The development of structured representations

Charles Kemp

July 25, IPAM Summer School

Acknowledgements

Josh Tenenbaum
Pooja Jotwani

Logical theories

Hierarchies

Euclidean spaces

Semantic networks

Context-free Grammars

Causal networks

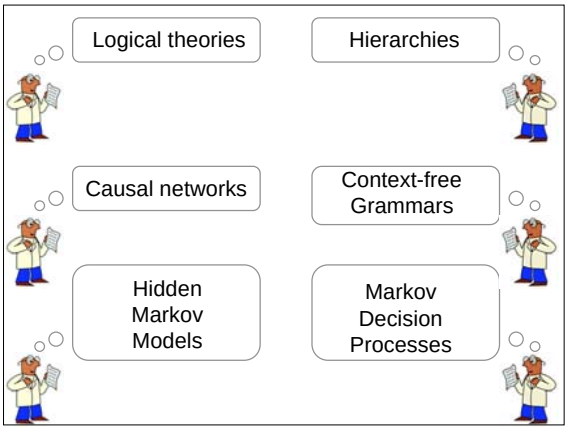
Feature vectors

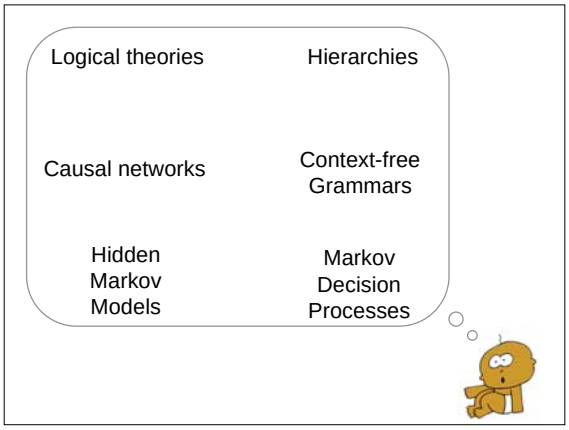
Lambda calculus

Hidden Markov Models

Markov Decision Processes

Relational schemata





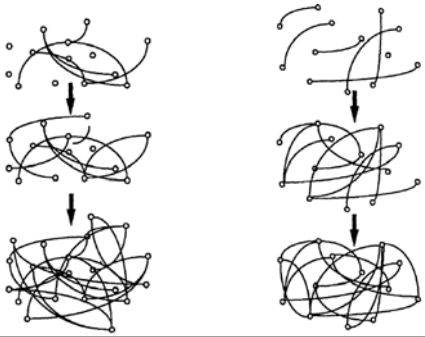
Hypotheses about representations

1. There is one kind of representation that will handle every domain.

Developmental Picture

Domain A

Domain B

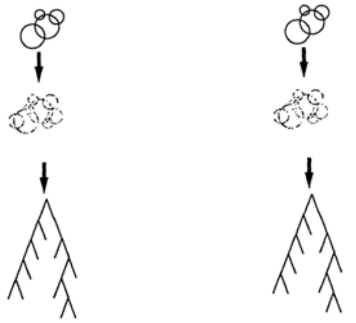


(Keil 91)

Developmental Picture

Domain A

Domain B



(Keil 91)

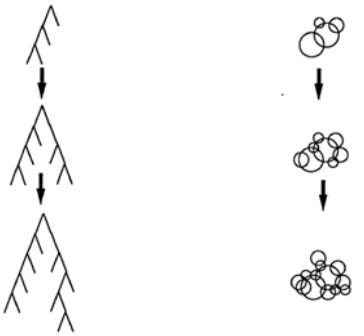
Hypotheses about representations

1. There is one kind of representation that will handle every domain.
2. Children begin with innate, domain-specific representational constraints.

