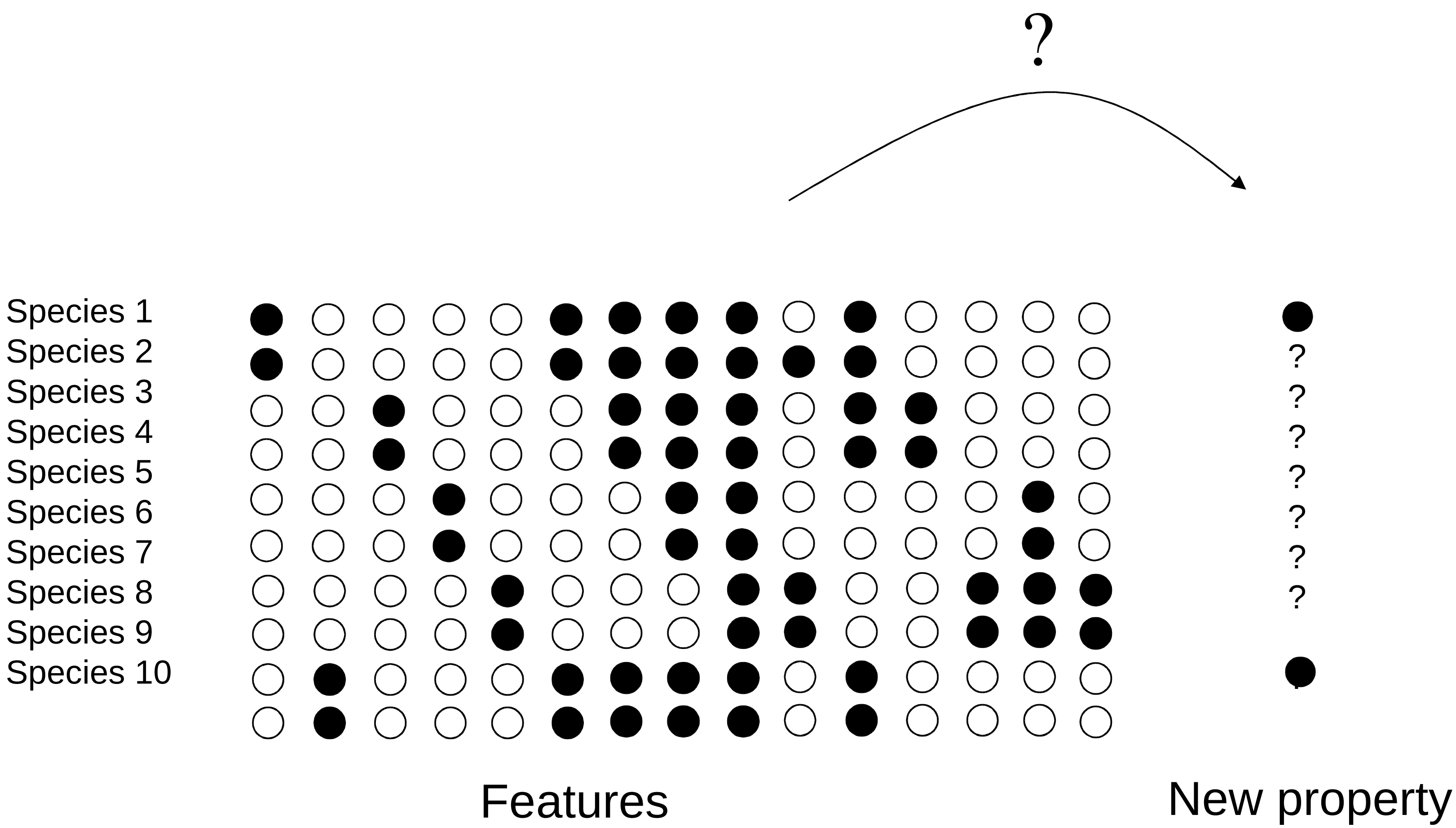
The computational problem



Feature rating data

(Osherson, D. N., et. al. "Category-based Induction." *Psychological Review* 197 (1990): 185-200.)

- People were given 48 animals, 85 features, and asked to rate whether each animal had each feature.
- E.g., elephant:

'gray' 'hairless' 'toughskin'
'big' 'bulbous' 'longleg'
'tail' 'chewteeth' 'tusks'
'smelly' 'walks' 'slow'
'strong' 'muscle’ 'quadrapedal'
'inactive' 'vegetation' 'grazer'
'oldworld' 'bush' 'jungle'
'ground' 'timid' 'smart'
'group'

Property Induction: Biology

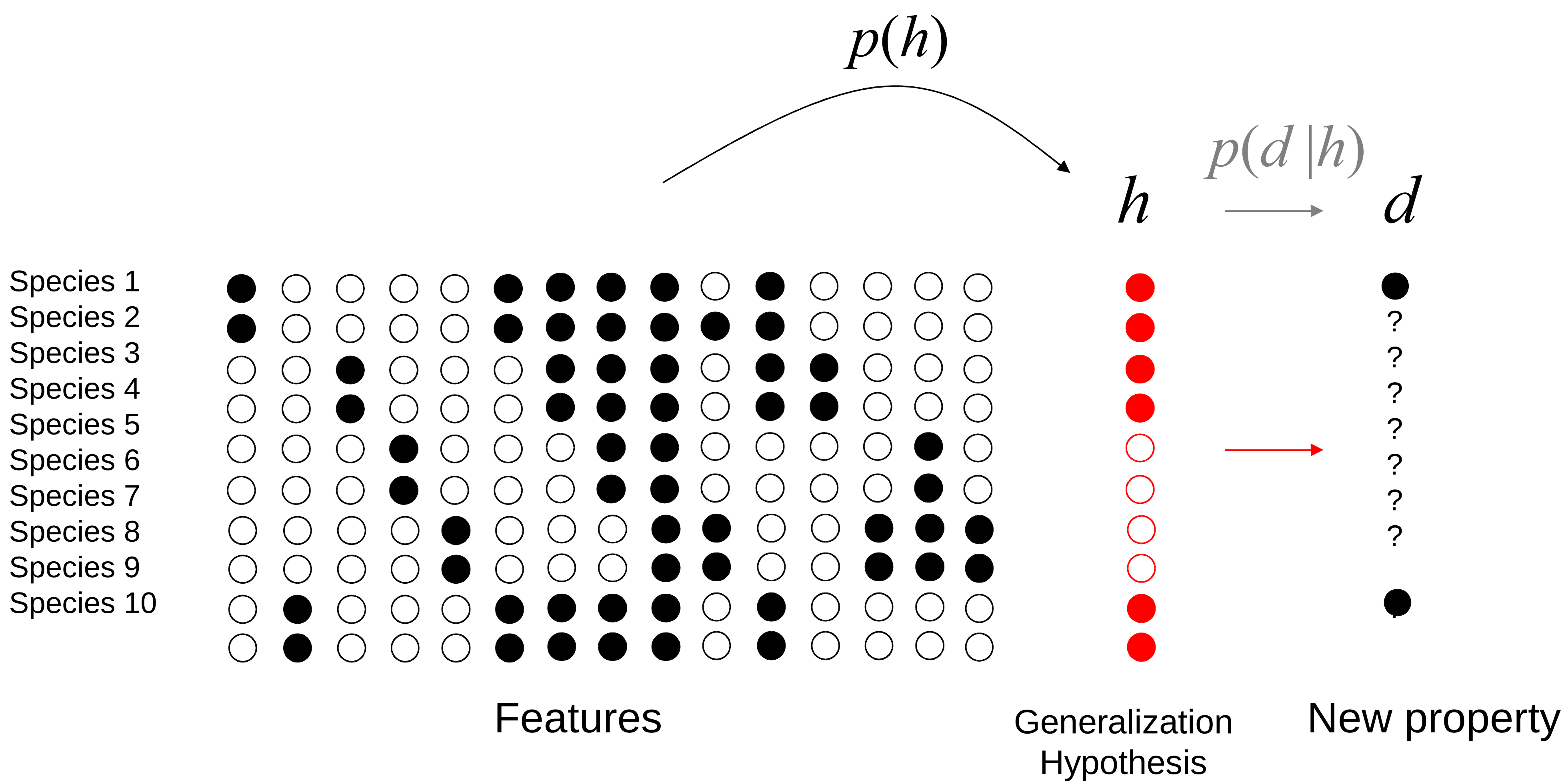
- Subjects rated two kinds of arguments:

Specific	General
Dolphins can catch Disease X Seals can catch Disease X ――――――――――― Horses can catch Disease X	Dolphins can catch Disease X Seals can catch Disease X ――――――――――― All mammals can catch Disease X

(Osherson, Smith, Wilkie, Lopez, & Shafir, 1990)

