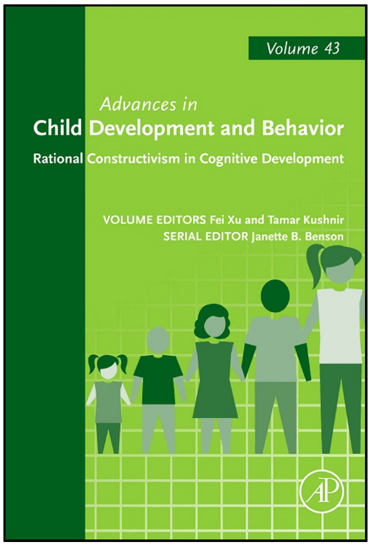
Volume 43
Advances in
Child Development and Behavior
Rational Constructivism in Cognitive Development
VOLUME EDITORS Fei Xu and Tamar Kushnir
SERIAL EDITOR Janette B. Benson
AP

From Schulz, L., (2012). Finding New Facts; Thinking New Thoughts. In J. B. Benson (Serial Ed.) & F. Xu & T. Kushnir (Vol. Eds.), *Rational Constructivism in Cognitive Development* (pp. 269–294). Elsevier Inc.: Academic Press.
ISBN: 9780123979193

Academic Press

CHAPTER TEN

Finding New Facts; Thinking New Thoughts

Laura Schulz

Department of Brain aS100nd Cognitive Sciences, Massachusetts Institute of Technology, Boston, Massachusetts E-mail: lschulz@mit.edu

Contents

Abstract

The idea of the child as an active learner is one of Piaget's enduring legacies. In this chapter, I discuss the ways in which contemporary computational models of learning do, and do not, address learning as an active, child-driven process. In Part 1, I discuss the problem of search and exploration. In Part 2, I discuss the (harder and more interesting) problem of hypothesis generation. I conclude by proposing some possible new directions for research.

Constructivism is a clunky word. Arguably, however, only such a ponderous term could stand up to those venerable pillars of epistemology: nativism and empiricism. In contrast to both, Piaget insisted that learning was driven by interactions between the child's representations and her experience of the environment. Today, we can express this insight with mathematical precision; prior hypotheses constrain our interpretation of evidence and affect whether and how we revise our beliefs from evidence (see Tenenbaum, Kemp, Griffiths, & Goodman, 2011, for exposition and review). In adding

Advances in Child Development and Behavior, Volume 43
ISSN 0065-2407,
http://dx.doi.org/10.1016/B978-0-12-397919-3.00010-1

Rational Constructivism in Cognitive Development, First Edition, 2012, 269–294

clarity and rigor to concepts like accommodation and assimilation, computational models introduced rational analyses to constructivism and motivated much of the enterprise to which this volume pays homage. A decade of empirical support speaks to the success of this approach (Gopnik & Wellman, 2012; Schulz, 2012). In this chapter, however, and mindful of the mandate to reflect critically on "needed theoretical, technological, and empirical advances," I will focus on an idea that is critical to the constructivist vision but largely missing from contemporary accounts of learning: the idea of the child as an active learner.

Consider this passage, written in 1937:

> *Such construction is not the act of an a priori deduction, nor is it due to purely empirical gropings. The sequence … testifies much more strongly to progressive comprehension than to haphazard achievements. If there is experimentation, the experiments are directed.*
>
> ***Piaget, The Construction of Reality in the Child.***

The excerpt refers to the development of the infants' object concept. Never mind the details (which are wrong; Baillargeon & Luo, 2002; Spelke, 1999; Spelke, Brinlinger, Macomber, & Jacobson, 1992). For my purposes, Piaget's central claim is not a less rigorous, less precise instantiation of the Bayesian idea that prior knowledge and evidence interact. The central claim is that the child actively seeks to understand the world.

What might *active learning* mean? It could mean what we sometimes mean by "hands-on learning": that children like to do things and that the things they do can sometimes generate evidence that supports new inferences. Such activities, however, presumably fall under the purview of "empirical gropings." Piaget's claim is stronger. He suggests that the child generates hypotheses about how the world works and that the child's actions—starting with literal manipulations of objects but ending in "cognitive acts" ranging from mental rotation to thought experiments—are systematic attempts to understand how the world works.

As someone who putatively works on exploration and active learning in early childhood, I find it hard to overstate the degree to which this vision of active learning is absent from current research on cognitive development (my own included). Even my writing betrays this; I find myself repeatedly opting for periphrastic locutions ("make inferences"; "distinguish hypotheses") over verbs that more clearly ascribe intentional activity to the child: "thought," "decided," "wondered," and "tried." To the degree that grammar is "the metaphysics of the people" (Nietzsche, 1882, 1974), I would seem to be an agnostic about active learning.

This grammatical timidity is due in part to the problematic nature of making claims about children's internal states. However, I think the phrasing also accurately reflects the current state of our theories of learning. It is one thing to talk about how learners draw rational inferences from data; it is another to suggest that children actively work to construct new knowledge. The latter commitment poses at least two problems that current accounts of cognitive development elide.