Developmental Picture

Domain A

Domain B



(Keil 91)

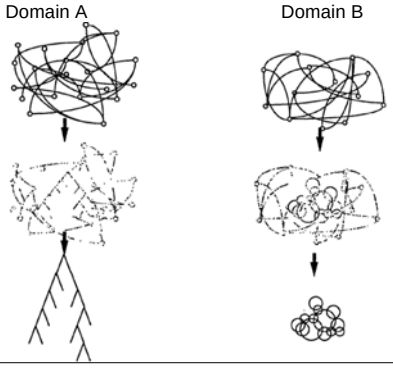
Hypotheses about representations

1. There is one kind of representation that will handle every domain.
2. Children begin with innate, domain-specific representational constraints.
3. Children discover which kind of representation is best for each domain.

Developmental Picture

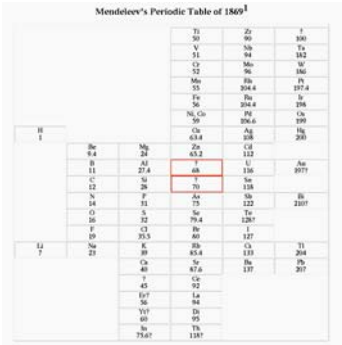
Domain A

Domain B



(Keil 91)

Scientists discover structural form

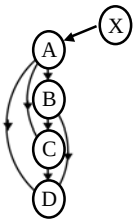


Children discover structural form

- Children may discover that
  - Social networks are often organized into cliques
  - The months form a cycle
  - "Heavier than" is transitive
  - Category labels can be organized into hierarchies

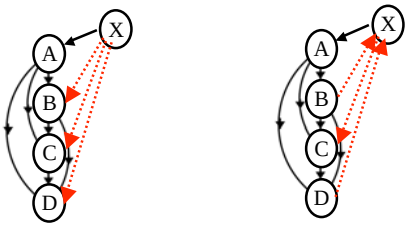
Why form discovery matters

- Structural forms provide inductive constraints

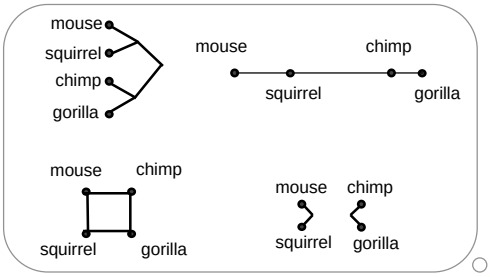


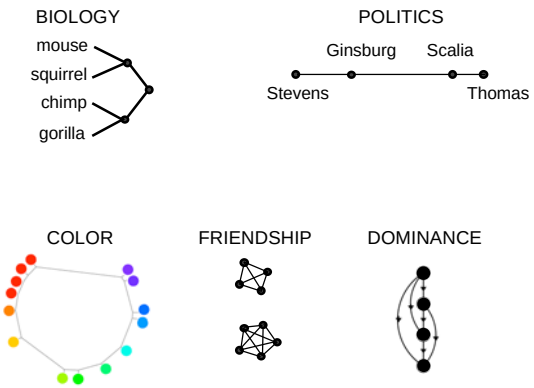
Why form discovery matters

- Structural forms provide inductive constraints



This talk: graph structures





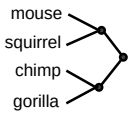
Outline

- Discovery of structural form
  - Feature data
  - Similarity
  - Relational data
- Form discovery in the lab

A hierarchical Bayesian framework

F: form → S: structure → D: data

Tree

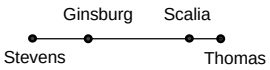


	whiskers	hands	tail	warm blooded
mouse				
squirrel				?
chimp				?
gorilla				?

A hierarchical Bayesian framework

F: form → S: structure → D: data

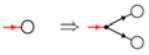
Chain

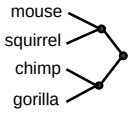


	case 1	case 2	case 3	case n
Stevens				
Ginsburg				?
Scalia				?
Thomas				?

