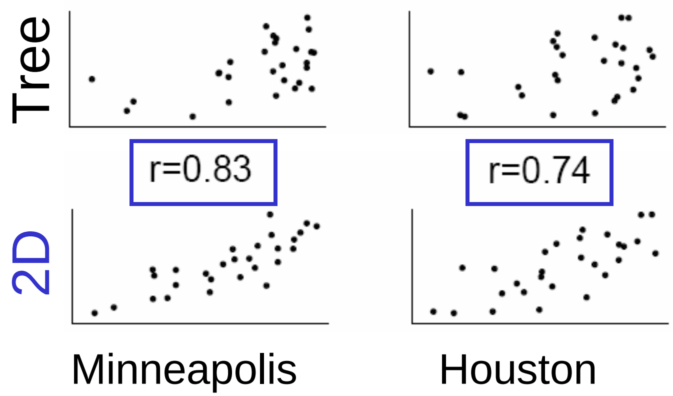




Property ~ $N\left(0, (\Delta + \frac{1}{\sigma^2} I)^{-1}\right)$

(Zhu, Lafferty & Ghahramani, 2003)





Learning structural forms

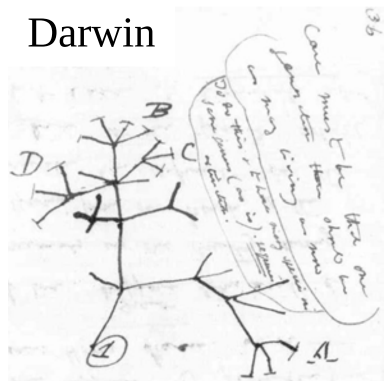
People can discover structural forms…

– Scientists

Linnaeus

Kingdom Animalia
Phylum Chordata
Class Mammalia
Order Primates
Family Hominidae
Genus Homo
Species *Homo sapiens*

Darwin



Mendeleev

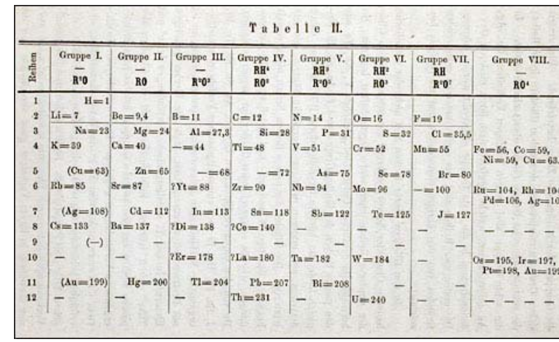


Tabelle II.

Reihen	Gruppe I. — R^2O	Gruppe II. — RO	Gruppe III. — R^2O^3	Gruppe IV. RH^4 RO^2	Gruppe V. RH^3 R^2O^5	Gruppe VI. RH^2 RO^3	Gruppe VII. RH R^2O^7	Gruppe VIII. — RO^4
1	H=1							
2	Li=7	Be=9,4	B=11	C=12	N=14	O=16	F=19	
3	Na=23	Mg=24	Al=27,3	Si=28	P=31	S=32	Cl=35,5	
4	K=39	Ca=40	—=44	Ti=48	V=51	Cr=52	Mn=55	Fe=56, Co=59, Ni=59, Cu=63.
5	(Cu=63)	Zn=65	—=68	—=72	As=75	Se=78	Br=80	
6	Rb=85	Sr=87	?Yt=88	Zr=90	Nb=94	Mo=96	—=100	Ru=104, Rh=104, Pd=106, Ag=108.
7	(Ag=108)	Cd=112	In=113	Sn=118	Sb=122	Te=125	J=127	
8	Cs=133	Ba=137	?Di=138	?Ce=140	—	—	—	— — — —
9	(—)	—	—	—	—	—	—	
10	—	—	?Er=178	?La=180	Ta=182	W=184	—	Os=195, Ir=197, Pt=198, Au=199.
11	(Au=199)	Hg=200	Tl=204	Pb=207	Bi=208	—	—	
12	—	—	—	Th=231	—	U=240	—	— — — —

– Children

e.g., hierarchical structure of category labels, cyclical structure of seasons or days of the week, clique structure of social networks.

… but standard learning algorithms assume fixed forms.

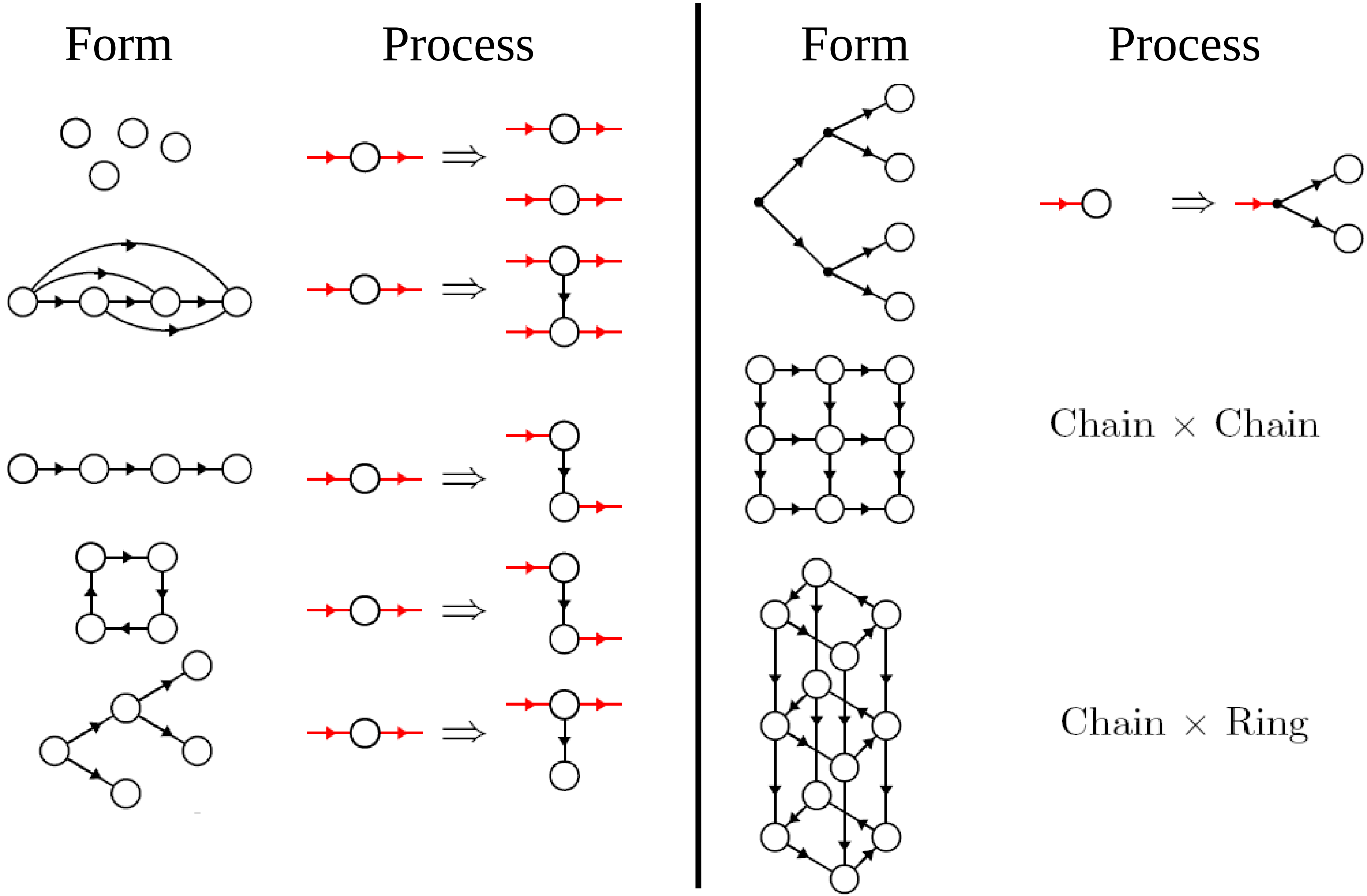
– Principal components analysis: low-dimensional spatial structure

