Cognitive Development 27 (2012) 455–480

Contents lists available at SciVerse ScienceDirect

Cognitive Development

Theory learning as stochastic search in the language of thought

Tomer D. Ullman[a,*], Noah D. Goodman[b], Joshua B. Tenenbaum[a]

[a] *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, USA*
[b] *Department of Psychology, Stanford University, Stanford, USA*

ARTICLE INFO

Keywords:
Conceptual change
Bayesian models
Algorithms
Language of thought
Intuitive theories

ABSTRACT

We present an algorithmic model for the development of children's intuitive theories within a hierarchical Bayesian framework, where theories are described as sets of logical laws generated by a probabilistic context-free grammar. We contrast our approach with connectionist and other emergentist approaches to modeling cognitive development. While their subsymbolic representations provide a smooth error surface that supports efficient gradient-based learning, our symbolic representations are better suited to capturing children's intuitive theories but give rise to a harder learning problem, which can only be solved by exploratory search. Our algorithm attempts to discover the theory that best explains a set of observed data by performing stochastic search at two levels of abstraction: an outer loop in the space of theories and an inner loop in the space of explanations or models generated by each theory given a particular dataset. We show that this stochastic search is capable of learning appropriate theories in several everyday domains and discuss its dynamics in the context of empirical studies of children's learning.

> *If a person should say to you "I have toiled and not found", don't believe. If they say "I have not toiled but found", don't believe. If they say "I have toiled and found", believe. - Rabbi Itz'hak, Talmud*

For the Rabbis of old, learning was toil, exhausting work – a lesson many scientists appreciate. Over recent decades, scientists have toiled hard trying to understand learning itself: what children know

---

[*] Corresponding author. Tel.: +1 617 452 3894.
E-mail address: tomeru@mit.edu (T.D. Ullman).

0885-2014/$ – see front matter
http://dx.doi.org/10.1016/j.cogdev.2012.07.005

when, and how they come to know it. How do children go from sparse fragments of observed data to rich knowledge of the world? From one instance of a rabbit to all rabbits, from occasional stories and explanations about a few animals to an understanding of basic biology, from shiny objects that stick together to a grasp of magnetism – children seem to go far beyond the specific facts of experience to structured interpretations of the world.

What some scientists found in their toil is themselves. It has been argued that children's learning is much like a kind of science, both in terms of the knowledge children create, its form, content, and function, and the means by which they create it. Children organize their knowledge into intuitive theories – abstract coherent frameworks that guide inference and learning within particular domains (Carey, 1985, 2009; Gopnik & Meltzoff, 1997; Murphy & Medin, 1985; Wellman & Gelman, 1992). Such theories allow children to generalize from given evidence to new examples, make predictions and plan effective interventions on the world. Children even construct and revise these intuitive theories using many of the same practices that scientists do (Schulz, 2012b): searching for theories that best explain the data observed, trying to make sense of anomalies, exploring further and even designing new experiments that could produce informative data to resolve theoretical uncertainty, and then revising their hypotheses in light of the new data.

Consider the following concrete example of theory acquisition, which we return to frequently. A child is given a bag of shiny, elongated, hard objects to play with and finds that some pairs seem to exert mysterious forces on each other, pulling or pushing apart when they are brought near enough. These are magnets, but she doesn't know what that would mean. This is her first exposure to the domain. To make matters more interesting, and more like the situation of early scientists exploring the phenomena of magnetism in nature, suppose that all of the objects have an identical metallic appearance, but only some of them are magnetic, and only a subset of those are actually magnets (permanently magnetized). She may initially be confused trying to figure out what interacts with what, but like a scientist developing a first theory, after enough exploration and experimentation, she might start to sort the objects into groups based on similar behaviors or similar functional properties. She might initially distinguish two groups, the magnetic objects (which can interact with each other) and the nonmagnetic ones (which do not interact). Perhaps then she will move on to subtler distinctions, noticing that this very simple theory doesn't predict everything she observes. She could distinguish three groups, separating the permanent magnets from the rest of the magnetic objects as well as from the nonmagnetic objects and recognizing that there will only be an interaction if at least one of the two magnetic objects brought together is a permanent magnet. With more time to think and more careful observation, she might even come to discover the existence of magnetic poles and the laws by which they attract or repel when two magnets are brought into contact. These are but three of a large number of potential theories, varying in complexity and power, that a child could entertain to explain her observations and make predictions about unseen interactions in this domain.

Our goal here is to explore computational models of how children might acquire and revise an intuitive theory such as this on the basis of domain experience. Any model of learning must address two kinds of questions: what and how? Which representations can capture the form and content of what the learner comes to know, and which principles or mechanisms can explain how the learner comes to know it, moving from one state of knowledge to another in response to observed data? Here we address the 'how' question. We build on much recent work addressing the 'what' question, which proposes to represent the content of children's intuitive theories as probabilistic generative models defined over hierarchies of structured symbolic representations (Kemp, Goodman, & Tenenbaum, 2008b; Tenenbaum, Griffiths, & Kemp, 2006; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Previously the 'how' question has been addressed only at a very high level of abstraction, if at all: The principles of Bayesian inference explain how an ideal learner can successfully identify an appropriate theory, based on maximizing the posterior probability of a theory given data (as given by Bayes' rule). But Bayes' rule says nothing about the processes by which a learner could construct such a theory or revise it in light of evidence. Here our goal is to address the 'how' of theory construction and revision at a more mechanistic, process level, exploring cognitively realistic learning algorithms. Put in terms of Marr's (1982) three levels of analysis, previous Bayesian accounts of theory acquisition have concentrated on the level of computational theory, while here we move to the algorithmic level of analysis,