# The discovery of structural form

**Charles Kemp*† and Joshua B. Tenenbaum‡**

*Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; and ‡Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved May 30, 2008 (received for review March 17, 2008)

**Algorithms for finding structure in data have become increasingly important both as tools for scientific data analysis and as models of human learning, yet they suffer from a critical limitation. Scientists discover qualitatively new forms of structure in observed data: For instance, Linnaeus recognized the hierarchical organization of biological species, and Mendeleev recognized the periodic structure of the chemical elements. Analogous insights play a pivotal role in cognitive development: Children discover that object category labels can be organized into hierarchies, friendship networks are organized into cliques, and comparative relations (e.g., ''bigger than'' or ''better than'') respect a transitive order. Standard algorithms, however, can only learn structures of a single form that must be specified in advance: For instance, algorithms for hierarchical clustering create tree structures, whereas algorithms for dimensionality-reduction create low-dimensional spaces. Here, we present a computational model that learns structures of many different forms and that discovers which form is best for a given dataset. The model makes probabilistic inferences over a space of graph grammars representing trees, linear orders, multidimensional spaces, rings, dominance hierarchies, cliques, and other forms and successfully discovers the underlying structure of a variety of physical, biological, and social domains. Our approach brings structure learning methods closer to human abilities and may lead to a deeper computational understanding of cognitive development.**

cognitive development | structure discovery | unsupervised learning

Discovering the underlying structure of a set of entities is a fundamental challenge for scientists and children alike (1–7). Scientists may attempt to understand relationships between biological species or chemical elements, and children may attempt to understand relationships between category labels or the individuals in their social landscape, but both must solve problems at two distinct levels. The higher-level problem is to discover the form of the underlying structure. The entities may be organized into a tree, a ring, a dimensional order, a set of clusters, or some other kind of configuration, and a learner must infer which of these forms is best. Given a commitment to one of these structural forms, the lower-level problem is to identify the instance of this form that best explains the available data.

The lower-level problem is routinely confronted in science and cognitive development. Biologists have long agreed that tree structures are useful for organizing living kinds but continue to debate which tree is best—for instance, are crocodiles better grouped with lizards and snakes or with birds (8)? Similar issues arise when children attempt to fit a new acquaintance into a set of social cliques or to place a novel word in an intuitive hierarchy of category labels. Inferences like these can be captured by standard structure-learning algorithms, which search for structures of a single form that is assumed to be known in advance (Fig. 1*A*). Clustering or competitive-learning algorithms (9, 10) search for a partition of the data into disjoint groups, algorithms for hierarchical clustering (11) or phylogenetic reconstruction (12) search for a tree structure, and algorithms for dimensionality reduction (13, 14) or multidimensional scaling (15) search for a spatial representation of the data.

Higher-level discoveries about structural form are rarer but more fundamental, and often occur at pivotal moments in the development of a scientific field or a child's understanding (1, 2, 4). For centuries, the natural representation for biological species was held to be the ''great chain of being,'' a linear structure in which every living thing found a place according to its degree of perfection (16). In 1735, Linnaeus famously proposed that relationships between plant and animal species are best captured by a tree structure, setting the agenda for all biological classification since. Modern chemistry also began with a discovery about structural form, the discovery that the elements have a periodic structure. Analogous discoveries are made by children, who learn, for example, that social networks are often organized into cliques, that temporal categories such as the seasons and the days of the week can be arranged into cycles, that comparative relations such as ''longer than'' or ''better than'' are transitive (17, 18) and that category labels can be organized into hierarchies (19). Structural forms for some cognitive domains may be known innately, but many appear to be genuine discoveries. When learning the meanings of words, children initially seem to organize objects into nonoverlapping clusters, with one category label allowed per cluster (20); hierarchies of category labels are recognized only later (19). When reasoning about comparative relations, children's inferences respect a transitive ordering by the age of 7 but not before (21). In both of these cases, structural forms appear to be learned, but children are not explicitly taught to organize these domains into hierarchies or dimensional orders.

Here, we show that discoveries about structural form can be understood computationally as probabilistic inferences about the organizing principles of a dataset. Unlike most structure-learning algorithms (Fig. 1*A*), the model we present can simultaneously discover the structural form and the instance of that form that best explain the data (Fig. 1*B*). Our approach can handle many kinds of data, including attributes, relations, and measures of similarity, and we show that it successfully discovers the structural forms of a diverse set of real-world domains.