– Hierarchical clustering: tree structure

– *k*-means clustering, mixture models: flat partition.

Hypothesis space of structural forms

(Kemp & Tenenbaum, PNAS 2008)



A hierarchical Bayesian approach

(Kemp & Tenenbaum, PNAS 2008)

$P(F)$

F: form

$P(S \mid F)$

Simplicity

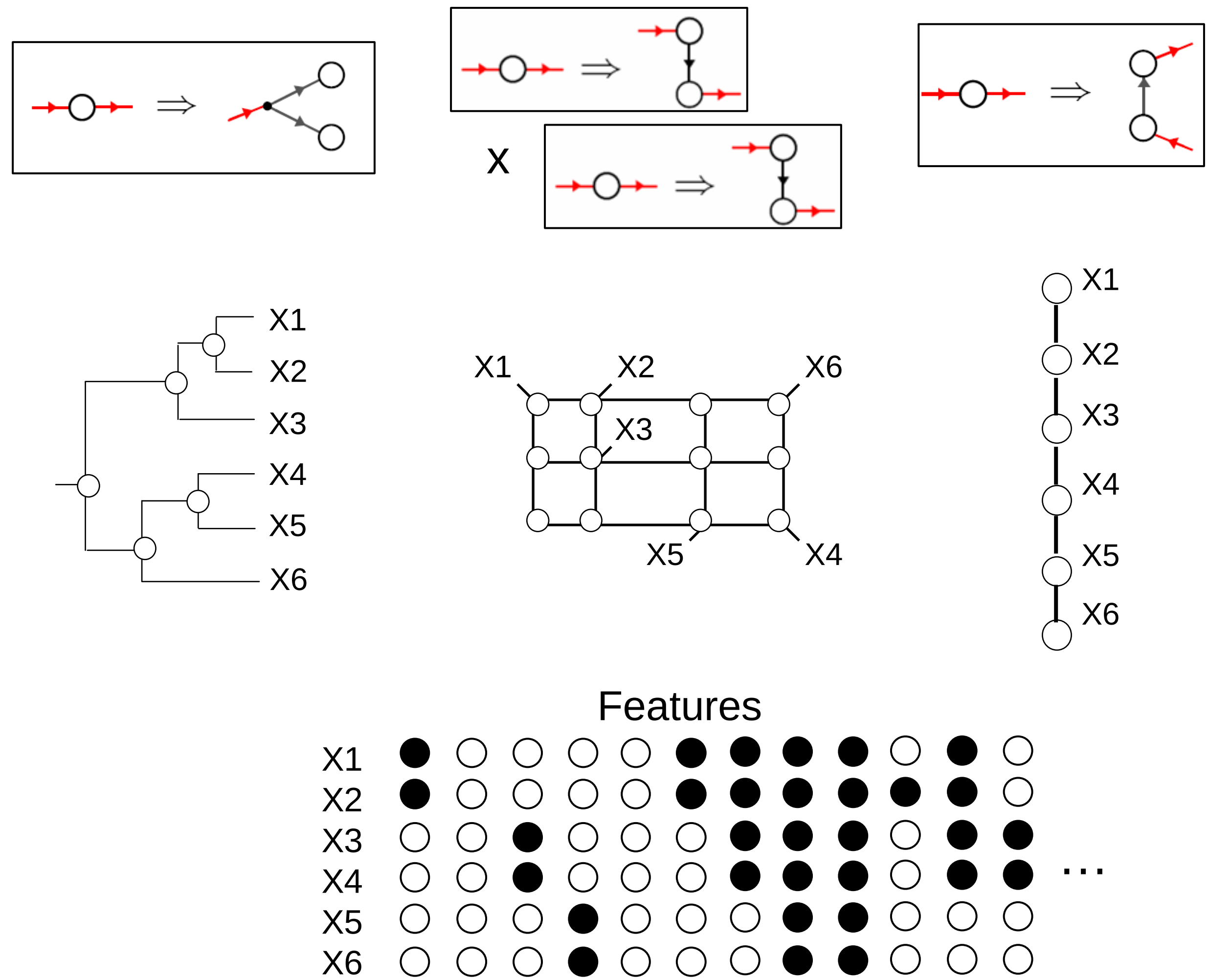
S: structure

$P(D \mid S)$

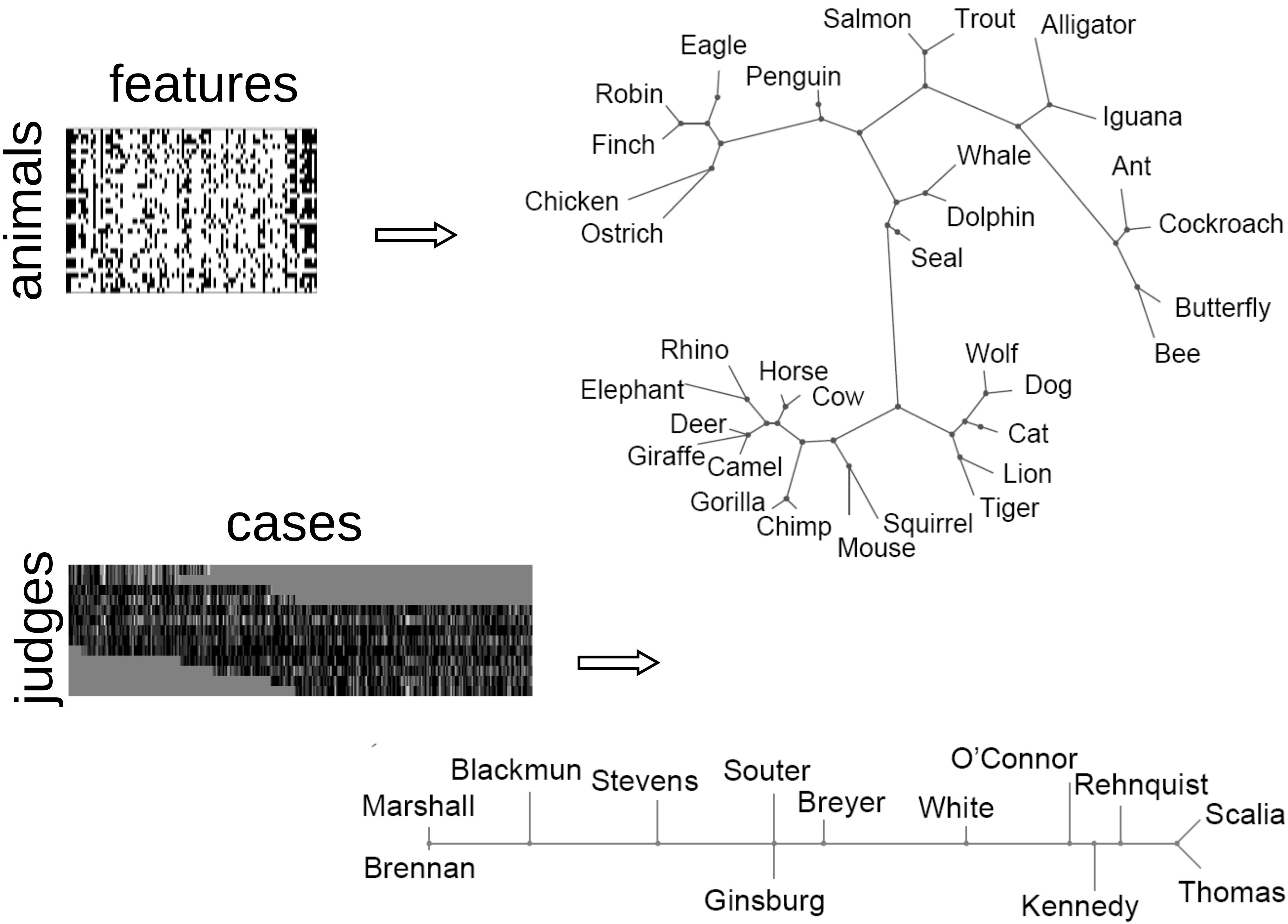
Smoothness

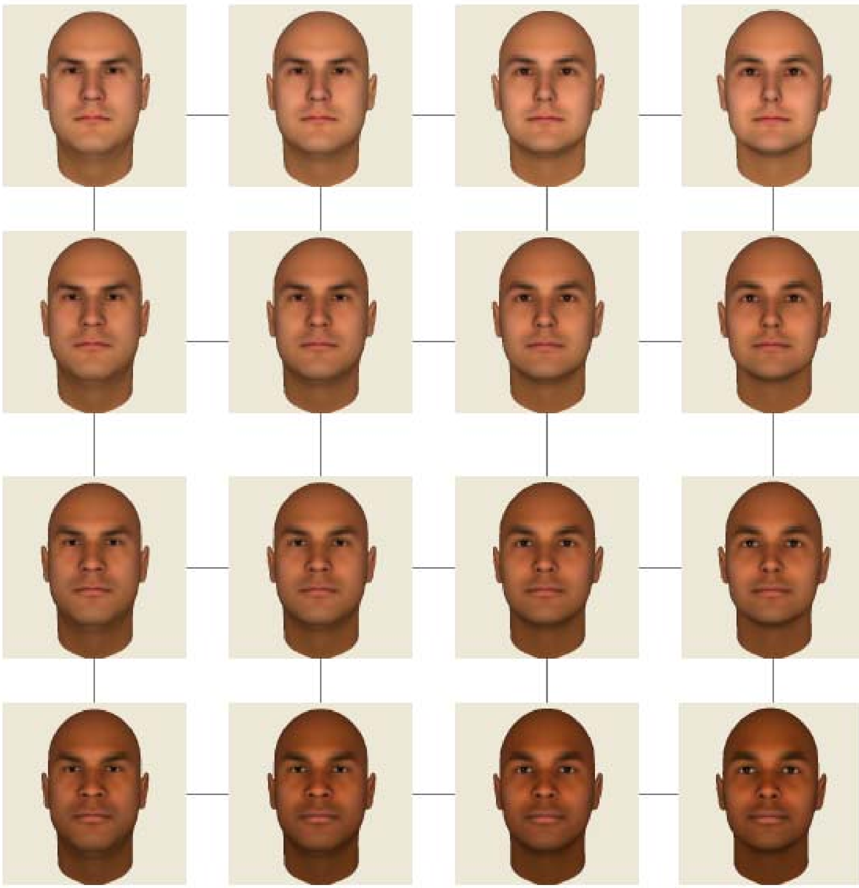
(Fit to data)