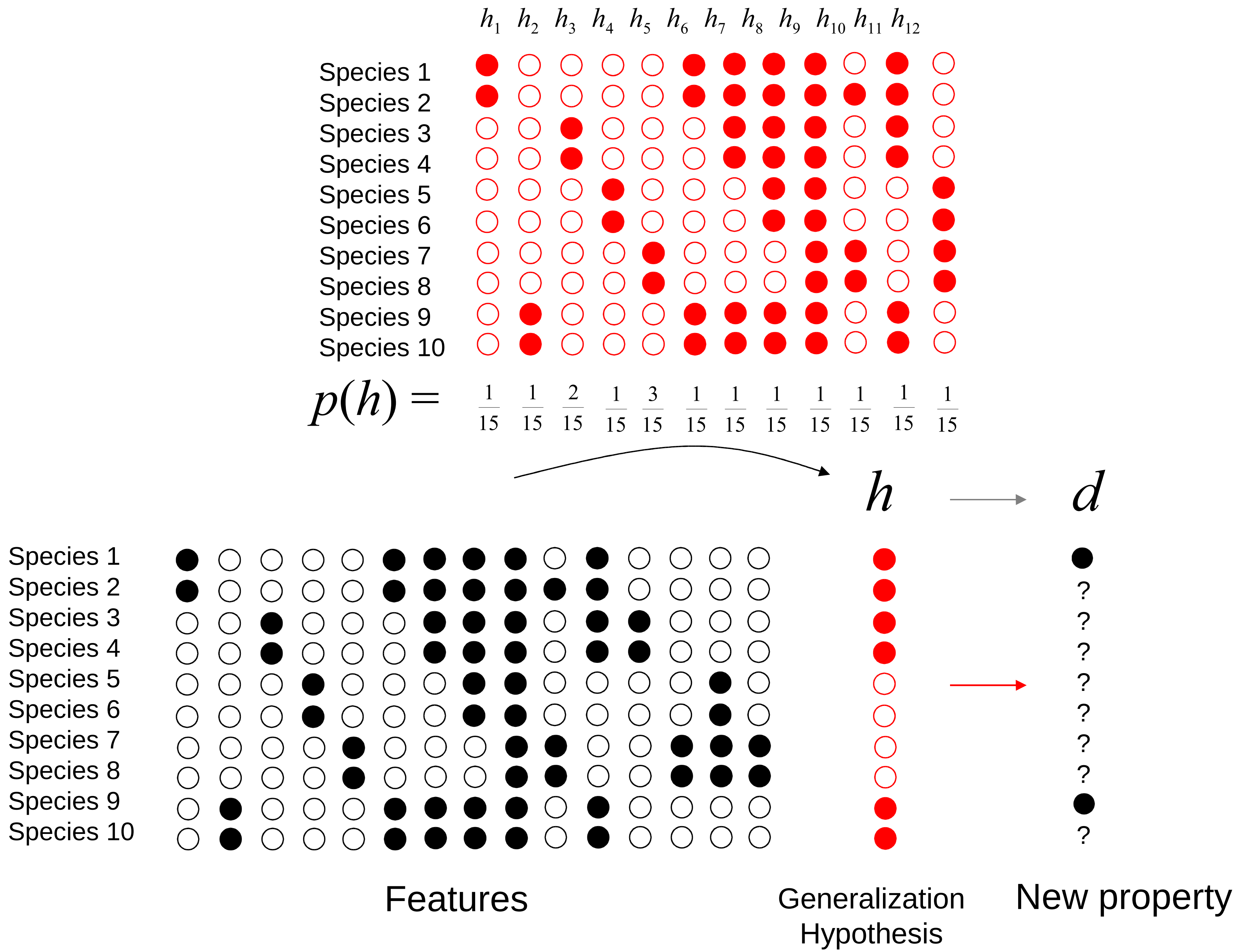
Probability that property Q holds for species x:

$$p(Q(x) \mid d) = \sum_{\substack{h \text{ consistent} \\ \text{with } Q(x),\, d}} p(h) \Bigg/ \sum_{\substack{h \text{ consistent} \\ \text{with } d}} p(h)$$



"Empiricist" Bayes:

(Heit, E. "A Bayesian Analysis of Some Forms of Inductive Reasoning." In *Rational Models of Cognition*. Edited by M. Oaksford and N. Chater. Oxford: Oxford University Press, 1998, pp. 248-274.)



“Theory-based” Bayes

Two principles

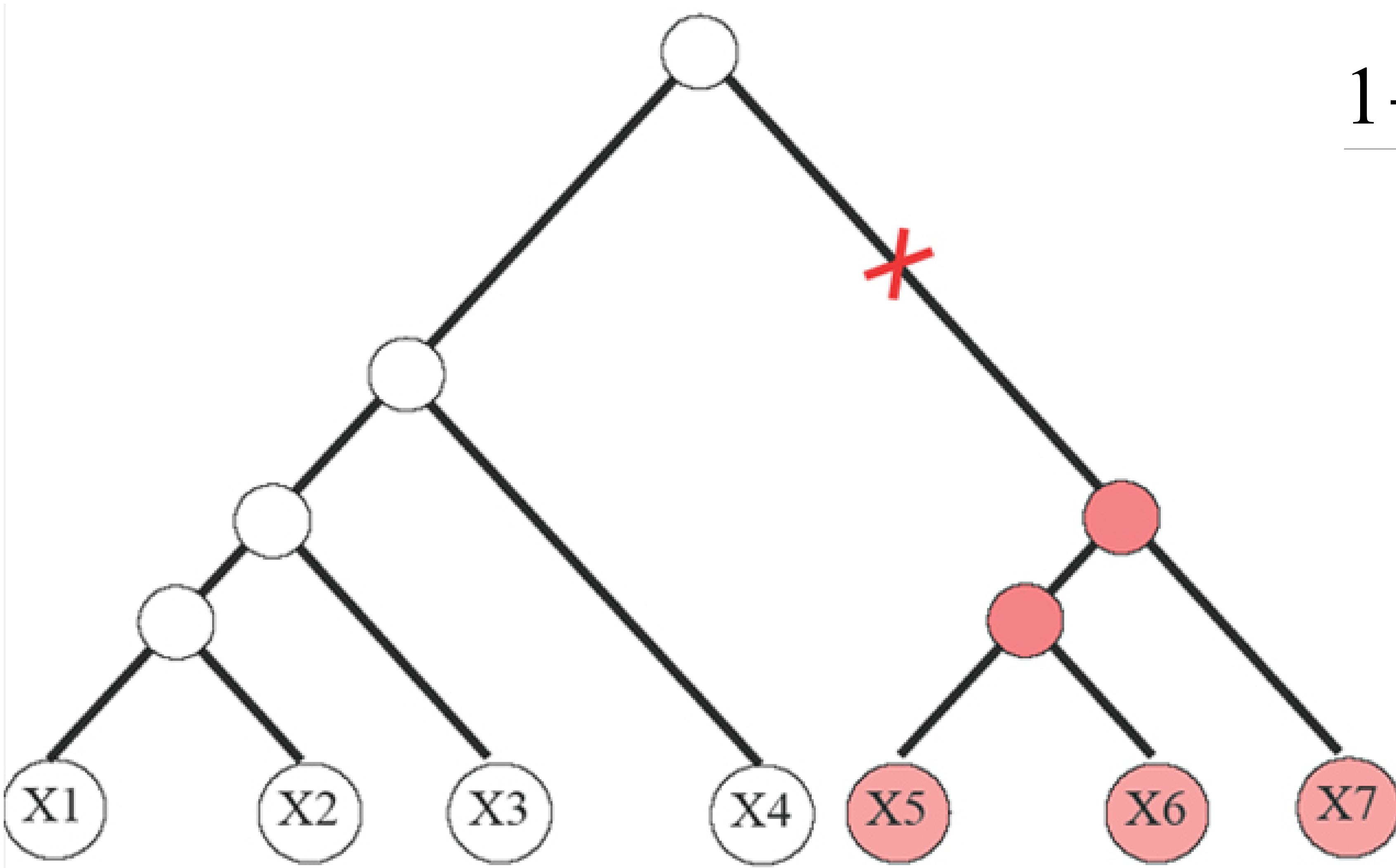
1. Species generated by an evolutionary branching process.

2. Features generated by stochastic mutation process over the tree

Generating Features: p(h|T)

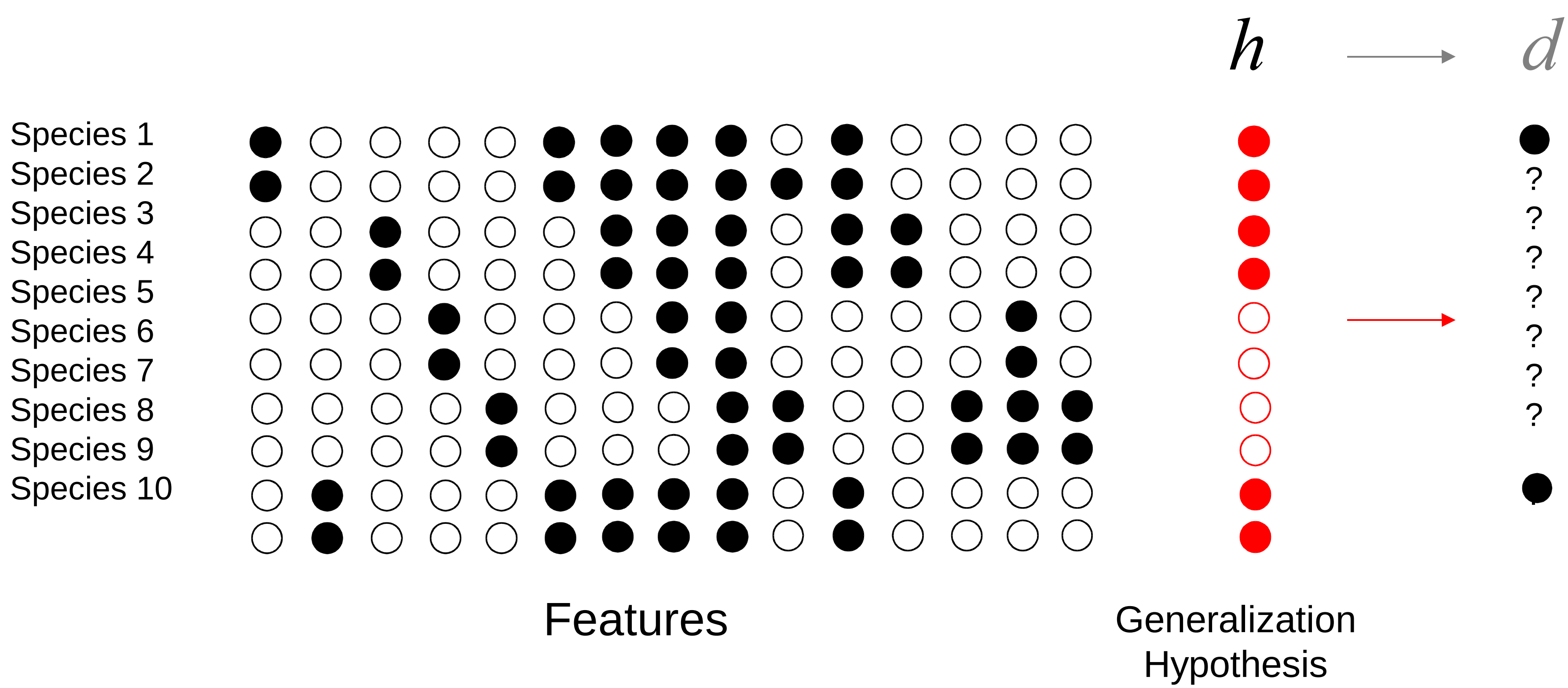
Probability of mutation along branch b :