Any model of form discovery must specify the space of structural forms it is able to discover. We represent structures using graphs and use graph grammars (22) as a unifying language for expressing a wide range of structural forms (Fig. 2). Of the many possible forms, we assume that the most natural are those that can be derived from simple generative processes (23). Each of the first six forms in Fig. 2*A* can be generated by using a single context-free production that replaces a parent node with two child nodes and specifies how to connect the children to each other and to the neighbors of

Author contributions: C.K. and J.B.T. designed research; C.K. performed research; and C.K. and J.B.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence should be addressed. E-mail: ckemp@cmu.edu.

Freely available online through the PNAS open access option.

See Commentary on page 10637.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802631105/DCSupplemental.

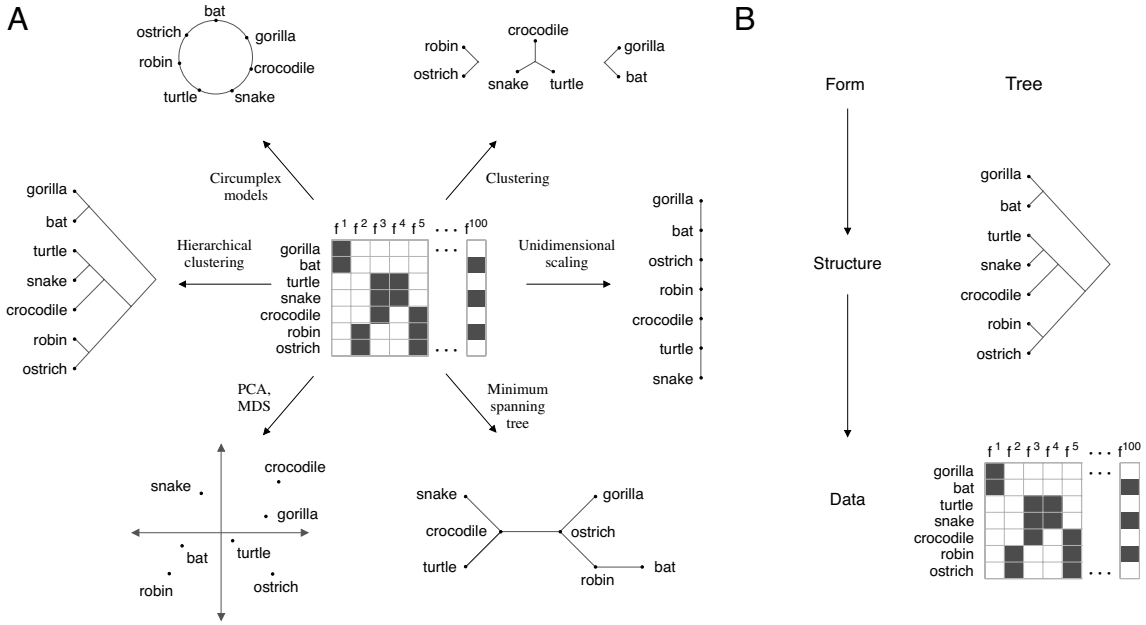


**Fig. 1.** Finding structure in data. (*A*) Standard structure learning algorithms search for representations of a single form that is fixed in advance. Shown here are methods that discover six different kinds of structures given a matrix of binary features. (*B*) A hierarchical model that discovers the form *F* and the structure *S* that best account for the data *D*. The model searches for the form and structure that jointly maximize $P(S, F|D) \propto P(D|S)P(S|F)P(F)$.

the parent node. Fig. 2 *B–D* shows how three of these productions generate chains, orders, and trees. More complex forms, including multidimensional spaces and cylinders, can be generated by combining these basic forms or by using more complex productions.

It is striking that the simple grammars in Fig. 2*A* generate many of the structural forms discussed by psychologists (24) and assumed by algorithms for unsupervised learning or exploratory data analysis. Partitions (9, 25), chains (26), orders (1, 25, 27), rings (28, 29), trees (1, 12, 30), hierarchies (31, 32) and grids (33) recur again and again in formal models across many different literatures. To highlight just one example, Inhelder and Piaget (1) suggest that the elementary logical operations in children's thinking are founded on two forms: a classification structure that can be modeled as a tree and a seriation structure that can be modeled as an order. The popularity of the forms in Fig. 2 suggests that they are useful for describing the world, and that they spring to mind naturally when scientists seek formal descriptions of a domain.