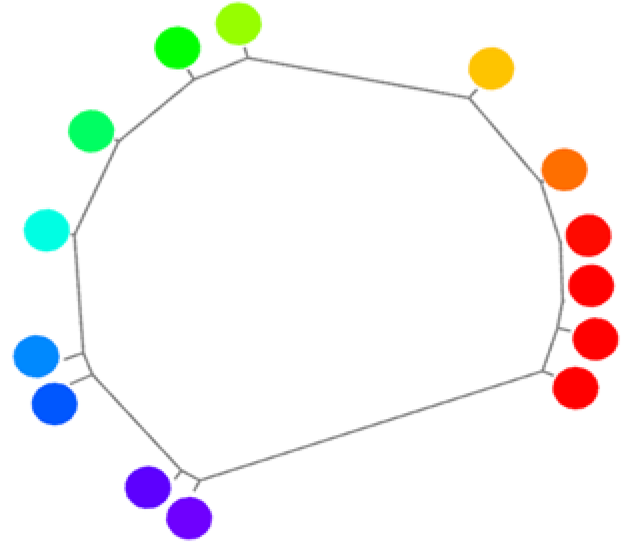
D: data

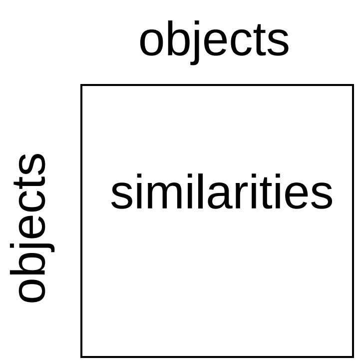


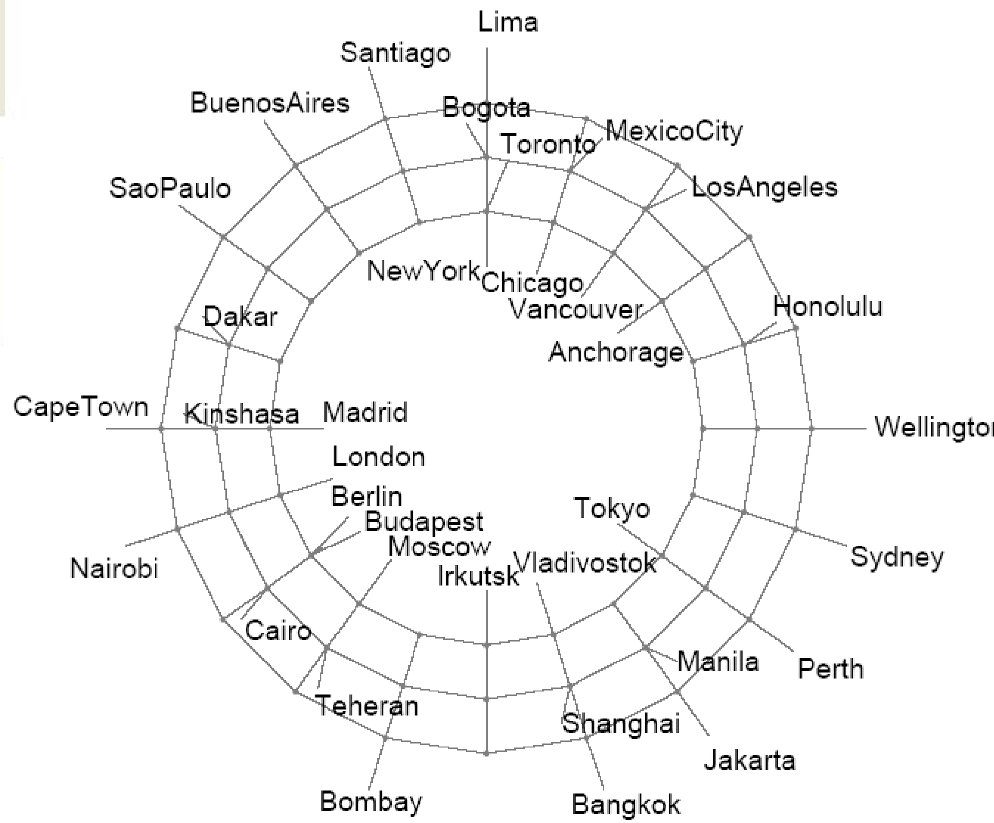
$$P(S, F|D) \propto \underline{P(D|S)}\,\underline{P(S|F)}\,P(F)$$


features
animals
Salmon
Trout
Alligator
Eagle
Penguin
Robin
Iguana
Finch
Whale
Ant
Chicken
Ostrich
Dolphin
Cockroach
Seal
Butterfly
Bee
Rhino
Wolf
Horse
Elephant
Cow
Dog
Deer
Cat
Giraffe
Camel
Lion
Gorilla
Tiger
Chimp
Squirrel
Mouse
cases
judges
O'Connor
Blackmun
Stevens
Souter
Rehnquist
Marshall
Breyer
White
Scalia
Brennan
Thomas
Ginsburg
Kennedy



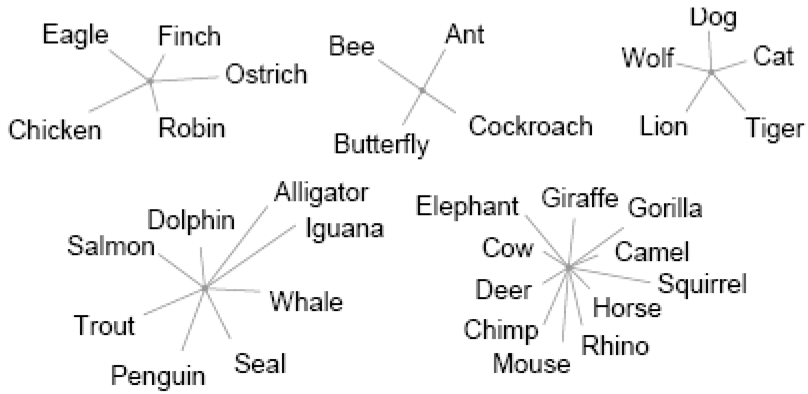