$$\frac{1-e^{-2\lambda|b|}}{2}$$

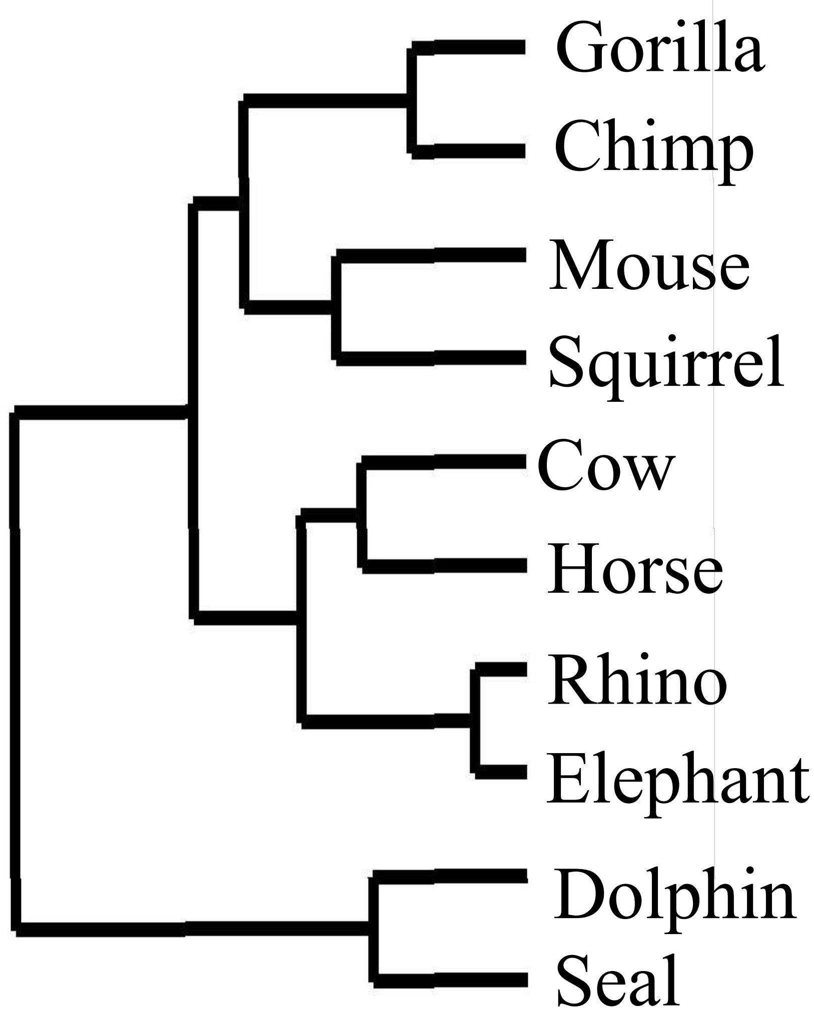


Choosing a tree

Find T that maximizes p(Features|T)



Best tree for Osherson data



Results

Theory-based Bayes

Bias is just right!

Image removed due to copyright considerations.

Taxonomic Bayes

Bias is too strong

"Empiricist" Bayes

Bias is too weak

An Unstructured PDP Approach

Image removed due to copyright considerations. Please see:

McClelland and Rogers. "The Parallel Distributed Processing Approach to Semantic Cognition."

Nature Reviews Neuroscience 4 (April 2003): 1-14.

Emergent Structure

Image removed due to copyright considerations.

Image removed due to copyright considerations. Please see:

Seidenberg, M. S., and J. L. Elman. "Do Infants Learn Grammar with Algebra or Statistics." *Science* 284 (1999a): 434435.

PDP simulations

- Architectures: 48-64-85, 40-20-20-85, 48-35-64-85, 48-100-100-85, 48-15-30-85
- Learning rates: 0.05, 0.005, 0.1
- Momentum: 0, 0.9
- Bias: 0, -2
- Training epochs: 2000, 4000, 8000, 12000, 20000

