A. Gopnik, & J. Tenenbaum (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science (special section on Bayesian and Bayes-Net approaches to development)* 10(3): 281-287.

Bayesian networks, Bayesian learning and cognitive development

Alison Gopnik

Department of Psychology

University of California, Berkeley

Joshua B. Tenenbaum

Department of Brain and Cognitive Sciences

Massachusetts Institute of Technology

Acknowledgments: The writing of this article was supported by the James S. McDonnell Foundation Causal Learning Research Collaborative, and the Paul E. Newton Career Development Chair (JBT). We are grateful to Denis Mareschal and Laura Schulz for encouraging and helpful discussions. Address correspondence to A. Gopnik, Department of Psychology, University of California at Berkeley, Berkeley, CA, 94720, gopnik@berkeley.edu, or J. B. Tenenbaum, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, jbt@mit.edu.

Developmental Science. Vol 10(3), May 2007, pp. 281-287

Bayesian networks, Bayesian learning, and cognitive development

Over the past thirty years we have discovered an enormous amount about what children know and when they know it. But the real question for developmental cognitive science is not so much what children know, when they know it or even whether they learn it. The real question is *how* they learn it and *why* they get it right. Developmental "theory theorists" (e.g., Carey, 1985; Gopnik & Meltzoff, 1997; Wellman & Gelman, 1997) have suggested that children's learning mechanisms are analogous to scientific theory-formation. However, what we really need is a more precise computational specification of the mechanisms that underlie both types of learning, in cognitive development and scientific discovery.

The most familiar candidates for learning mechanisms in developmental psychology have been variants of associationism, either the mechanisms of classical and operant conditioning in behaviorist theories (e.g., Rescorla & Wagner 1972) or more recently, connectionist models (e.g., Rumelhart & McClelland, 1986; Elman, Bates, Johnson & Karmiloff-Smith, 1996; Shultz, 2003; Rogers & McClelland 2004). Such theories have had difficulty explaining how apparently rich, complex, abstract, rule-governed representations, such as we see in everyday theories, could be derived from evidence. Typically, associationists have argued that such abstract representations do not really exist, and that children's behavior can be just as well explained in terms of more specific learned associations between task inputs and outputs. Connectionists often qualify this denial by appealing to the notion of distributed representations in hidden layers of units that relate inputs to outputs (Rogers & McClelland, 2004; Colunga & Smith, 2005). On this view however, the representations are not explicit, task-independent models of the world structure that is responsible for the input-output relations. Instead, they are implicit summaries of the input-output relations for a specific set of tasks that the connectionist network has been trained to perform.

Conversely, more nativist accounts of cognitive development endorse the existence of abstract rule-governed representations but deny that their basic structure is learned. Modularity or "core knowledge" theorists, for example, suggest that there are a small number of innate causal schemas designed to fit particular domains of knowledge, such as a belief-desire schema for intuitive psychology or a generic object schema for intuitive physics. Development is either a matter of enriching those innate schemas, or else involves quite sophisticated and culture-specific kinds of learning like those of the social institutions of science (e.g., Spelke, Breinlinger, Macomber, & Jacobson, 1992).

This has left empirically-minded developmentalists, who seem to see both abstract representation *and* learning in even the youngest children, in an unfortunate theoretical bind. There appears to be a vast gap between the kinds of knowledge that children learn and the mechanisms that could allow them to learn that knowledge. The attempt to bridge this gap dates back to Piagetian ideas about constructivism, of course, but simply saying that there are constructivist learning mechanisms is a way of restating the problem rather than providing a solution. Is there a more precise computational way to bridge this gap?

Recent developments in machine learning and artificial intelligence suggest the answer may be yes. These new approaches to inductive learning are based on sophisticated and rational mechanisms of statistical inference operating over explicitly structured representations. They allow abstract, coherent, theory-like knowledge to be derived from patterns of evidence, and show how that knowledge provides constraints on future inductive inferences that a learner might make. These computational accounts take the kinds of evidence that have been considered in traditional associative learning accounts – such as evidence about contingencies among events or evidence about the consequences of actions – and use that data to learn structured knowledge representations of the kinds that have been proposed in traditional nativist accounts, such as causal networks, generative grammars, or ontological hierarchies.