objects
objects
similarities

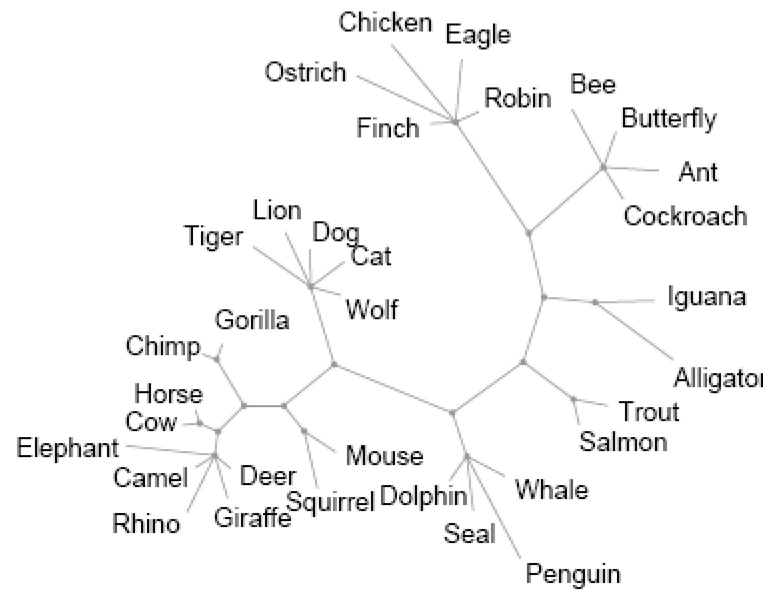

Lima
Santiago
BuenosAires
Bogota
MexicoCity
Toronto
SaoPaulo
LosAngeles
NewYork
Chicago
Vancouver
Honolulu
Dakar
Anchorage
CapeTown
Kinshasa
Madrid
Wellington
London
Berlin
Budapest
Moscow
Tokyo
Sydney
Nairobi
Irkutsk
Vladivostok
Cairo
Manila
Perth
Teheran
Shanghai
Jakarta
Bombay
Bangkok

Development of structural forms as more data are observed

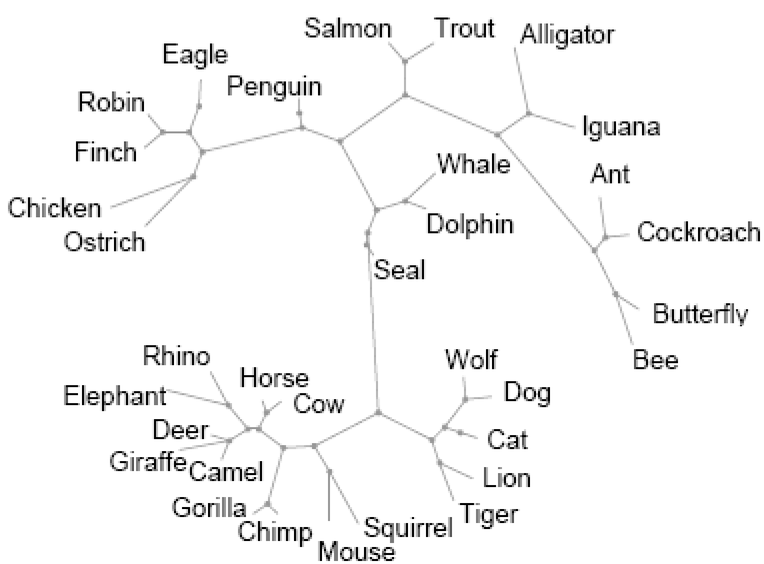
5 features



20 features



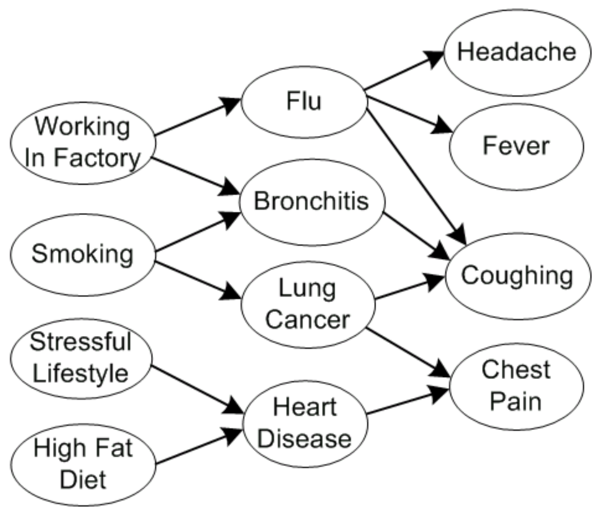
110 features



“blessing of abstraction”

