TREE-STRUCTURED STICK BREAKING FOR HIERARCHICAL DATA

Ryan Prescott Adams
University of Toronto

Zoubin Ghahramani
University of Cambridge

Michael I. Jordan
UC Berkeley

http://www.cs.toronto.edu/~rpa

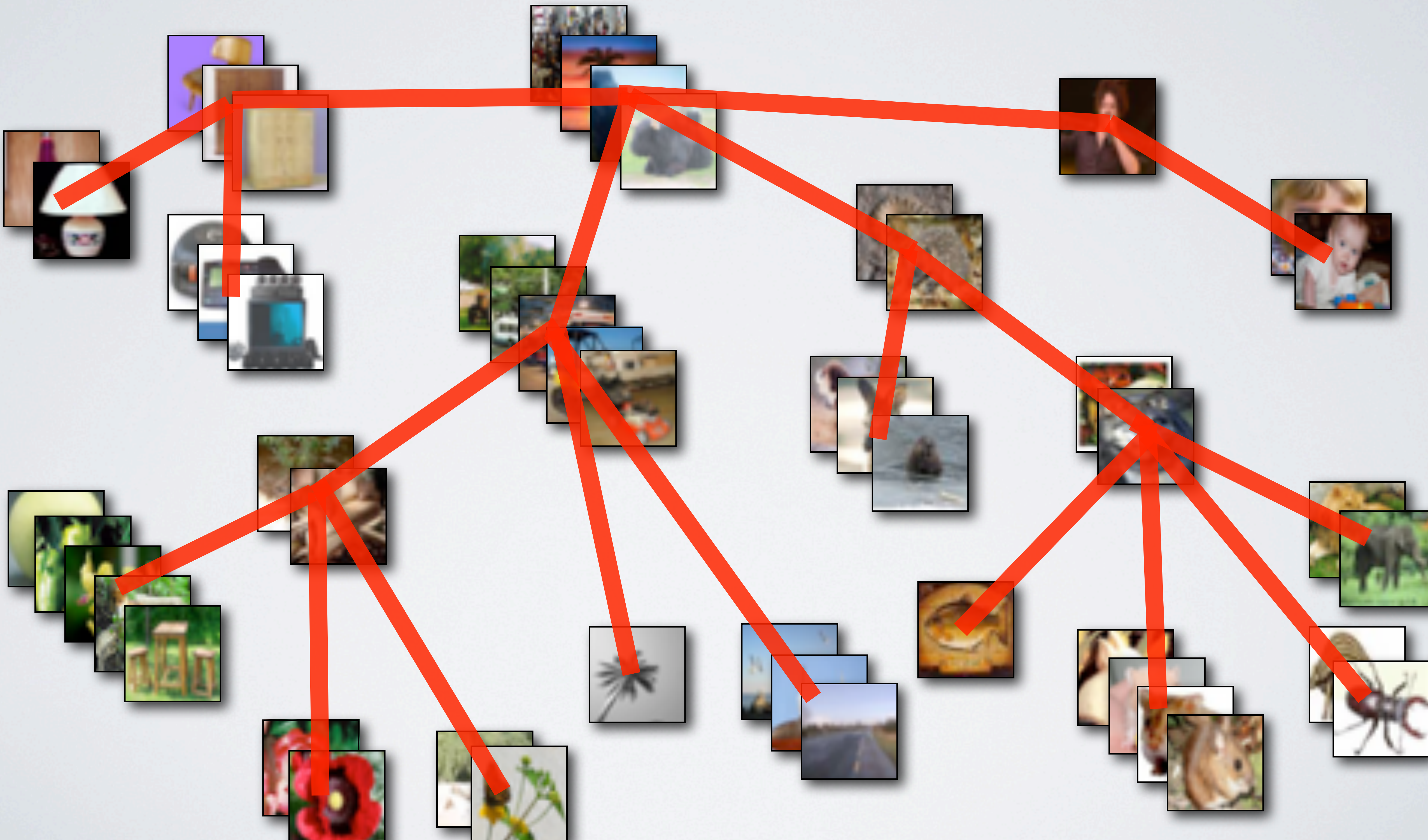
CIFAR

THE BIG PICTURE

Discover a latent tree structure over data.

THE BIG PICTURE

Discover a latent tree structure over data.



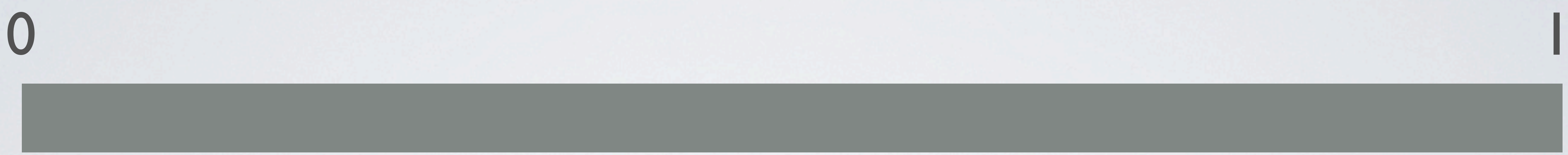
HIERARCHICAL CLUSTERING

The main idea of our approach:
Construct a mixture model in which the components have a tree-structured topology.

- Bayesian nonparametric approach
- Unbounded width and depth.
- Data live at internal nodes of the tree
- Infinitely exchangeable urn scheme
- MCMC inference
- Applied to image and text

DIRICHLET PROCESS MIXTURES

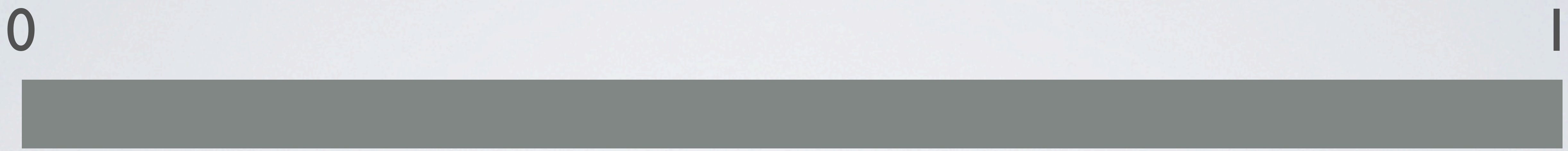
$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$



Standard stick breaking: natural numbers as index set

DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \boxed{\pi_k} f_X(x \mid \theta_k)$$



Standard stick breaking: natural numbers as index set

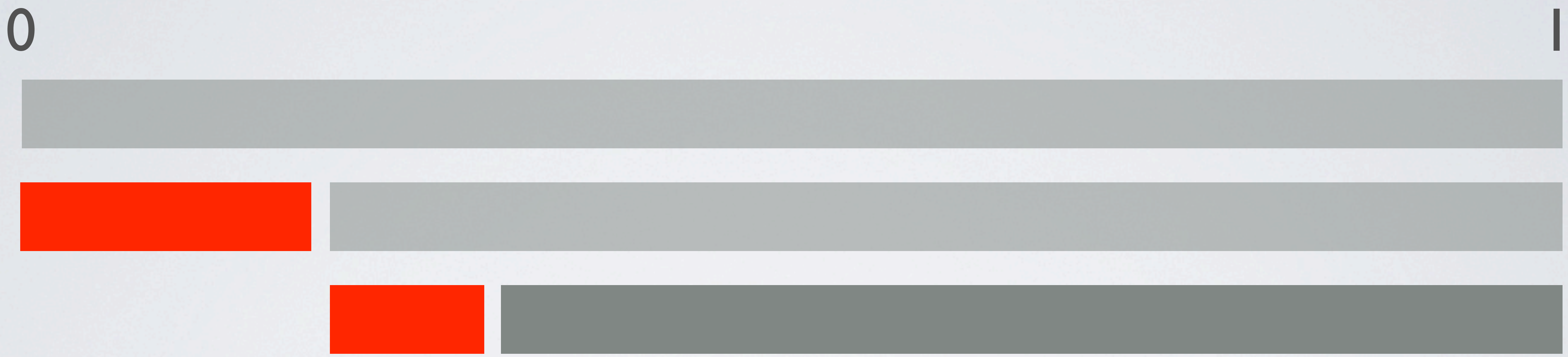
DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Standard stick breaking: natural numbers as index set

DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$



Standard stick breaking: natural numbers as index set

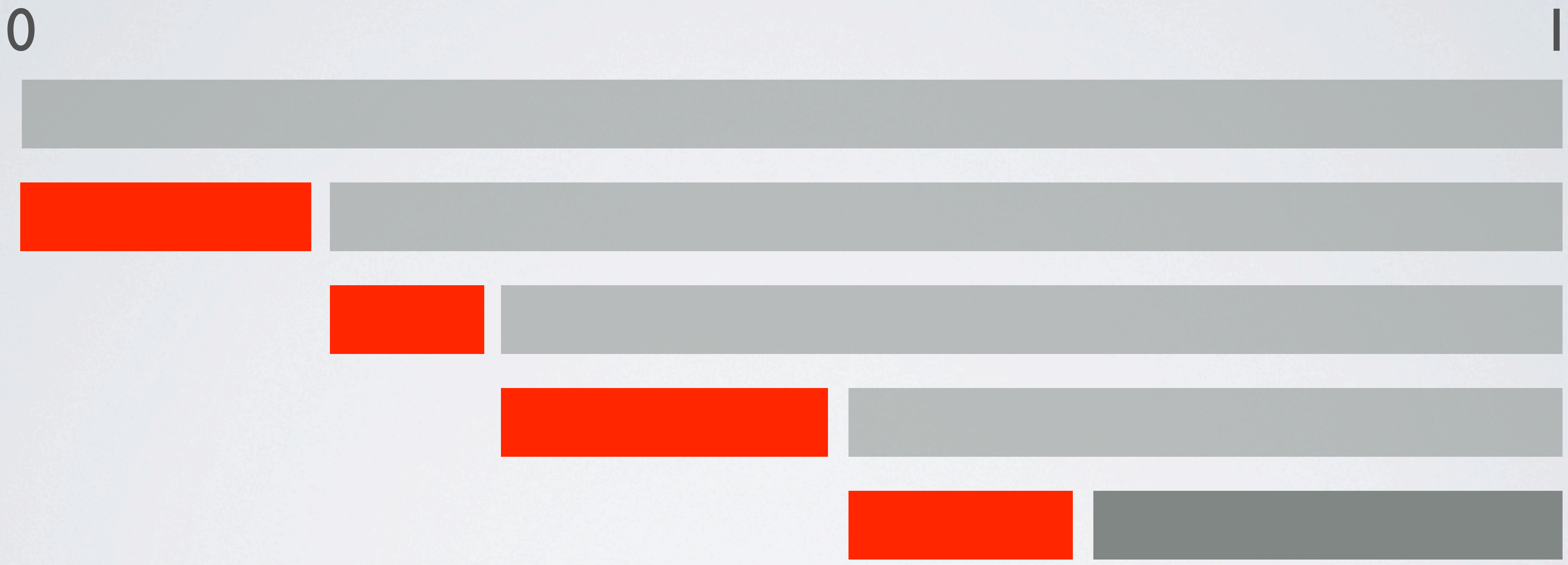
DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Standard stick breaking: natural numbers as index set

DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$



Standard stick breaking: natural numbers as index set

DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Standard stick breaking: natural numbers as index set

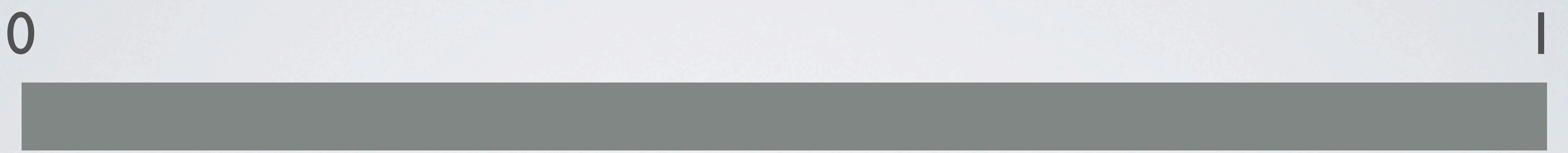
DIRICHLET PROCESS MIXTURES

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Standard stick breaking: natural numbers as index set

TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \boxed{\pi_k} f_X(x \mid \theta_k)$$



Tree-structured stick breaking: finite-length strings as index set

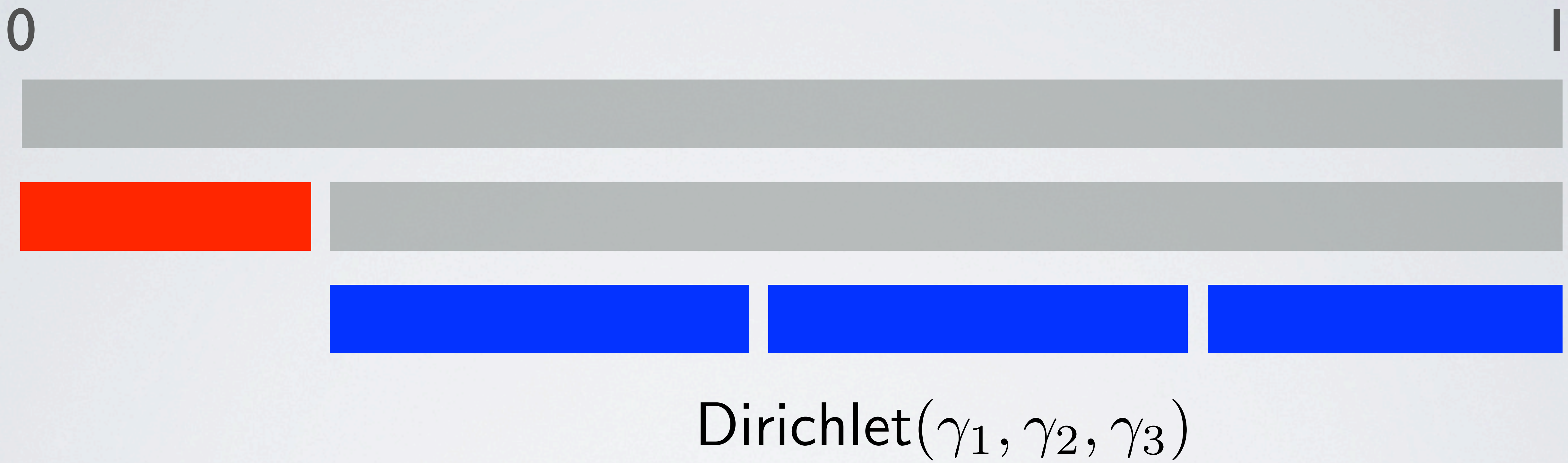
TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Tree-structured stick breaking: finite-length strings as index set

TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$



Tree-structured stick breaking: finite-length strings as index set

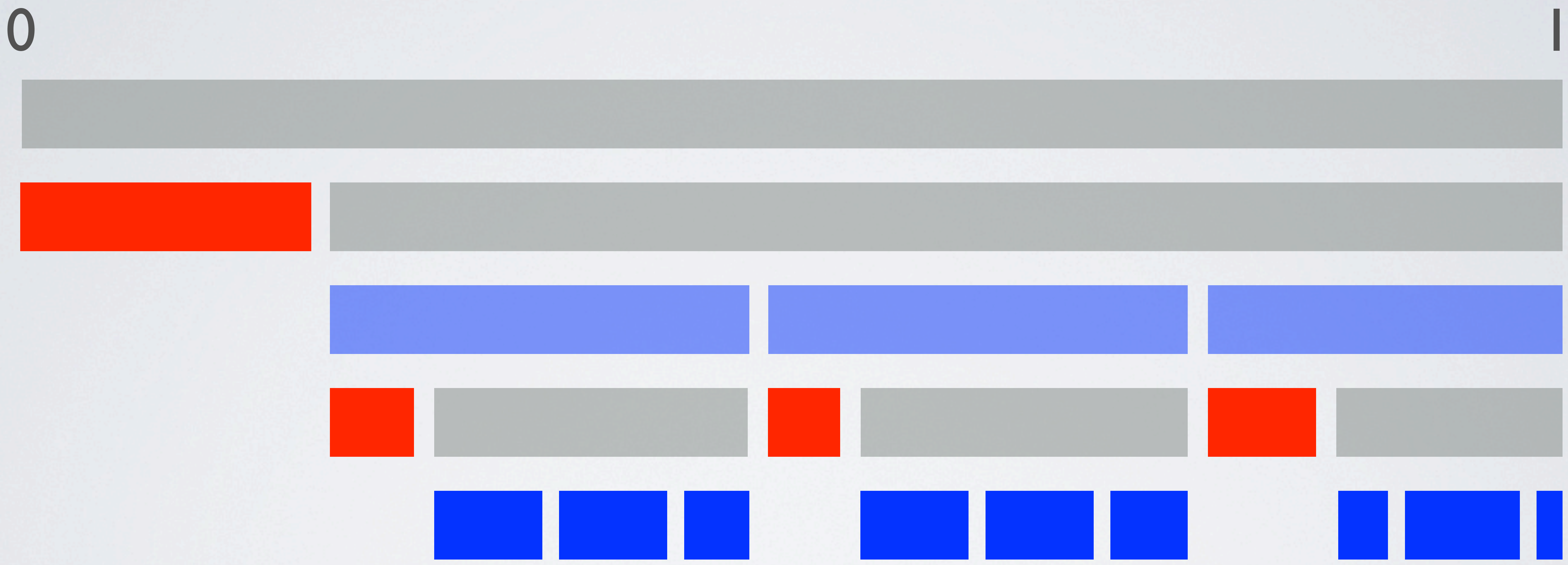
TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Tree-structured stick breaking: finite-length strings as index set

TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$



Tree-structured stick breaking: finite-length strings as index set

TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

Tree-structured stick breaking: finite-length strings as index set

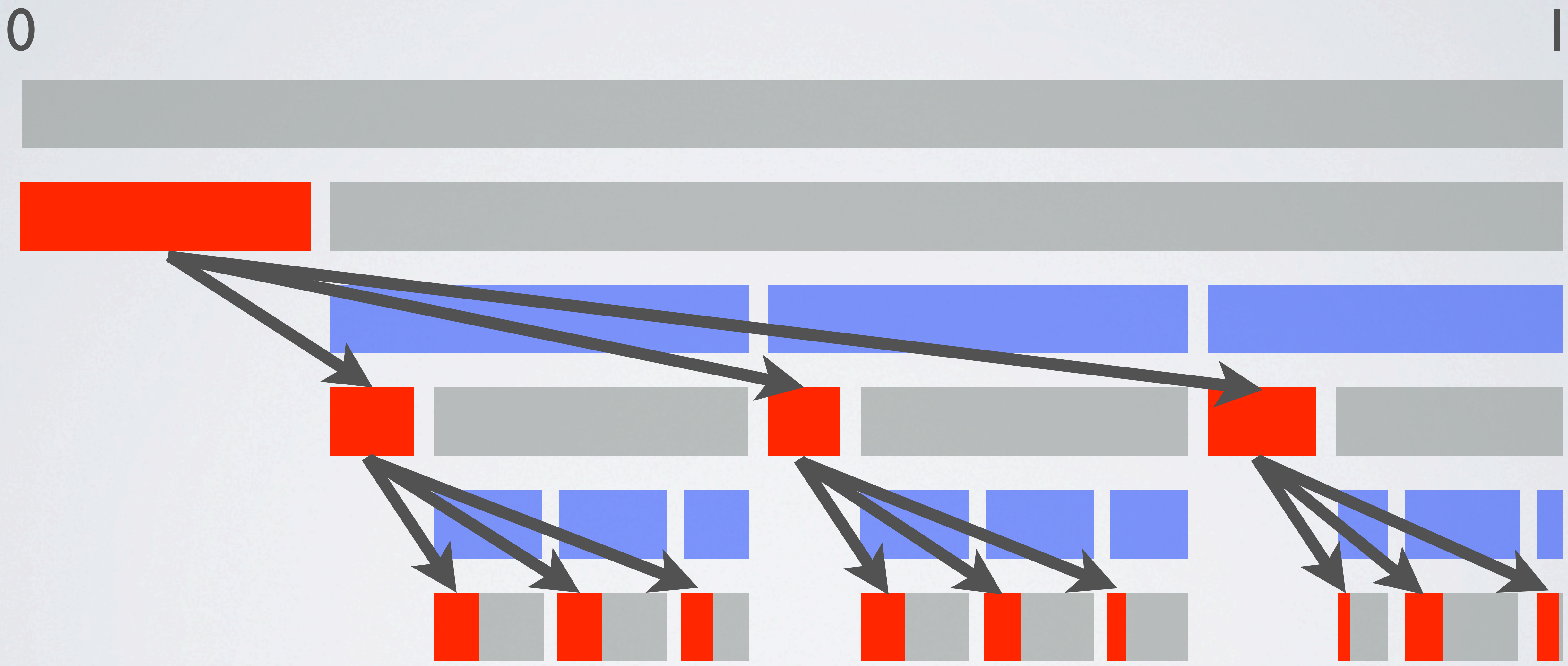
TREE STRUCTURED PARTITIONS

$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

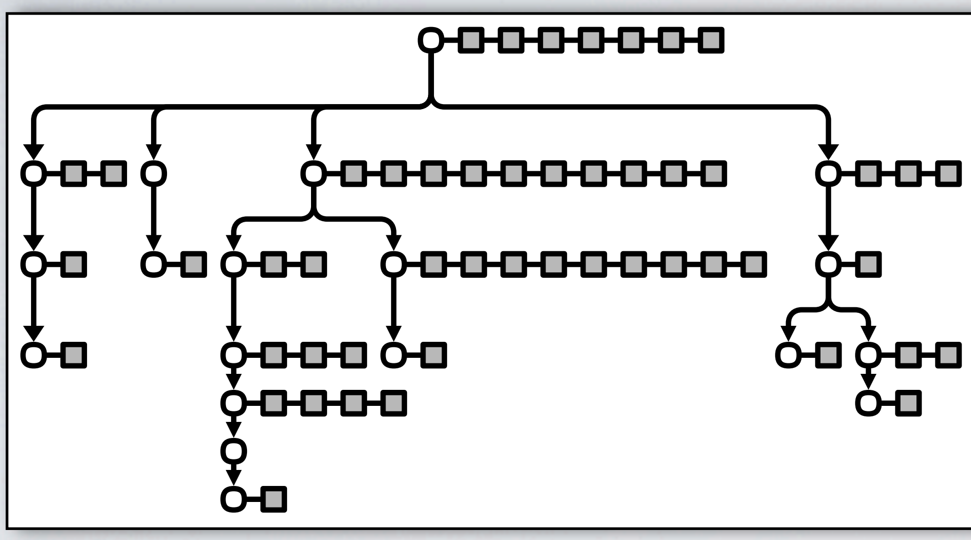
Tree-structured stick breaking: finite-length strings as index set

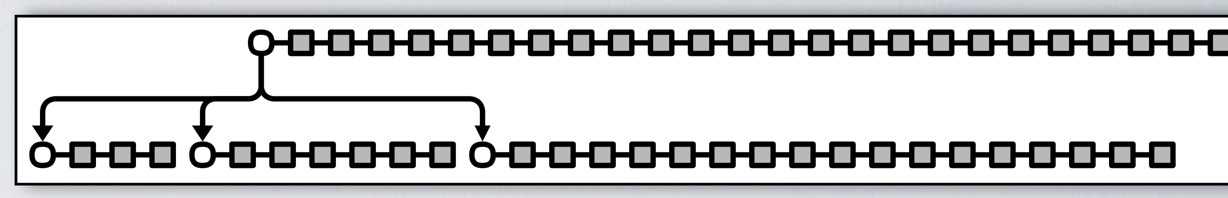
TREE STRUCTURED PARTITIONS

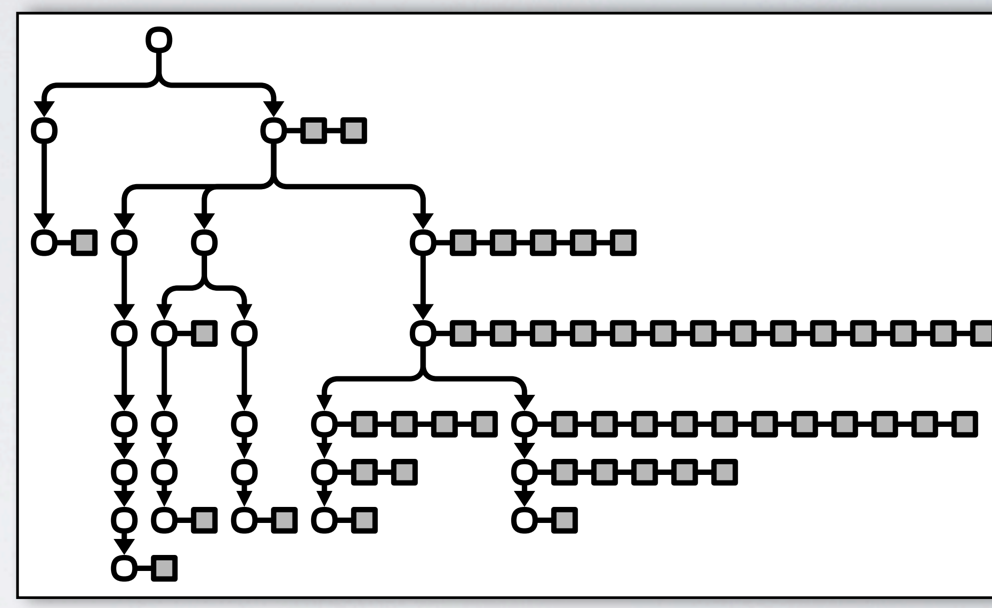
$$p(x \mid \{\pi_k, \theta_k\}_{k=1}^{\infty}) = \sum_{k=1}^{\infty} \pi_k f_X(x \mid \theta_k)$$

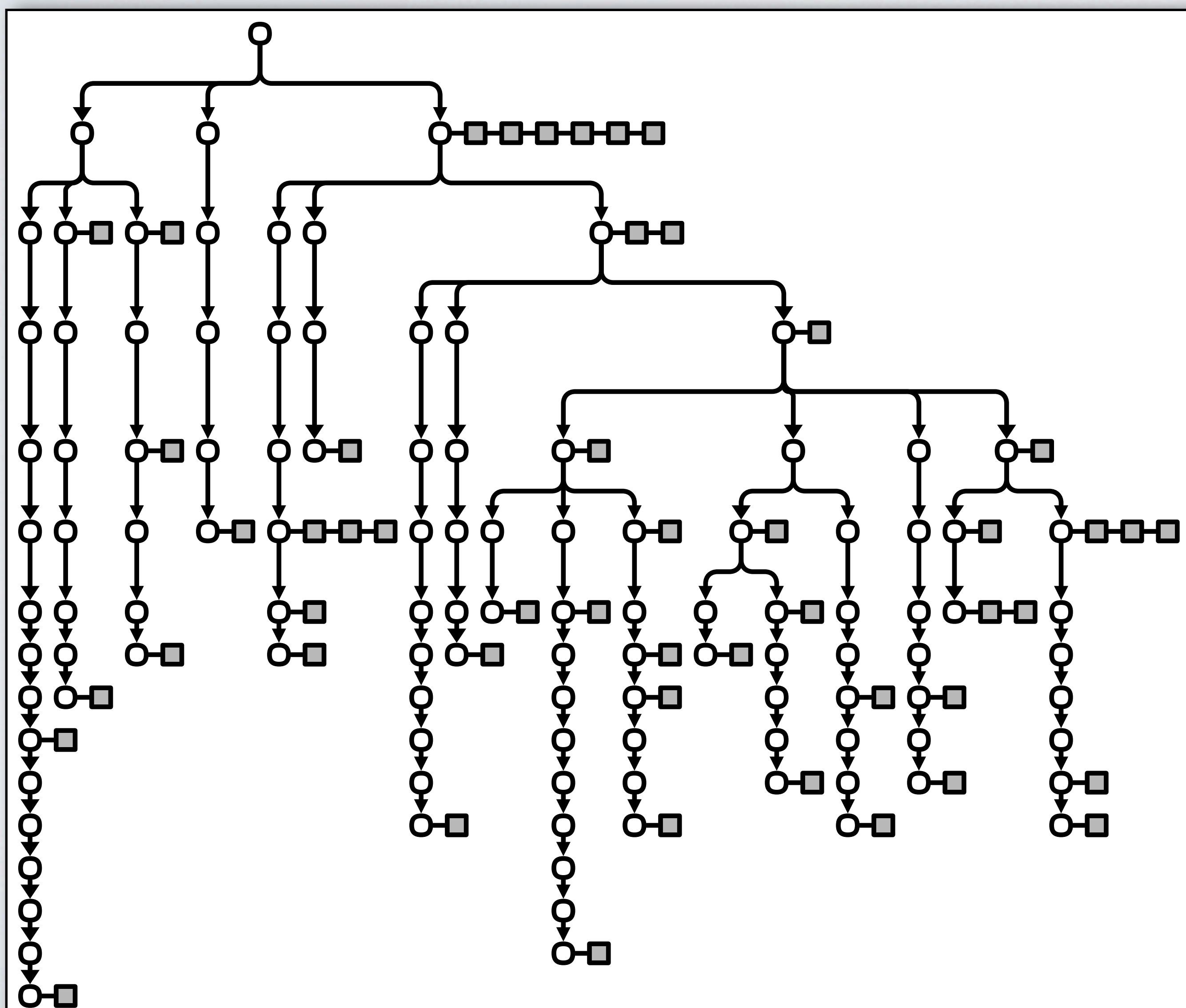


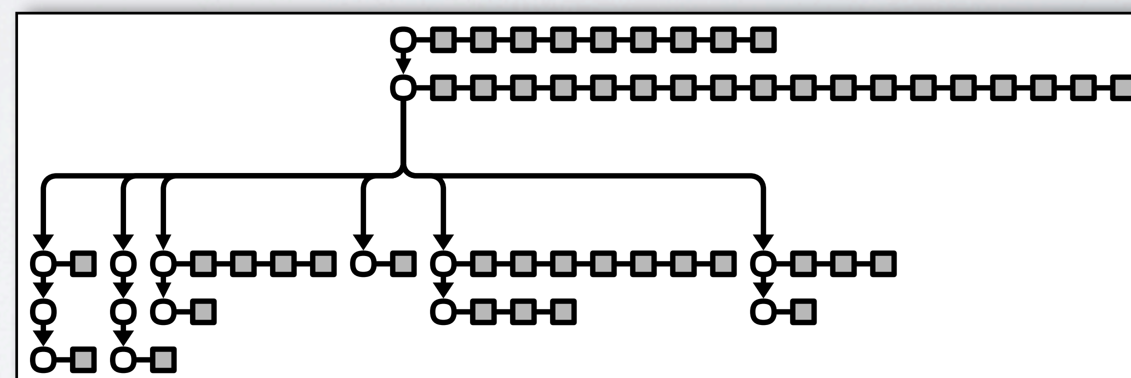
Tree-structured stick breaking: finite-length strings as index set

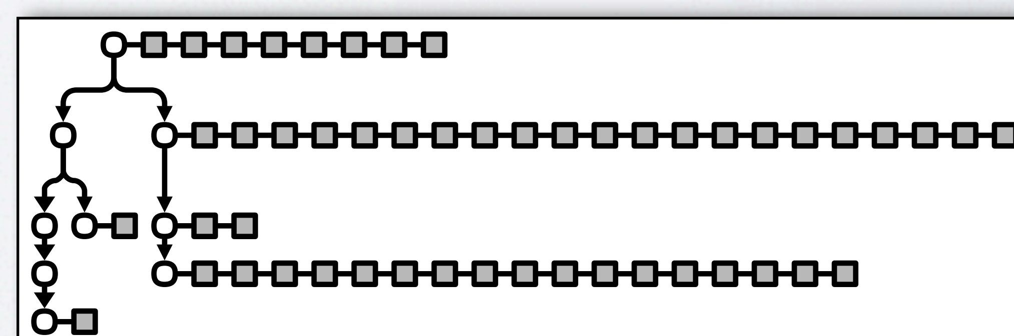










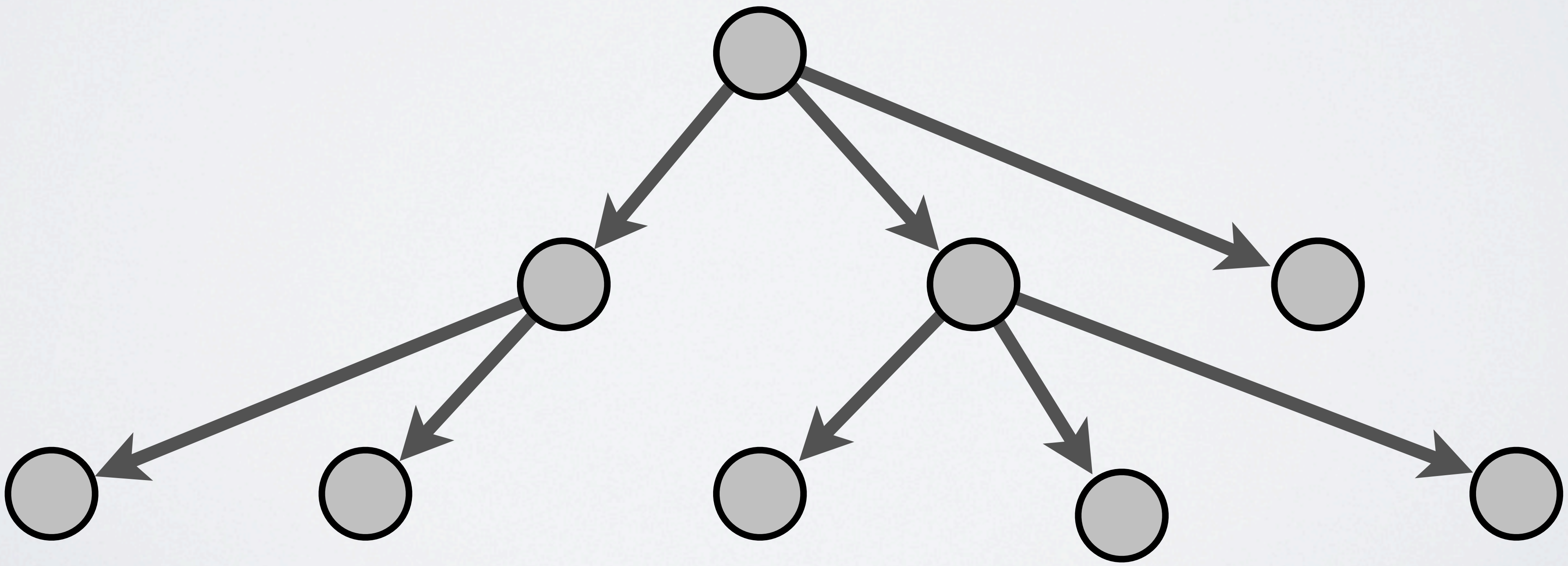


PARAMETER DEPENDENCE

A fancy topology over the partitions is not useful unless the parameters are linked.

Tree topology = directed graphical model

Parents then provide the prior for the children.



INFERENCE VIA MCMC

Simulate from the posterior over structure and parameters

Unknown quantities:

1. Assignments of data to nodes
2. Stick lengths
3. Node parameters
4. Stick-breaking hyperparameters

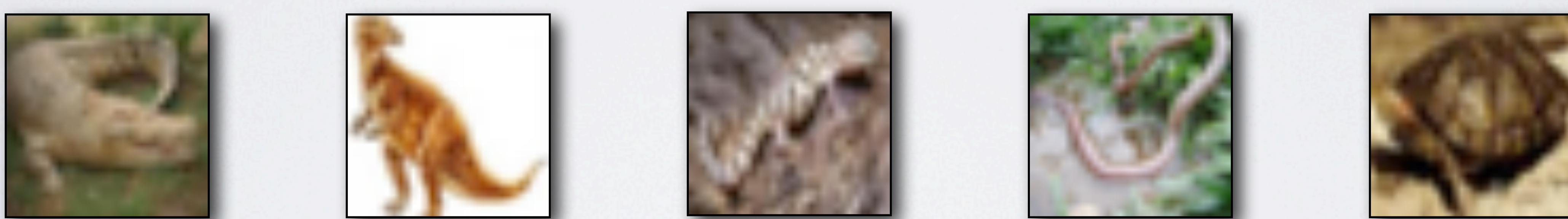
All moves are relatively standard, except #1.

To choose a "best" tree, we keep the assignments which maximized the complete data log likelihood.