The papers in this special section show how these sophisticated statistical inference frameworks can be applied to problems of longstanding interest in cognitive development. The papers focus on two classes of learning problems: learning causal relations, from observing co-occurrences among events and active interventions; and learning how to organize the world into categories and map word labels onto categories, from observing examples of objects in those categories. Causal learning, category learning and word learning are all problems of *induction*, in which children form representations of the world's abstract structure that extend qualitatively beyond the data they observe and that support generalization to new tasks and contexts. While philosophers have long seen inductive inference as a source of great puzzles and paradoxes, children solve these natural problems of induction routinely and effortlessly. Through a combination of new computational approaches and empirical studies motivated by those models, developmental scientists may now be on the verge of understanding how they do it.

Learning causal Bayesian networks

Three of the five papers in this section focus on children's causal learning. This work is inspired by the development of causal Bayesian networks, a rational but cognitively appealing formalism for representing, learning, and reasoning about causal relations (Pearl, 2000; Glymour, 2001; Gopnik et al., 2004; Gopnik and Schulz, 2007). "Theory theorists" in cognitive development point to an analogy between learning in children and learning in science. Causal Bayesian networks provide a computational account of a kind of inductive inference that should be familiar from everyday scientific thinking: testing hypotheses about the causal structure underlying a set of variables by observing patterns of correlation and partial correlation among these variables, and by examining the consequences of interventions (or experiments) on these variables.

Bayesian networks represent causal relations as directed acyclic graphs. Nodes in the graph represent variables in a causal system, and edges (arrows) represent direct causal relations between those variables. Variables can be binary or discrete sets of propositions (e.g., a person's eye color or a student's grade) or continuous quantities (e.g., height or weight). They can be observable or hidden. The direct causal relations can also take on many different functional forms: deterministic or probabilistic, generative or inhibitory, linear or non-linear.

The graph structure of a causal Bayesian network is used to define a joint probability distribution over the variables in the network – thereby specifying how likely is any joint setting of all the variables. These probabilistic models can be used to reason and make predictions about the variables when the graph structure is known, and also to learn the graph structure when it is unknown, by observing which settings of the variables tend to occur together more or less often. The probability distribution specified by a causal Bayesian network is a product of many local components, each corresponding to one variable and its direct causes in the graph. Two variables may be correlated – or probabilistically dependent – even if they are not directly connected in the graph, but if they are not directly connected, their correlation will be mediated by one or more other variables.

As a consequence of how the graph structure of a causal Bayesian network is used to define a probabilistic model, the graph places constraints on the probabilistic relations that may hold among the variables in that network, regardless of what the variables represent or how the causal mechanisms operate. In particular, the graph structure constrains the conditional independencies among those variables.[1] Given a certain causal structure, only some patterns of conditional independence will be expected to occur generically among the variables. The precise constraints are captured by the *causal Markov condition*: conditioned on its direct causes, any variable will be independent of all other variables in the graph except for its own direct and indirect effects. For example, in the causal chain $A \rightarrow B \rightarrow C \rightarrow D$, the variables C and A are normally dependent, but they become independent conditioned on C's one direct cause B; C remains probabilistically dependent on its direct effect D under all conditions. The same patterns of dependence and conditional dependence would hold if the chain runs the other way, $A \leftarrow B \leftarrow C \leftarrow D$; these two networks are said to be *Markov equivalent*. A graph that appears only slightly different, such as $A \rightarrow B \leftarrow C \rightarrow D$, may imply quite different dependencies: here, the variable C is independent of A, but becomes dependent on A when we condition on B.

Causal Bayesian networks also allow us to reason about the effects of outside interventions on variables in a causal system.[2] Interventions on a particular variable X induce predictable changes in the probabilistic dependencies over all variables in the network. Formally, these dependencies are still governed by the Markov condition but now applied to a "mutilated" graph in which all incoming arrows to X are cut. Two networks that would otherwise imply identical patterns of probabilistic dependence may become distinguishable under intervention. For example, if we intervene to set the value of C in the above graphs, then the structures $A \rightarrow B \rightarrow C \rightarrow D$ and $A \leftarrow B \leftarrow C \leftarrow D$ predict distinct patterns of probabilistic dependence, given by these "mutilated" graphs respectively: $A \rightarrow B \quad C \rightarrow D$ and $A \leftarrow B \leftarrow$

C   *D* . For the first graph, *B* should now be independent of *C*, but *C* and *D* should remain dependent; for the second graph, the opposite pattern should hold.