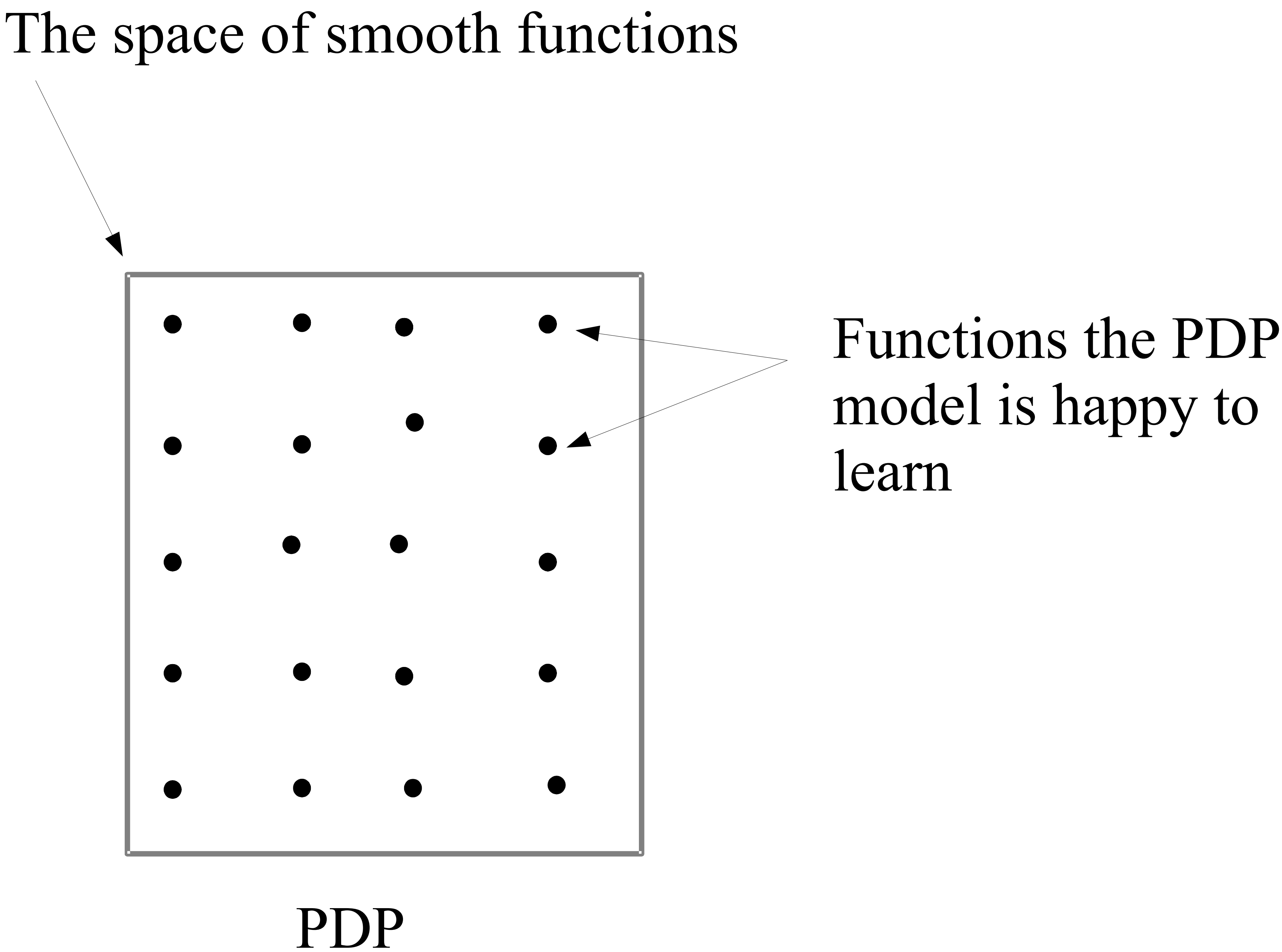
Structured

PDP

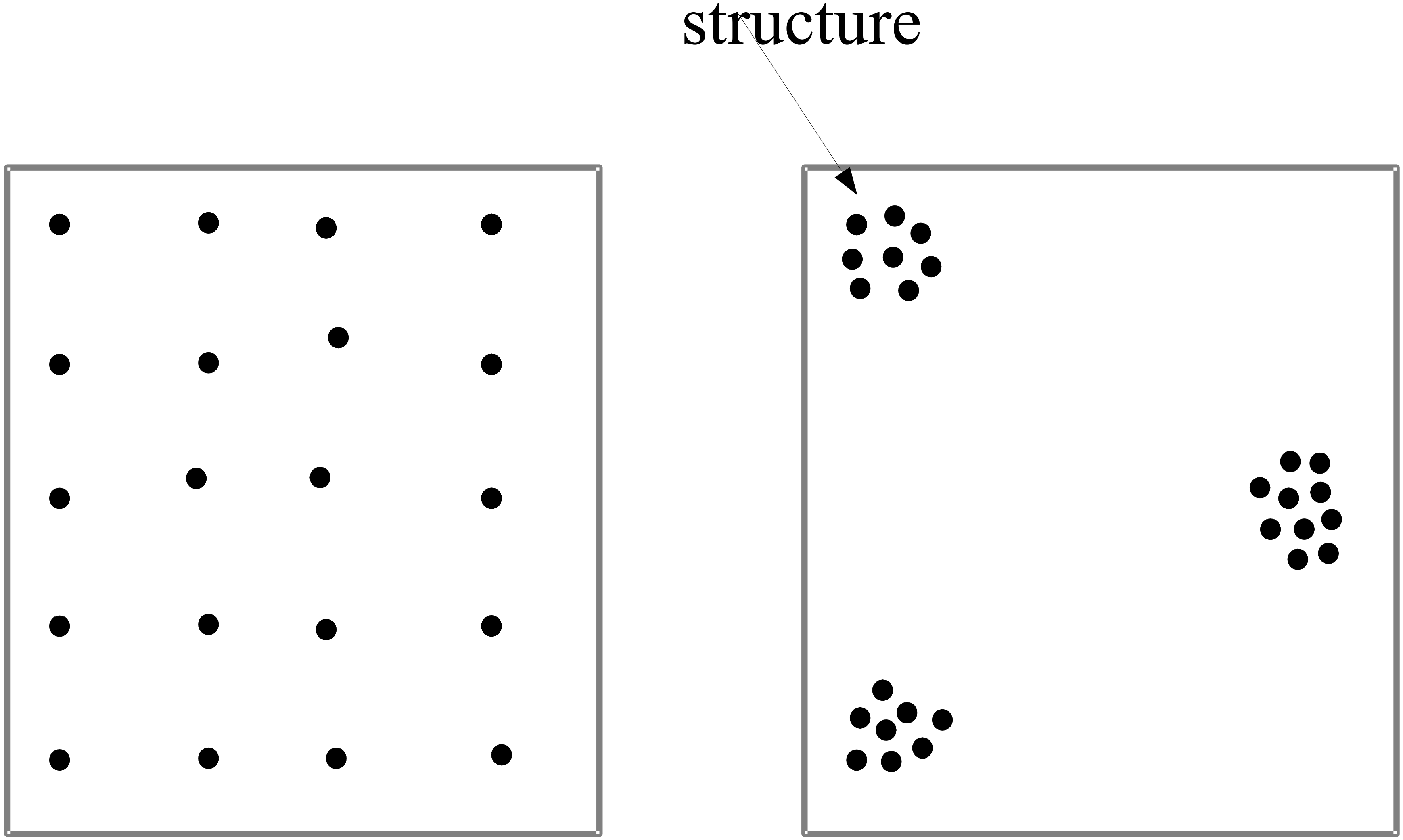
Specific

Image removed due to copyright considerations.