CIFAR-100 IMAGE DATA

- 50,000 32x32 color images of 100 classes[1]
- Subset of the "80 Million Tiny Images" data set of Torralba, Fergus and Freeman [2,3]
- Sorted and labeled by Krizhevsky, Nair and Hinton
- The 100 classes aggregate into 20 "super classes"

reptiles: crocodile, dinosaur, lizard, snake, turtle



furniture: bed, chair, couch, table, wardrobe

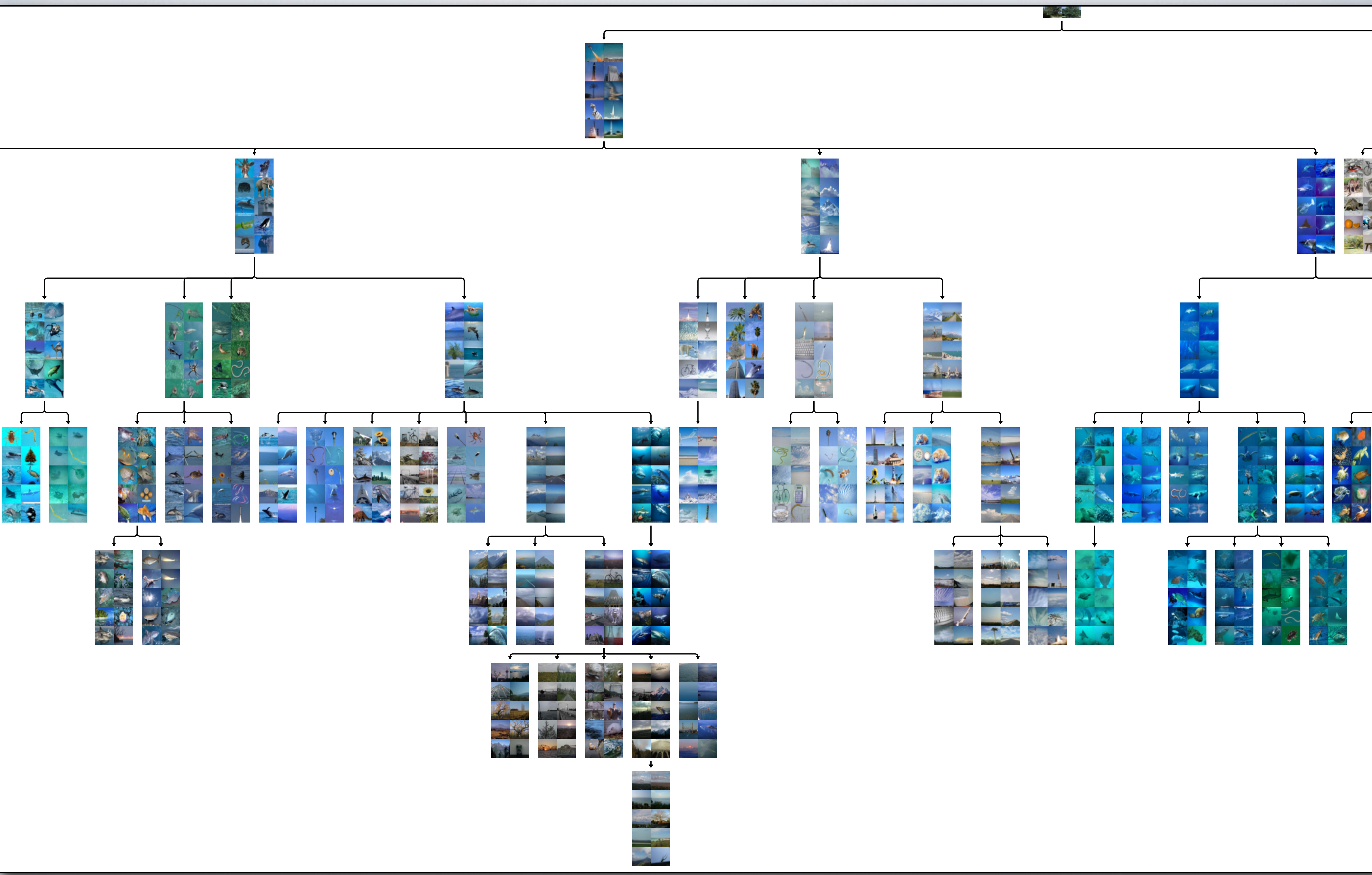
1. http://www.cs.toronto.edu/~kriz/cifar.html
2. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008
3. http://groups.csail.mit.edu/vision/TinyImages/

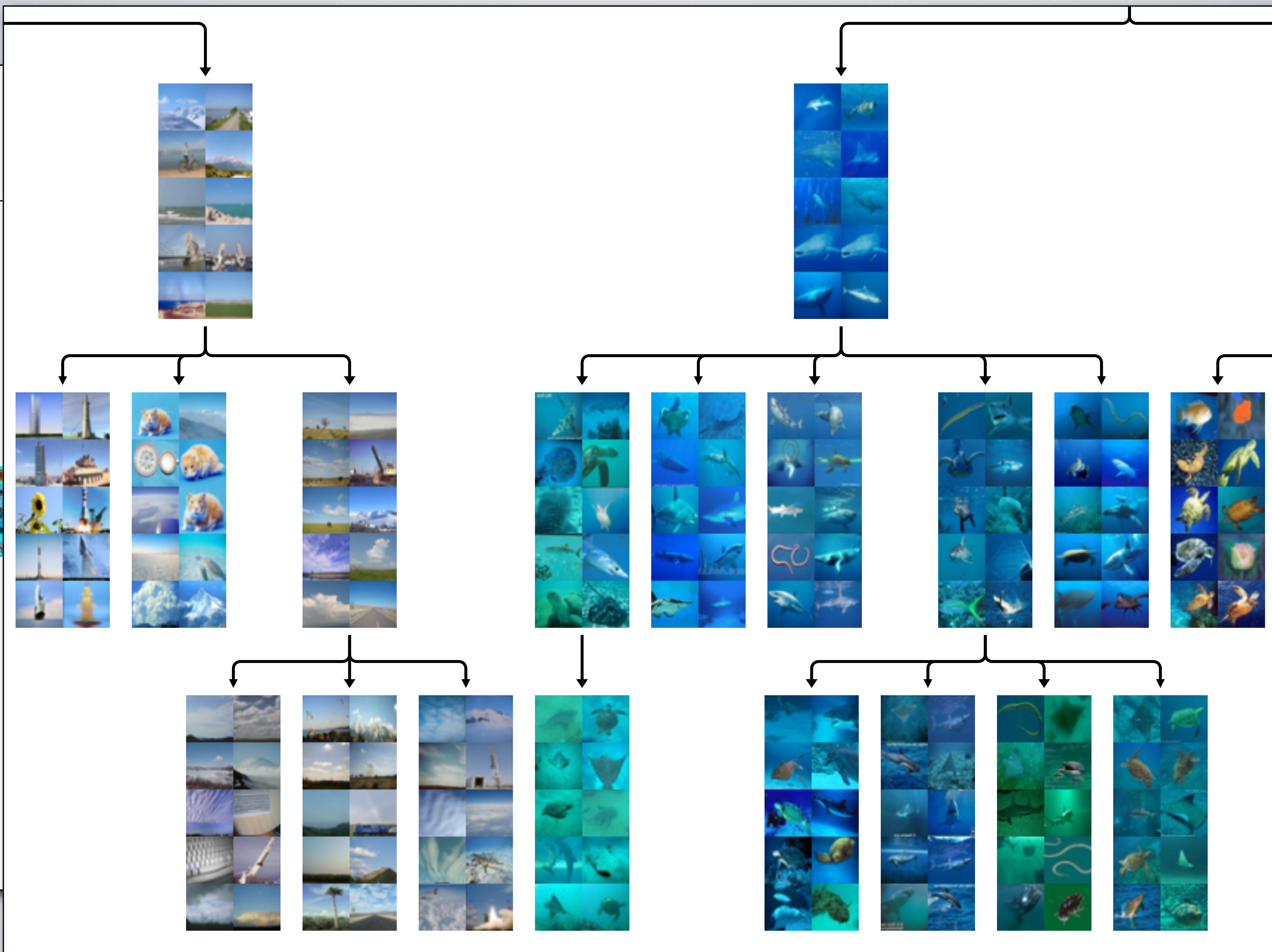
CLUSTERING IMAGES

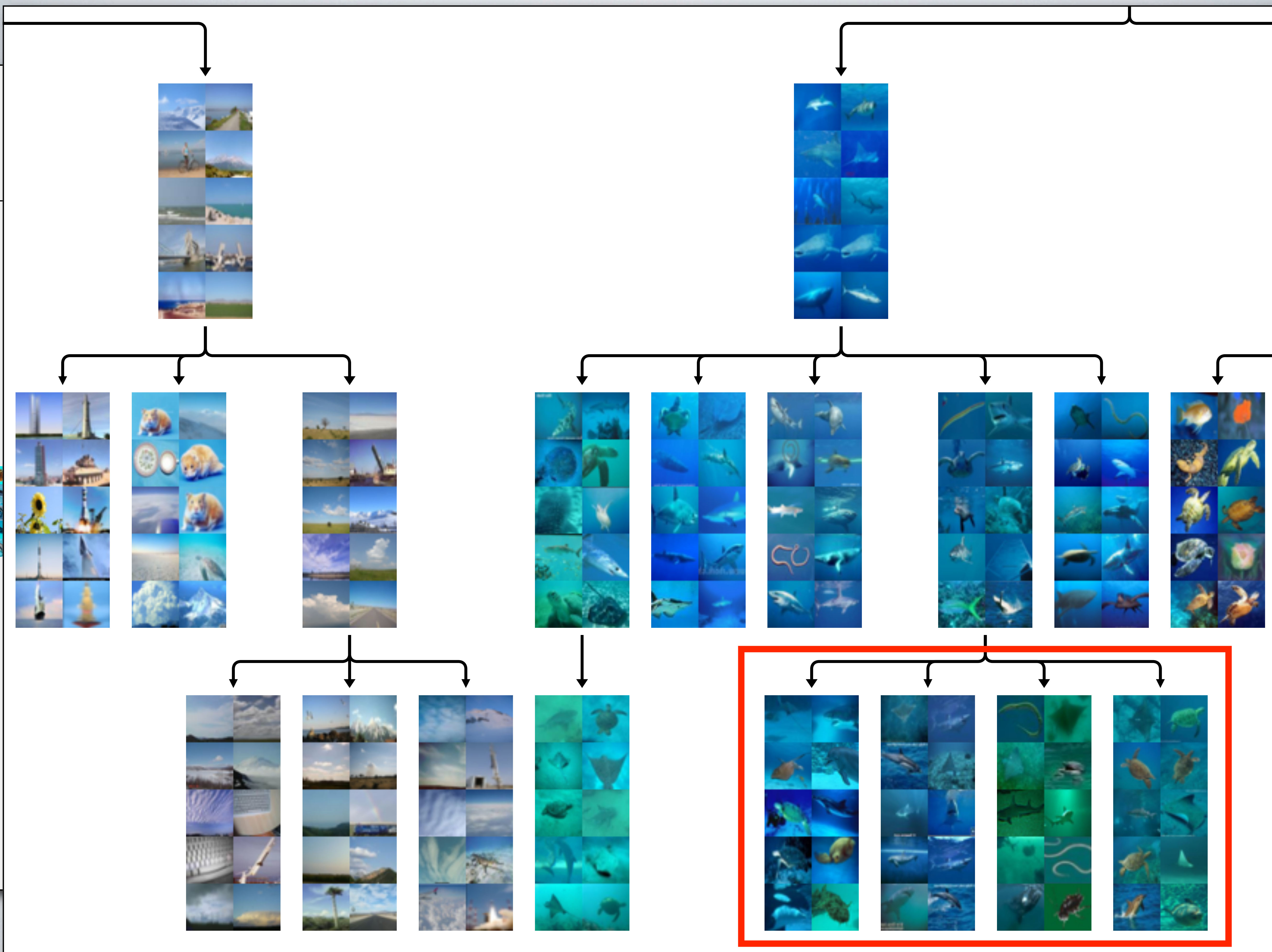
- Letting a mixture model see high-dimensional real-valued data (e.g., 1024 RGB pixels) can be a disaster.
- Used 256-dimensional binary feature vectors.
- Features were from a deep autoencoder trained for separate work on image retrieval by Alex Krizhevsky
- Alex's codes are great!
- Check out his tech report[1]

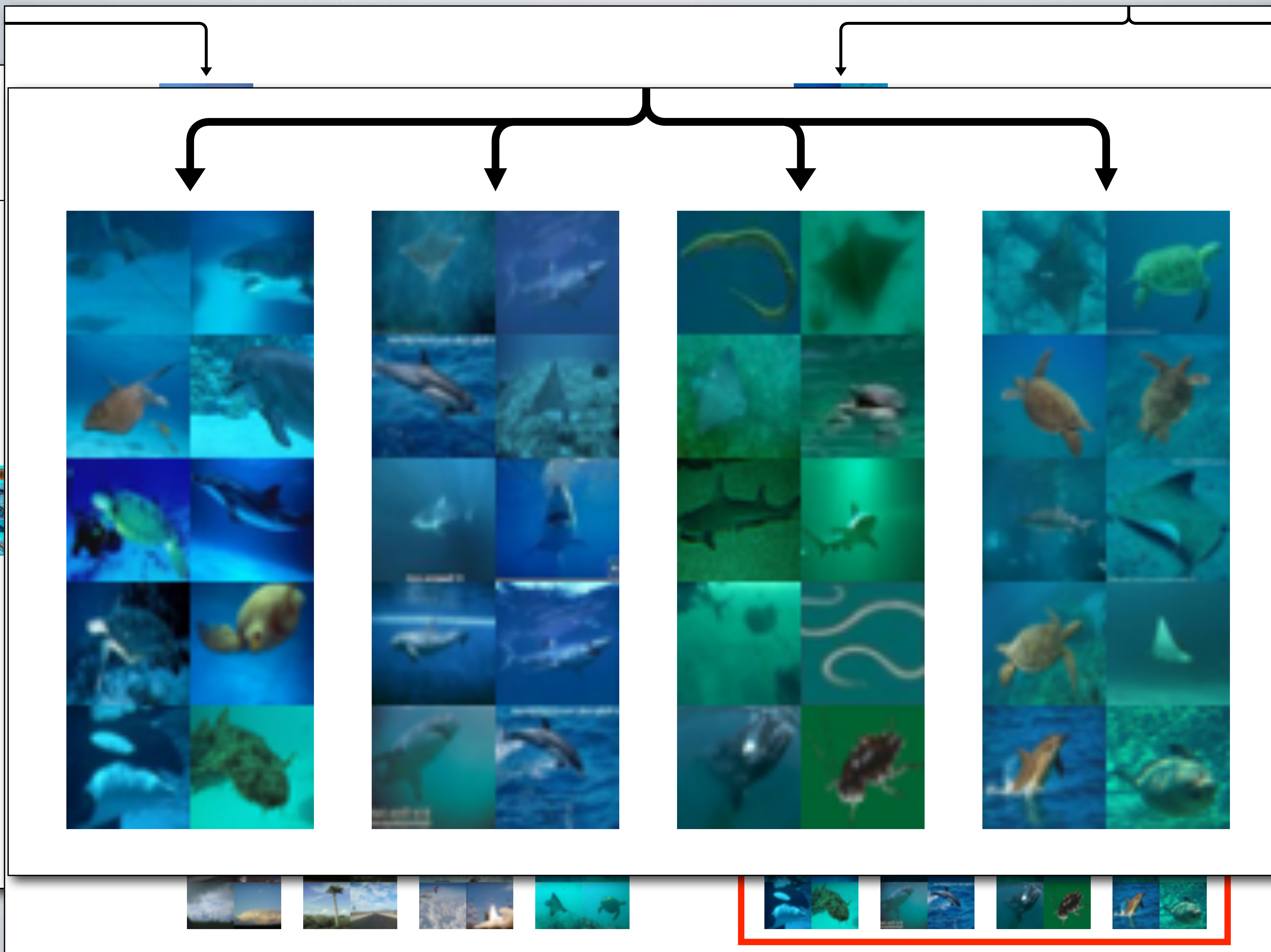
- Each node owns a product of Bernoulli distributions, parameterized by logistic-transformed real values.
- A child's parameters have a Gaussian prior with a mean determined by the parent's parameters.
- The prior variances are shared and also inferred.