First, accounts of cognitive development have focused primarily on problems of inductive inference. Principles of induction have implications for but do not directly address problems equally critical to learning: problems of search, exploration, or decision making. Thus, our current accounts of constructivism tend to stop precisely where the active part of active learning begins.

The second and to my mind more fascinating problem (impatient readers should skip directly to Problem Two) is that, we routinely generate new ideas without having access to new data. With all due respect to the innovative proposals currently in play, we still do not understand how learners think of new ideas. The hard part of this problem goes beyond a search problem. I think a precise formal solution to this problem is a ways off but I will talk about what I think is missing from the current proposals and suggest possible new directions for research.

1. PROBLEM 1: EXPLORING

Why do we explore? Intuitively, we explore either when we encounter something surprising or when we encounter something (even a perfectly ordinary something) that we cannot explain. Bayes' law can illustrate the common principle underlying these two seemingly quite different motivational states.

Bayes' law states that the learner's belief in a hypothesis after observing evidence, the *posterior probability of the hypothesis*, P(h|e), is proportional to both to its *likelihood*, P(e|h), the probability that the hypothesis, if true, would have generated the observed evidence, and its *prior probability*, P(h), the probability that the hypothesis is generated by the learner's background theories. Formally: $P(h|e) \propto P(e|h)P(h)$.

If the posterior probability (the probability of the hypothesis given the evidence) of two or more hypotheses is approximately equivalent ($P(h1|e) \approx P(h2|e)$), the learner will be uncertain which hypothesis is true. This can occur either if the prior probability favors one hypothesis and the likelihood another

($P(e|h1) < P(e|h2)$ and $P(h1) > P(h2)$) or if the prior probability and the likelihood of multiple hypothesis are equivalent (e.g. $P(h1) \approx P(h2)$ and $P(e|h1) \approx P(e|h2)$). The first is a formalization of what it means for evidence to be surprising; the second is a formalization of what it means for evidence to be confounded. Thus Bayes' law provides an intuitive account of why exploration in the face of surprise and confounding derive from a common principle.

So far, so good, and numerous empirical studies attest to the fact that children selectively explore when confronted either with theory-violating evidence (Bonawitz, van Schijndel, Friel, & Schulz, 2012) or with evidence that is confounded (Schulz & Bonawitz, 2007). Patently, however, both children and adults can experience inductive uncertainty without engaging in exploration. Understanding when and why children do or do not engage in exploration will require understanding both how children decide when exploration is valuable and how children know what exploratory actions to take. These processes are not independent (e.g. the learner's assessment of the value of exploration depends on her assessment of the availability of potentially informative actions). However, progress has been made on each of these fronts across quite different disciplines, suggesting the possibility that an integrated approach to understanding exploration could predict and explain more of children's behavior.

1.1. Knowing When to Explore

When children explore, there are other things they are not doing. Children have to decide when the potential advantages of exploration exceed the costs. A number of fields, including machine learning (Gittens, 1979; Kaebling, 1993; Kaelbling, Littman, & Moore, 1996; Kaelbling, Littman, & Cassandra, 1998; Tong & Koller, 2001), decision making (Sutton & Barto, 1998), neuroscience (Daw, Niv, & Dayan, 2005; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; McClure, Daw, & Read Montague, 2003), and ethology (Charnov, 1976, 2006; Krebs, Kacelnik, & Taylor, 1978; Stephens & Krebs, 1996), have proposed resolutions to such exploration/exploitation dilemmas. These accounts suggest search strategies that consider various reward functions and constraints on the organism and maximize the expected cost to benefit ratio of staying in a given state relative to transitioning to a new one (predicting, for instance, that organisms should stay longer at a food patch as the distance between patches increases; Charnov, 1976).

If applied to problems of learning in early childhood, comparable approaches might help predict and explain children's exploratory behavior beyond what can be explained by epistemic uncertainty alone. It seems intuitive, for instance, that children will be more likely to explore in contexts where there are just a few plausible hypotheses than when there are many. However, formalizing even such simple intuitions requires integrating problems of inductive inference with a consideration of the relative value of exploratory actions (e.g. if the same action can eliminate a single hypothesis from consideration across cases, the potential for information gain is greater when hypotheses are fewer).

Typically, however, solutions to exploration/exploitation dilemmas have been proposed, not for challenging learning problems but for cases where the epistemic component is relatively straightforward. In particular, optimal search processes have been developed to maximize rewards when the distribution of rewards is uncertain but arbitrary (e.g. rewards distributed among slot machines with different payoffs (Daw et al., 2006; Gittins, 1989; Strehl et al., 2006), decks of cards with different values (Bechara, Damasio, Tranel, & Damasio, 1997; Sang, Todd, & Goldstone, 2011), or food patches with different caloric and nutritional worth (Kacelnik & Bateson, 1996; Stephens & Krebs, 1986).[1] Although the organism's search may be affected by rational considerations of expected costs and benefits, such search processes nonetheless arguably remain closer to "empirical groping" than to constructivism. The search process is not random but neither is it guided by an abstract theory of the domain. Arbitrary distributions of rewards are unlikely to lend themselves to "progressive comprehension."