Causal learning and reasoning

Causal model



↓

Event data

Patient 1: Stressful lifestyle
Chest Pain

Patient 2: Smoking
Coughing

Patient 3: Working in factory
Chest Pain

. . .

(Griffiths & Tenenbaum; Kemp, Goodman, Tenenbaum)

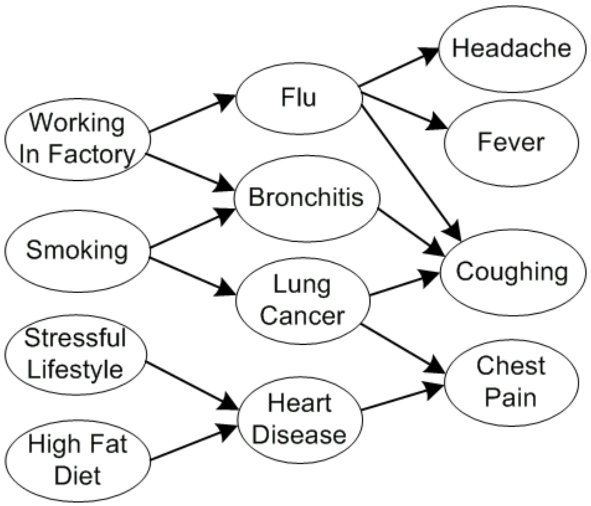
Causal learning and reasoning

Causal schema

Behaviors → Diseases → Symptoms

Behaviors	Diseases	Symptoms
high-fat diet working in factory …	heart disease lung cancer …	coughing chest pain …

Causal model



Cut down hypothesis space from size 521,939,651,343,829,405,020,504,063 to 131,072

Event data

Patient 1: Stressful lifestyle
Chest Pain

Patient 2: Smoking
Coughing

Patient 3: Working in factory
Chest Pain

. . .

(Griffiths & Tenenbaum; Kemp, Goodman, Tenenbaum)

Causal schema

Causal model

Event data

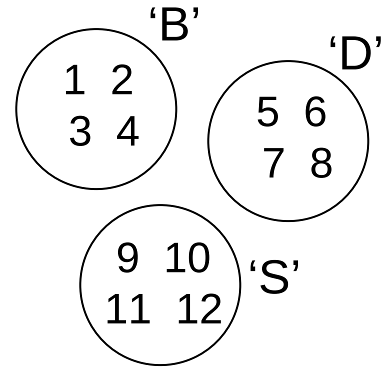
Laws

η

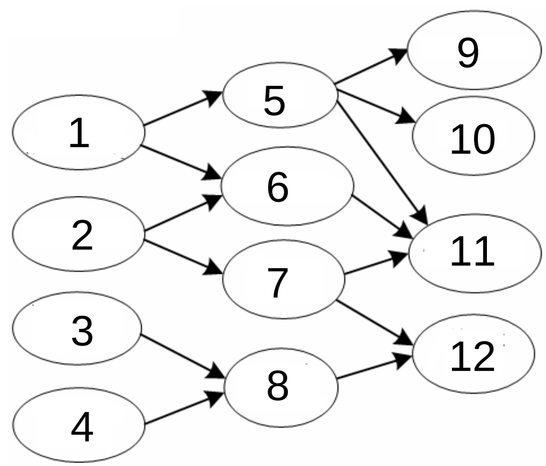
	'B'	'D'	'S'
'B'	0.0	0.3	0.01
'D'	0.0	0.0	0.25
'S'	0.0	0.0	0.0

Classes

z



Infinite relational model



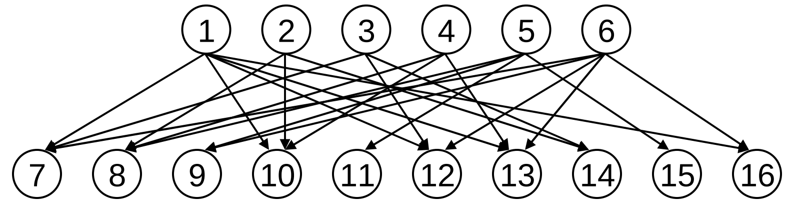
Belief net with Dirichlet-multinomial parameterization

	1	2	3	4	5	6	7	8	9	10	11	12
Patient 1	●	○	○	○	○	●	●	●	●	○	●	○
Patient 2	●	○	○	○	○	●	●	●	●	●	●	○
Patient 3	○	○	●	○	○	○	●	●	●	○	●	●
Patient 4	○	○	●	○	○	○	●	●	●	○	●	●
Patient 5	○	○	○	●	○	○	○	●	●	○	○	○

…

(Mansinghka, Kemp, Tenenbaum, Griffiths, UAI 2006)

Ground-truth causal network



20 80 1000 # samples

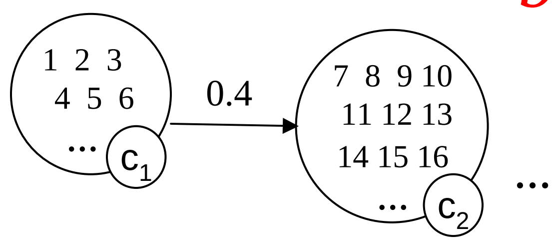
Causal model

↓

Event data

recovered model

"blessing of abstraction"



recovered schema

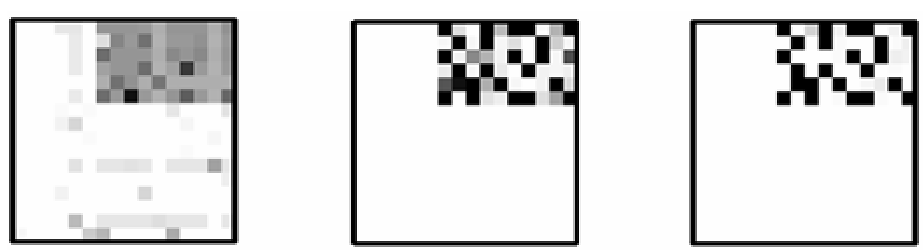
Causal schema

↓

Causal model

↓

Event data



recovered model

(Mansinghka, Kemp, Tenenbaum, Griffiths, UAI 2006)

Conclusions

How does the mind get so much from so little, in learning about objects, classes, causes, scenes, sentences, thoughts, social systems?