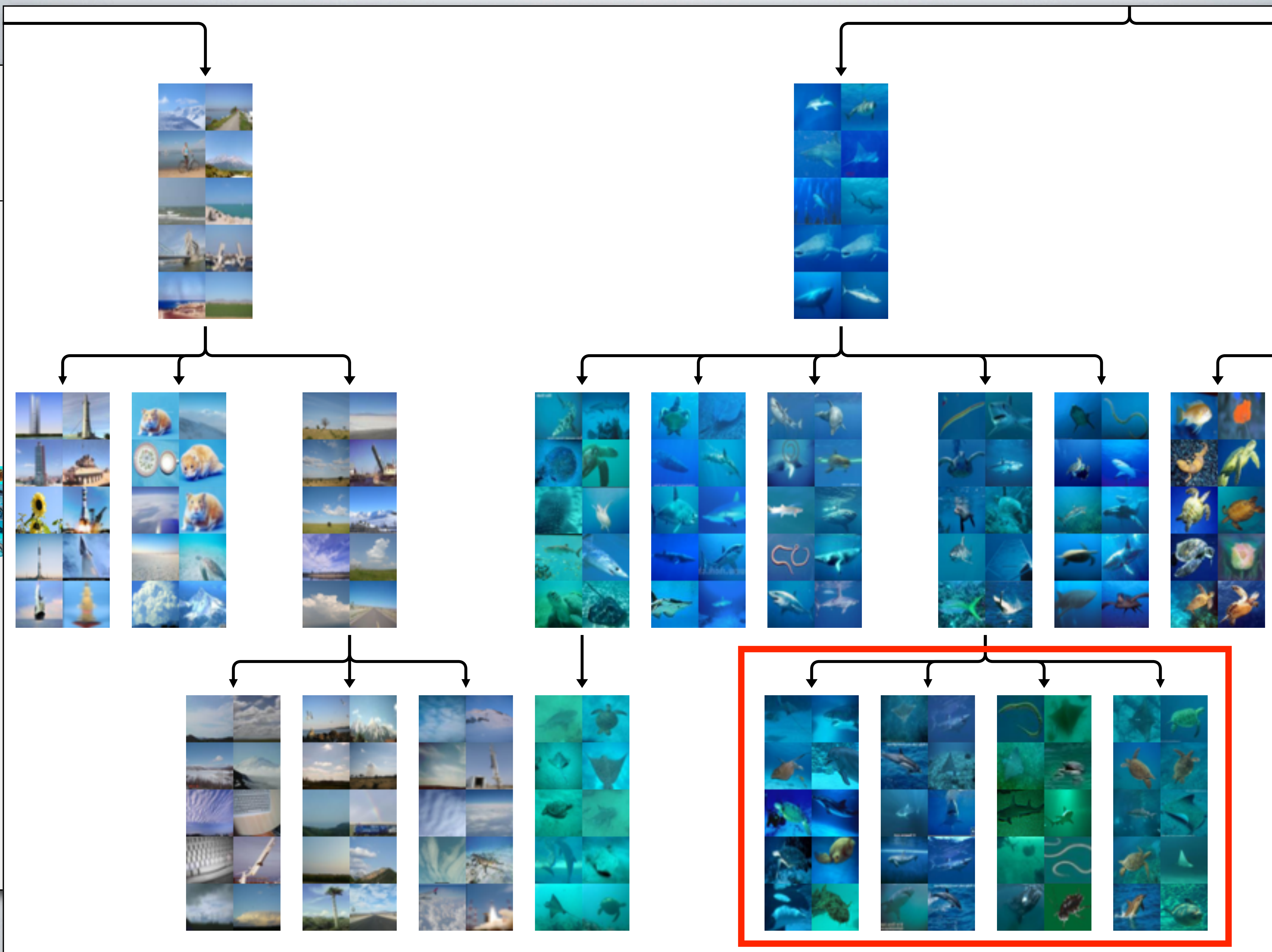
1. http://www.cs.toronto.edu/~kriz

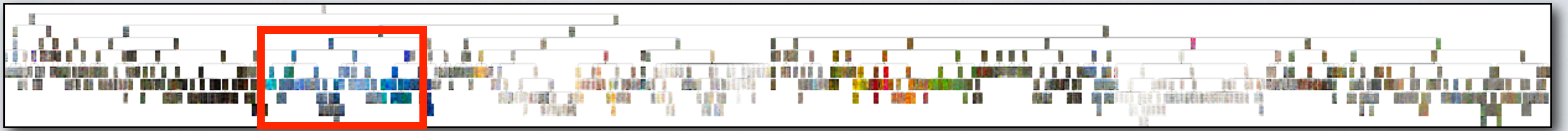


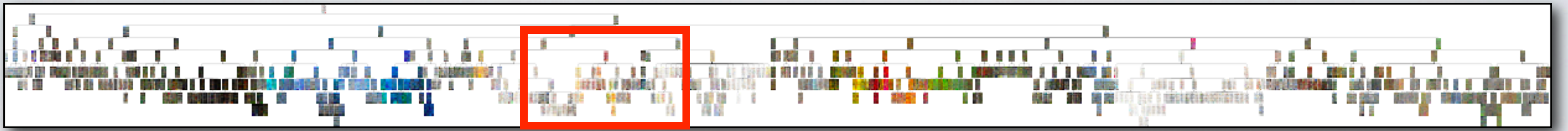


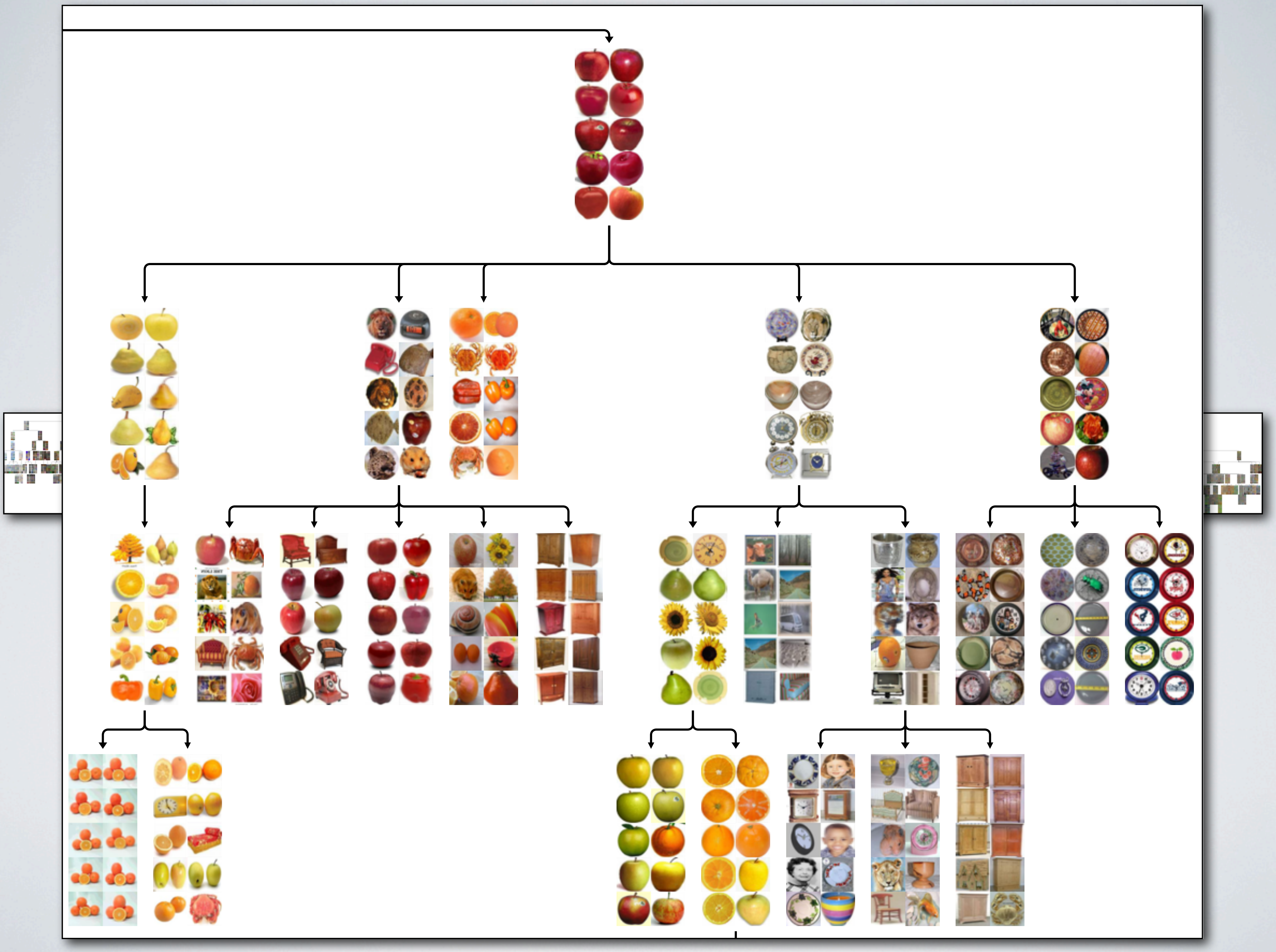


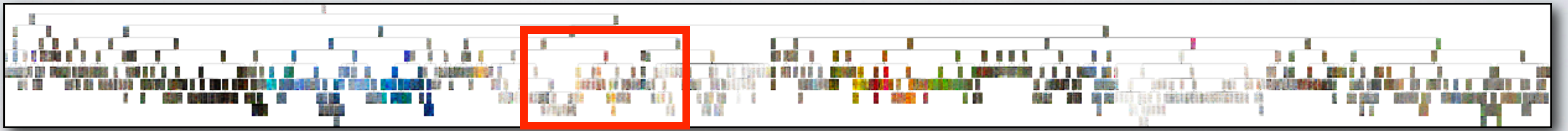


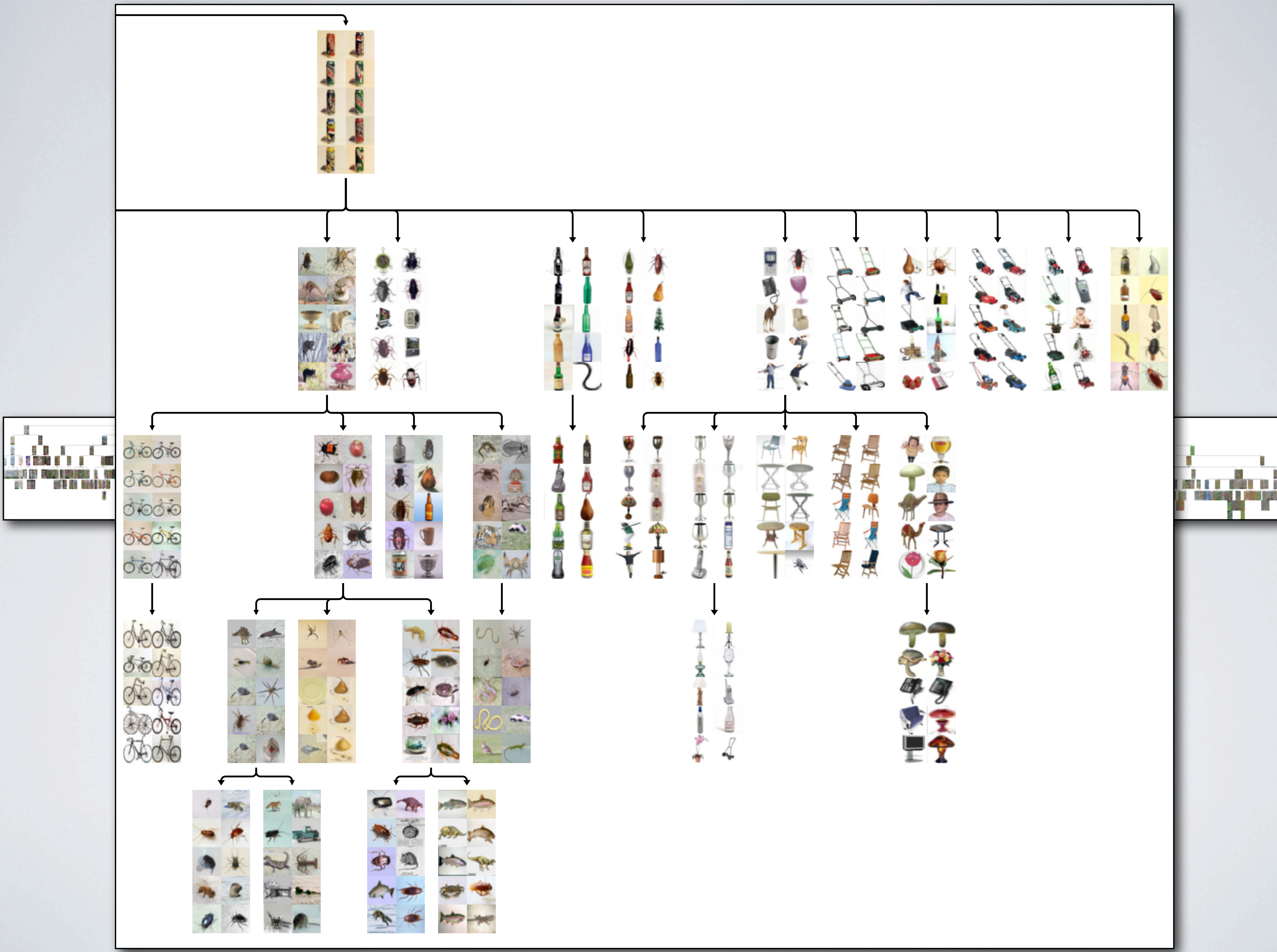


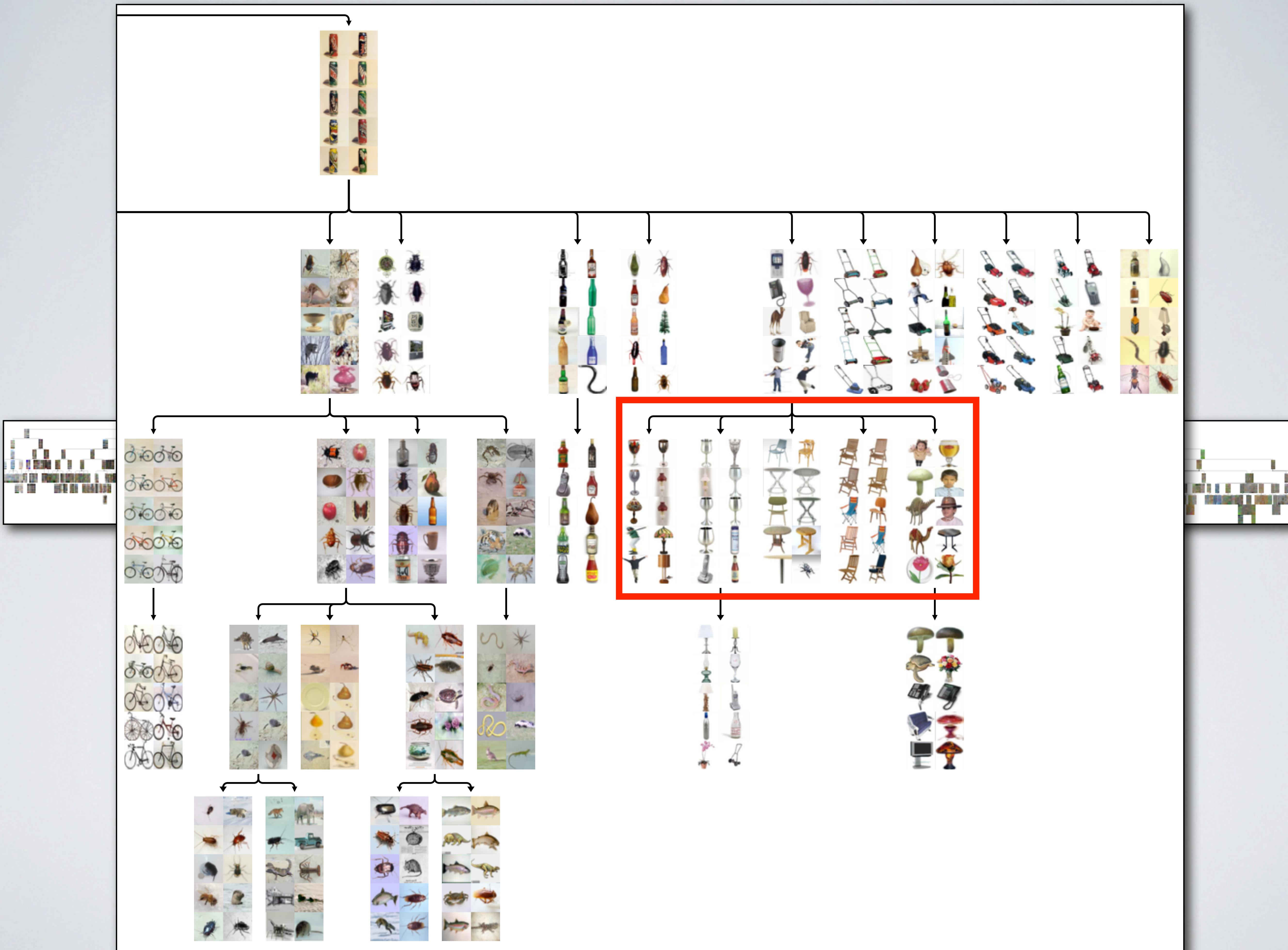


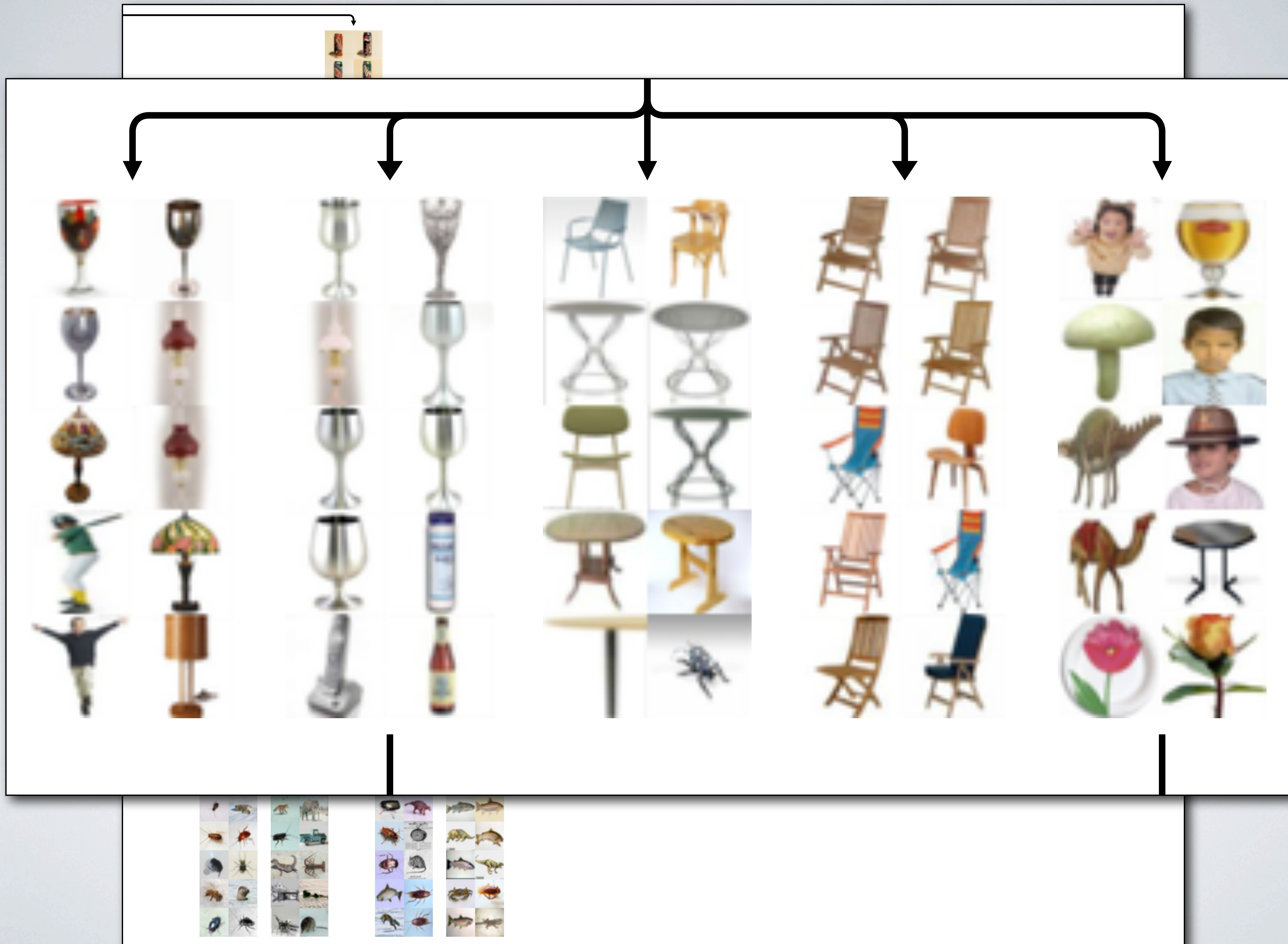


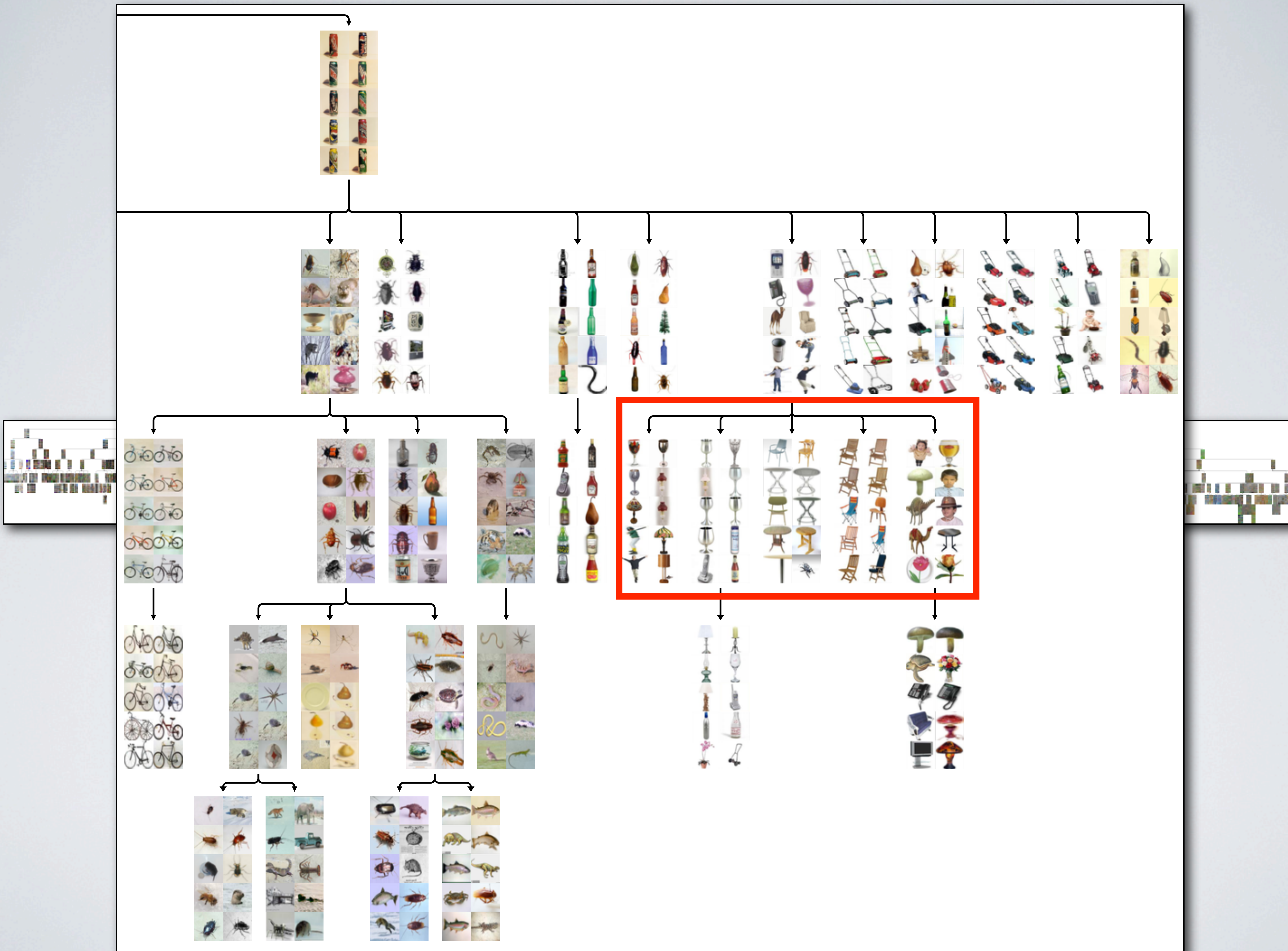


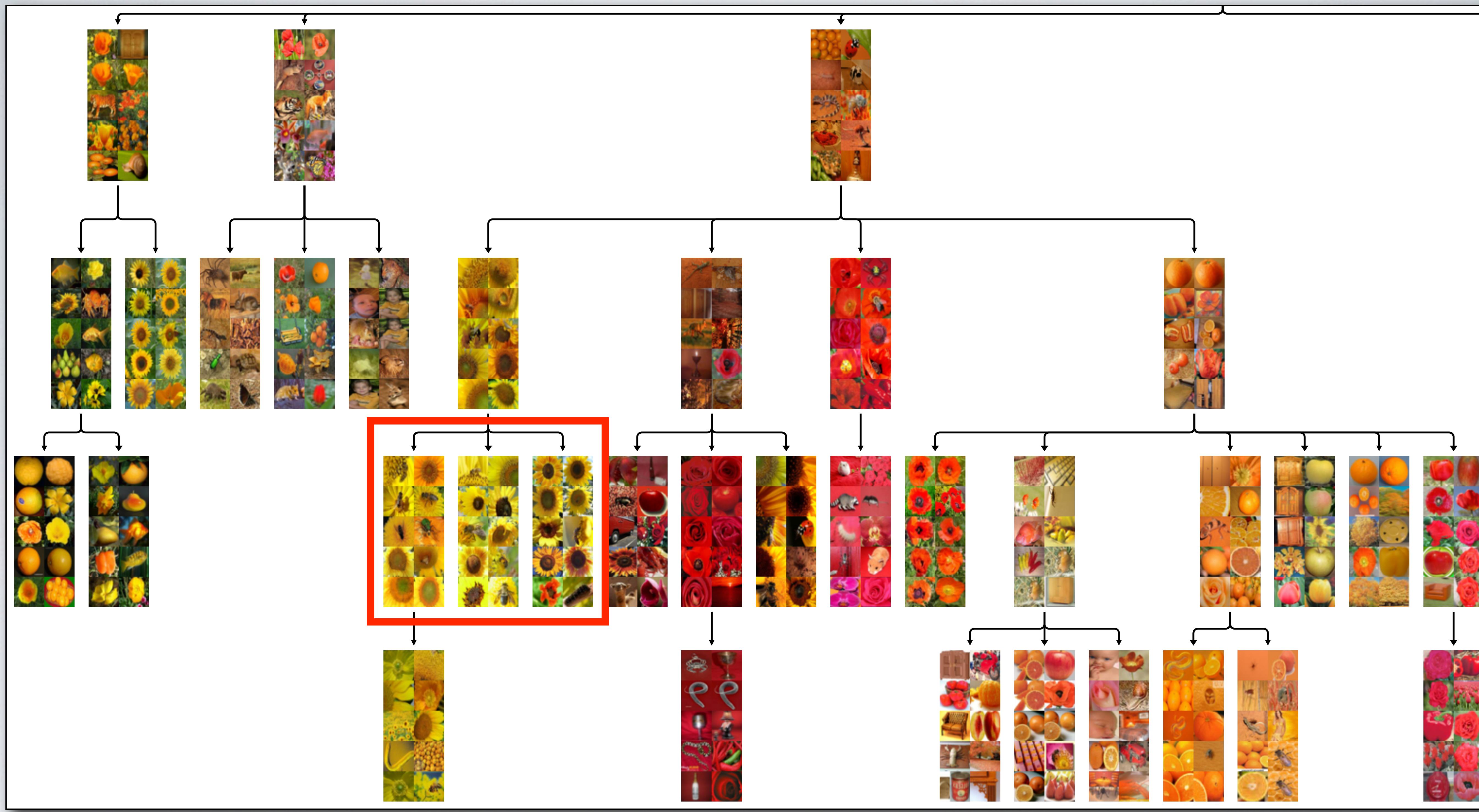