The Markov condition and the logic of intervention together form the basis for one popular approach to learning causal Bayesian networks from data, known as *constraint-based learning* (Spirtes, Glymour & Schienes, 2001; Pearl, 2000). Given observed patterns of independence and conditional independence among a set of variables, perhaps under different conditions of intervention, these algorithms work backwards to figure out the set of causal structures compatible with the constraints of that evidence. Computationally tractable algorithms can search for or construct the subset of possible network structures that is compatible with the evidence, and have been extensively applied in a range of disciplines (eg, Glymour and Cooper, 1999). It is even possible to infer the existence of new unobserved variables that are common causes of the observed variables (Silva, Scheines, Glymour & Spirtes, 2003).

Bayesian learning

Despite the impressive accomplishments of these constraint-based learning algorithms, human causal learning often goes beyond their capacities. People – even young children – can correctly infer causes from only one or a small number of examples, far too little data to compute reliable measures of correlation as these algorithms require. Such rapid inferences may depend on more articulated causal hypotheses than can be captured simply by a causal graph. For instance, people may have ideas about the kinds of causal mechanisms at work, which would allow more specific predictions about the patterns of data that are likely to be observed under different hypothesized causal structures. People are also inclined to judge certain causal structures as more likely than others, rather than simply asserting some causal structures as consistent with the data and others inconsistent. These degrees of belief may be strongly influenced by prior expectations about which kinds of causal relations are more or less likely to hold in particular domains.

These aspects of human causal learning may be best captured by an alternative computational approach that explicitly formalizes the learning problem as a Bayesian probabilistic inference. The learner constructs a hypothesis space *H* of possible causal models, and given some data *d* – observations of the states of one or more variables in the causal system for different cases, individuals or situations – computes a *posterior probability* distribution $P(h|d)$ representing a degree of belief that each causal-model hypothesis *h* corresponds to the true causal structure. These posteriors depend on two more primitive quantities: *prior probabilities* $P(h)$, measuring the plausibility of each causal hypothesis independent of the data, and *likelihoods* $P(d|h)$, expressing how likely we would be to observe the data *d* if the hypothesis *h* were correct. Bayes' rule dictates how these quantities are related: posterior probabilities are proportional to the product of priors and likelihoods, normalized by the sum of these scores over all alternative hypotheses h',

$$P(h \mid d) = \frac{P(d \mid h)\, P(h)}{\sum_{h'} P(d \mid h')\, P(h')} \quad . \qquad (1)$$

Like constraint-based learning algorithms, Bayesian algorithms for learning causal networks have been successfully applied across many tasks in machine learning, artificial intelligence and related disciplines (Heckerman, 1999). A distinctive strength of Bayesian learning comes from the ability to use informative, highly structured priors and likelihoods that draw on the learner's background knowledge. This knowledge can often be expressed in the form of abstract conceptual frameworks or schemas, specifying what kinds of entities or variables exist, and what kinds of causal relations are likely to exist between entities or variables as a function of these types (Pasula & Russell, 2001; Milch, Marthi, Russell, Sontag, Ong & Kolobov, 2005; Mansinghka, Kemp, Tenenbaum & Griffiths, 2006; Griffiths & Tenenbaum, 2007). These frameworks are much like the "framework theories" that cognitive developmentalists have identified as playing a key role in children's learning (Wellman & Gelman, 1992). They can be formalized as systems for generating a constrained space of causal Bayesian networks and a prior distribution over that space to support Bayesian learning.

Bayesian principles can be applied not only to causal learning, but to a much broader class of cognitively important inductive inference tasks (Chater, Tenenbaum & Yuille, 2006). These tasks include learning concepts and properties (Anderson, 1991; Mitchell, 1997; Tenenbaum, 2000; Tenenbaum and Griffiths, 2001; Tenenbaum, Kemp & Shafto, in press), learning systems of categories that characterize domains (Kemp, Perfors & Tenenbaum, 2004; Kemp, Tenenbaum, Griffiths, Yamada & Ueda, 2006; Shafto, Kemp, Mansinghka, Gordon & Tenenbaum, 2006; Navarro, 2006), syntactic parsing in natural language and learning the rules of syntax (Chater & Manning, 2006), and parsing visual images of natural scenes (Yuille & Kersten, 2006). These approaches use various structured frameworks to represent the learner's knowledge, such as generative grammars, tree-structured hierarchies, or predicate logic. They thus show how a diverse range of abstract representations of the external world – not only directed causal graphs – can be rationally inferred from sparse, ambiguous data, and used to guide subsequent predictions and actions. For instance, Kemp et al. (2004) show how a Bayesian learner can discover that a system of categories and their properties is best organized according to a taxonomic tree structure, and can then use this structure to generate priors for inferring how a novel property is distributed over categories, given only a few examples of objects with that property.