A toolkit for studying the nature, use and acquisition of abstract knowledge:

- *Bayesian inference* in probabilistic generative models.
- Probabilistic models defined over a range of *structured representations*: spaces, graphs, grammars, predicate logic, schemas, programs.
- *Hierarchical models*, with inference at multiple levels of abstraction.
- *Nonparametric models*, adapting their complexity to the data and balancing constraint with flexibility.

An alternative to classic "either-or" dichotomies: Nature versus Nurture, Logic (Structure, Rules, Symbols) versus Probability (Statistics).

- How can domain-general mechanisms of learning and representation build domain-specific abstract knowledge?
- How can structured symbolic knowledge be acquired by statistical learning?

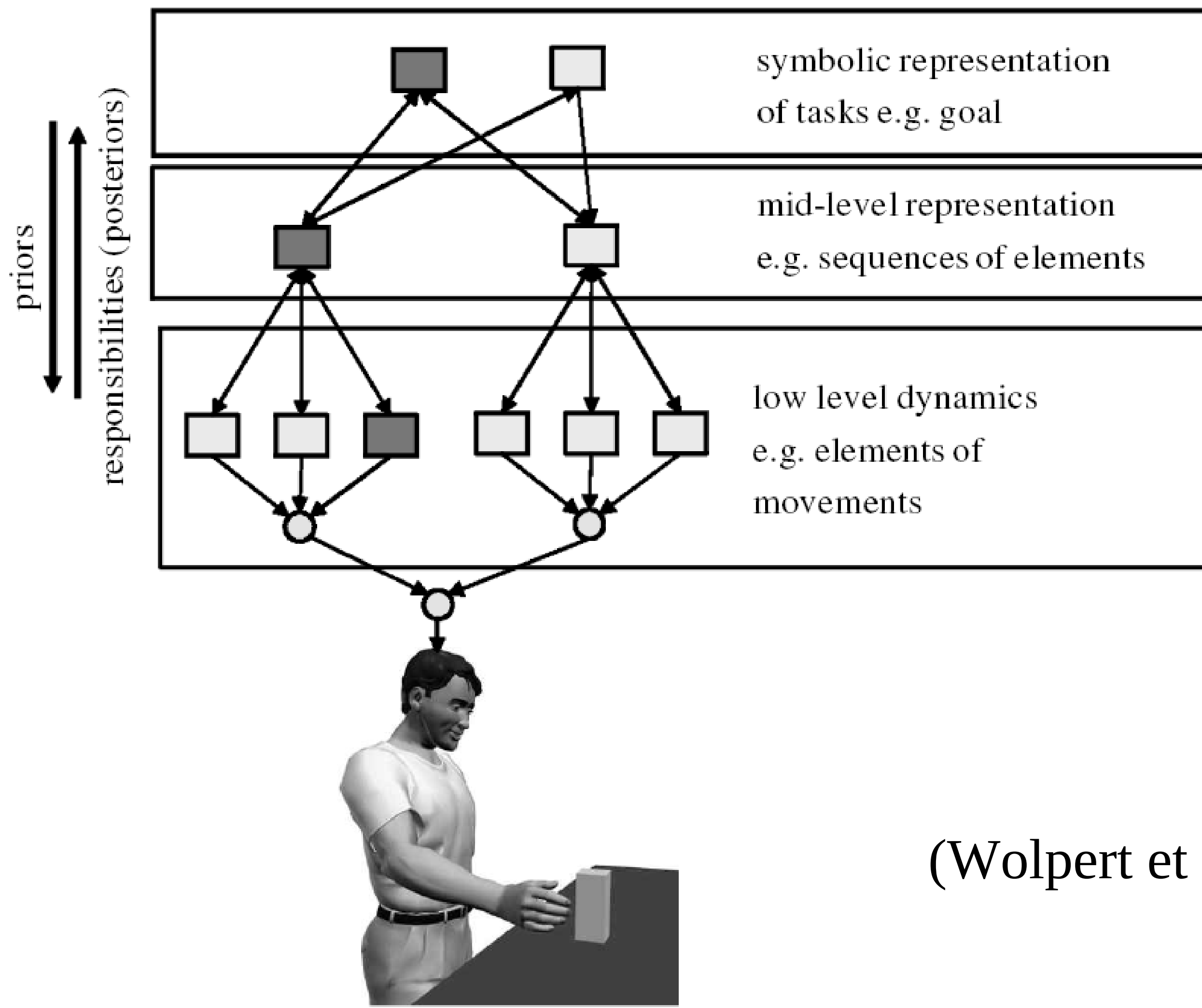
A different way to think about the development of a cognitive system.

- Powerful abstractions can be learned surprisingly quickly, together with or prior to learning the more concrete knowledge they constrain.
- Structured representations need not be rigid, static, hand-wired, brittle. Embedded in a probabilistic framework, they can grow dynamically and robustly in response to the sparse, noisy data of experience.

Open directions and challenges

- More precise relation to psychology
  - How does human cognitive processing perform approximate probabilistic inference (i.e., approximately implement rational methods of approximate inference)?
- Relation to the brain
  - How to implement structured probabilistic models in neural architectures?
- Probabilistic models for richly structured knowledge
  - How to formalize an intuitive theory of physics or psychology?
- Effective learning of structured probabilistic models
  - How to balance expressiveness/learnability tradeoff?

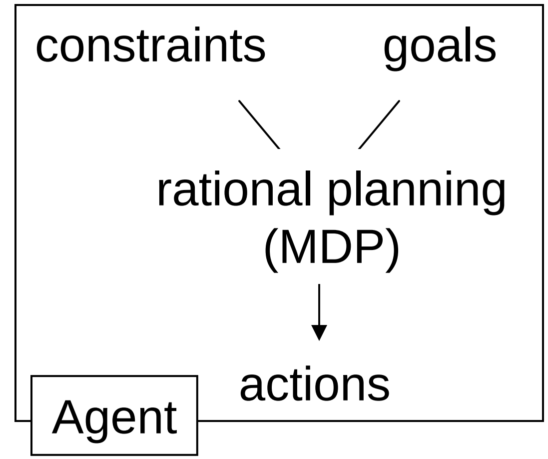
Goal-directed action (production and comprehension)



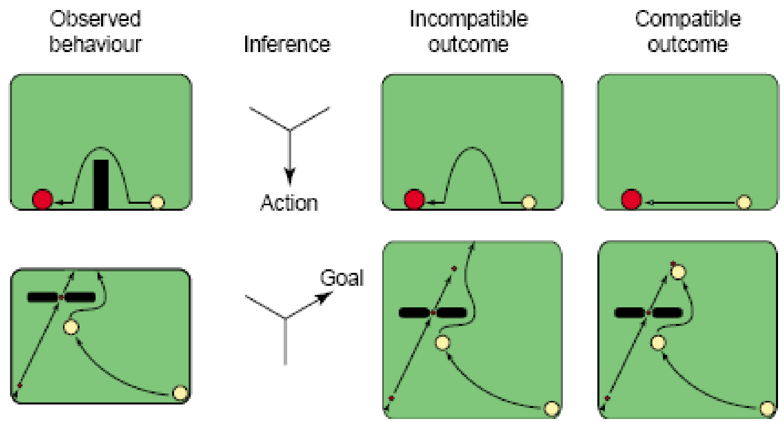
(Wolpert et al., 2003)