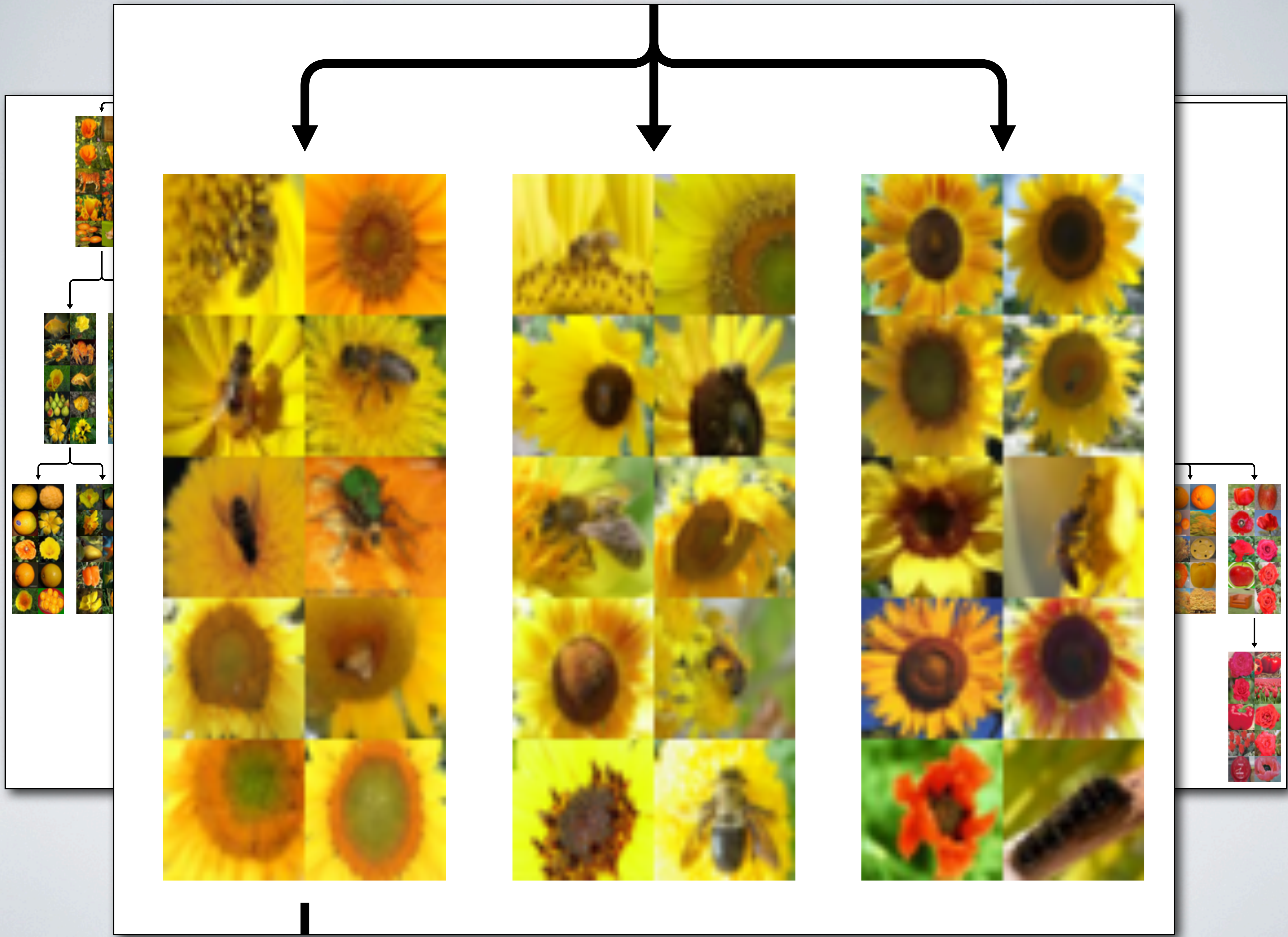


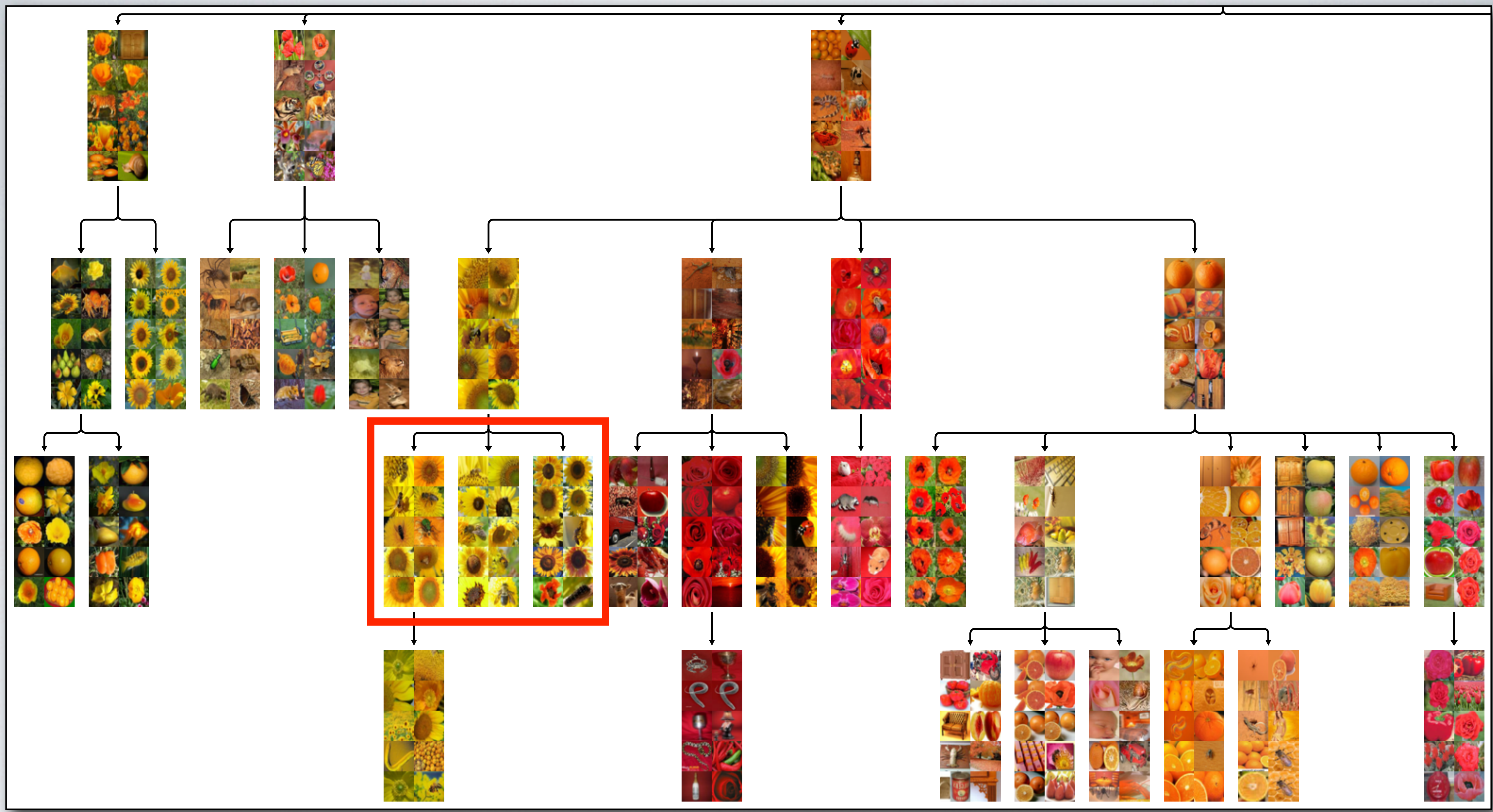


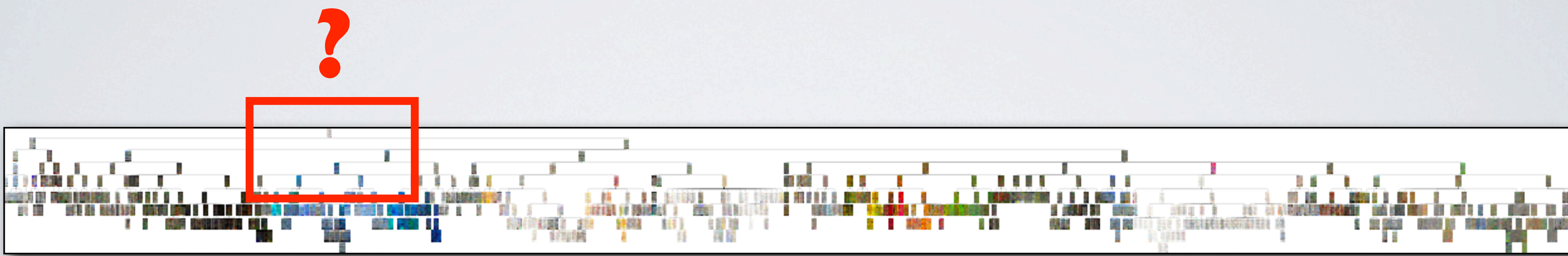


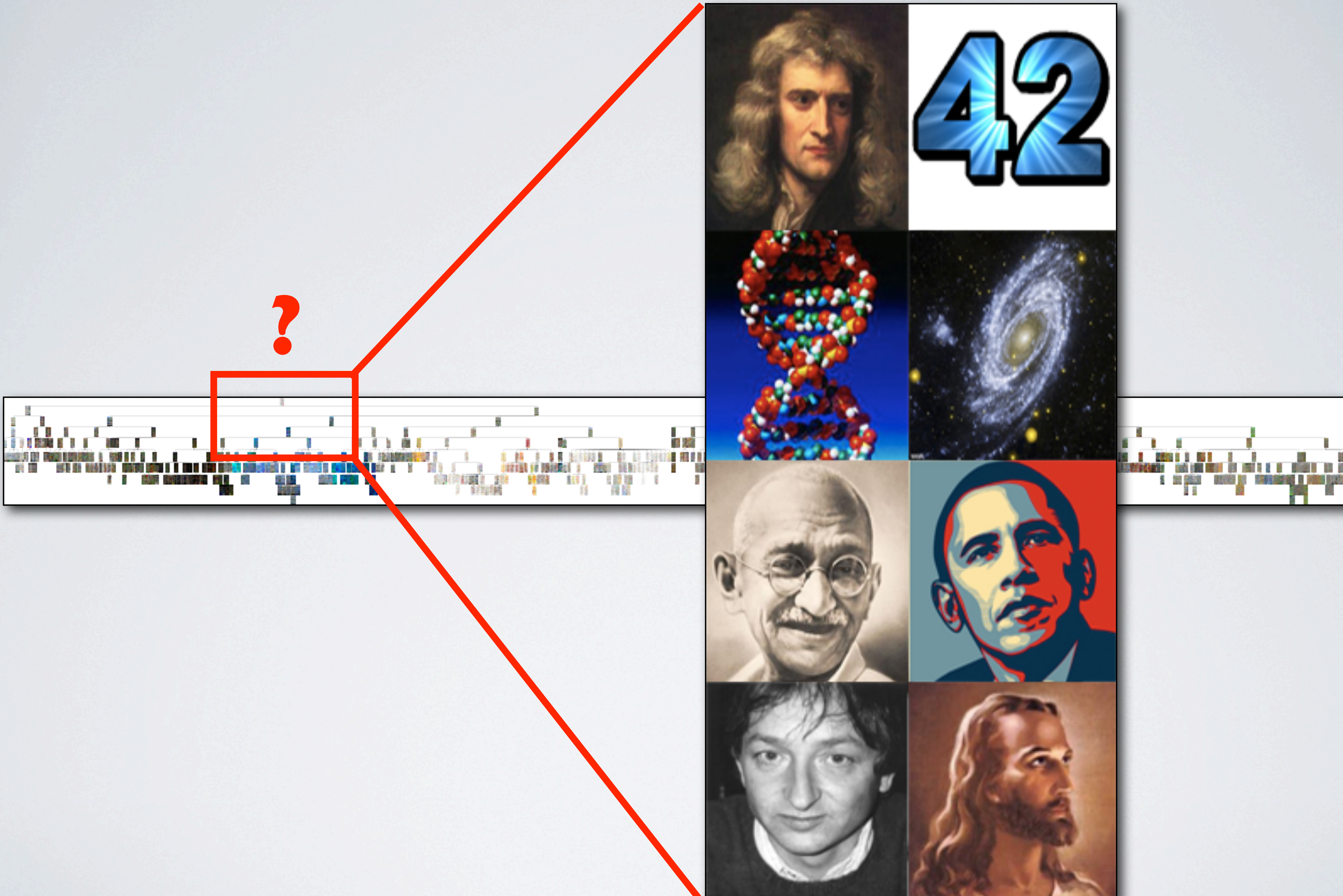


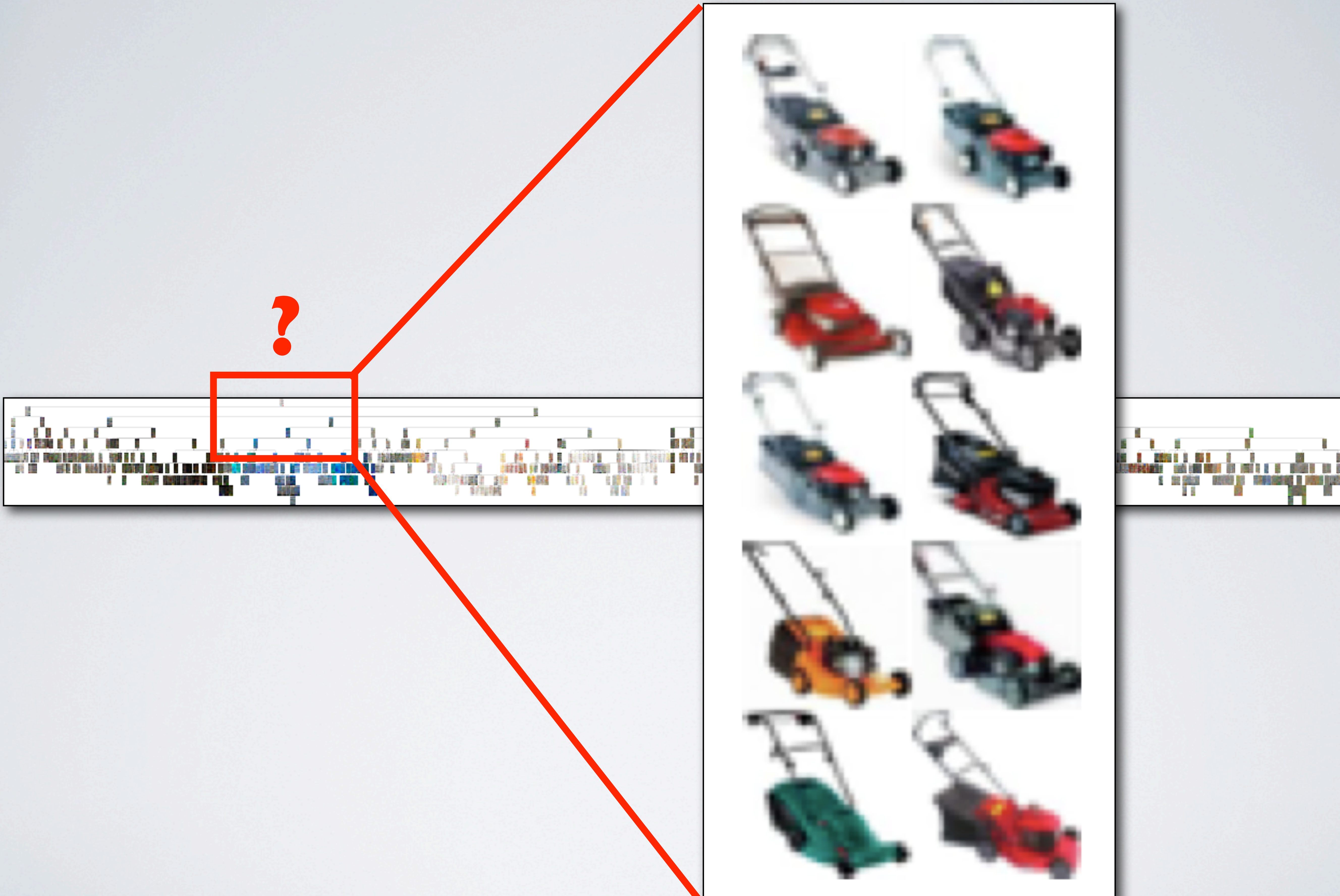









?
42



DOCUMENT MODELING

- Hierarchical clustering with LDA-style topic model.
- A "topic" is a distribution over words.
- A document has a distribution over topics.
- Words are exchangeable in documents.
- Documents at a node share a single topic distribution.