A research program for cognitive development

Over the last few years, several groups of researchers have explored the hypothesis that children implicitly use representations and computations similar to those discussed above to learn about the structure of the world. The papers in this special section represent some of their latest efforts. "Implicitly" and "similar to" are key words here. These formal approaches are framed at an abstract level of analysis, what Marr (1982) called the level of computational theory. They specify ideal inferential relations between structured hypotheses and patterns of data. Hence the developmental research focuses on comparing children's behavior with the output of Bayesian network or Bayesian learning models, rather than testing precise correspondences between children's cognitive processes and the algorithmic operations of these models.

One line of work, inspired by constraint-based algorithms for learning causal Bayesian networks, has looked at children's abilities to make causal inferences from patterns of conditional probabilities and interventions. Eight-month-olds can calculate elementary conditional independence relations and can use these relations to make predictions (Sobel & Kirkham, in press; this issue). Two-year-olds can combine conditional independence and intervention information appropriately to choose among candidate causes of an effect. Four-year-olds not only do this but also design novel and appropriate interventions on these causes (Gopnik et al, 2001; Sobel & Kirkham, in press), and do so across a wide variety of domains (Schulz & Gopnik, 2004). They can use different patterns of evidence to determine the direction of causal relations – whether A causes B or B causes A – and to infer unobserved causes (Gopnik et al., 2004; Kushnir et al. 2003; Schulz & Somerville 2006). They discriminate between observations and interventions appropriately (Gopnik et al 2004; Kushnir & Gopnik, 2005) and use probabilities to calculate causal strength (Kushnir & Gopnik, 2005, in press).

Despite this wealth of results, many interesting empirical questions remain. Most existing studies of children's causal learning have involved only very simple networks (e.g, two variables) and learning from passively presented data. Schulz et al (this issue) advance along both of these fronts, by studying inferences about more complex network structures based on conditional interventions, and examining the inferential power of the spontaneous interventions that children natural make in their everyday play. Most studies have involved deterministic causal relations – eg., wheel A always makes wheel B spin – rather than the probabilistic relations for which causal Bayesian networks were originally intended. The developmental trajectory of causal learning abilities, from infancy up to the preschool years where most existing studies focus, is only beginning to be probed. Sobel and Kirkham (this issue) are pushing this frontier in their work with five-month-olds.

Another line of work has explored Bayesian learning models as accounts of how children infer causal structure, as well as word meanings and other kinds of world structure. By age four, children appear able to combine prior knowledge about hypotheses and new evidence in a Bayesian fashion. In causal learning, children can learn the prior probability that a particular causal relation is likely to hold for a particular kind of object, and use that knowledge together with potentially ambiguous data to judge the causal efficacy of a new object of that kind (Sobel et al, 2004; Tenenbaum, Sobel, Griffiths & Gopnik, submitted). Four-year-olds can use new evidence to override causal hypotheses that are favored a priori: for example, they can conclude that a biological effect has a psychological cause (Bonawitz, Griffiths & Schulz, 2006) or that a physical object can act at a distance on another object (Kushnir & Gopnik, in press). However, they are less likely to accept those hypotheses than hypotheses that are consistent with prior knowledge. Tenenbaum and Xu (2000; Xu and Tenenbaum, 2005, in press) have developed a Bayesian model of word learning, and shown that it accounts for how preschoolers learn words from one or a few examples – the "fast mapping" behavior first studied by Carey and Bartlett (1978). This model encodes traditional principles of word learning, such as the assumption that kind labels pick out whole objects and map onto taxonomic categories (Markman, 1989), as constraints on the hypothesis space of possible word meanings. Fast-mapping is then explained as a Bayesian inference over that hypothesis space. Related Bayesian models have been proposed to explain how children learn the meanings of other aspects of language, including verbs (Niyogi, 2002), adjectives (Dowman, 2002), and anaphoric constructions (Regier & Gahl, 2004).