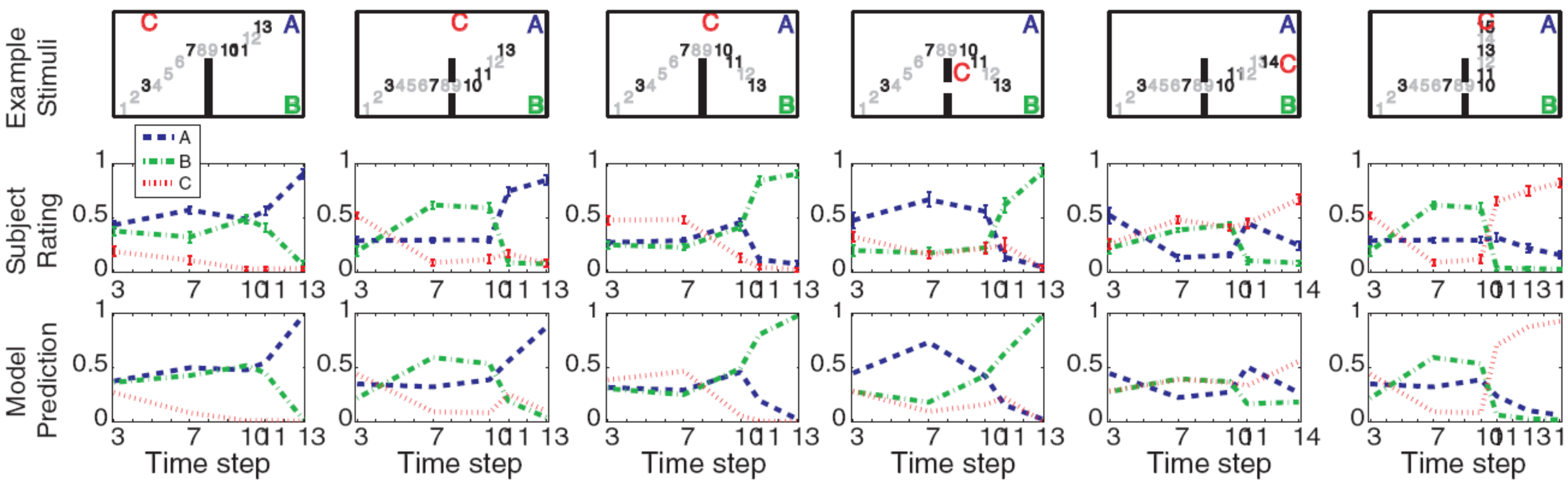
Goal inference as inverse probabilistic planning

(Baker, Tenenbaum & Saxe, *Cognition*, in press)



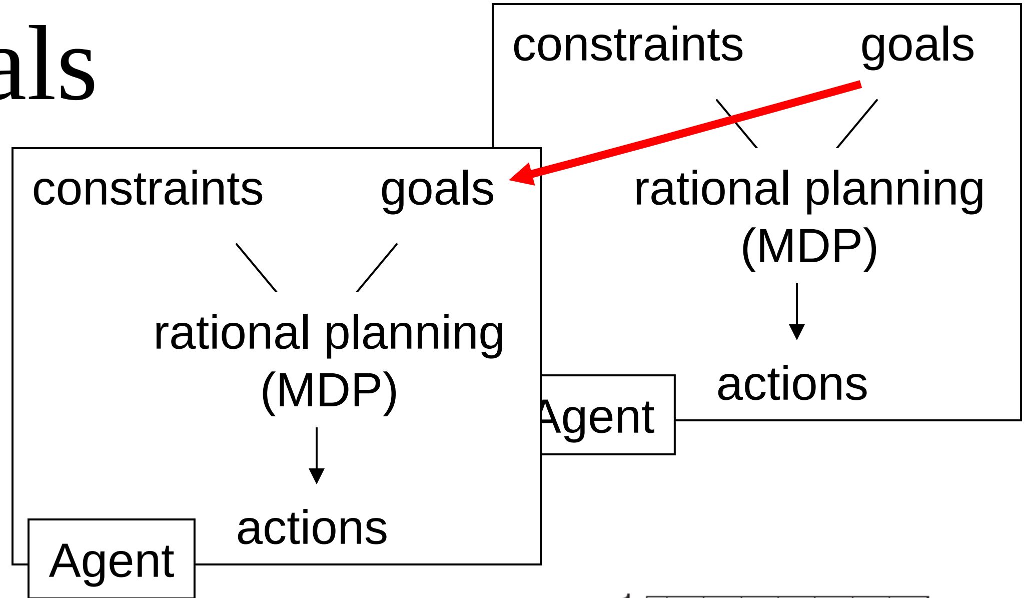
Gergely, Csibra et al.:



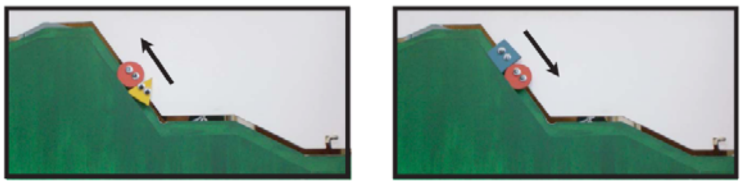


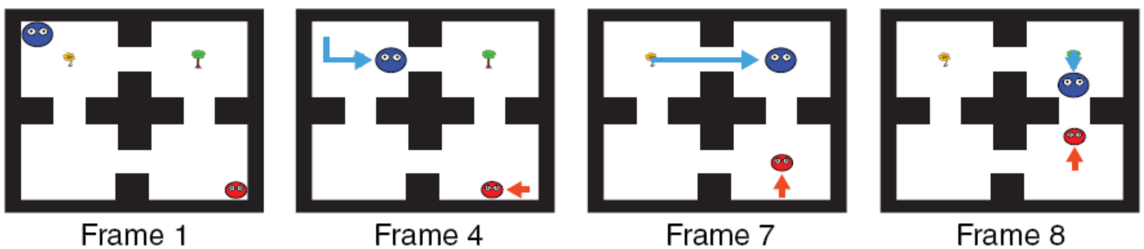
Inferring social goals

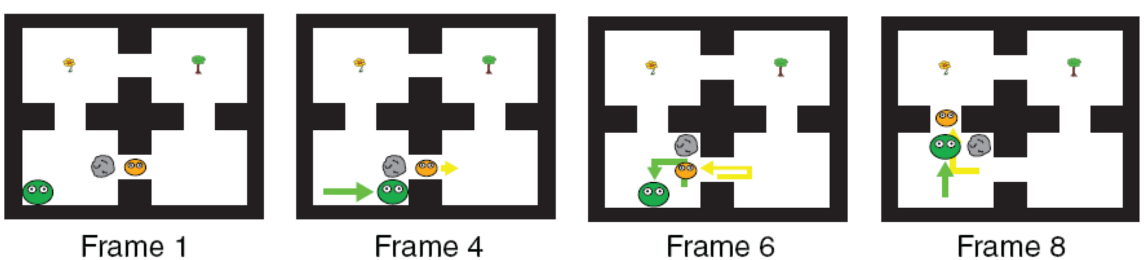
(Baker, Goodman & Tenenbaum, *Cog Sci* 2008; Ullman, Baker, Evans, Macindoe & Tenenbaum, submitted)