The problem of form discovery can now be posed. Given data *D* about a finite set of entities, we want to find the form *F* and the structure *S* of that form that best capture the relationships between these entities. We take a probabilistic approach, and define a hierarchical generative model (34) that specifies how the data are generated from an underlying structure, and how this structure is generated from an underlying form (Fig. 1*B*). We then search for the structure *S* and form *F* that maximize the posterior probability

$$P(S, F|D) \propto P(D|S)P(S|F)P(F). \quad [1]$$

$P(F)$ is a uniform distribution over the forms under consideration (Fig. 2). Structure *S* is a cluster graph, an instance of one of the forms in Fig. 2, where the nodes represent clusters of entities (Fig. 4*A* shows a cluster graph with the form of an order). The prior $P(S|F)$ favors graphs where *k*, the number of clusters, is small: $P(S|F) \propto \theta^k$ if *S* is compatible with *F*, and $P(S|F) = 0$ otherwise [see supporting information (SI) *Appendix* for the definition of compatibility]. The parameter $\theta$ determines the

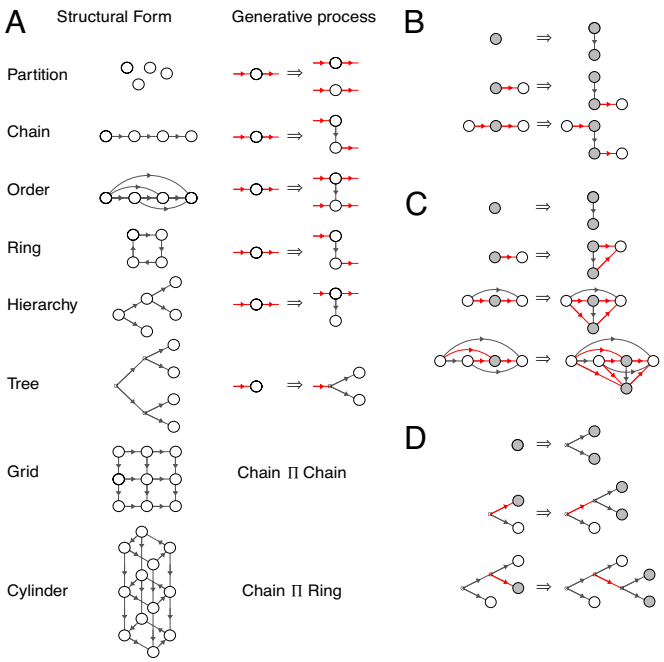


**Fig. 2.** A hypothesis space of structural forms. (*A*) Eight structural forms and the generative processes that produce them. Open nodes represent clusters of objects: A hierarchy has objects located internally, but a tree may only have objects at its leaves. The first six processes are node-replacement graph grammars. Each grammar uses a single production, and each production specifies how to replace a parent node with two child nodes. The seed for each grammar is a graph with a single node (in the case of the ring, this node has a self-link). (*B–D*) Growing chains, orders, and trees. At each step in each derivation, the parent and child nodes are shown in gray. The graph generated at each step is often rearranged before the next step. In *B*, for instance, the right side of the first step and the left side of the second step are identical graphs. The red arrows in each production represent *all* edges that enter or leave a parent node. When applying the order production, all nodes that previously sent a link to the parent node now send links to both children.