Two frontiers of this research program are explored in papers in the special section. Most Bayesian analyses to date have focused on the child in isolation, without considering the central role of the social and intentional context in which the child's learning and reasoning are embedded (Gopnik & Meltzoff, 1997; Bloom, 2000). Xu and Tenenbaum (this issue) extend Bayesian models of word learning to account for the different inferences children make depending on whether the examples they observe are labeled ostensively by a teacher, or actively chosen by the learners themselves. Related Bayesian models are being developed to explain children's intuitive psychological reasoning – how they infer the beliefs and goals of other agents from observations about their behavior (Goodman et al., 2006; Baker, Saxe and Tenenbaum, 2006).

Bayesian models have also traditionally been limited by a focus on learning representations at only a single level of abstraction. In contrast, children can learn in parallel at multiple levels: for example, they can learn new causal relations from sparse data, guided by priors from larger-scale framework theories of a domain, but over time they will also change their framework theories as they observe how causal structures in that domain tend to operate. Tenenbaum, Griffiths and Niyogi (2007) have suggested how multiple levels of causal learning can be captured within a hierarchical Bayesian framework. Kemp, Perfors and Tenenbaum (this issue) develop a hierarchical Bayesian model for learning overhypotheses about categories and word labels – principles such as the shape bias, specifying the kinds of categories that labels tend to pick out, which are learned by children (around age two) at the same time that they are learning specific word-category mappings (Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002). Similar analyses have been proposed for how children can acquire other kinds of abstract knowledge, such as a tree-structured system of ontological classes (Schmidt, Kemp & Tenenbaum, 2006) or the hierarchical structure of syntactic phrases in natural language (Perfors, Tenenbaum & Regier, 2006).

Perhaps the greatest open question about Bayesian network and Bayesian learning models is how they might be implemented in the brain. The appeal of connectionist models of development comes partly from their relatively straightforward mapping onto known neural mechanisms; that is certainly not true for Bayesian networks and Bayesian learning. Some first steps have recently been made, however. Computational neuroscientists have begun studying how Bayesian updating may be implemented in neural circuitry or population codes (Knill & Pouget, 2004). McClelland and Thompson (this issue) suggest how several experimental studies of children's causal learning can be modeled in a brain-inspired connectionist architecture, which also approximates the relevant Bayesian inferences. Their proposal includes several elements that go beyond traditional associative or connectionist models, including complementary "cortical" and "hippocampal" learning systems, the capacity to interleave different kinds of trials during training, and a "backpropagation to representation" mechanism to infer the hidden-layer representations of novel stimuli necessary for causal attribution. How far such models can go towards capturing the structure of children's intuitive theories remains an important question for future work.

Conclusion

Developing rigorous theories that generate testable experimental predictions is, of course, a holy grail of developmental science, and like all grails tends to shimmer on the horizon rather than to be firmly and indubitably grasped. But certainly the papers in this special section pass a more realistic test. To use a technical term from adolescence they are really cool – cool experimental results and cool computational findings. We hope and believe that the interaction of probabilistic models and developmental experiments will keep generating lots of cool work for years to come.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. Psychological Review, 98, 409-429.

Baker, C., Saxe, R., & Tenenbaum J. B. (2006). A Bayesian framework for human action understanding. In Y. Weiss, B. Scholkopf, and J. Platt (eds.), Advances in Neural Information Processing Systems 18, 99-106.

Bloom, P. (2000). How children learn the meanings of words. Cambridge, MA: MIT Press.

Bonawitz, E. B., Griffiths, T. L., & Schulz, L. E. (2006). Modeling cross-domain causal learning in preschoolers as Bayesian inference. In R. Sun & N. Miyake (Eds.), Proceedings of the 28th annual conference of the cognitive science society (pp. 89-94). Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. & Bartlett, E. (1978) Acquiring a single new word. Papers and Reports on Child Language Development, 15, 17-29.

Carey, S. (1985). Conceptual Change in Childhood. Cambridge, MA: MIT Press/Bradford Books.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. Trends in Cognitive Sciences, 10, 335-344.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences, 10, 287-291.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: a role for associative learning. Psychological Review, 112 (2).