A hierarchical Bayesian framework

F: form → S: structure → D: data





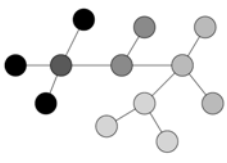
	whiskers	hands	tail
mouse			
squirrel			
chimp			
gorilla			

$$P(S, F|D, n) \propto P(D|S)P(S|F, n)P(F)$$

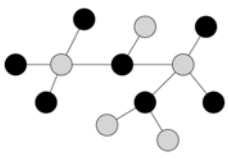
p(D|S): Generating feature data

- Intuition: features should be smooth over graph S

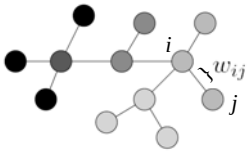
Relatively smooth



Not smooth



p(f|W σ): Generating a single feature



Let f_i be the feature value at node *i*

$$p(f) \propto \exp\left(-\frac{1}{4}\sum_{i,j}\frac{(f_i - f_j)^2}{w_{ij}} - \frac{1}{2\sigma}f^\top f\right)$$

$$= \exp(-\frac{1}{2}f^\top \Sigma^{-1} f)$$

(Zhu, Lafferty, Ghahramani 03)

p(D|S,W,σ): Generating feature data

- The log likelihood for the feature model is

$$\log(p(D|W,\sigma)) = -\frac{mn}{2}\log(2\pi) - \frac{m}{2}\log|\Sigma| - \frac{1}{2}\mathbf{tr}(\Sigma^{-1}DD^{\mathsf{T}})$$

where

– D is a matrix of objects (n) by features (m)

– W is a weighted graph

– σ specifies a prior on the variance of each feature value

p(D|S): Generating feature data

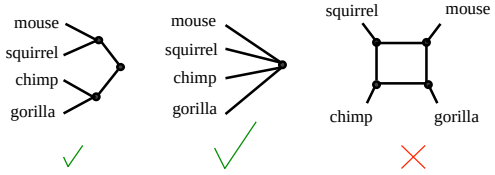
- Generating a weighted graph

$$w_{ij}|S,\beta \sim \text{Exponential}(\beta) \text{ if } s_{ij} = 1$$

- We integrate out W and σ using the Laplace approximation

$$p(D|S,\beta) = \int p(D|W,\sigma)p(W|S,\beta)p(\sigma)dW\,d\sigma$$

P(S|F,n): Generating structures



• Each structure is weighted by the number of nodes it contains:

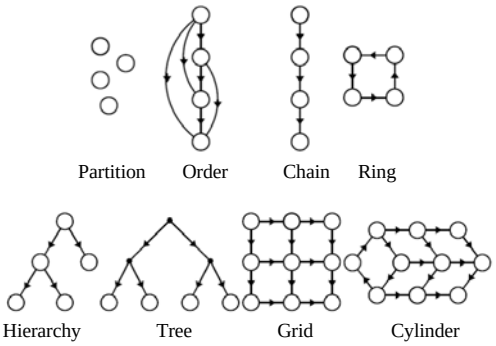
$$P(S|F) \propto \begin{cases} 0 & \text{if S inconsistent with F} \\ \theta(1-\theta)^{|S|} & \text{otherwise} \end{cases}$$

where $|S|$ is the number of nodes in *S*

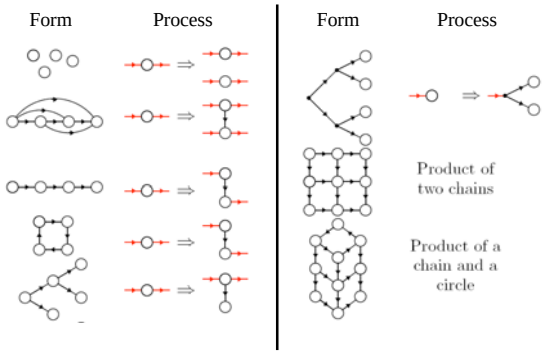
P(S|F,n): Generating structures

- Simpler forms generate fewer structures, and are therefore preferred by the prior