Recently, however, researchers in machine learning have begun to consider how search might proceed in domains that support more structured representations. Robots, for instance, may start with a map of the terrain and search for efficient routes to a goal within the terrain (Leonard & Durrant-Whyte, 1991). Researchers have also begun to address chicken-and-egg problems of exploration: designing robots that can simultaneously use a map to evaluate the expected utility of various state transitions and use the information gained during exploration to revise the map (Durrant-Whyte & Bailey, 2006; Thrun, Burgard, & Fox, 2005). Although these particular approaches only solve these problems for finite two-dimensional spaces, they offer a hint as to how we might begin to formalize the idea of theory-guided and theory-shaping exploration in higher dimensional spaces.

[1] Of course, the actual distribution of calories across patches is not arbitrary; in the case of food patches, it is more accurate to say that the foraging animal is presumably ignorant of the biological and ecological factors that affect the distribution of rewards.

Such approaches would seem to lend themselves well to an understanding of the exploratory aspect of constructivism. If we can consider the relative utility of competing courses of action in the context of hierarchical representations of the domain being explored, we might be able to better explain both when exploration is likely to occur and how the child's exploration is likely to transform the child's knowledge. A full analysis of the relative costs and benefits of different actions may be intractable, such an account would have to consider not only the value of non-exploratory behaviors—playing, eating, daydreaming—to the child but also the effect of the child's culture, temperament, upbringing, and individual interests on how she perceives the value of the information she might gain through exploration. Nonetheless, advancing our understanding of how children assess the relative value of information gain seems critical, given that competing utilities can have determinative effects on learning.

1.2. Knowing How to Explore

Understanding the expected utility of exploratory actions helps answer the question of *when* the child should engage in exploration. However, the child must know not only that there is information to be gained but also *how* precisely to gain it. Even the simplest forms of exploratory behavior raise questions about how humans (Adolph, Eppler, & Gibson, 1993; Berger, Adolph, & Lobo, 2005; Brown, 1990; Gibson, 1977; Lockman, 2000; Norman, 1988, 1999) and other animals (Brauer, Kaminski, Reidel, Call, & Tomasello, 2006; Emery & Clayton, 2004; Hood, Carey, & Prasada, 2000; Mendes, Hanus, & Call, 2007; Stulp, Emery, Verhulst, & Clayton, 2009) learn to recognize the possible actions that the environment affords. Nonetheless, in cases where there is a direct mapping between an action and information gain, the question of how to explore has a relatively straightforward answer: act on the entity with greatest uncertainty (e.g. by pulling a lever, putting a block on a machine, or lifting a card to learn its value; see Oaksford & Chater, 1994).

Sometimes, however, no single action available to the learner will support information gain. The learner may have to plan a complex series of actions in order to isolate variables or may have to resign herself to the fact that isolating the relevant variables is impractical or impossible. Effectively generating informative evidence requires combining an understanding of the probability of information gain together with an understanding of the affordances that might permit it.

In principle, children might both learn from informative evidence (see Gopnik & Wellman, 2012; Schulz, 2012, for review) and engage in exploration when they observe uninformative evidence (Schulz & Bonawitz, 2007) without understanding either what it is about evidence that makes it informative or how to generate such evidence. However, we now know that at least in very simple contexts, children have both of these abilities. Preschoolers, for instance, not only selectively explore when they are uncertain which of two connected beads activate a toy, they also separate the beads and test each one individually. Moreover, if the beads cannot be detached, children orient the connected bead so that only a single bead makes contact with the toy at a time (Cook, Goodman, & Schulz, 2011). Thus, preschoolers seem to understand not only when there is potential for information gain but also the probability that particular actions will generate data relevant to a particular hypothesis (see also Sodian, Zaitchik, & Carey, 1991).

Needless to say, the real world rarely makes it so easy. Part of what distinguishes science from cognition more broadly is the cultural accumulation of tools and knowledge that can support information gain in ways that go well beyond naive exploration. However, at least in simple contexts, researchers have made progress in analyzing and formalizing the cognitive processes involved in optimizing information gain (Cook et al., 2011; Klayman, 1988; Klayman & Ha, 1987; Oaksford & Chater, 1994; Sobel & Kushnir, 2006; Steyvers, Tenenbaum, Wagonmakers, & Blum, 2003). To the degree that we can show how these abilities manifest in infants and young children, we may come closer to understanding of what it means for children to actively construct knowledge.

2. PROBLEM 2: THINKING

Thus far I have focused on how children might construct new knowledge by gathering more data. Arguably, however, the distinguishing attribute of human cognition is that we can arrive at new ideas—some of which turn out to be true—merely by thinking of them. How is this possible?

As noted, hierarchical Bayesian inference models provide an elegant account of how learners integrate theories on different levels of abstraction with the interpretation of new evidence. They thus provide a way of thinking about some key points of comparison between scientific inference and cognition in early childhood (see Gopnik, in press; Gopnik, & Wellman, 2012; Schulz, 2012). In light of the ways that probabilistic inference models have revolutionized cognitive science (Tenenbaum et al., 2011), it might seem churlish to suggest that they do not really address the core issue at the heart of constructivism. However, these models explain how learners select among competing hypotheses; for the most part, they do not attempt to explain how learners construct hypotheses in the first place. They tell us how we might choose the best idea from the ideas we have but they do not tell us how we might think of something new.