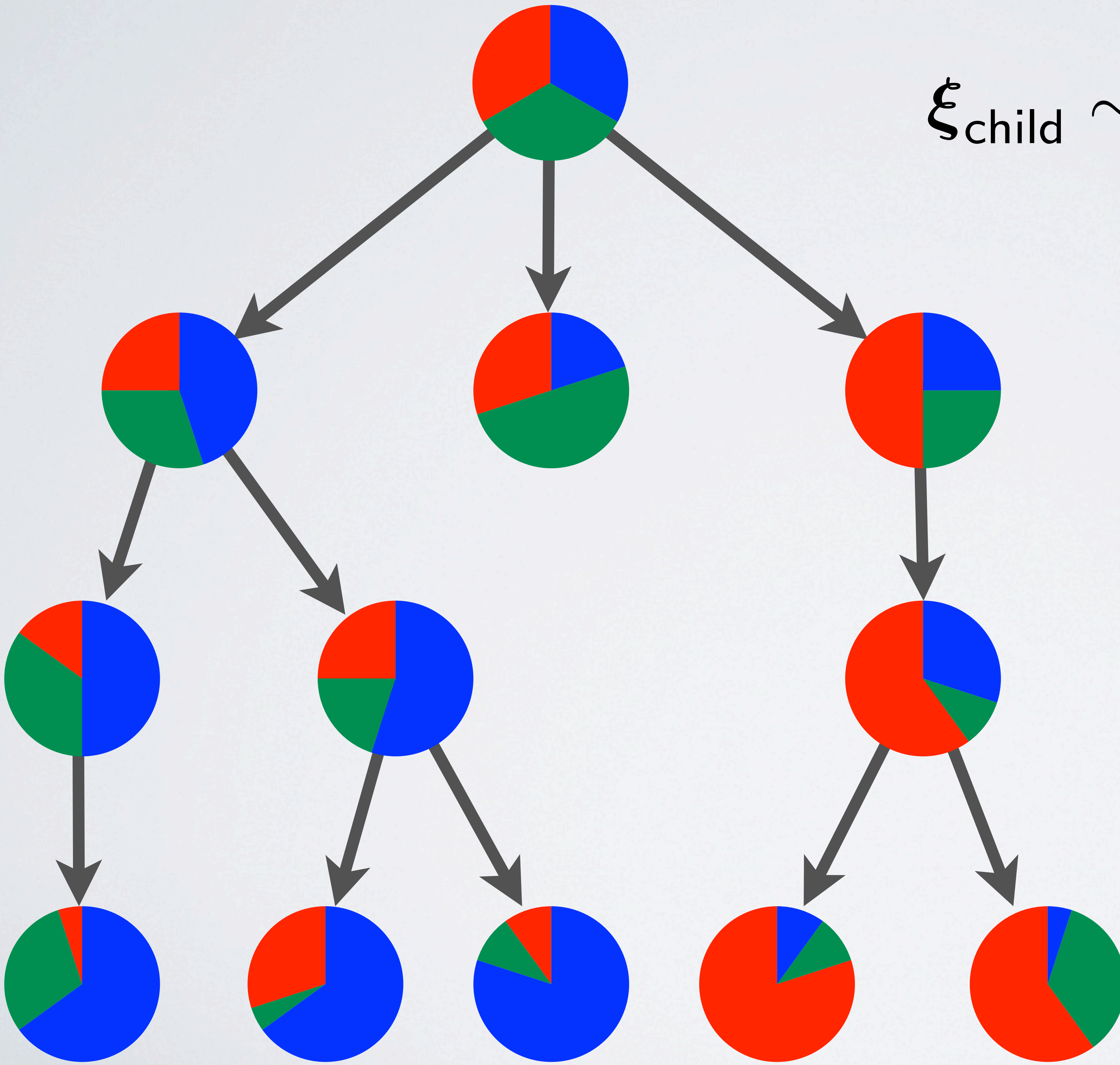
Difference with nested Chinese restaurant process[1]: nCRP has a *tree over topics*, and a document is an infinitely-long path down the tree.

We applied the model and inference to NIPS 1-12.

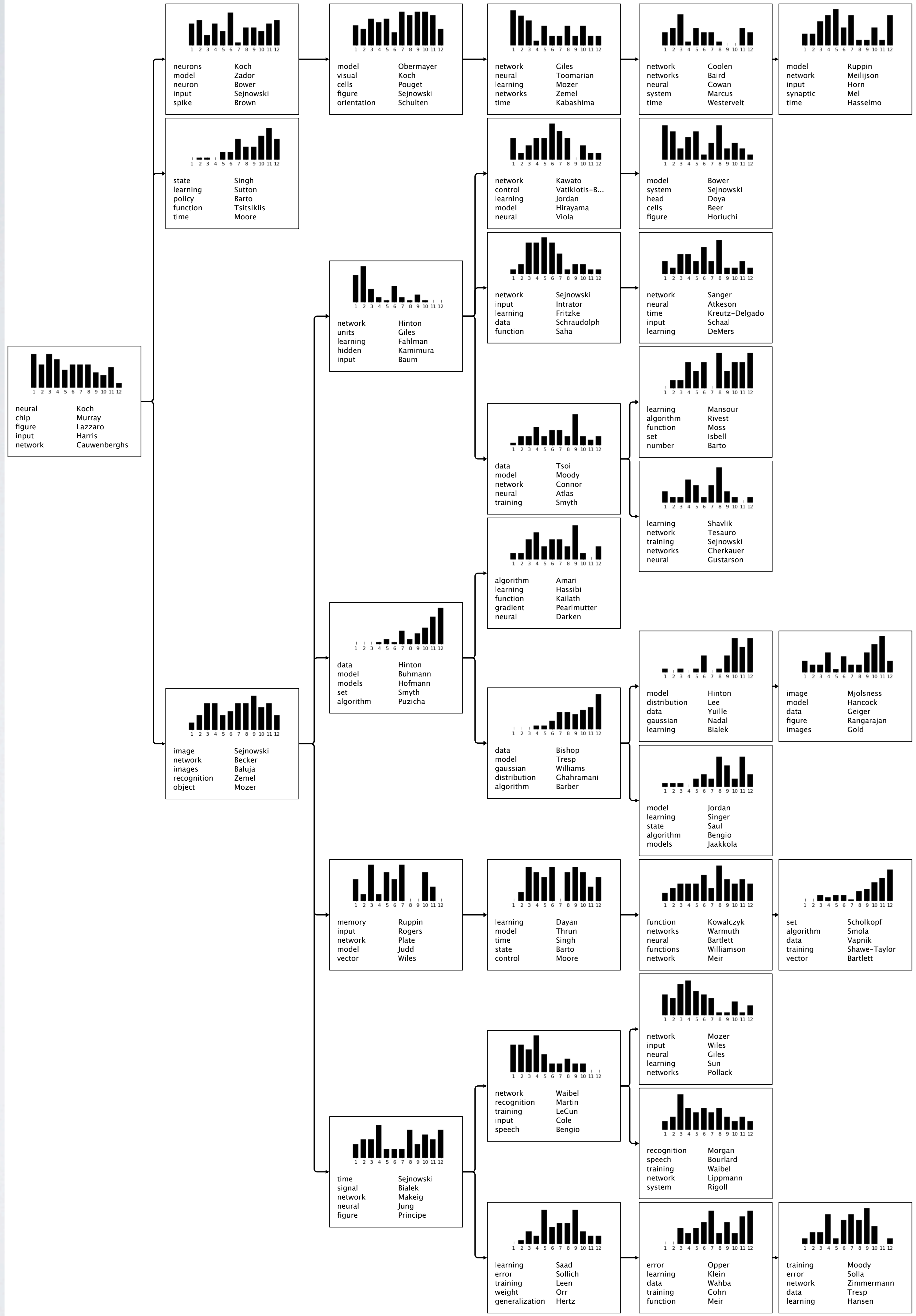
1. Blei, Griffiths and Jordan. Journal of the ACM, 2010.

DOCUMENT MODELING

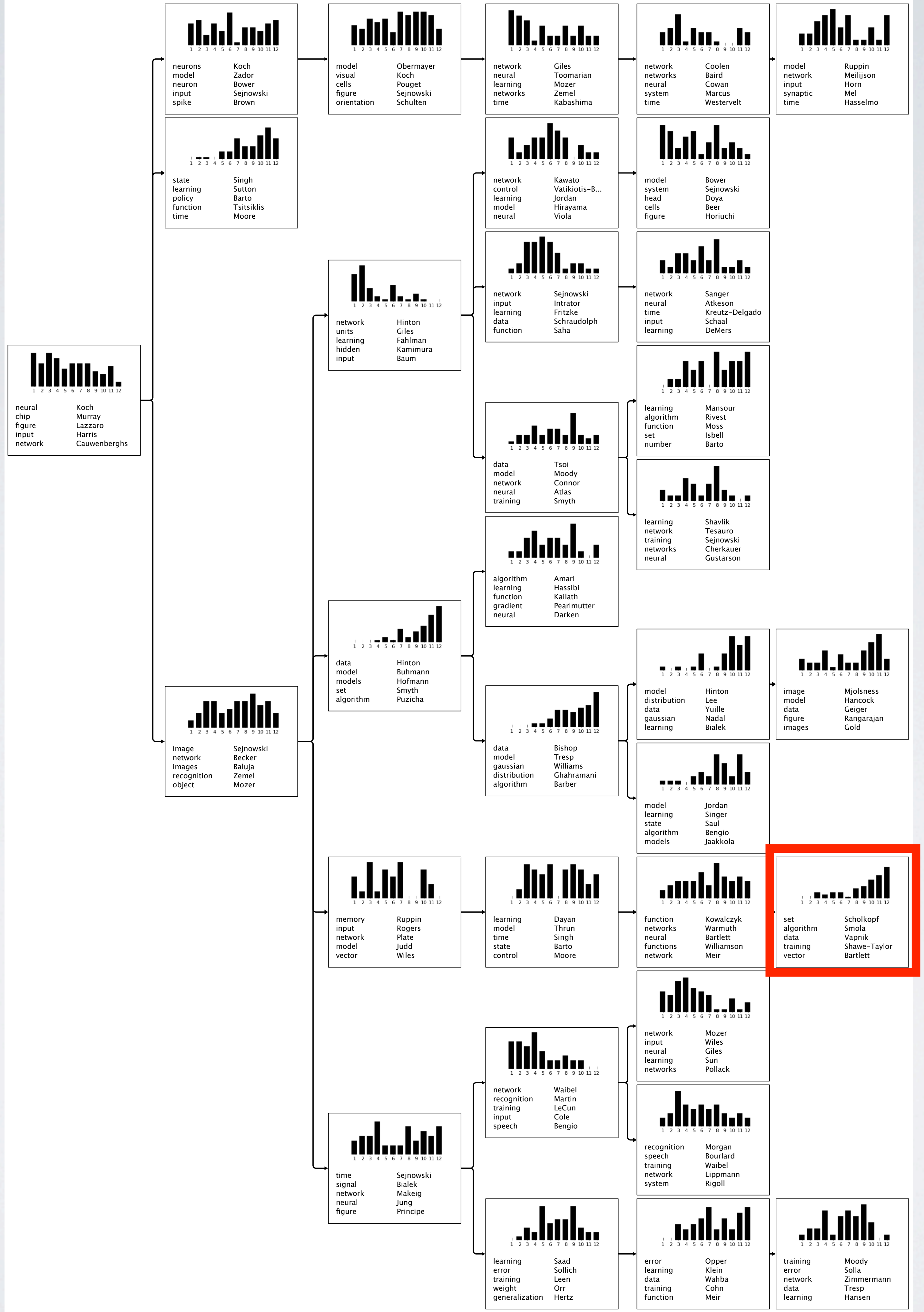
Each node has a distribution over topics ξ