General

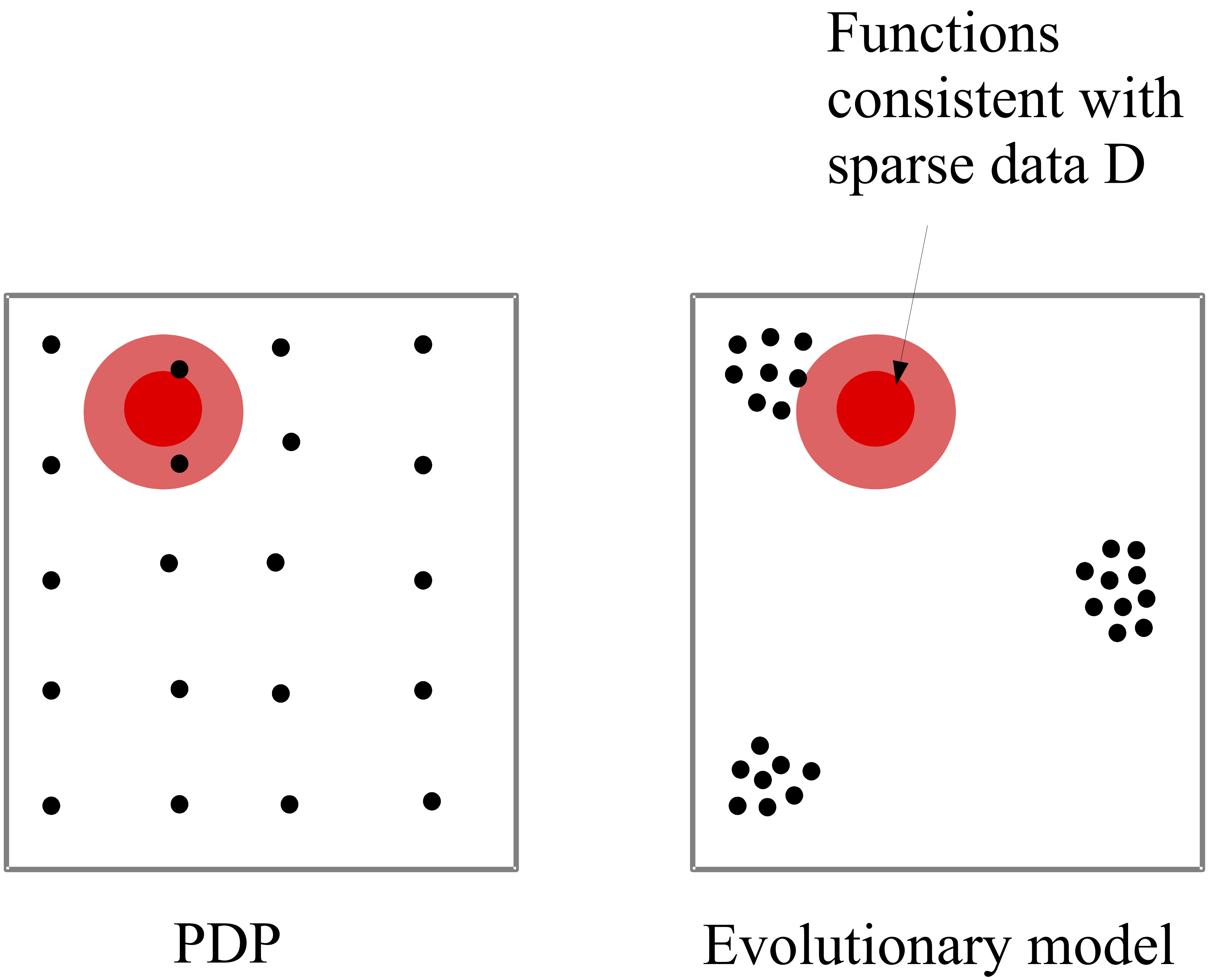

The space of smooth functions
Functions the PDP model is happy to learn
PDP

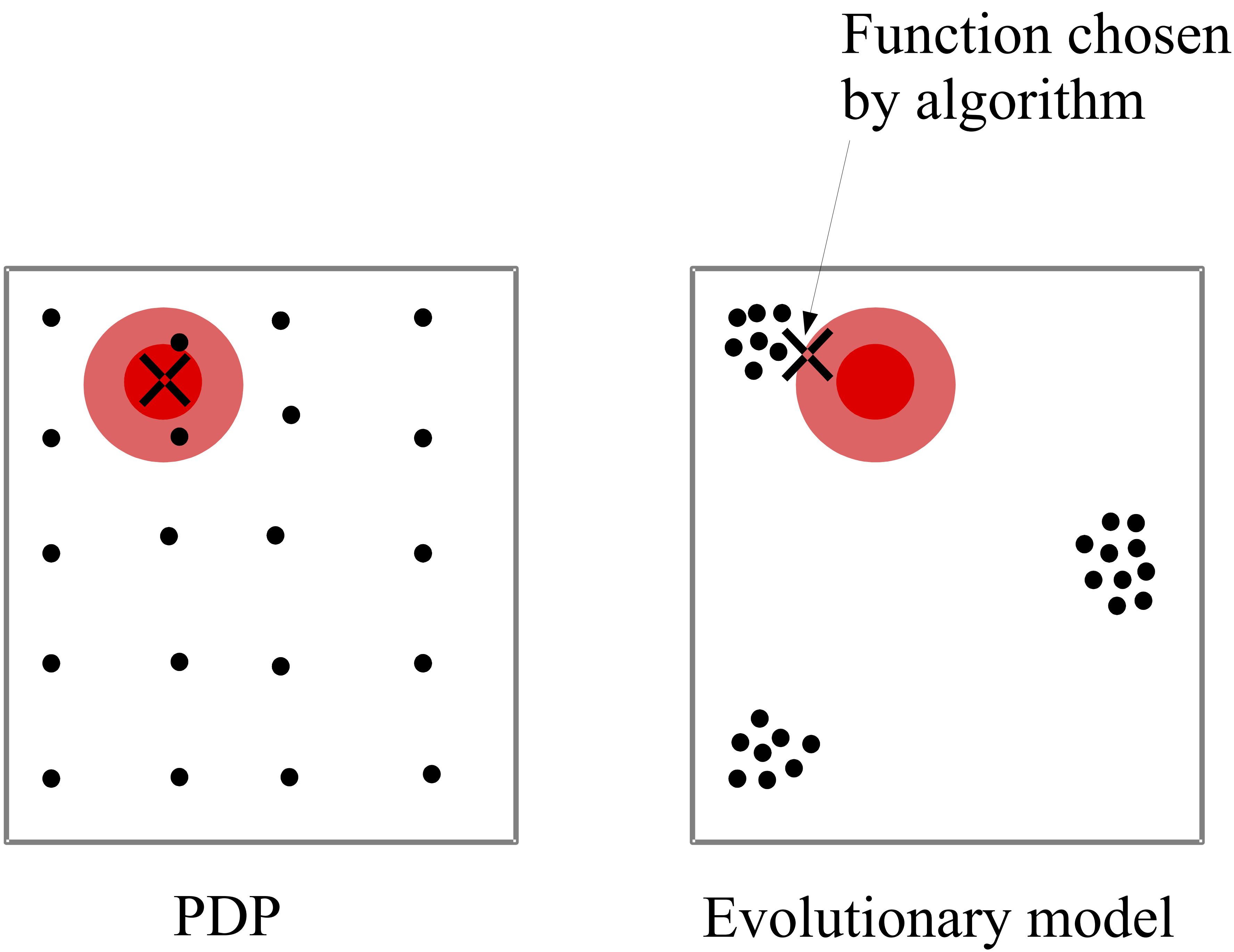
Functions consistent with a specific tree structure



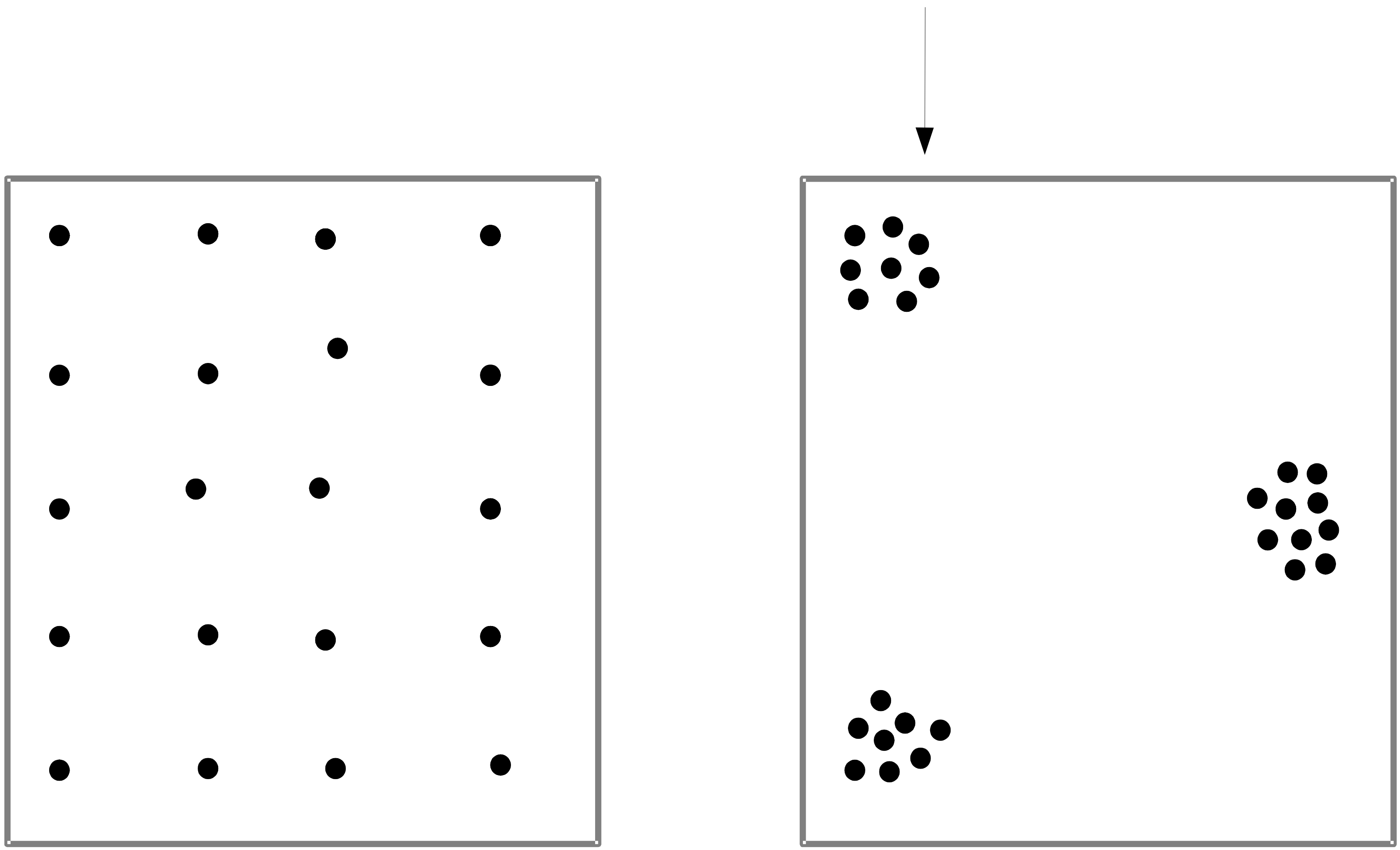
PDP

Evolutionary model


Functions consistent with sparse data D
PDP
Evolutionary model


Function chosen
by algorithm
PDP
Evolutionary model

Structured model is better than PDP approach if the distribution of true functions looks like this



Some connectionists agree

- Geman, 1992
  - “.. strong a priori representations are unavoidable”
  - “the paradigm of near *tabula rasa* learning, which has been so much emphasized in the neural computing literature of the last decade, may be of relatively minor biological importance”

Inductive bias

- The prior matters:
  - a prior that matches the world does better than a prior that doesn't
  - a tree-based prior is good for biological induction because the world is actually structured that way
  - generic smoothness priors are not sufficient!
- Where does the prior come from?
  - How do people know to use a tree representation for biology?

Innate Biological Knowledge ?

Atran, 1998

'Universal Taxonomy' is a core module -- an 'innately determined cognitive structure'

Keil:

“Those who argue for the importance of constraints almost invariably share the assumption that there are domain-specific or autonomous cognitive subsystems”

Structure Learning

structure grammar G

↓

structure S

↓

object-property matrix D

A domain-general framework for learning structured, domain-specific representations

Telluric Screw

- Beguyer de Chancourtois, 1862

Image removed due to copyright considerations.

Chemistry

Benfey's periodic spiral, 1960.

Image removed due to copyright considerations.

(Benfey, 1960)

Structure discovery and induction

- "One can predict the discovery of many new elements, for example analogues of Si and Al with atomic weights of 65-75."

- "A few atomic weights will probably require correction; for example Te cannot have the atomic weight 128, but rather 123-126."

(Mendeleev)

Structure Discovery

- Cultures all over the world group animals into hierarchies
- Children learn the properties of the integers
- Primates discover dominance hierarchies
- Time is cyclic on many levels (days, seasons)
- Children learn kinship systems

Structure Learning

structure grammar G

↓

structure S

↓

object-property matrix D

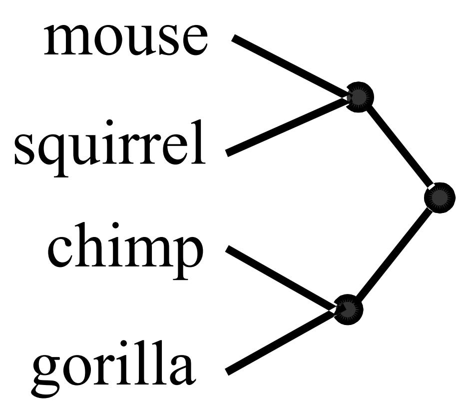
	tail	hands	smart
mouse	●	○	○
squirrel	●	○	○
chimp	○	●	●
gorilla	○	●	●