Note that the problem of how we think of new ideas is not reducible to the (also interesting) problem of how to make Bayesian inference algorithmically tractable (Bonawitz chapter, this volume; Bonawitz & Griffiths, 2010; Sanborn, Griffiths, & Navarro, 2010; Shi, Feldman, & Griffiths, 2008).[2] As Bonawitz et al. note, in simple cases, all the relevant hypotheses may be available to the learner in principle if not at any given moment. Given, for instance, three colors of chips that activate a machine, the learner might believe the machine is activated only by red chips but can also entertain the possibility that it is activated only by blue chips, only by green chips, only by red and blue chips, etc. Monte Carlo (randomized, sampling-based) algorithms provide an efficient way to search among the hypotheses. The idea of inference by rationally randomized sampling (i.e. sampling from hypotheses with the highest posterior probability) also reconciles the empirical evidence of variability in children's learning with ideal Bayesian analysis; the learner can entertain only a single hypothesis at a time and nonetheless converge on the correct hypothesis. However, efficiently sampling hypotheses is not the same as constructing them.

3. THEORY-GUIDED STOCHASTIC SEARCH

A closer approximation to constructivist learning comes from a family of computational models, which suggest that learners have a "grammar" for generating potentially infinitely many hypotheses (Goodman, Ullman, &

[2] The problem of how we think of new ideas is also distinct from the so-called "old evidence" problem (Glymour, 1980). The supposed problem is that you can't learn anything from old evidence because once evidence is known, it has a prior probability of $1(p(e) = 1)$, therefore also a likelihood of $1(p(e|h) = 1)$ and a posterior probability of $1(p(h|e) = 1)$. This would contradict our intuition that, for instance, well-known anomalies in Mercury's orbit provided evidential support for Einstein's theory of relativity. There are many responses to this (Eels, 1982; Garber, 1983; Howson, 1985), most of which dispute the grounds for assuming that p(e) and $p(e|h) = 1$. However, the point here is simply that the old evidence problem is different from the problem of how we generate new hypotheses in the first place. Thanks to the editors for suggesting that I draw this distinction.

Tenenbaum, 2011; Tenenbaum, Griffiths, & Niyogi, 2007; Ullman, Goodman, & Tenenbaum, 2012). If the learner's initial theory fails to account for the data in some respect, she can engage in a sampling-based stochastic search process, proposing randomized changes to the hypothesis, constrained by the prior probability that the grammar will generate the new hypothesis. Efficient learning is further enabled by "templates": grammatical predicates that encode common logical, causal, or constitutive relations (e.g. transitive relationships among variables). Rather than only constructing new hypotheses piecemeal, the learner can sample from the space of existing templates. The stochastic search through the "outer loop" of the hypothesis space is then grounded out by a search through an "inner loop," testing how well the new candidate hypothesis applies to the learner's observations. If the new hypothesis predicts the data better than the previous hypothesis, the new hypothesis is likely to be accepted. In simulated experiments, this approach shows dynamic features commensurate with what we know about children's learning: individual learning curves are variable but learning on average is predictable, often following characteristic sequences of transitions and typically proceeding from simpler to more complex hypotheses.

This approach represents an exciting and welcome development. Rather than simply describing data-driven learning, the account of theory acquisition as stochastic search explains how structured representations can change in the absence of new evidence. In this, it seems to capture some of what we mean by "thinking." And may be search by means of random variation and selection is enough. Certainly, evolution testifies to the power of random variation, together with the re-use of components that have functioned well in the past. Perhaps, thought does not require any more intelligence in its design than life itself.

Perhaps. However, with only the minimal constraints of simplicity, grammaticality, and previously productive templates, changes to hypotheses generated by random variation seems at best inefficient. More importantly, our minds seem to have access to rich sources of information that could better constrain the process of hypothesis generation and that current approaches do not exploit. I will discuss these additional possible constraints on hypothesis generation in the hopes that they might inspire new directions for both computational and empirical research. As will be obvious, all the ideas to follow are shamelessly speculative. I am taking advantage of the genre of "chapter" rather than "journal paper" to advance ideas that are just at their inception. However, begging the reader's indulgence for the paucity of detail, I hope this attempt to think new thoughts about thinking new thoughts might at least spark conversation.

4. ABSTRACT ERROR MAPS AS CONSTRAINTS ON HYPOTHESIS GENERATION

The first kind of knowledge that could help guide a learner's search for new ideas is an abstract representation of the flaws in her current theory, what we might call an "error map." Here, I mean not a record of individual prediction errors but a more abstract inference drawn from them: a representation of the kind of errors being made and the relationship among those errors. That is, much as the learner can draw inferences from individual events to abstract categories and relationships, the learner might draw inductive inferences from specific prediction errors to general kinds of errors and the relations among them.