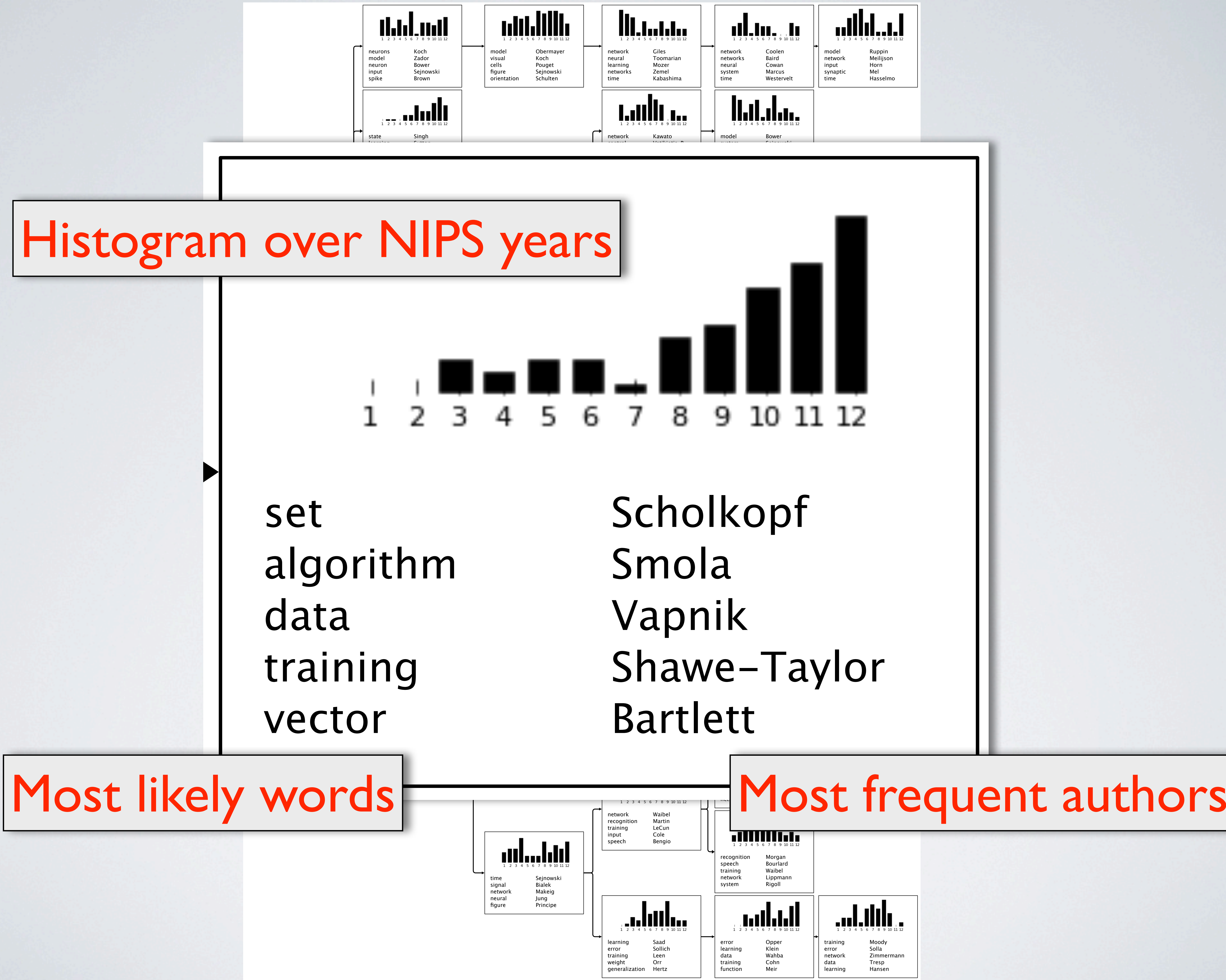


$$\xi_{\text{child}} \sim \text{Dirichlet}(\gamma\,\xi_{\text{parent}})$$

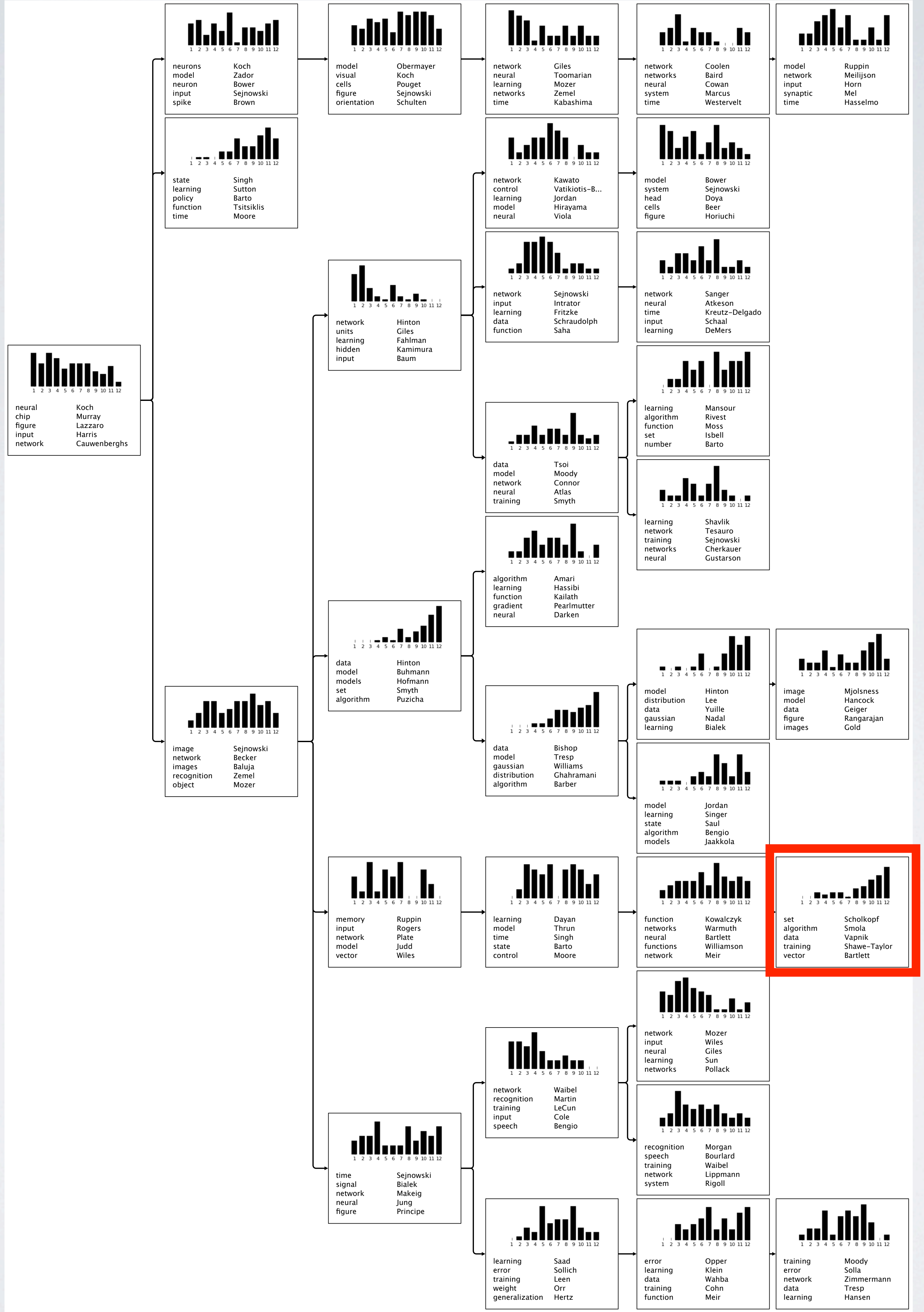


neural chip figure input network
Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike
Koch Zador Bower Sejnowski Brown
state learning policy function time
Singh Sutton Barto Tsitsiklis Moore
image network images recognition object
Sejnowski Becker Baluja Zemel Mozer
model visual cells figure orientation
Obermayer Koch Pouget Sejnowski Schulten
network units learning hidden input
Hinton Giles Fahlman Kamimura Baum
data model models set algorithm
Hinton Buhmann Hofmann Smyth Puzicha
memory input network model vector
Ruppin Rogers Plate Judd Wiles
time signal network neural figure
Sejnowski Bialek Makeig Jung Principe
network neural learning networks time
Giles Toomarian Mozer Zemel Kabashima
network control learning model neural
Kawato Vatikiotis-B... Jordan Hirayama Viola
network input learning data function
Sejnowski Intrator Fritzke Schraudolph Saha
data model network neural training
Tsoi Moody Connor Atlas Smyth
algorithm learning function gradient neural
Amari Hassibi Kailath Pearlmutter Darken
data model gaussian distribution algorithm
Bishop Tresp Williams Ghahramani Barber
learning model time state control
Dayan Thrun Singh Barto Moore
network recognition training input speech
Waibel Martin LeCun Cole Bengio
learning error training weight generalization
Saad Sollich Leen Orr Hertz
network networks neural system time
Coolen Baird Cowan Marcus Westervelt
model system head cells figure
Bower Sejnowski Doya Beer Horiuchi
network neural time input learning
Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number
Mansour Rivest Moss Isbell Barto
learning network training networks neural
Shavlik Tesauro Sejnowski Cherkauer Gustarson
model distribution data gaussian learning
Hinton Lee Yuille Nadal Bialek
model learning state algorithm models
Jordan Singer Saul Bengio Jaakkola
function networks neural functions network
Kowalczyk Warmuth Bartlett Williamson Meir
network input neural learning networks
Mozer Wiles Giles Sun Pollack
recognition speech training network system
Morgan Bourlard Waibel Lippmann Rigoll
error learning data training function
Opper Klein Wahba Cohn Meir
model network input synaptic time
Ruppin Meilijson Horn Mel Hasselmo
image model data figure images
Mjolsness Hancock Geiger Rangarajan Gold
set algorithm data training vector
Scholkopf Smola Vapnik Shawe-Taylor Bartlett
training error network data learning
Moody Solla Zimmermann Tresp Hansen

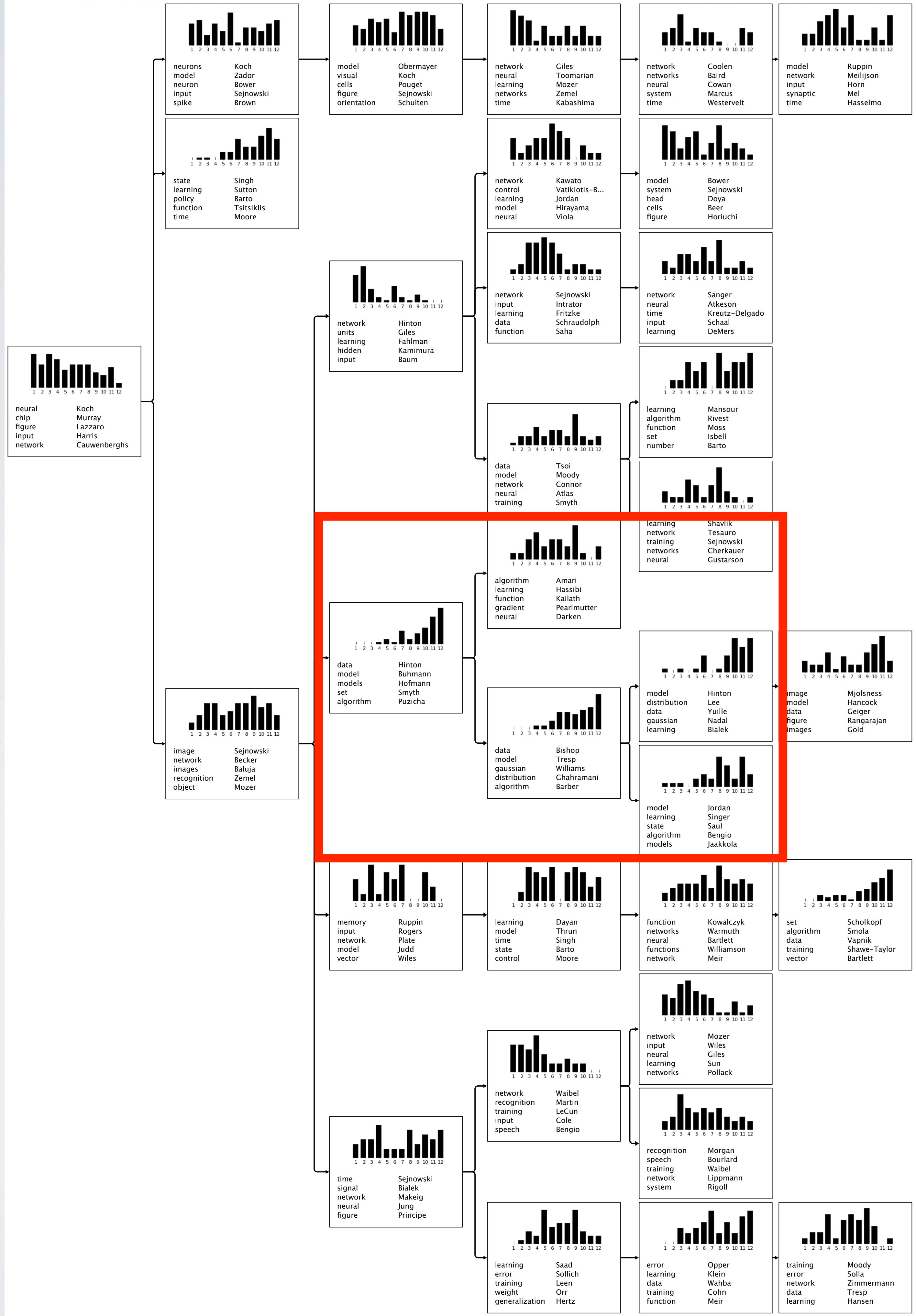

neural chip figure input network
Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike
Koch Zador Bower Sejnowski Brown
state learning policy function time
Singh Sutton Barto Tsitsiklis Moore
image network images recognition object
Sejnowski Becker Baluja Zemel Mozer
model visual cells figure orientation
Obermayer Koch Pouget Sejnowski Schulten
network units learning hidden input
Hinton Giles Fahlman Kamimura Baum
data model models set algorithm
Hinton Buhmann Hofmann Smyth Puzicha
memory input network model vector
Ruppin Rogers Plate Judd Wiles
time signal network neural figure
Sejnowski Bialek Makeig Jung Principe
network neural learning networks time
Giles Toomarian Mozer Zemel Kabashima
network control learning model neural
Kawato Vatikiotis-B... Jordan Hirayama Viola
network input learning data function
Sejnowski Intrator Fritzke Schraudolph Saha
data model network neural training
Tsoi Moody Connor Atlas Smyth
algorithm learning function gradient neural
Amari Hassibi Kailath Pearlmutter Darken
data model gaussian distribution algorithm
Bishop Tresp Williams Ghahramani Barber
learning model time state control
Dayan Thrun Singh Barto Moore
network recognition training input speech
Waibel Martin LeCun Cole Bengio
learning error training weight generalization
Saad Sollich Leen Orr Hertz
network networks neural system time
Coolen Baird Cowan Marcus Westervelt
model system head cells figure
Bower Sejnowski Doya Beer Horiuchi
network neural time input learning
Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number
Mansour Rivest Moss Isbell Barto
learning network training networks neural
Shavlik Tesauro Sejnowski Cherkauer Gustarson
model distribution data gaussian learning
Hinton Lee Yuille Nadal Bialek
model learning state algorithm models
Jordan Singer Saul Bengio Jaakkola
function networks neural functions network
Kowalczyk Warmuth Bartlett Williamson Meir
network input neural learning networks
Mozer Wiles Giles Sun Pollack
recognition speech training network system
Morgan Bourlard Waibel Lippmann Rigoll
error learning data training function
Opper Klein Wahba Cohn Meir
model network input synaptic time
Ruppin Meilijson Horn Mel Hasselmo
image model data figure images
Mjolsness Hancock Geiger Rangarajan Gold
set algorithm data training vector
Scholkopf Smola Vapnik Shawe-Taylor Bartlett
training error network data learning
Moody Solla Zimmermann Tresp Hansen

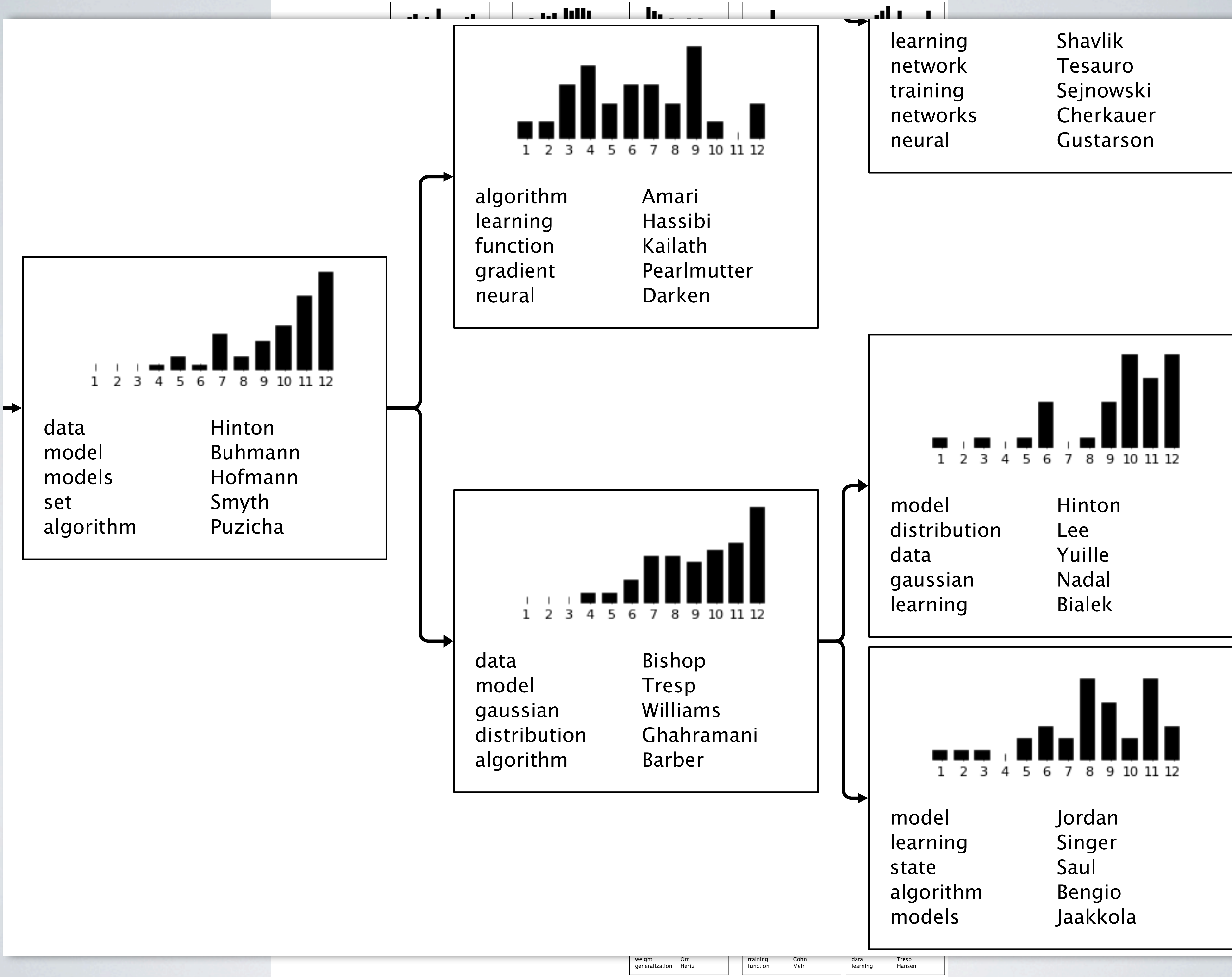


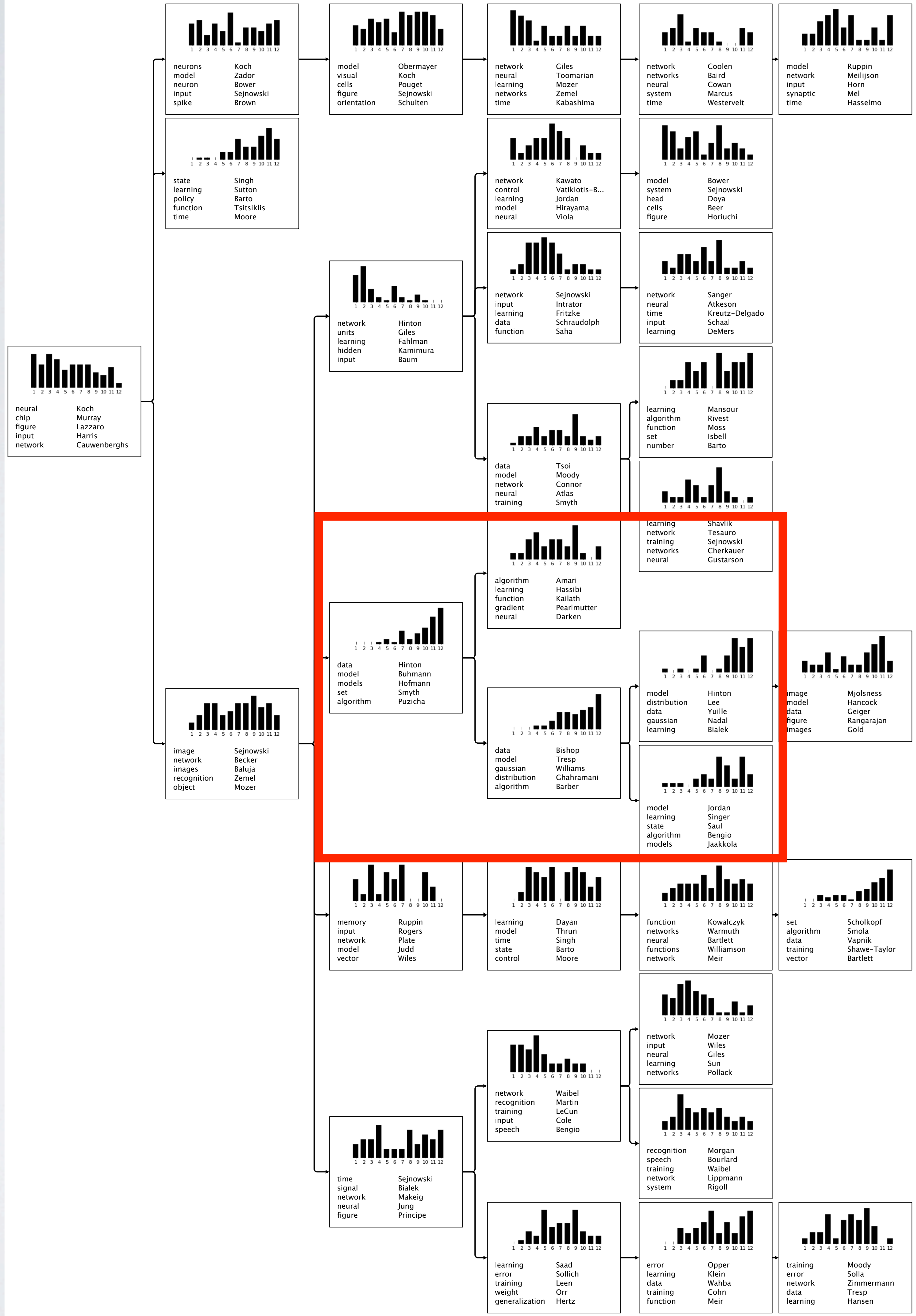
Histogram over NIPS years
1 2 3 4 5 6 7 8 9 10 11 12
set
algorithm
data
training
vector
Scholkopf
Smola
Vapnik
Shawe-Taylor
Bartlett
Most likely words
Most frequent authors