Dowman, M. (2002). Modelling the acquisition of colour words. In Proceedings of the15th Australian. Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence, pp. 259–271, Springer-Verlag.

Elman, J. L., Bates, E. A., Johnson, M. H., & Karmiloff-Smith, A. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: MIT Press.

Glymour, C. (2001). The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology. Cambridge, MA: MIT Press.

Glymour, C, & Cooper, G. (1999). Computation, Causation, and Discovery. Menlo Park, CA: AAAI/MIT Press.

Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., Schulz, L., and Tenenbaum, J. B. (2006). Intuitive theories of mind: A rational approach to false belief. Proceedings of the Twenty-Eigth Annual Conference of the Cognitive Science Society.

Gopnik, A. Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T, & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 111, 1-30.

Gopnik, A., & Meltzoff, A. (1997). Words, Thoughts and Theories. Cambridge, MA: MIT Press.

Gopnik, A. & Schulz, L. (2007). Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Gopnik, A., Sobel, D. M., Schulz, L. & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental Psychology, 37, 5, 620–629.

Griffiths, T. L. and Tenenbaum, J.B. (2007). Two proposals for causal grammar. In A. Gopnik and L. Schulz (eds.), Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Kemp, C., Perfors, A. F., & Tenenbaum, J. B. (2004). Learning domain structure. Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society.

Kemp, C., Perfors, A. F., & Tenenbaum, J. B. (this issue). Learning overhypotheses with hierarchical Bayesian models. Developmental Science.

Kemp, C., Tenenbaum J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. Twenty-First National Conference on Artificial Intelligence (AAAI 2006).

Knill, D., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neuroscience, 27(12):712-9.

Kushnir, T., Gopnik, A., Schulz, L. E., & Danks, D. (2003). Inferring hidden causes. Proceedings of the 25th Annual Meeting of the Cognitive Science Society.

Kushnir T. & Gopnik, A., (2005). Children infer causal strength from probabilities and interventions. Psychological Science, 16(9), 678-683.

Mansinghka, V. K., Kemp, C., Tenenbaum, J. B., and Griffiths, T. L. (2006). Structured priors for structure learning. Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006).

Markman, E.M. (1989) Categorization and naming in children. Cambridge, MA: MIT Press.

Marr, D. (1982). Vision. San Francisco: W.H. Freeman.

Milch, B., Marthi, B., Russell, S., Sontag, D., Ong, D. L., & Kolobov, A. (2005). BLOG: Probabilistic Models with Unknown Objects. Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI), 1352-1359.

Mitchell, T. (1997). Machine Learning. McGraw Hill.

McClelland, J. L. & Thompson, R. M. (this issue). Using domain-general principles to explain children's causal reasoning abilities. Developmental Science.

Navarro, D. J. (2006). From natural kinds to complex categories. In R. Sun & N. Miyake (Eds), Proceedings of the 27th Annual Conference of the Cognitive Science Society (pp. 621-626). Mahwah, NJ: Lawrence Erlbaum.

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In Proceedings of the 24th Annual Conference of the Cognitive Science Society (Gray, W. and Schunn, C., eds), pp. 697–702, Erlbaum.

Pasula, H. & Russell, S. (2001). Approximate inference for first-order probabilistic languages. Proceedings of the International Joint Conference on Artificial Intelligence 2001 (IJCAI-2001).

Pearl, J. (2000). Causality. New York: Oxford University Press.

Perfors, A., Tenenbaum J. B., & Regier, T. (2006). Poverty of the stimulus? A rational approach. Proceedings of the Twenty-Eigth Annual Conference of the Cognitive Science Society.

Regier, T. & Gahl, S. (2004). Learning the unlearnable: the role of missing evidence. Cognition, 93, 147–155.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical Conditioning II: Current theory and research (pp. 64-99). New York: Appleton-Century-Crofts.

Rogers, T., & McLelland, J. (2004). Semantic cognition: A parallel distributed approach. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA: MIT Press.

Schmidt, L., Kemp, C., & Tenenbaum J. B. (2006). Nonsense and sensibility: Inferring unseen possibilities. Proceedings of the Twenty-Eigth Annual Conference of the Cognitive Science Society.