I will spell this idea out in the examples to follow. First, however, I want to stress that my use of the phrase "error map" is provisional. As I discuss below, the gaps that matter can occur not just between hypotheses and data but also between hypotheses and explanatory desiderata or between hypotheses and functional goals. Thus, the notion of error here refers to something less like a prediction error and more like the learner's subjective error signal and her abstract representation of why her current hypothesis is unsatisfactory. It might be better to think of these as "gap maps," "goodness-of-fit maps," or simply "maps of our discontents." Nonetheless, in many cases, these gaps present as prediction errors (including both events that the learner predicted wrongly and those she failed to predict at all), so for the time being I will stick with the notion of error maps.

In almost any conventional approach to learning (whether connectionist or Bayesian; McClelland, 1988; Munakata & McClelland, 2003; Ullman et al., 2012), prediction errors inform hypothesis selection. All else being equal, learners will retain new hypotheses that improve the fit to the data and reject those that do not. However, even in theory-guided stochastic search (Ullman et al., 2012), prediction errors are put to use only at the stage of hypothesis selection, *after* a new hypothesis has been generated. Suppose instead that learners could use a representation of the gaps between their current hypotheses and the evidence to constrain the process by which they generate new ideas in the first place. If so, many simple, grammatical changes to the current hypothesis that would be randomly generated only to be rejected will not even be attempted; the constraints imposed by such an abstract error map might mean that learners could recognize a priori that most of the new ideas she could generate are unlikely to solve her problem.

5. AN EXAMPLE: MAKING SENSE OF NOISE

Imagine, for instance, a child who knows many things about airplanes and many things about phones. However, nothing in her intuitive theory of airplanes or phones predicts that passengers will be asked to turn off their phones on take-off. She does not have the concept of radio interference and (like many of us, albeit for rather different reasons) she wants to understand why you have to turn off your phone on the airplane. Now imagine that the child's only option is to randomly add or delete simple, logical predicates that have a high probability given her prior beliefs about airplanes and phones. She could connect the two artifacts with infinitely many simple constitutive, causal, or other relational claims consistent with her prior beliefs and expressible in the grammar of her theories: perhaps airplanes, like phones, have push buttons; perhaps airplanes and phones are both manufactured in Ohio; perhaps airplanes fly over the earth, and the earth has phones, so airplanes sometimes fly over phones. On any account, these hypotheses, once generated, will be swiftly rejected because none predicts that you should shut off your phone when flying. But the odds of converging on a valuable new idea through this kind of process seem, prima facie, low.

Suppose instead that the child can constrain the space in which she generates hypotheses by availing herself of an abstract representation of the problem: the unanticipated but evident incompatibility between planes and phones. She need not bother hypothesizing that planes and phones have infinitely many commensurable features. She can randomly generate only hypotheses in which some feature of planes is in conflict with some feature of phones. In this way, she might selectively generate hypotheses that are recognizably "good" hypotheses (in that, if true, they would solve the problem), even though they might not be "good" with respect to their truth value. Consider, for instance, the following (true) anecdote:

> Adele (age 4): "Mommy, I know why they make you turn off your phone when the plane is taking off."
>
> Me: "Oh really? Why?"

Adele: "Because when the plane takes off it's too noisy to talk on the phone."

6. GOOD WRONG IDEAS AND BAD ONES

This is of course wrong; it is wrong even about the direction of causality. At the same time, it is recognizably a good hypothesis. And it is a remarkable feature of human cognition that we can simultaneously recognize the "goodness" of an idea and its falsity. I suggest that we can do this is because we evaluate a new idea first on the extent to which it is consistent with the constraints of our abstract error map and only subsequently on its truth value or fit to the data.

The idea that a learner first generates and evaluates hypotheses through the constraints of an abstract error map (and only secondarily through an "inner-loop" checking the degree to which the hypothesis accounts for the data) predicts many features of human cognition that seem intuitively to be true. For instance, we seem to know that we are on the right (or wrong) track in thinking about a problem well before we know whether our ideas generate better fits to the data. Arguably this is because the ideas we generate may have (or lack) key features of the abstract form of the solution to a problem well before they are in fact solutions to the problem. Similarly, we seem to have an internal "stopping function" that lets us know we have arrived at a good idea (or lets us dismiss an idea out of hand) well before we have tested its predictions. Our "ah ha" moments (see Gopnik, 2000) can come months, even years, before we have any evidence that our great new breakthrough idea is true; indeed, even when our great new breakthrough turns out to be false, it might only slightly diminish our sense of its brilliance. This is reasonable if our criterion for the elegance of an idea is its congruence with an error map, rather than with the world. Finally, we seem to have an intuitive sense of how tractable problems are, even in cases where tractability does not reduce to technological or resource limitations. In such cases what it might mean for a problem to be tractable is that the representation of the problem—the abstract error map—sufficiently constrains the search space for new hypotheses. If this account is correct then learning might sometimes be facilitated not by changing our prior beliefs or the evidence but merely by changing how we represent the gap between them. As one outstanding generator of new hypotheses put it: "The formulation of the problem may be more important than the solution, which may simply be a matter of mathematical or experimental skill" (Einstein, in Chang, 2006, p. 179).