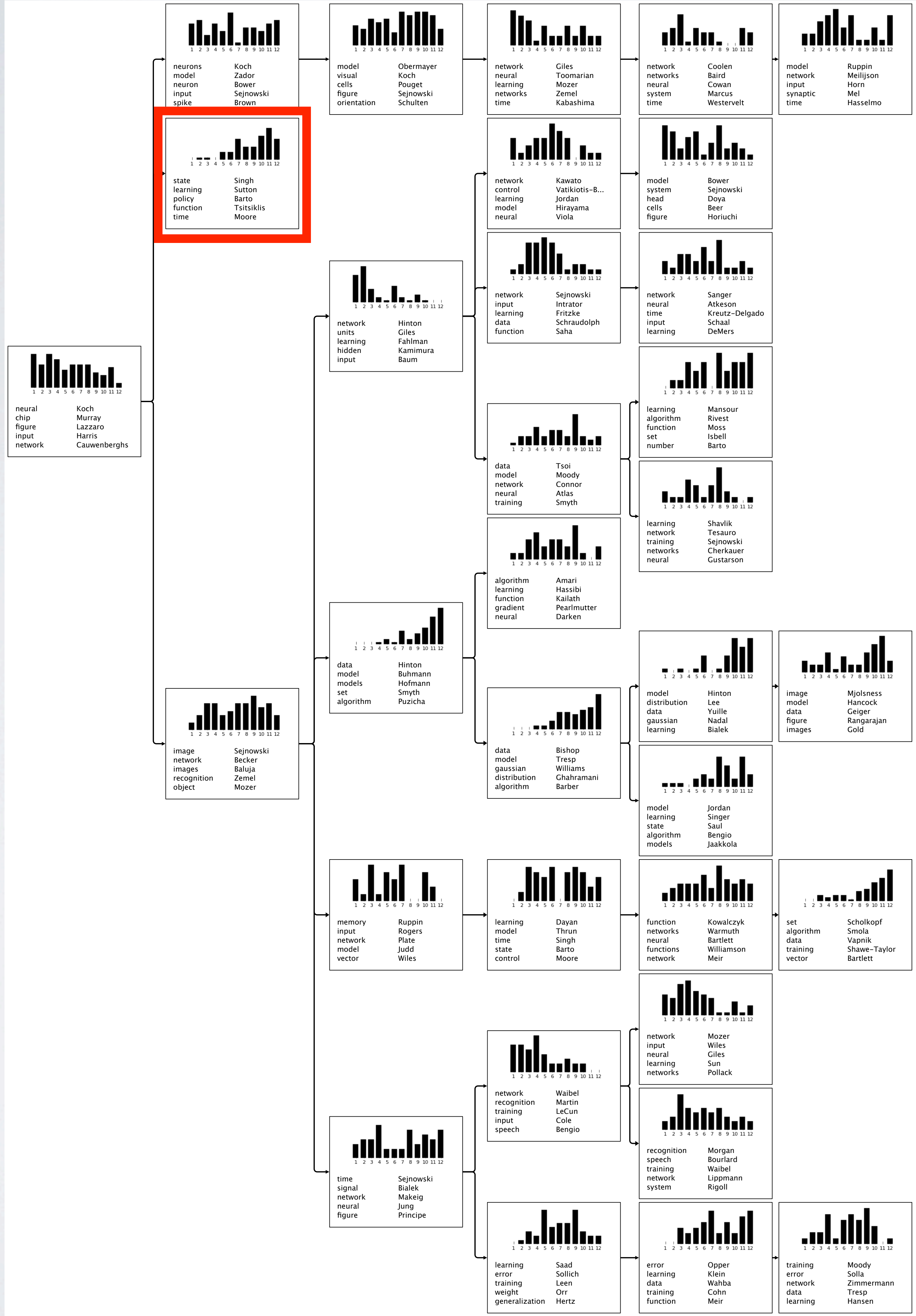
neural chip figure input network
Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike
Koch Zador Bower Sejnowski Brown
state learning policy function time
Singh Sutton Barto Tsitsiklis Moore
image network images recognition object
Sejnowski Becker Baluja Zemel Mozer
model visual cells figure orientation
Obermayer Koch Pouget Sejnowski Schulten
network units learning hidden input
Hinton Giles Fahlman Kamimura Baum
data model models set algorithm
Hinton Buhmann Hofmann Smyth Puzicha
memory input network model vector
Ruppin Rogers Plate Judd Wiles
time signal network neural figure
Sejnowski Bialek Makeig Jung Principe
network neural learning networks time
Giles Toomarian Mozer Zemel Kabashima
network control learning model neural
Kawato Vatikiotis-B... Jordan Hirayama Viola
network input learning data function
Sejnowski Intrator Fritzke Schraudolph Saha
data model network neural training
Tsoi Moody Connor Atlas Smyth
algorithm learning function gradient neural
Amari Hassibi Kailath Pearlmutter Darken
data model gaussian distribution algorithm
Bishop Tresp Williams Ghahramani Barber
learning model time state control
Dayan Thrun Singh Barto Moore
network recognition training input speech
Waibel Martin LeCun Cole Bengio
learning error training weight generalization
Saad Sollich Leen Orr Hertz
network networks neural system time
Coolen Baird Cowan Marcus Westervelt
model system head cells figure
Bower Sejnowski Doya Beer Horiuchi
network neural time input learning
Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number
Mansour Rivest Moss Isbell Barto
learning network training networks neural
Shavlik Tesauro Sejnowski Cherkauer Gustarson
model distribution data gaussian learning
Hinton Lee Yuille Nadal Bialek
model learning state algorithm models
Jordan Singer Saul Bengio Jaakkola
function networks neural functions network
Kowalczyk Warmuth Bartlett Williamson Meir
network input neural learning networks
Mozer Wiles Giles Sun Pollack
recognition speech training network system
Morgan Bourlard Waibel Lippmann Rigoll
error learning data training function
Opper Klein Wahba Cohn Meir
model network input synaptic time
Ruppin Meilijson Horn Mel Hasselmo
image model data figure images
Mjolsness Hancock Geiger Rangarajan Gold
set algorithm data training vector
Scholkopf Smola Vapnik Shawe-Taylor Bartlett
training error network data learning
Moody Solla Zimmermann Tresp Hansen


neural chip figure input network
Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike
Koch Zador Bower Sejnowski Brown
model visual cells figure orientation
Obermayer Koch Pouget Sejnowski Schulten
network neural learning networks time
Giles Toomarian Mozer Zemel Kabashima
network networks neural system time
Coolen Baird Cowan Marcus Westervelt
model network input synaptic time
Ruppin Meilijson Horn Mel Hasselmo
state learning policy function time
Singh Sutton Barto Tsitsiklis Moore
network control learning model neural
Kawato Vatikiotis-B... Jordan Hirayama Viola
model system head cells figure
Bower Sejnowski Doya Beer Horiuchi
network units learning hidden input
Hinton Giles Fahlman Kamimura Baum
network input learning data function
Sejnowski Intrator Fritzke Schraudolph Saha
network neural time input learning
Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number
Mansour Rivest Moss Isbell Barto
data model network neural training
Tsoi Moody Connor Atlas Smyth
learning network training networks neural
Shavlik Tesauro Sejnowski Cherkauer Gustarson
algorithm learning function gradient neural
Amari Hassibi Kailath Pearlmutter Darken
data model models set algorithm
Hinton Buhmann Hofmann Smyth Puzicha
image network images recognition object
Sejnowski Becker Baluja Zemel Mozer
model distribution data gaussian learning
Hinton Lee Yuille Nadal Bialek
image model data figure images
Mjolsness Hancock Geiger Rangarajan Gold
data model gaussian distribution algorithm
Bishop Tresp Williams Ghahramani Barber
model learning state algorithm models
Jordan Singer Saul Bengio Jaakkola
memory input network model vector
Ruppin Rogers Plate Judd Wiles
learning model time state control
Dayan Thrun Singh Barto Moore
function networks neural functions network
Kowalczyk Warmuth Bartlett Williamson Meir
set algorithm data training vector
Scholkopf Smola Vapnik Shawe-Taylor Bartlett
network input neural learning networks
Mozer Wiles Giles Sun Pollack
network recognition training input speech
Waibel Martin LeCun Cole Bengio
recognition speech training network system
Morgan Bourlard Waibel Lippmann Rigoll
time signal network neural figure
Sejnowski Bialek Makeig Jung Principe
learning error training weight generalization
Saad Sollich Leen Orr Hertz
error learning data training function
Opper Klein Wahba Cohn Meir
training error network data learning
Moody Solla Zimmermann Tresp Hansen

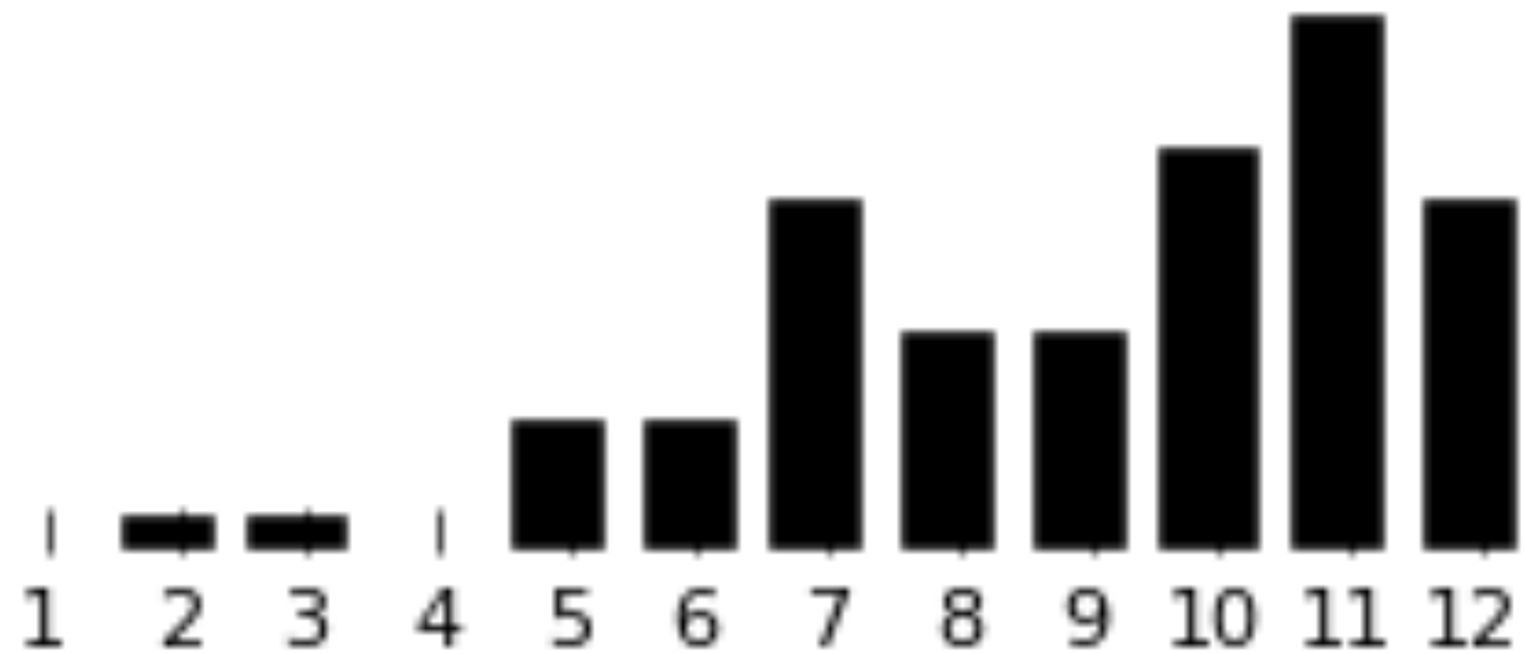

learning Shavlik
network Tesauro
training Sejnowski
networks Cherkauer
neural Gustarson
algorithm Amari
learning Hassibi
function Kailath
gradient Pearlmutter
neural Darken
data Hinton
model Buhmann
models Hofmann
set Smyth
algorithm Puzicha
model Hinton
distribution Lee
data Yuille
gaussian Nadal
learning Bialek
data Bishop
model Tresp
gaussian Williams
distribution Ghahramani
algorithm Barber
model Jordan
learning Singer
state Saul
algorithm Bengio
models Jaakkola
weight Orr
generalization Hertz
training Cohn
function Meir
data Tresp
learning Hansen



neural chip figure input network — Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike — Koch Zador Bower Sejnowski Brown
state learning policy function time — Singh Sutton Barto Tsitsiklis Moore
image network images recognition object — Sejnowski Becker Baluja Zemel Mozer
model visual cells figure orientation — Obermayer Koch Pouget Sejnowski Schulten
network units learning hidden input — Hinton Giles Fahlman Kamimura Baum
data model models set algorithm — Hinton Buhmann Hofmann Smyth Puzicha
memory input network model vector — Ruppin Rogers Plate Judd Wiles
time signal network neural figure — Sejnowski Bialek Makeig Jung Principe
network neural learning networks time — Giles Toomarian Mozer Zemel Kabashima
network control learning model neural — Kawato Vatikiotis-B... Jordan Hirayama Viola
network input learning data function — Sejnowski Intrator Fritzke Schraudolph Saha
data model network neural training — Tsoi Moody Connor Atlas Smyth
algorithm learning function gradient neural — Amari Hassibi Kailath Pearlmutter Darken
data model gaussian distribution algorithm — Bishop Tresp Williams Ghahramani Barber
learning model time state control — Dayan Thrun Singh Barto Moore
network recognition training input speech — Waibel Martin LeCun Cole Bengio
learning error training weight generalization — Saad Sollich Leen Orr Hertz
network networks neural system time — Coolen Baird Cowan Marcus Westervelt
model system head cells figure — Bower Sejnowski Doya Beer Horiuchi
network neural time input learning — Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number — Mansour Rivest Moss Isbell Barto
learning network training networks neural — Shavlik Tesauro Sejnowski Cherkauer Gustarson
model distribution data gaussian learning — Hinton Lee Yuille Nadal Bialek
model learning state algorithm models — Jordan Singer Saul Bengio Jaakkola
function networks neural functions network — Kowalczyk Warmuth Bartlett Williamson Meir
network input neural learning networks — Mozer Wiles Giles Sun Pollack
recognition speech training network system — Morgan Bourlard Waibel Lippmann Rigoll
error learning data training function — Opper Klein Wahba Cohn Meir
model network input synaptic time — Ruppin Meilijson Horn Mel Hasselmo
image model data figure images — Mjolsness Hancock Geiger Rangarajan Gold
set algorithm data training vector — Scholkopf Smola Vapnik Shawe-Taylor Bartlett
training error network data learning — Moody Solla Zimmermann Tresp Hansen



neural chip figure input network
Koch Murray Lazzaro Harris Cauwenberghs
neurons model neuron input spike
Koch Zador Bower Sejnowski Brown
state learning policy function time
Singh Sutton Barto Tsitsiklis Moore
image network images recognition object
Sejnowski Becker Baluja Zemel Mozer
model visual cells figure orientation
Obermayer Koch Pouget Sejnowski Schulten
network units learning hidden input
Hinton Giles Fahlman Kamimura Baum
data model models set algorithm
Hinton Buhmann Hofmann Smyth Puzicha
memory input network model vector
Ruppin Rogers Plate Judd Wiles
time signal network neural figure
Sejnowski Bialek Makeig Jung Principe
network neural learning networks time
Giles Toomarian Mozer Zemel Kabashima
network control learning model neural
Kawato Vatikiotis-B... Jordan Hirayama Viola
network input learning data function
Sejnowski Intrator Fritzke Schraudolph Saha
data model network neural training
Tsoi Moody Connor Atlas Smyth
algorithm learning function gradient neural
Amari Hassibi Kailath Pearlmutter Darken
data model gaussian distribution algorithm
Bishop Tresp Williams Ghahramani Barber
learning model time state control
Dayan Thrun Singh Barto Moore
network recognition training input speech
Waibel Martin LeCun Cole Bengio
learning error training weight generalization
Saad Sollich Leen Orr Hertz
network networks neural system time
Coolen Baird Cowan Marcus Westervelt
model system head cells figure
Bower Sejnowski Doya Beer Horiuchi
network neural time input learning
Sanger Atkeson Kreutz-Delgado Schaal DeMers
learning algorithm function set number
Mansour Rivest Moss Isbell Barto
learning network training networks neural
Shavlik Tesauro Sejnowski Cherkauer Gustarson
model distribution data gaussian learning
Hinton Lee Yuille Nadal Bialek
model learning state algorithm models
Jordan Singer Saul Bengio Jaakkola
function networks neural functions network
Kowalczyk Warmuth Bartlett Williamson Meir
network input neural learning networks
Mozer Wiles Giles Sun Pollack
recognition speech training network system
Morgan Bourlard Waibel Lippmann Rigoll
error learning data training function
Opper Klein Wahba Cohn Meir
model network input synaptic time
Ruppin Meilijson Horn Mel Hasselmo
image model data figure images
Mjolsness Hancock Geiger Rangarajan Gold
set algorithm data training vector
Scholkopf Smola Vapnik Shawe-Taylor Bartlett
training error network data learning
Moody Solla Zimmermann Tresp Hansen



state
learning
policy
function
time

Singh
Sutton
Barto
Tsitsiklis
Moore

SUMMARY

- A generative prior on tree-structured measures:
  - Unbounded width and depth
  - Data live at internal nodes at finite depth
  - Data are infinitely exchangeable
  - Corresponds to a Blackwell-MacQueen urn model

- We perform inference via MCMC.
- Generates an interesting visualization of images.
- Compares well with LDA in modeling NIPS text.

- Thanks to: Alex Krizhevsky, Kurt Miller, Iain Murray, Yee Whye Teh, Hanna Wallach, Sinead Williamson