Schulz, L. E. and Gopnik, A. (2004). Causal learning across domains. Developmental Psychology, 40, 162-176.

Schulz, L. E. & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. Child Development, 77 (2), 427-442.

Schulz, L. E. , Gopnik, A., & Glymour, C. (this issue). Preschool children learn about causal structure from conditional interventions. Developmental Science.

Shafto, P., Kemp, C., Mansinghka, V., Gordon, M., and Tenenbaum J. B. (2006). Learning cross-cutting systems of categories. Proceedings of the Twenty-Eigth Annual Conference of the Cognitive Science Society.

Shultz, T.R (2003). Computational Developmental Psychology. Cambridge, MA: MIT Press.

Silva, R., Scheines, R., Glymour, C. & Spirtes, P. (2003). Learning measurement models for unobserved variables. Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (AAAI-2003). AAAI Press.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. Psychological Science, 13 (1), 13–19.

Sobel, D. M. & Kirkham, N. Z. (in press). Blickets and babies: The development of causal reasoning in toddlers and infants. Developmental Psychology.

Sobel, D. M. & Kirkham, N. Z. (this issue). Bayes nets and Babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. Developmental Science.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive Science, 28, 303-333.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. Psychological Review, 99, 605-632.

Spirtes, P., Glymour, C., & Scheines, R. (2001). Causation, Prediction, and Search (Springer Lecture Notes in Statistics, 2nd edition, revised). Cambridge, MA: MIT Press.

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. Trends in Cognitive Sciences, 10, 327-334.

Tenenbaum, J. B. (2000). Rules and similarity in concept learning. In S. Solla, T. Leen, and K. R. Mueller (eds.), Advances in Neural Information Processing Systems 12. Cambridge, MA: MIT Press, 59-65.

Tenenbaum, J. B. & Xu, F. (2000). Word learning as Bayesian inference. Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, 517-522

Tenenbaum, J. B. & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. Behavioral and Brain Sciences, 24(4), 629-64.

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In S. Becker, S. Thrun & K. Obermayer (Eds.), Advances in neural information processing systems 15 (pp. 35-42). Cambridge, MA: The MIT Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10, 309-318

Tenenbaum, J.B., Griffiths, T. L., and Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik and L. Schulz (eds.), Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Tenenbaum, J. B., Kemp, C., Shafto, P. (in press). Theory-based Bayesian models for inductive reasoning. In A. Feeney and E. Heit (eds.), Induction. Cambridge University Press.

Tenenbaum, J. B., Sobel, D. M., Griffiths, T. L., & Gopnik, A. (submitted). Bayesian reasoning in adults' and children's causal inferences.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. Annual Review of Psychology, 43, 337-375.

Wellman, H. M., & Gelman, S. A. (1997). Knowledge acquisition in foundational domains. In D. Kuhn & R. Siegler (Eds.), Handbook of child psychology (5th Ed). New York: Wiley.

Xu, F. & Tenenbaum, J. B. (2005). Word learning as Bayesian inference: Evidence from preschoolers. In Proceedings of the 27th annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum Associates.

Xu, F. & Tenenbaum, J. B. (in press). Word learning as Bayesian inference. Psychological Review.

Xu, F. & Tenenbaum, J. B. (this issue). Sensitivity to sampling in Bayesian word learning. Developmental Science.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? Trends in Cognitive Sciences, 10, 301-308.

Notes

[1] Conditional and unconditional dependence and independence are defined as follows. Two variables *X* and *Y* are unconditionally independent in probability if and only if for every value *x* of *X* and *y* of *Y* the probability of *x* and *y* occurring together equals the unconditional probability of *x* multiplied by the unconditional probability of *y*: $P(x, y) = P(x)\,P(y)$. Two variables *X* and *Y* are independent in probability conditional on some third variable *Z* if and only if for every value *x*, *y*, and *z* of those variables, the probability of *x* and *y* given *z* equals the conditional probability of *x* given *z* multiplied by the conditional probability of *y* given *z*: $P(x, y \mid z) = P(x \mid z)\,P(y \mid z)$.

[2] An outside intervention on a variable *X* can be captured formally by adding a new variable *I* to the network that obeys these conditions: it is exogenous (not caused by other variables in the graph); it directly fixes *X* to some value; and it does not affect the values of any other variables in the graph except through its influence on *X*.