7. ABSTRACT ERROR MAPS AND VARIABILITY IN LEARNING

As researchers have noted (see Bonawitz chapter, this volume), sampling-based approaches to Bayesian inference can help account for individual differences in children's learning. It is interesting to consider how adding error maps as a top–down constraint on hypothesis generation might affect individual variability in learning.

Imagine, for instance, two children, Jane and Michael. Borrowing and extending the Ullman et al. (2012) example, suppose both children have an incorrect theory of magnetism. The children are playing with magnets, paperclips, and pennies but believe that they are playing with magnets and non-magnets; they have failed to consider the possibility that paperclips belong in a third category: ferrous non-magnets. Suppose further, that the children have mis-categorized the paperclips in different ways. Jane (who has noticed some magnetized paperclips sticking together) has classified the paperclips as magnets; Michael has classified them with the pennies, as non-magnets. The children's different subtheories generate slightly different prediction errors. Jane wrongly predicts that any paperclip will interact with any other paperclip. When she gets data to the contrary, she will have to explain how magnetism might sometimes disappear. Michael wrongly predicts that no paperclip will interact with any other paperclip; he will have to explain how magnetism could sometimes appear.

This asymmetry may lead the children to generate quite different abstract representations of the problem and different abstract criteria that constrain their search for new ideas. Jane is trying to generate new hypotheses that satisfy the (perceived) desiderata of including a variable whose value can diminish over time. Thus, Jane may come up, for instance, with the idea that magnetism is a kind of energetic force that (like the energy in batteries) sometimes runs out. Michael, by contrast, represents the problem as a problem of explaining the unexpected appearance of a rare property (rather than the less surprising problem of explaining its diminishment or disappearance). Michael may thus be faster to recognize that only objects made of specific materials can become magnetized and that they can be magnetized only immediately after the relatively rare event of contacting a magnet.

Each of these different error maps may lead to new hypotheses that get different things right and wrong. Jane may correctly think of magnetism as an energetic force but overlook the role of the particular materials; Michael may recognize that the property of magnetism can be transferred from some materials to others but fail to subsume magnetism into the more general category of a kind of energy. In short, because learners' theories constrain what they represent as errors or gaps in their understanding, learners with even subtly different theories may generate different abstract representations of the problems they are trying to solve. To the degree that these different abstract error maps constrain the generation of new hypotheses, differences in the ways that learners represent the problem they are solving could lead to quite different learning outcomes.

8. ABSTRACT ERROR MAPS, QUINIAN BOOTSTRAPPING, AND ANALOGICAL REASONING

The ideas behind theory grammars and stochastic search (Goodman et al., 2011; Ullman et al., 2012) were themselves partly motivated by another account of how learners might move beyond hypothesis selection: Quinian bootstrapping (Carey, 2009). Quinian bootstrapping is a proposal for how learners might generate genuinely new representational resources. Two key ideas are critical to the account. Quinian bootstrapping depends first on the learner having access to explicit symbols (e.g. through language or mathematics). These enable the learner to develop representations whose meaning is genuinely novel in that it inheres in the relationship among the symbols rather than only in earlier concepts (see also Block, 1986). The learner can then use these to construct "placeholder" representations that support inductive inferences about the specific role and meaning of the new concepts.[3] For instance, a child may notice the similarity between the order of words in the count list and words corresponding to larger analog magnitudes. The words then serve as a placeholder representation allowing the child to bootstrap an explicit representational system in which she infers the meaning of the number words.

[3] The computational proposals echo this insofar as variables get their meaning from their relationship to other terms in the theory grammar and serve as placeholder concepts (Goodman et al., 2011; Ullman et al., 2012).

This brief description in no way does justice to the work (see Carey, 2009, for exposition and review; see also Herme & Spelke, 1996 and Spelke, 2003). Here I merely want to note that the current proposal is both indebted to and compatible with these ideas. In the number example, for instance, neither the child's earliest understanding of the count list nor her representation of analog magnitudes predicts any similarity between the two. Insofar as the child is able to constrain the new ideas, she generates to just those that posit a commonality between the count list and analog magnitudes, one could think of that as a constraint on hypothesis generation imposed by an abstract error map.

Arguably, however, the current proposal is more general than the case of Quinian bootstrapping in two respects. First, learning need not depend on the learner's exposure to explicit linguistic or mathematical symbols nor the learner happening to notice analogical mappings between representations. To the degree that the learner can formulate an abstract representation of the gaps between her current hypotheses and the evidence (i.e. by categorizing the problem as one involving an unexpected conflict between two variables, a diminishing property, an appearing property), she might constrain the new hypotheses she generates to those that might fill the gap.

Second, real discontinuities in development (e.g. manifest in the development of the child's understanding of number or the development of the child's ability to differentiate weight and density; see Carey, 2009, for exposition and review) are compelling case studies in hypothesis generation. However, there are many more mundane instances of hypothesis generation (e.g. as manifest in Adele's explanation of airplane regulations) that do not involve radical conceptual change but are also not merely data driven. Even in such ordinary cases, there is a real puzzle about how learners think of new ideas. Nothing in the current account depends on the incommensurability of earlier and later ideas or requires the construction of altogether new mental symbols. Constraints imposed by abstract error maps might support the generation of new ideas quite broadly.