Structure Learning

structure grammar **G**

↓

structure **S**



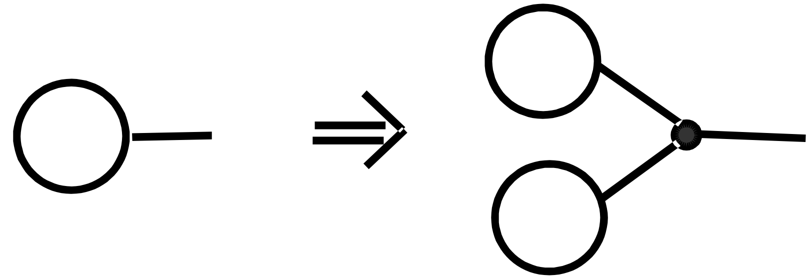
↓

object-property matrix **D**

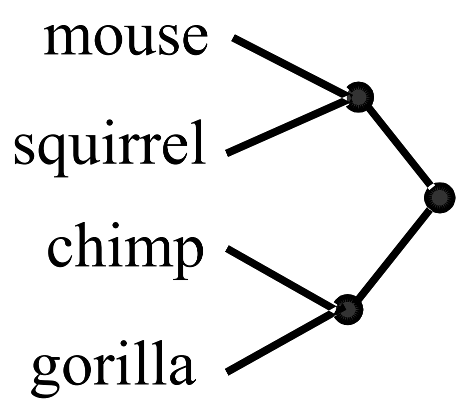
	tail	hands	smart
mouse	●	○	○
squirrel	●	○	○
chimp	○	●	●
gorilla	○	●	●

Structure Learning

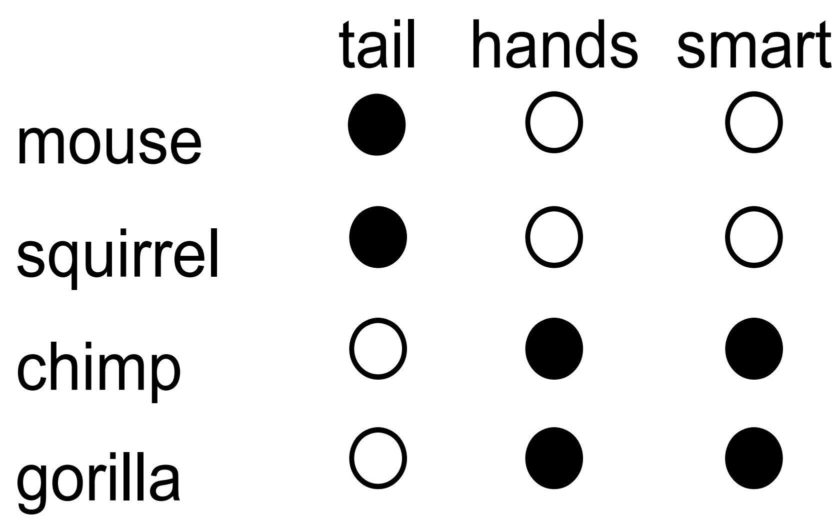
structure grammar **G**



structure **S**



object-property matrix **D**



Structure Learning

structure grammar — G

↓

structure — S

↓

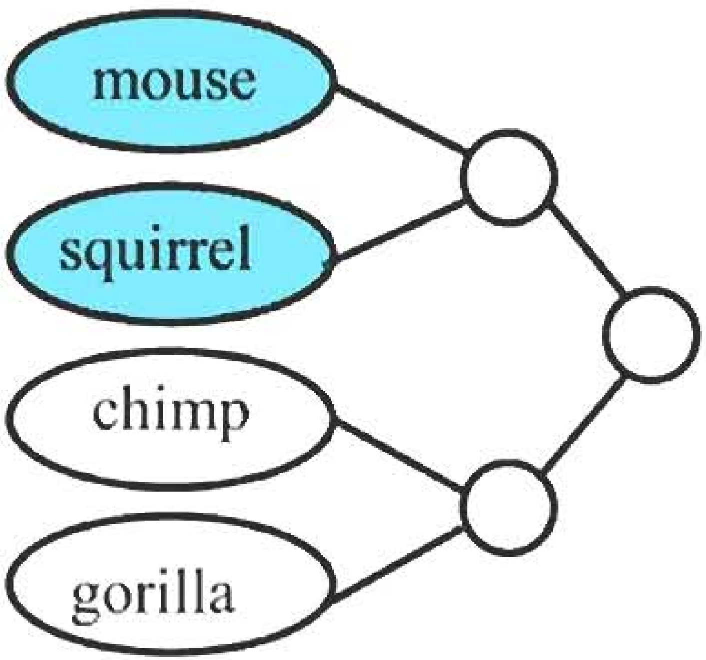
object-property matrix — D

Given D, choose S and G that maximize :

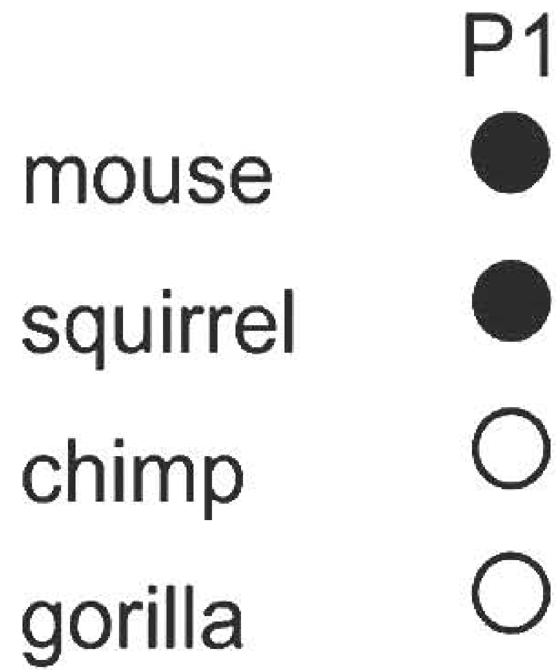
$$p(S, G|D) \alpha p(D|S)p(S|G)p(G)$$

p(D|S): Generating properties

S:

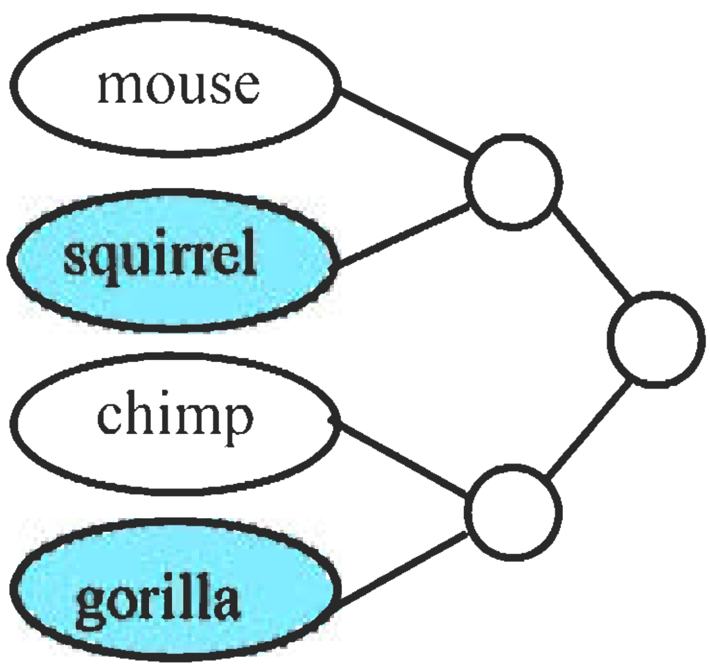


D:



p(D|S): Generating properties

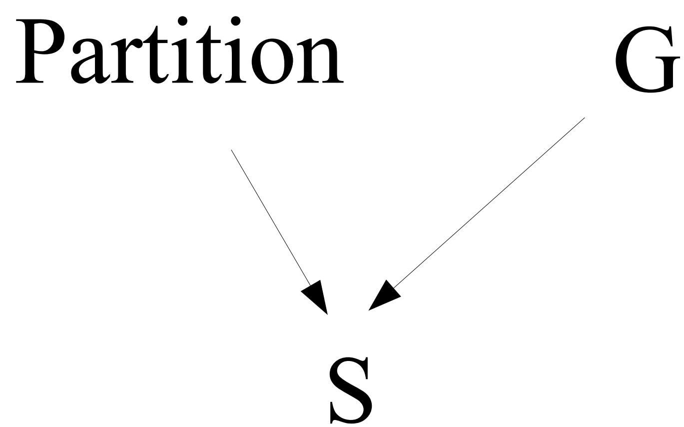
S:



D:

	P1	P2
mouse	●	○
squirrel	●	●
chimp	○	○
gorilla	○	●

p(S|G): Generating structures



p(G): Generating structure grammars

Image removed due to copyright considerations.

Characterizing the space of structures

- Grammars with multiple productions
- Probabilistic productions
- Ways of combining structures
  (eg cartesian product)

Structure Learning

structure grammar G

↓

structure S

↓

object-property matrix D

Given D, choose S and G that maximize :

$$p(S, G|D) \; \alpha \; p(D|S)p(S|G)p(G)$$

Biological Data

- 50 mammals, 85 properties

	Tail	Hands	Smart
Mouse	●	○	○
Squirrel	●	○	○
Chimp	○	●	●
Gorilla	○	●	●

(Osherson, Stern, Wilkie, Stob & Smith, 1992)

Supreme Court Data

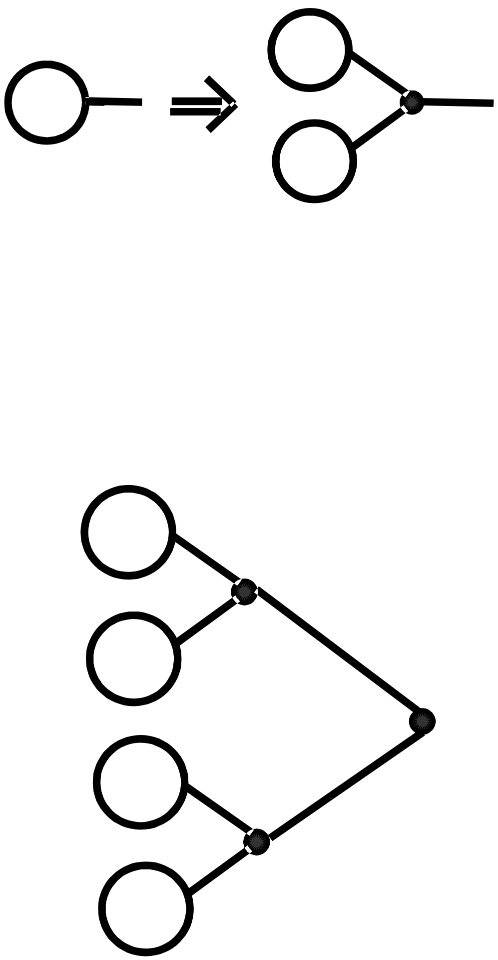
- Judgments from 1981 to 1985
- 9 judges
- 637 cases

Images removed due to copyright considerations.