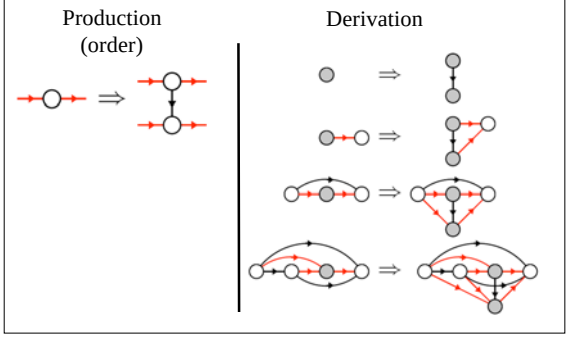
P(F): Structural forms



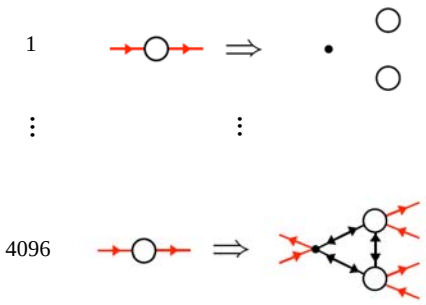
Generating Graphs



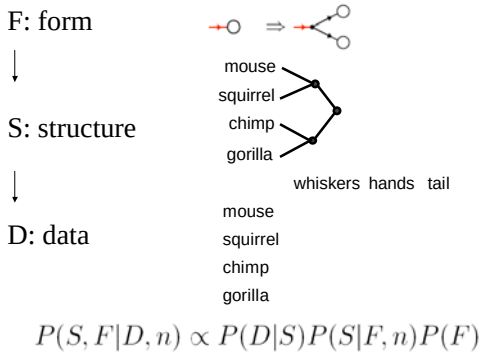
Node-replacement graph grammars



The complete space of grammars



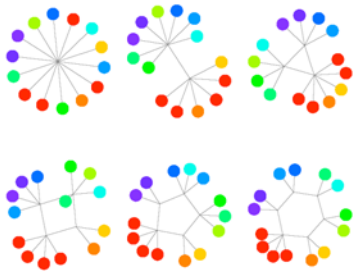
A hierarchical Bayesian framework



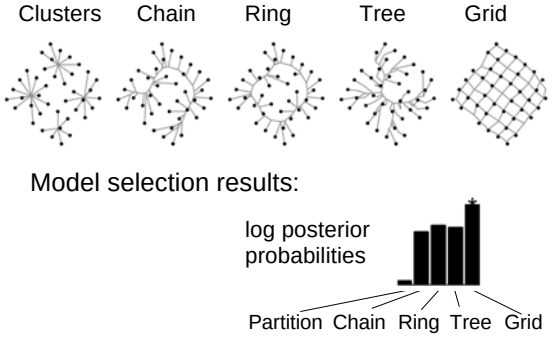
$$P(S, F|D, n) \propto P(D|S)P(S|F, n)P(F)$$

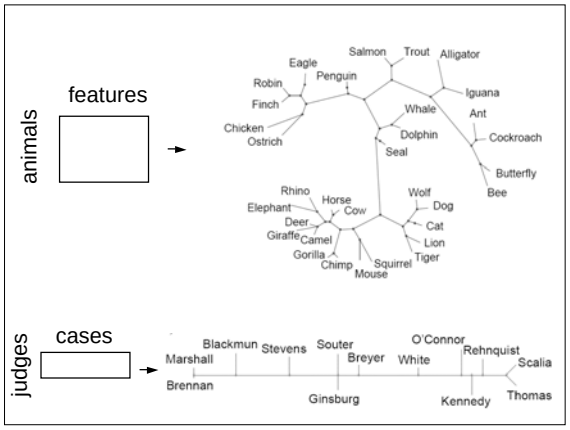
Model fitting

- Run greedy, grammar-based searches for each form in parallel



Synthetic data: Grid





Similarity data

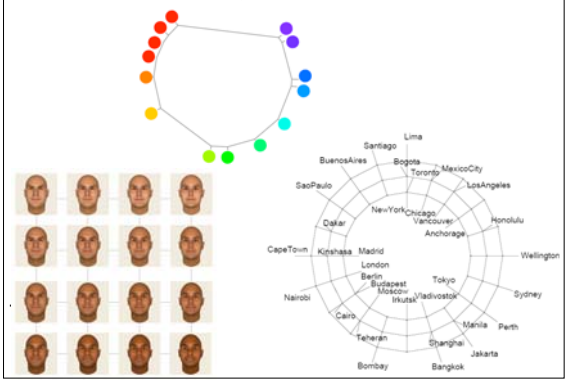
- The log likelihood for the feature model is