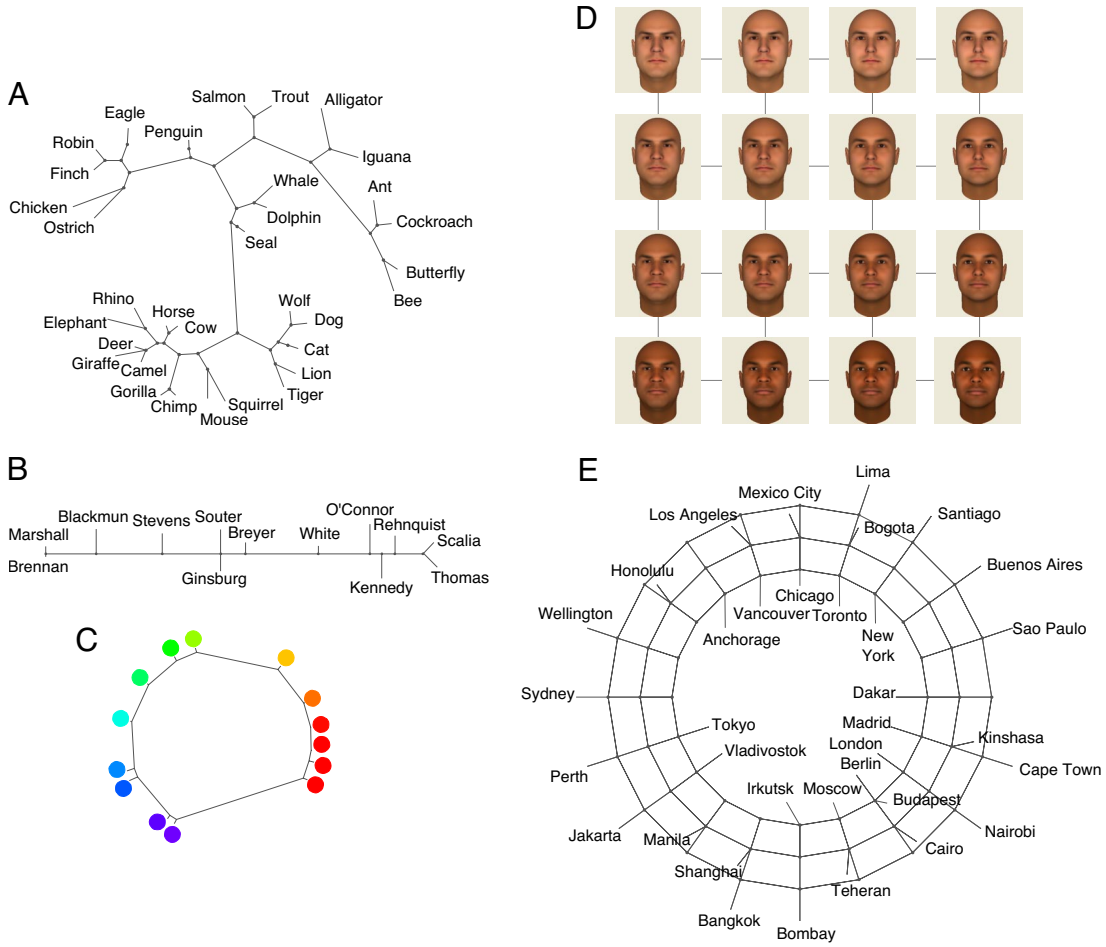


**Fig. 3.** Structures learned from biological features (*A*), Supreme Court votes (*B*), judgments of the similarity between pure color wavelengths (*C*), Euclidean distances between faces represented as pixel vectors (*D*), and distances between world cities (*E*). For *A–C*, the edge lengths represent maximum *a posteriori* edge lengths under our generative model.

extent to which graphs with many clusters are penalized, and is fixed for all of our experiments. The normalizing constant for $P(S|F)$ depends on the number of structures compatible with a given form, and ensures that simpler forms are preferred whenever possible. For example, any chain $S_c$ is a special case of a grid, but $P(S_c|F = \text{chain}) > P(S_c|F = \text{grid})$ because there are more possible grids than chains given a fixed number of entities. It follows that $P(S_c, F = \text{chain}|D) > P(S_c, F = \text{grid}|D)$ for any

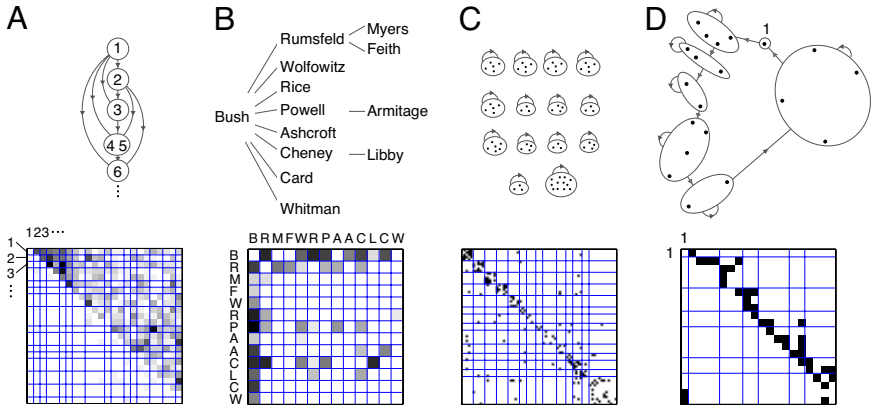


**Fig. 4.** Structures learned from relational data (*Upper*) and the raw data organized according to these structures (*Lower*). (*A*) Dominance relationships among a troop of sooty mangabeys. The sorted data matrix has most of its entries above the diagonal, indicating that animals tend to dominate only the animals below them in the order. (*B*) A hierarchy representing interactions between members of the Bush administration. (*C*) Social cliques representing friendship relations between prisoners. The sorted matrix has most of its entries along the diagonal, indicating that prisoners tend only to be friends with prisoners in the same cluster. (*D*) The Kula ring representing armshell trade between New Guinea communities. The relative positions of the communities correspond approximately to their geographic locations.

dataset $D$, and that the grid form will only be chosen if the best grid is substantially better than the best chain.