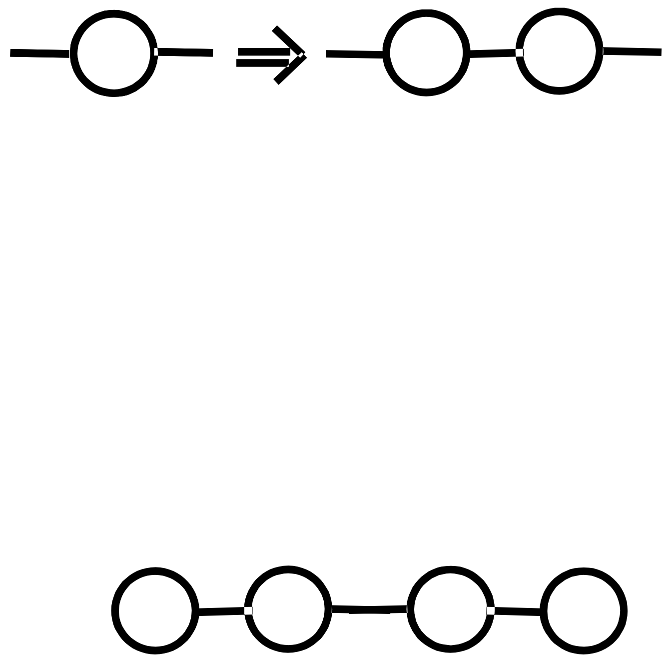
(Spaeth, 1998)

Three Grammars

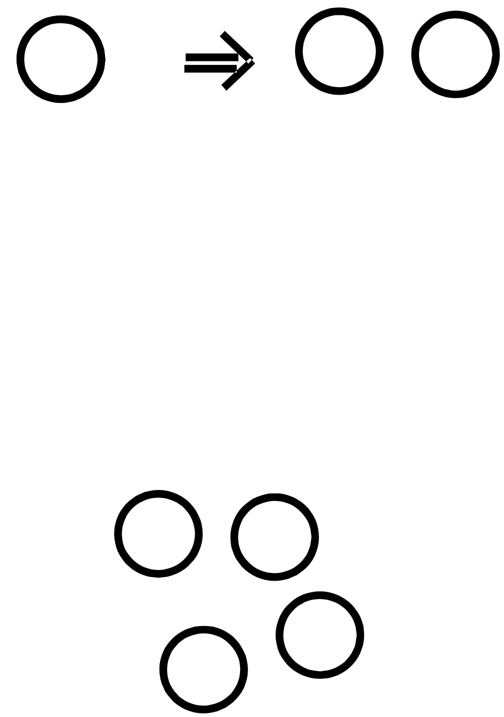
G_{tree}:



G_{linear}:



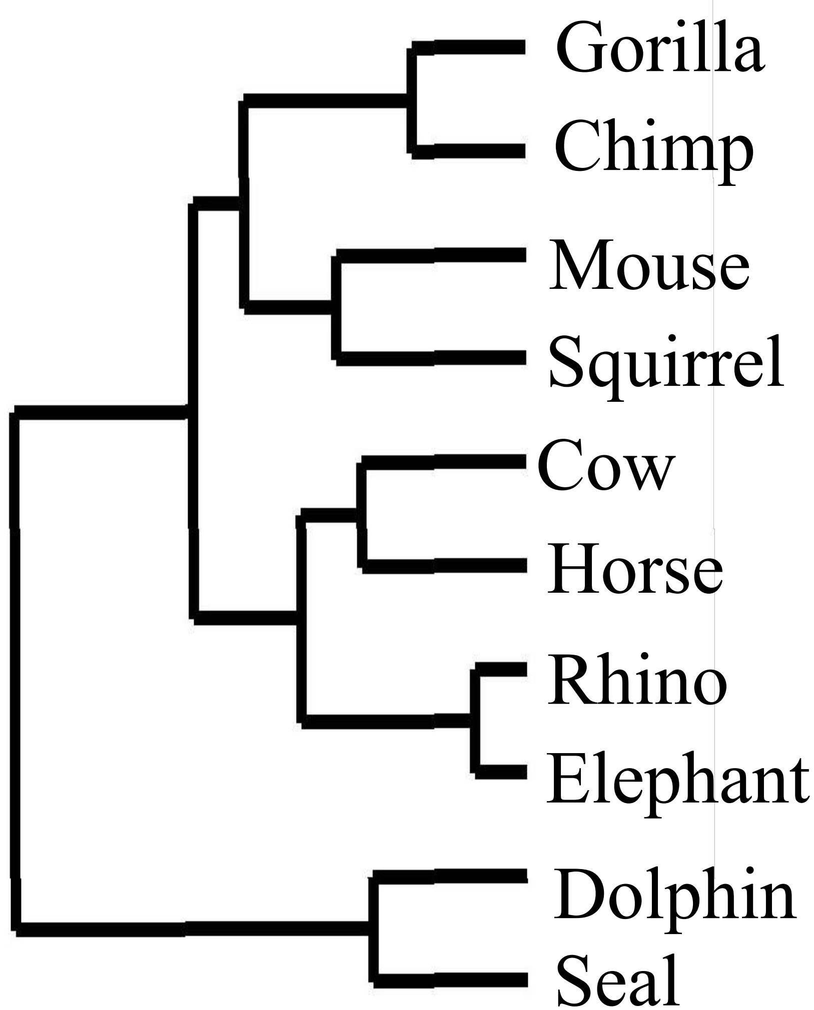
$G_{disconnected}$:



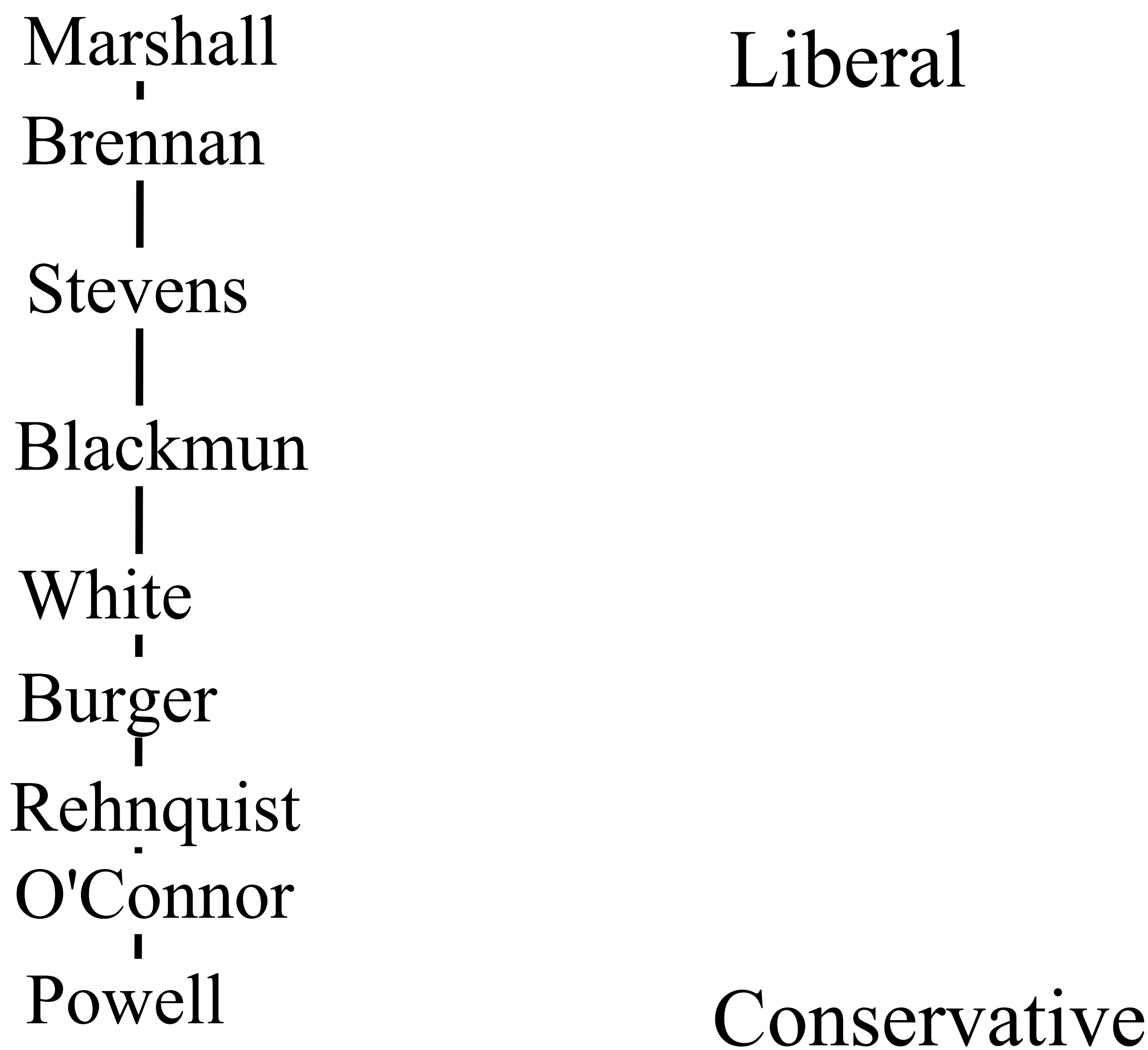
$$\log p(G|D_{obs})$$

Data	G_{tree}	G_{linear}	$G_{disconnected}$
Biology	339	230	0
Supreme Court	883	1312	0
Scrambled Biology	0	74	138

Best Structure: Biological Data



Best Structure: Supreme Court Data



Why learn structural constraints?