$$\log(p(D|W,\sigma)) = -\frac{mn}{2}\log(2\pi) - \frac{m}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}DD^{\mathsf{T}})$$

where D is a matrix of objects (n) by features (m)

- The kernel trick: replacing DD^{T} with a similarity matrix lets us learn structural forms from similarity data

Similarity data: results

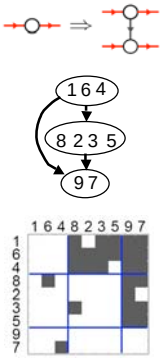


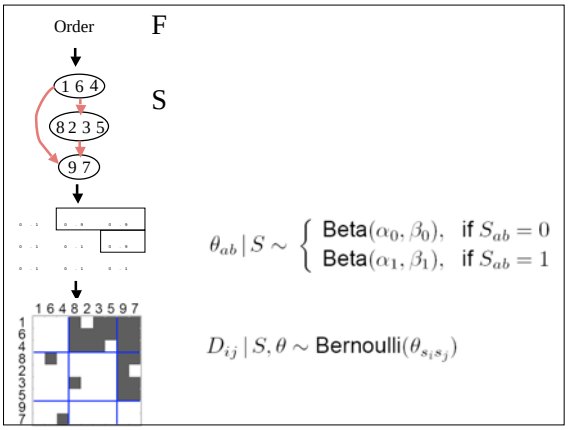
Relational Data

F: form

S: structure

D: data





$$\theta_{ab}\,|\,S \sim \begin{cases} \text{Beta}(\alpha_0,\beta_0), & \text{if } S_{ab}=0 \\ \text{Beta}(\alpha_1,\beta_1), & \text{if } S_{ab}=1 \end{cases}$$

$$D_{ij}\,|\,S,\theta \sim \text{Bernoulli}(\theta_{s_i s_j})$$

P(D|S): Relational data

$$P(D|S) = \sum_{(\alpha_0,\beta_0,\alpha_1,\beta_1)} P(D|S,\alpha_0,\beta_0,\alpha_1,\beta_1)P(\alpha_0,\beta_0,\alpha_1,\beta_1)$$

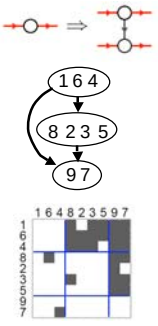
- The hyperparameters are drawn from a 4D grid where
  - $\alpha_0+\beta_0$ and $\alpha_1+\beta_1$ belong to $\{\frac{1}{16},\frac{1}{8},\frac{1}{4},\frac{1}{2},1,2,4,8,16,32\}$
  - $\frac{\beta_0}{\alpha_0+\beta_0}$ and $\frac{\beta_1}{\alpha_1+\beta_1}$ belong to $\{0.05,0.15,\ldots,0.95\}$
  - $\frac{\beta_0}{\alpha_0+\beta_0} \leq \frac{\beta_1}{\alpha_1+\beta_1}$