Finally, considering more narrowly just the role of analogical reasoning (Christie & Gentner, 2010; Gentner et al., 1997; Gentner, 2002; Gentner, Holyoak, & Kokinov, 2001; Gentner & Markman, 1997; Gentner & Namy, 1999; Gentner & Smith, 2012; Holyoak & Thagard, 1996), it seems clear that once we have an analogy—either because the relevant relationships are given to us pedagogically or because we ourselves notice a surprising coincidence in structural relations across events[4]—it constrains the hypotheses we generate (Christie & Gentner, 2010). What makes learning difficult, however, is that fruitful analogical relations are not always obvious; the critical question is how we know what kinds of events can be meaningfully compared. In principle, abstract error maps might serve as higher order constraints, constraining even the kinds of analogies we generate. If, for instance, you represent a problem as a problem involving dissipating properties, you can then consider other kinds of events that involve dissipating properties. In this way, you might arrive at an analogy, for instance, between paperclips losing their sticking power and batteries losing their charge.

Importantly, however, even when our abstract representation of a problem does *not* generate meaningful analogies, it can still effectively constrain hypothesis generation. If, for instance, we represent being asked to turn off our phone on the plane as a problem of an unexpected incompatibility between events, we can restrict our hypotheses to potential incompatibilities between phones and planes without analogical reasoning per se. Similarly, if we represent the problem of sticking paperclips as the problem of the unexpected appearance or transfer of a property, we can constrain our hypotheses to those involving specific materials or rare events without comparing these events to structurally similar ones. Thus, I suggest that analogical reasoning is an effective constraint on hypothesis generation insofar as it constrained by the more general ability to come up with an abstract representation of problems in the first place—and this representation can constrain hypothesis generation in ways that extend beyond analogical reasoning.

9. FUNCTIONAL ROLES AS CONSTRAINTS ON HYPOTHESIS GENERATION

So far I have discussed constraints on hypothesis generation that are, so to speak, epistemically respectable. Constraints imposed by abstract error maps

[4] I suggest that the learner starts with an abstract representation of a problem and this can constrain the kinds of analogies she generates. Sometimes, however, the learner may observe an unexpected structural alignment between events and register this alignment as a coincidence in need of explanation (Griffiths & Tenenbaum, 2007). If the learner happens to start with an analogy she is looking to explain (rather than starting with a problem and looking for analogies that might elucidate it), the analogy may itself support the construction of an abstract error map, constraining the learner's generation of new hypotheses to those that might explain the otherwise surprising relational alignment.

plausibly increase the probability that we will get at least some aspects of the world right. Intuitively, however, there are less truth-preserving, but arguably no less advantageous, considerations that seem to constrain the ideas we generate.

Specifically, we have goals for our ideas. We want our ideas to do things: to persuade, cajole, impress, explain, deceive, entertain, or instruct. We can readily distinguish "good" and "bad" ideas on prudential grounds independent of the extent to which they get the facts right. Given that we can evaluate ideas with respect to our goals, it seems plausible that we might also use our goals to constrain the ways we construct knowledge in the first place. To the degree that we propose randomized changes to our current hypotheses subject to the constraint of an abstract representation of what function we want our ideas to fulfill, we may not only be more likely to select but also to generate ideas that are in fact likely to do what we want them to do.

Indeed, the possibility that hypothesis generation is constrained by our goals may go a long way to explaining the diversity of ideas that human beings entertain. In science, for instance, it is a desiderata of our hypotheses that they be falsifiable. This is a functional constraint, not a normative one. It is in no way incumbent on the truth that it be falsifiable. However, if we specifically and selectively generate hypotheses that meet the goal of being falsifiable, we can substantially constrain the space of new ideas. In other disciplines, by contrast, the functional constraint on generating a hypothesis is that the idea be plausible within the social, political, and economic conditions of the day. Ideas are dismissed not for being unfalsifiable but for being "ahistorical." In the same vein, novelists may generate new ideas in proportion to the probability that they are "in character," engineers to the extent that the new idea is feasible, business executives to the degree that the new idea is profitable, and divinity students to the extent that the new idea might provide spiritual guidance or inspiration.

These of course are merely the functional desiderata of our professions. As human beings, we look for ideas to fill an even broader range of goals. Even supposing we were confronted with the same problem in all cases, we would generate different solutions depending on whether we wanted the new idea to impress a superior, entertain a crowd, teach a child, win an election, or woo a lover. Such constraints may or may not serve the function of getting the world right but they allow us to constrain the space in which we generate new ideas beyond merely the limit of whatever might be lawfully expressed in the grammar of our current theories.

If our goals are the top of our hierarchy of constraints on hypothesis generation, these goals may probabilistically generate a constraint one level lower: the criteria for fulfilling those goals. Suppose, for instance, you have the goal of wanting to get from point A to point B. Having the goal of navigation might generate a set of subordinate desiderata (e.g. to find variously, the shortest distance between two points, the fastest route between two points, the most scenic route between two points, or the route between two points most likely to run into a certain someone). These criteria might in turn be more likely to generate some hypotheses for abstract structural forms than others (e.g. two-dimensional maps may be more probable than tree structures; see Kemp & Tenenbaum, 2008).