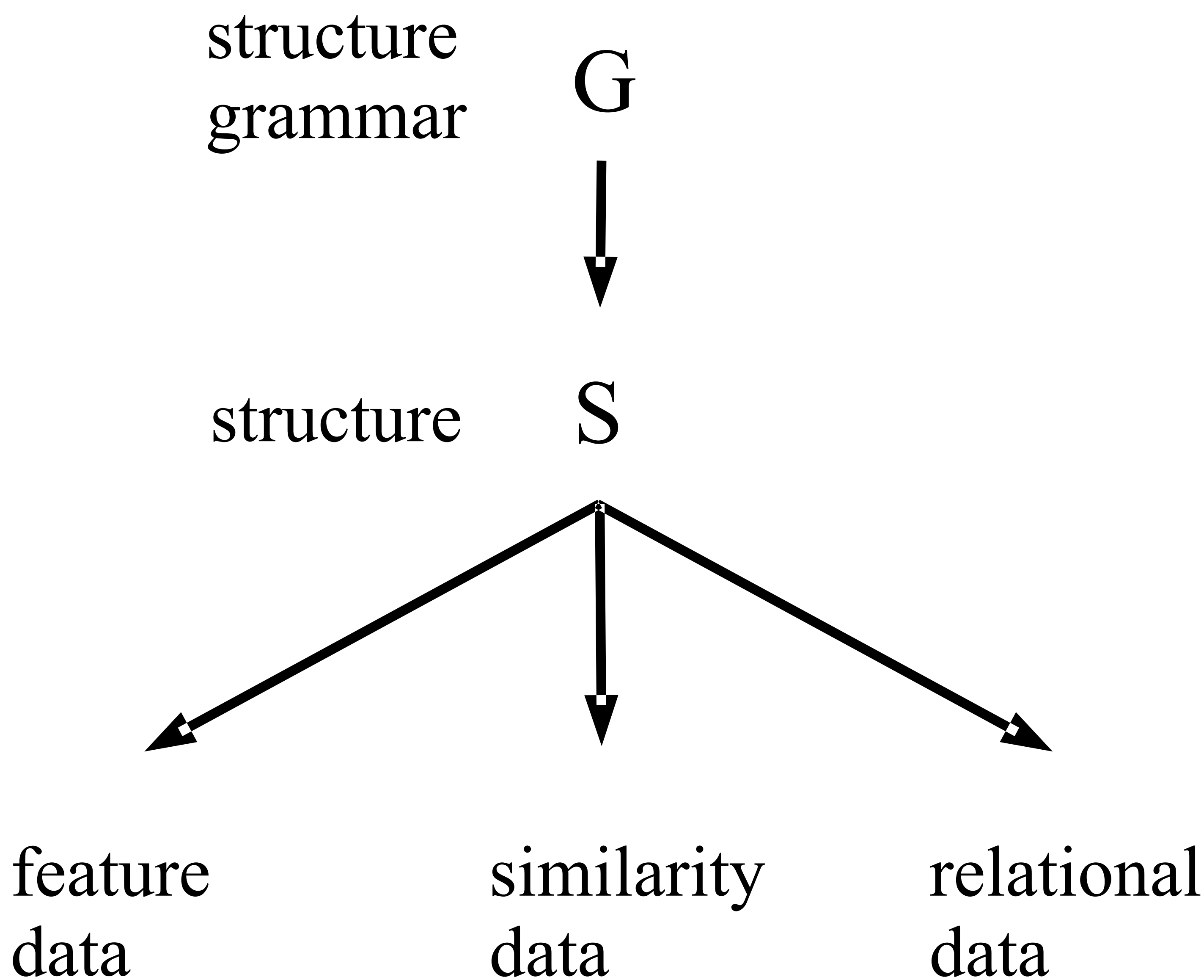
- More explanatory than assuming innate, domain-specific constraints
- Allows structure-discovery in novel domains
- Allows developmental shifts within a single domain

- Keil: innate, domain-specific structural constraints

Image removed due to copyright considerations.

Developmental Shift

Image removed due to copyright considerations.


structure
grammar
G
structure
S
feature
data
similarity
data
relational
data

Similarity Data

Image removed due to copyright considerations.

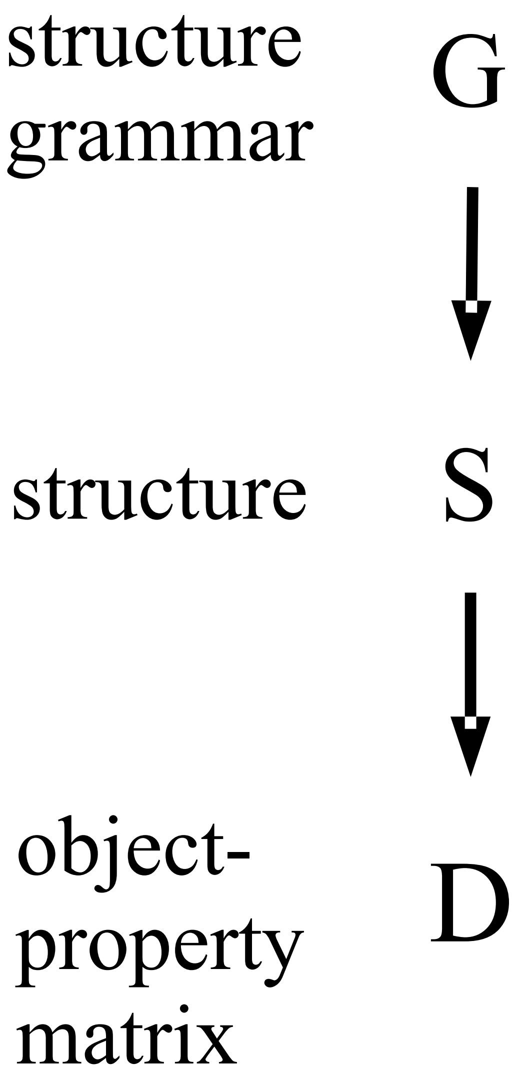
Image removed due to copyright considerations.

Relational Data

Image removed due to copyright considerations.

Relational Data

Images removed due to copyright considerations.



Why learn structure grammars?

- Allow representations to grow as new objects are encountered
- Transfer across related sub-domains

Why learn structures?

- Structured representations provide an inductive bias that matches a structured world

Knowledge Transfer

Image removed due to copyright considerations.

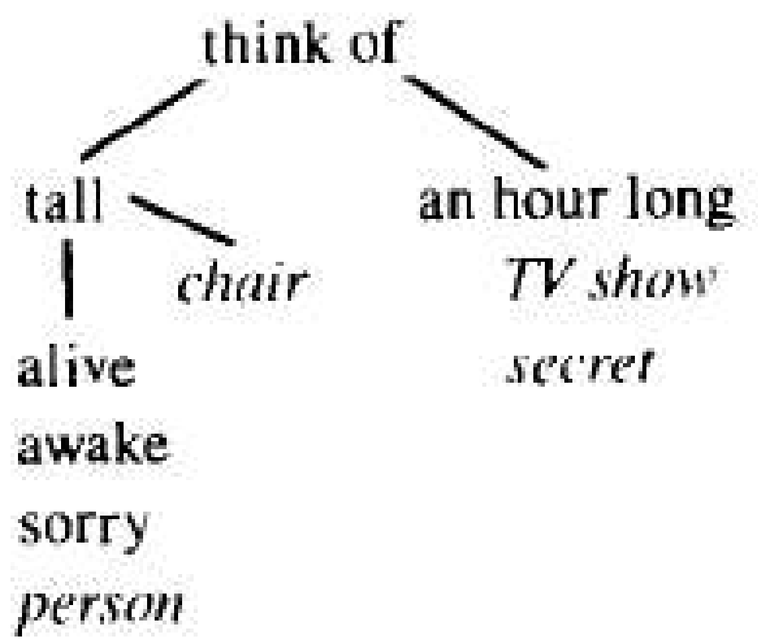
Sir Joseph Banks

Issues

- Cultural transmission is often important
- Even in cases where cultural transmission is vital there's still something to explain. Consider a child learning the properties of the natural numbers.

Issues

- Can we learn the constraints discussed by Keil?



Keil, Frank C. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: The MIT Press, 1989. (Courtesy of MIT press. Used with permission.)