A hierarchical Bayesian framework

F: form

S: structure

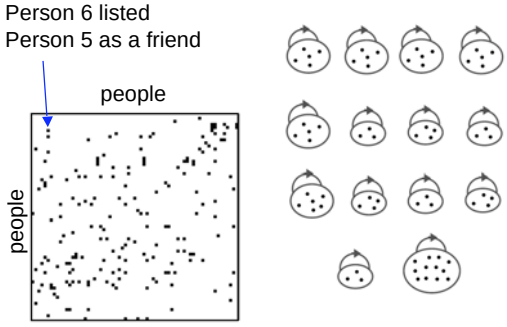
D: data



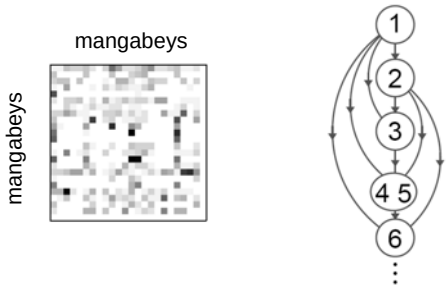
$$P(S,F|D,n) \propto P(D|S)P(S|F,n)P(F)$$

Friendship (MacRae & Gagnon)

Person 6 listed
Person 5 as a friend

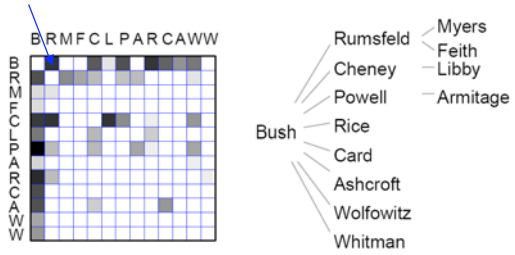


Sooty Mangabeys (Range and Noë)



Bush Cabinet

"Bush told Rumsfeld"

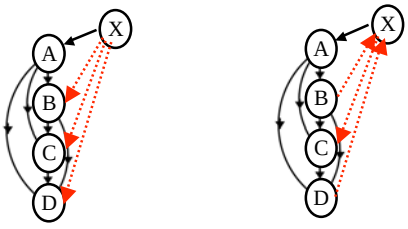


Outline

- Discovery of structural form
  - Feature data
  - Similarity
  - Relational data
- Form discovery in the lab

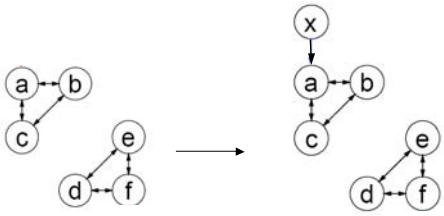
Why form discovery matters

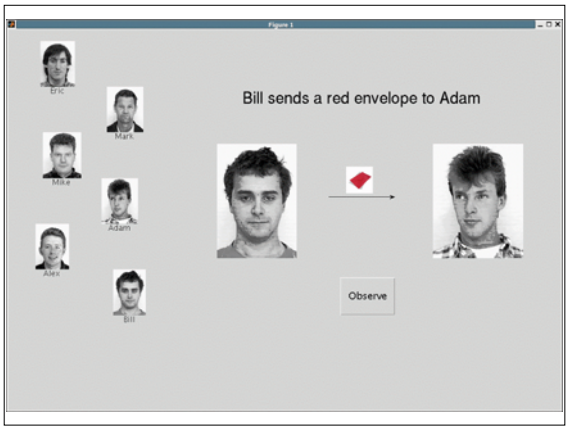
- Structural forms provide inductive constraints



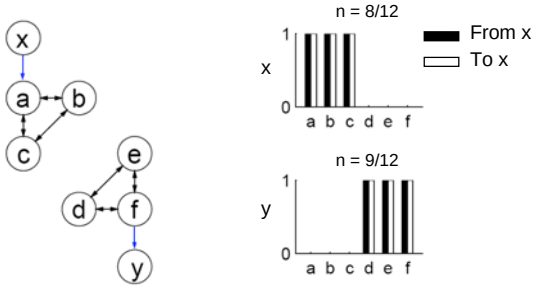
Experiment 1: Form discovery

- Structural forms support predictions about new or sparsely-observed entities.