The remaining term in Eq. **1**, $P(D|S)$, measures how well the structure $S$ accounts for the data $D$. Suppose that $D$ is a feature matrix like the matrix in Fig. 1. $P(D|S)$ will be high if the features in $D$ vary smoothly over the graph $S$, that is, if entities nearby in $S$ tend to have similar feature values. For instance, feature $\mathbf{f}^{1}$ is smooth over the tree in Fig. 1*B*, but $\mathbf{f}^{100}$ is not. Even though Fig. 1 shows binary features, we treat all features as continuous features and capture the expectation of smoothness by assuming that these features are independently generated from a multivariate Gaussian distribution with a dimension for each node in graph $S$. As described in *SI Appendix*, the covariance of this distribution is defined in a way that encourages nearby nodes in graph $S$ to have similar feature values, and the term $P(D|S)$ favors graphs that meet this condition.

In principle, our approach can be used to identify the form $F$ that maximizes $P(F|D)$, but we are also interested in discovering the structure $S$ that best accounts for the data. We therefore search for the structure $S$ and form $F$ that jointly maximize the scoring function $P(S, F|D)$ (Eq. **1**). To identify these elements, we run a separate greedy search for each candidate form. Each search begins with all entities assigned to a single cluster, and the algorithm splits a cluster at each iteration, using the production for the current form (Fig. 2). After each split, the algorithm attempts to improve the score, using several proposals, including proposals that move an entity from one cluster to another and proposals that swap two clusters. The search concludes once the score can no longer be improved. A more detailed description of the search algorithm is provided in *SI Appendix*.

We generated synthetic data to test this algorithm on cases where the true structure was known. The *SI Appendix* shows graphs used to generate five datasets, and the structures found by fitting five different forms to the data. In each case, the model recovers the true underlying form of the data.

Next, we applied the model to several real-world datasets, in each case considering all forms in Fig. 2. The first dataset is a matrix of animal species and their biological and ecological properties. It consists of human judgments about 33 species and 106 features and amounts to a larger and noisier version of the dataset shown schematically in Fig. 1. The best scoring form for this dataset is the tree, and the best tree (Fig. 3*A*) includes subtrees that correspond to categories at several levels of resolution (e.g., mammals, primates, rodents, birds, insects, and flying insects). The second dataset is a matrix of votes from the United States Supreme Court, including 13 judges and their votes on 1,596 cases. Some political scientists (35) have argued that a unidimensional structure best accounts for variation in Supreme Court data and in political beliefs more generally, although other structural forms [including higher-dimensional spaces (36) and sets of clusters (37)] have also been proposed. Consistent with the unidimensional hypothesis, our model identifies the chain as the best-scoring form for the Supreme Court data. The best chain (Fig. 3*B*) organizes the 13 judges from liberal (Marshall and Brennan) to conservative (Thomas and Scalia).

If similarity is assumed to be a measure of covariance, our model can also discover structure in similarity data. Under our generative model for features, the expression for $P(D|S)$ includes only two components that depend on $D$: the number of features observed and the covariance of the data. As long as both components are provided, Eq. **1** can be used even if none of the features is directly observed. We applied the model to a matrix containing human judgments of the similarity between all pairs of 14 pure-wavelength hues (38). The ring in Fig. 3*C* is the best structure for these data and corresponds to the color circle described by Newton. Next, we analyzed a similarity dataset where the entities are faces that vary along two dimensions:

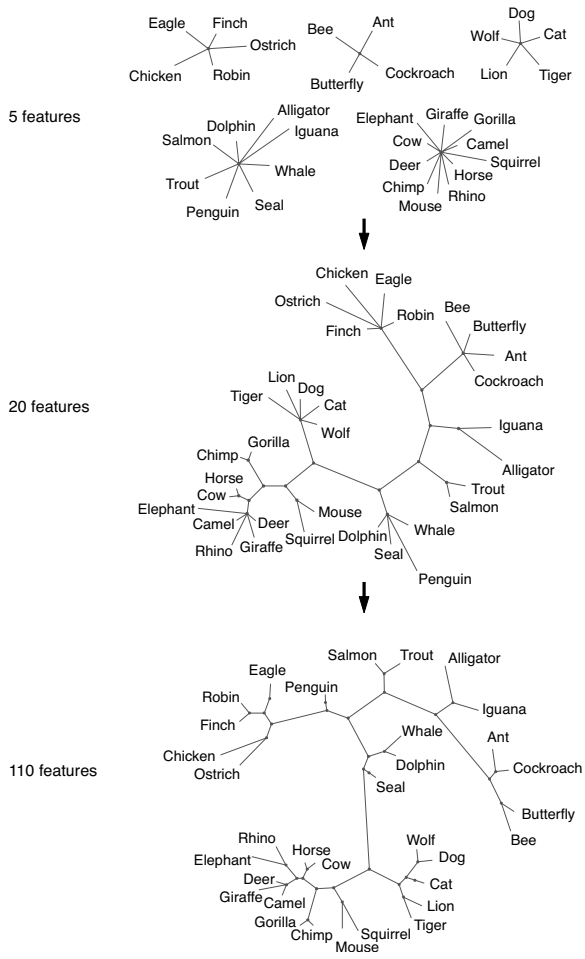


**Fig. 5.** Developmental changes as more data are observed for a fixed set of objects. After observing only five features of each animal species, the model chooses a partition, or a set of clusters. As the number of observed features grows from 5 to 20, the model makes a qualitative shift between a partition and a tree. As the number of features grows even further, the tree becomes increasingly complex, with subtrees that correspond more closely to adult taxonomic intuitions: For instance, the canines (dog, wolf) split off from the other carnivorous land mammals; the songbirds (robin, finch), flying birds (robin, finch, eagle), and walking birds (chicken, ostrich) form distinct subcategories; and the flying insects (butterfly, bee) and walking insects (ant, cockroach) form distinct subcategories. More information about these simulations can be found in *SI Appendix*.

masculinity and race. The model chooses a grid structure that recovers these dimensions (Fig. 3*D*). Finally, we applied the model to a dataset of distances between 35 world cities. Our model chooses a cylinder where the chain component corresponds approximately to latitude, and the ring component corresponds approximately to longitude.

The same algorithm can be used to discover structure in relational data, but we must modify the distribution $P(D|S)$. Suppose that $D$ is a square frequency matrix, where $D(i, j)$ indicates the number of times a certain relation has been observed between entities $i$ and $j$ (Fig. 4). We define a model where $P(D|S)$ is high if the large entries in $D$ correspond to edges in the graph $S$. A similar model can be defined if $D$ is a binary relation rather than a frequency matrix. Given a relation $D$, it is important to discover whether the relation tends to hold between elements in the same cluster or only between different clusters, and whether the relation is directed or not. The forms in Fig. 2*A* all have directed edges and nodes without self-links, and we

expanded this collection to include forms with self-links, forms with undirected edges, and forms with both of these properties.

First, we applied the model to a matrix of interactions among a troop of sooty mangabeys. The model discovers that the order is the most appropriate form, and the best order found (Fig. 4*A*) is consistent with the dominance hierarchy inferred by primatologists studying this troop (39). Hierarchical structure is also characteristic of human organizations, although tree-structured hierarchies are perhaps more common than full linear orders. We applied the model to a matrix of interactions between 13 members of George W. Bush's first-term administration (40). The best form is an undirected hierarchy, and the best hierarchy found (Fig. 4*B*) closely matches an organizational chart built by connecting individuals to their immediate superiors. Next, we analyzed social preference data (41) that represent friendships between prison inmates. Clique structures are often claimed to be characteristic of social networks (42), and the model discovers that a partition (a set of cliques) gives the best account of the data. Finally, we analyzed trade relations between 20 communities in New Guinea (43). The model discovers the Kula ring, an exchange structure first described by Malinowski (44).

We have presented an approach to structure discovery that provides a unifying description of many structural forms, discovers qualitatively different representations for a diverse range of datasets, and can handle multiple kinds of data, including feature data, relational data, and measures of similarity. Our hypothesis space of forms (Fig. 2) includes some of the most common forms, but does not exhaust the set of cognitively natural or scientifically important forms. Ultimately, psychologists should aim to develop a "Universal Structure Grammar" (compare with ref. 45) that characterizes more fully the representational resources available to human learners. This universal grammar might consist of a set of simple principles that generate all and only the cognitively natural forms. We can only speculate about how these principles might look, but one starting place is a metagrammar (46) for generating graph grammars. For instance, all of the forms shown in Fig. 2*A* can be generated by a metagrammar shown in *SI Appendix*.

Our framework may be most readily useful as a tool for data analysis and scientific discovery, but should also be explored as a model of human learning. Our model helps to explain how adults discover structural forms in controlled behavioral experiments (40), and is consistent with previous demonstrations that adults can choose the most appropriate representation for a given problem (47). Our model may also help to explain how children learn about the structure of their world. The model shows developmental shifts as more data are encountered, and often moves from a simple form to a more complex form that more faithfully represents the structure of the domain (Fig. 5 and *SI Appendix*). Identifying the best form for a domain provides powerful constraints on inductive inference, constraints that may help to explain how children learn new word meanings, concepts, and relations so quickly and from such sparse data (48–51). Discovering the clique structure of social networks can allow a child to predict the outcome of interactions between individuals who may never have interacted previously. Discovering the hierarchical structure of category labels allows a child to predict that a creature called a "chihuahua" might also be a dog and an animal, but cannot be both a dog and a cat.