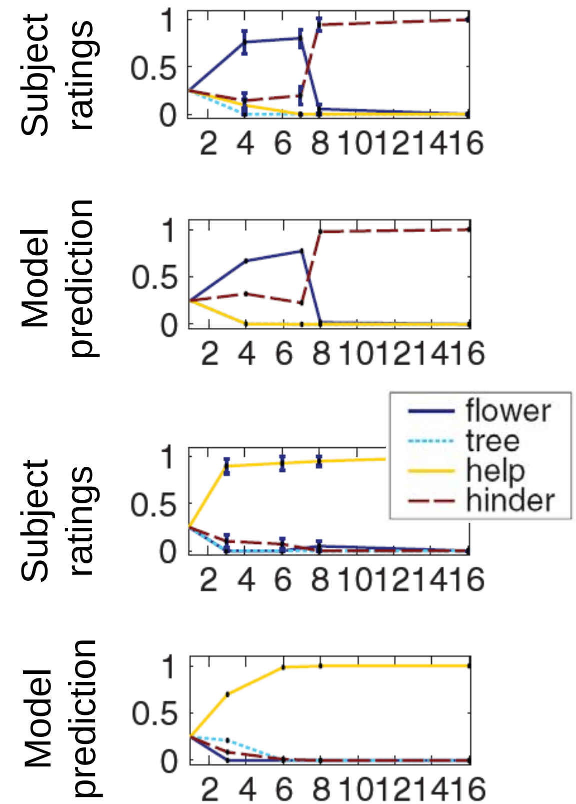


Hamlin, Kuhlmeier, Wynn & Bloom:









The *really* big questions

Where does it all end?

Across different domains and tasks, and different levels of abstraction, our probabilistic models are starting to look increasingly complex and to differ in almost arbitrarily many ways.   This seems unsatisfying…

The brain appears to have a uniform circuitry. Other cognitive modeling paradigms adopt a single unifying representational primitive (production rules, predicate logic, synaptic strengths, tensors). Is there a single universal Bayesian primitive?

How does it all begin?

Can all these different kinds of representations be learned? What is the ultimate hypothesis space of innate primitives – or is it simply “turtles all the way up”? Could a universal hypothesis space for all probabilistic models possibly be searched effectively?

(C.f. Kolmogorov complexity theory, Chater & Vitanyi)

Church: a universal probabilistic language

(Goodman et al., UAI 2008)

Church: a universal probabilistic language

(Goodman et al., UAI 2008)

```
(define cause (mem (lambda (a b)  (flip 0.5))))
(define spontaneous (mem (lambda (a t) (flip 0.01))))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a))))))
(define (parents a) (filter (lambda (y) (cause y a)) variables))

;;a noisy-or version:
(define strength (mem (lambda (a b)
                        (if (cause a b)
                            (beta 1 1) ;;or some other prior on strengths
                            0.0))))

(define (occurs a t)
  (or (spontaneous a t)
        (do a t)
        (fold (lambda (x y) (noisy-or (occurs x t) (strength x a) y 1.0))
              false
              (parents a))))
```

Causal networks

Church: a universal probabilistic language

(Goodman et al., UAI 2008)

Causal networks

```
(define cause (mem (lambda (a b)  (flip 0.5))))
(define spontaneous (mem (lambda (a t) (flip 0.01))))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a))))))
(define (parents a)

;;a noisy-or version
(define strength (me

(define (occurs a t)
  (or (spontaneous a
      (do a t)
      (fold (lambd
            false
            (paren
```

Relational schema

```
(define drawclass (DPmem 1.0 gensym))
(define class (mem (lambda (obj) (drawclass))))

(define irm-mean
  (mem (lambda (obj-class1 obj-class2)
          (normal 0.0 10.0) )))
(define irm-value
  (mem (lambda (obj1 obj2)
          (normal (irm-mean (class obj1) (class obj2))
                  1.0 ))))
```

Church: a universal probabilistic language

(Goodman et al., UAI 2008)

Causal networks

```
(define cause (mem (lambda (a b)  (flip 0.5))))
(define spontaneous (mem (lambda (a t) (flip 0.01))))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a))))))
(define (parents a)

;;a noisy-or version
(define strength (mem

(defi
  (or
```

Relational schema

```
(define drawclass (DPmem 1.0 gensym))
(define class (mem (lambda (obj) (drawclass))))
```

Physical objects

```
(define objects (repeat (poisson 1.0) gensym))

(define depth (mem (lambda (object time) (depth object (- time 1)))))

(define location (mem (lambda (object time)
                        (+ (drift) (location object (- time 1))))))
(define (drift) (uniform-draw (list 0 1 -1)))

(define extent (mem (lambda (object) (uniform-draw (list 1 2 3)))))

(define (object-seen location time)
               (argmin depth
                       (map (lambda (o) (intersects o location time)) objects)))

(define (view location time) (object-properties (object-seen location time)))
```

Church: a universal probabilistic language

(Goodman et al., UAI 2008)

Causal networks

```
(define cause (mem (lambda (a b)  (flip 0.5))))
(define spontaneous (mem (lambda (a t) (flip 0.01))))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a))))))
(define (parents a)

;;a noisy-or version
(define strength (me

(defi
  (or
```

Relational schema

```
(define drawclass (DPmem 1.0 gensym))
(define class (mem (lambda (obj) (drawclass))))
```

Physical objects

```
(define objects (repeat (poisson 1.0) gensym))

(define depth (mem (lambda (object time) (depth object (- time 1)))))

(define location (mem (lambda (ob
                        (+ (drift

(define (drift) (uniform-draw (li

(define extent (mem (lambda (obje

(define (object-seen location tim
              (argmin depth
                      (map (lamb

(define (view location time) (obj
```

Rational agents

```
(define (choose-action state)
  (lex-query
   '((action (action-prior)))
   'action
   '(flip (normalize-reward
            (sample-reward action state)))))

(define (sample-reward action state)
  (let ((next-state (state-transition state action)))
      (+ (reward next-state)
         (if (terminal? next-state)
             0
             (sample-reward
              (choose-action next-state)
              next-state)))))
```

Open directions and challenges

- More precise relation to psychology
  - How does human cognitive processing perform approximate probabilistic inference (i.e., approximately implement rational methods of approximate inference)?
- Relation to the brain
  - How to implement structured probabilistic models in neural architectures?
- Probabilistic models for richly structured knowledge
  - How to formalize an intuitive theory of physics or psychology?
- Effective learning of structured probabilistic models
  - How to balance expressiveness/learnability tradeoff?