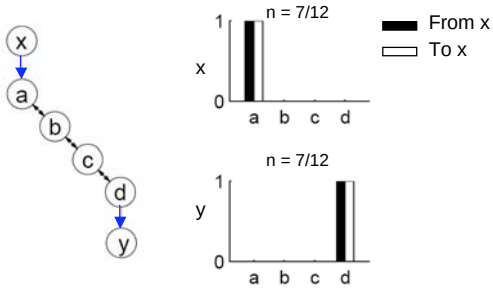




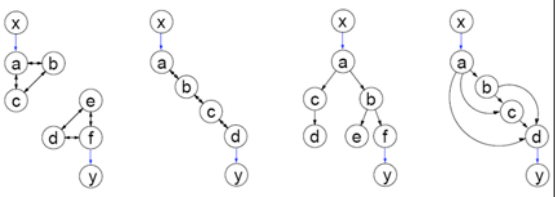
Experiment 1: Form discovery



Experiment 1: Form discovery



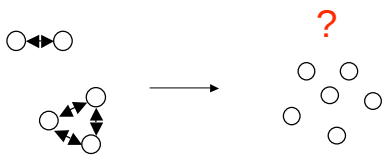
Experiment 1: Form discovery



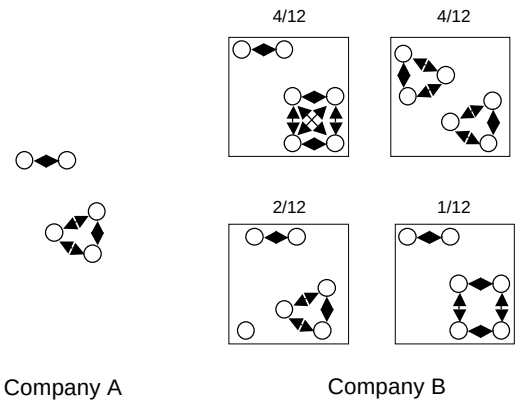
Cliques (8/12) | Chain (8/12) | Hierarchy (8/12) | Order (5/12)

Experiment 2: Form discovery

- Structural forms support predictions about novel systems of entities.

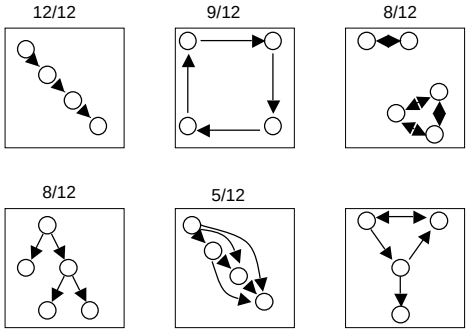


Company A | Company B

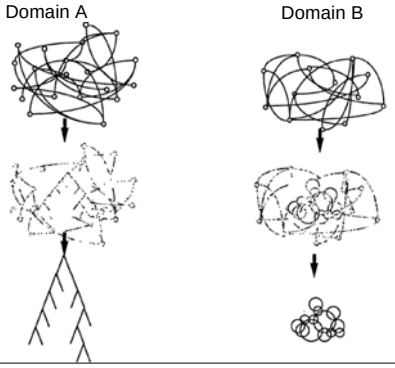


Company A | Company B

Experiment 2: Form discovery

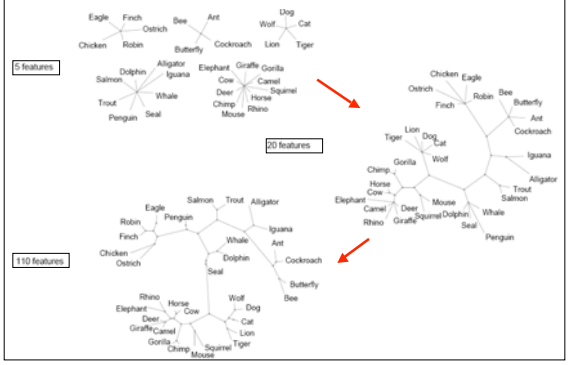


Developmental Picture



(Keil 91)

Developmental predictions



Issues and questions

- How can we work with richer collections of structure grammars?
- Where do these structure grammars come from?
- What about representations other than graphs?

Conclusions

- Hierarchical models can help to explain:
  - how people acquire mental representations
  - How people learn what *kind* of representation is best for a domain