As examples like these suggest, form discovery provides a way of acquiring domain-specific constraints on the structure of mental representations, a possibility that departs from two prominent views of cognition. A typical nativist view recognizes that inductive inference relies on domain-specific constraints but assumes that these constraints are innately provided (52–54). Chomsky (52), for instance, has suggested that "the belief that various systems of mind are organized along quite different principles leads to the natural conclusion that these systems are intrinsically determined, not simply the result of common mechanisms of learning or growth." A typical empiricist view emphasizes learning but assumes no domain-specific representational structure. Standard methods for learning associative networks (55) and neural networks (56) use the same generic class of representations for every task, instead of attempting to identify the distinctive kinds of structures that characterize individual domains. Without these constraints, empiricist methods can require unrealistically large quantities of training data to learn even very simple concepts (57). Our framework offers a third view that combines insights from both these approaches and shows how domain-specific structural constraints can be acquired by using domain-general probabilistic inference. As children learn about the structure of different domains, they make discoveries as impressive as those of Linnaeus and Mendeleev, and approaches like ours may help to explain how these discoveries are possible.

**ACKNOWLEDGMENTS.** We thank P. Gunkel, E. Newport, A. Perfors, and W. Richards for valuable discussions and D. Casasanto, M. Frank, N. Goodman, V. Mansinghka, R. Saxe, J. M. Tenenbaum, D. Tymoczko, the editor, and two anonymous reviewers for helpful suggestions. This work was supported in part by the William Asbjornsen Albert memorial fellowship (to C.K.), the Paul E. Newton career development chair (J.B.T.), the James S. McDonnell Foundation Causal Learning Research Collaborative, Air Force Office of Scientific Research Grant FA9550-07-1-0075, and the NTT Communication Sciences Laboratory.

1. Inhelder B, Piaget J (1964) *The Early Growth of Logic in the Child* (Routledge & Kegan Paul, London).
2. Carey, S (1985) *Conceptual Change in Childhood* (MIT Press, Cambridge, MA).
3. Gopnik, A, Meltzoff, AN (1997) *Words, Thoughts, and Theories* (MIT Press, Cambridge, MA).
4. Kuhn, TS (1970) *The Structure of Scientific Revolutions* (Univ of Chicago Press, Chicago), 2nd Ed.
5. Whewell W (1840) *The Philosophy of the Inductive Sciences: Founded Upon Their History* (J. W. Parker, London).
6. Jaynes ET (2003) *Probability Theory: The Logic of Science* (Cambridge Univ Press, Cambridge, UK).
7. Spirtes P, Glymour C, Scheines R (1993) *Causation, Prediction and Search* (Springer, New York).
8. Purves WK, Sadava D, Orians GH, Heller HC (2001) *Life: The Science of Biology* (Sinauer, Sunderland, MA), 6th Ed.
9. Anderson JR (1991) The adaptive nature of human categorization. *Psychol Rev* 98:409–429.
10. Rumelhart D, Zipser D (1986) Feature discovery by competitive learning. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds Rumelhart D, McClelland J, and the PDP research group (MIT Press, Cambridge, MA), Vol 1.
11. Duda RO, Hart PE, Stork DG (2000) *Pattern Classification* (Wiley, New York), 2nd Ed.
12. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
13. Pearson K (1901) On lines and planes of closest fit to systems of points in space. *Philos Mag* 2:559–572.
14. Spearman CE (1904) "General intelligence" objectively determined and measured. *Am J Psychol* 5:201–293.
15. Torgeson WS (1965) Multidimensional scaling of similarity. *Psychometrika* 30:379–393.
16. Lovejoy AO (1970) *The Great Chain of Being* (Harvard Univ Press, Cambridge, MA).
17. Piaget J (1965) *The Child's Conception of Number* (Norton, New York).
18. Shultz TR (2003) *Computational Developmental Psychology* (MIT Press, Cambridge, MA).
19. Rosch E (1978) Principles of categorization. *Cognition and Categorization*, eds Rosch E, Lloyd BB (Lawrence Erlbaum, New York), pp 27–48.
20. Markman, E (1989) *Naming and Categorization in Children* (MIT Press, Cambridge, MA).
21. Shultz, TR, Vogel, A (2004) A connectionist model of the development of transitivity. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, eds Forbus K, Gentner D, Regier T (Lawrence Erlbaum, Cambridge, MA) pp 1243–1248.
22. Engelfriet J, Rozenberg G (1997) Node replacement graph grammars. *Handbook of Graph Grammars and Computing by Graph Transformation*, ed Rozenberg G (World Scientific, Singapore), Vol 1.
23. Leyton M (1992) *Symmetry, Causality, Mind* (MIT Press, Cambridge, MA).
24. Shepard RN (1980) Multidimensional scaling, tree-fitting, and clustering. *Science* 210:390–398.
25. Fiske AP (1992) The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychol Rev* 99:689–723.
26. Guttman L (1944) A basis for scaling qualitative data. *Am Soc Rev* 9:139–150.

27. Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs. 1. The method of paired comparisons. *Biometrika* 39:324–345.
28. Guttman L (1954) A new approach to factor analysis: The radex. *Mathematical Thinking in the Social Sciences*, ed Lazarsfeld PF (Free Press, Glencoe, IL), pp 258–348.
29. Wiggins JS (1996) An informal history of the interpersonal circumplex tradition. *J Personality Assessment* 66:217–233.
30. Sneath PH, Sokal RR (1973) *Numerical Taxonomy: The Principles and Practice of Numerical Classification* (Freeman, San Francisco).
31. Collins AM, Quillian MR (1969) Retrieval time from semantic memory. *J Verbal Learn Verbal Behav* 8:240–247.
32. Carroll JD (1976) Spatial, nonspatial and hybrid models for scaling. *Psychometrika* 41:439–463.
33. Kohonen T (1997) *Self-Organizing Maps* (Springer, New York).
34. Gelman, A, Carlin JB, Stern HS, Rubin DB (2003) *Bayesian Data Analysis* (Chapman & Hall, New York), 2nd Ed.
35. Grofman B, Brazill T (2002) Identifying the median justice on the Supreme Court through multidimensional scaling: Analysis of ''natural courts'' 1953–1991. *Public Choice* 112:55–79.
36. Wilcox C, Clausen A (1991) The dimensionality of roll-call voting reconsidered. *Legislative Stud Q* 16:393–406.
37. Fleishman JA (1986) Types of political attitude structure: Results of a cluster analysis. *Public Opin Q* 50:371–386.
38. Ekman G (1954) Dimensions of color vision. *J Psychol* 38:467–474.
39. Range F, Noë R (2002) Familiarity and dominance relations among female sooty mangabeys in the Tai national park. *Am J Primatol* 56:137–153.
40. Kemp C (2008) PhD Thesis (Massachusetts Institute of Technology, Cambridge, MA).
41. MacRae J (1960) Direct factor analysis of sociometric data. *Sociometry* 22:360–371.
42. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99:7821–7826.
43. Hage, P, Harary, F (1991) *Exchange in Oceania: A Graph Theoretic Analysis* (Oxford Univ Press, Oxford).
44. Malinowski, B (1922) *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea* (G. Routledge & Sons, London).
45. Chomsky, N (1965) *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA).
46. Nagl M (1986) Set theoretic approaches to graph grammars. *Proceedings of the 3rd International Workshop on Graph-Grammars and their Application to Computer Science* (Springer, London, UK), pp 41–54.
47. Novick LR, Hurley SM (2001) To matrix, network or hierarchy: That is the question. *Cognit Psych* 42:158–216.
48. Carey S, Bartlett E (1978) Acquiring a single new word. *Pap Rep Child Lang Dev* 15:17–29.
49. Keil FC (1979) *Semantic and Conceptual Development* (Harvard Univ Press, Cambridge, MA).
50. Keil FC (1981) Constraints on knowledge and cognitive development. *Psychol Rev* 88:197–227.
51. Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends Cognit Sci* 10:309–318.
52. Chomsky N (1980) *Rules and Representations* (Basil Blackwell, Oxford).
53. Atran S (1998) Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences* 21:547–609.
54. Kant I, trans Smith NK (2003) *Critique of Pure Reason* (Palgrave Macmillan, New York).
55. Rescorla, RA, Wagner, AR (1972) A theory of pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning II: Current Research and Theory*, eds Black, AH, Prokasy, WF (Appleton-Century-Crofts, New York), pp 64–99.
56. Rogers, TT, McClelland, JL (2004) *Semantic Cognition: A Parallel Distributed Processing Approach* (MIT Press, Cambridge, MA).
57. Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias-variance dilemma. *Neural Computation* 4:1–58.