By contrast, if our goal is explanation, then our constraints on hypothesis generation might include all the criteria that psychologists and philosophers have proposed for hypothesis selection (e.g. simplicity, non-circularity, the ability to subsume specific relations under an abstract kind of relation, an appeal to plausible causal, mechanistic relations; Bonawitz & Lombrozo, in press; Hempel & Oppenheim, 1948; Keil & Wilson, 2000; Keleman, 1999; Kitcher, 1989; Legare, Gelman, & Wellman, 2010; Lombrozo, 2006, 2007, 2012; Lombrozo & Carey, 2006; Salmon, 1984; Strevens, 2004; Woodward, 2009). Again, these criteria might be more likely to generate some hypotheses for abstract structural forms than others (e.g. in this case, tree structures may be more probable than two-dimensional maps; Kemp & Tenenbaum, 2008). It is a well-established, if somewhat mysterious, fact that explaining something to oneself can affect learning and discovery, even in the absence of new data (Amsterlaw & Wellman, 2006; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1994; Roscoe & Chi, 2007, 2008; Siegler, 2002; Williams & Lombrozo, 2009). If we do have abstract representations of what might count as good explanations with respect to a problem, and we could condition randomized changes to current hypotheses on these desiderata, we might avoid generating any number of hypotheses that are simple, plausible, grammatically lawful, and unsatisfying. One way in which explanatory desiderata may support learning is by constraining the generation of new ideas to those new ideas to those that have a chance of being, in fact, explanatory.

In pursuing the goals of navigation or explanation, we are arguably trying to get the world right. However, even when our goals are more venal or more frivolous, constraining our ideas by the extent to which they serve a functional goal need not *lessen* our sensitivity to the facts of the matter.

Once generated, any hypothesis can be subject to a "fact-checking" process that assesses the extent to which the new idea predicts observed data. Still, our willingness to accept a new hypothesis may be a function jointly of its fit to the criteria set by a desired functional role and its fit to the data. If a new hypothesis succeeds admirably at the former, it might be accepted despite substantial difficulties with the latter.

Our all-too-human ability to admire ideas for reasons other than their truth can provoke considerable and sometimes comedic, hand wringing about human irrationality. In the latter vein, Stephen Colbert coined the term "truthiness" to poke fun of what we might value in false, or at the very least unsubstantiated, ideas. However, if by truthiness we mean something like an idea's ability to fill explanatory (or other) criteria generated by a functional goal, our predilection for truthiness may be a feature, not a bug, of human cognition.

Indeed, in order to be the kind of organism who can think of new ideas at all, it may be critical that we are not overly wedded to the facts. The state of having no new ideas and no new data might be rather like gambling on a single-armed bandit machine or foraging in a landscape with a single berry patch; even if the existing payoff is low, there is little reason to explore. No idea that we do not have will, at the time of not yet having it, fit the data better than whatever we currently believe. Given the low odds that random gropings will improve our lot, if we were committed only to maximizing our best fit to the data, an idea in the hand might always be preferred to the two not yet even in the bush. If instead we have truth-independent criteria for hypothesis generation, we might be motivated to generate ideas that payoff in other ways, by being explanatory, entertaining, provocative, or useful. We can find out later if they are true.

Indeed, it is a curious feature of human cognition that the kinds of goals that lead us to pursue new ideas are often neither here nor there with respect to the significance of the ideas themselves. The colonialists did not profit any less from the Americas because Columbus discovered them in a misguided search for the West Indies. Similarly, whatever you think of the medieval monks' quest for incontrovertible proof of the existence of God, it did not diminish the magnitude of their contributions to analytic logic. It is not merely that the merit of our discoveries is independent of the merit of our motivations but that the merit of our motivations may be precisely in imposing critical constraints on our search processes and enabling discovery at all.

10. CONSTRUCTIVISM AND IMAGINATION

I'll end with a speculation about imagination. There are many ways in which human beings interact with the world that seem peculiarly divorced from reality. We confabulate explanations for our behavior, both in sickness (Gazzaniga, 1998; Phelps & Gazzaniga, 1992) and in health (Carruthers, 2009; Nichols & Stich, 2000). We develop elaborate, coherent autobiographical narratives that are false and misleading (Kopelman, 1987; Kopelman, Ng, & Van den Brooke, 1997). We fret over imagined events. We engage in pretend play as children and we daydream as adults. We report our actual dreams as stories. We create and enjoy cultural artifacts ranging from myths to movies.

These phenomena are united primarily in being puzzling. Given our considerable aptitude for exploring the real world, why is so much of human cognition devoted to the construction and contemplation of fictional ones? What advantage does unreality confer that we should find it so compelling? Researchers have long pointed to the value of being able to reason counterfactually for planning, for causal reasoning, and for novel interventions (Gopnik, 1990; Harris, 2000; Harris, German, & Mills, 1996; Weisberg & Bloom, 2009; Buchsbaum, Bridgers, Weisber, & Gopnik, 2012; Walker & Gopnik, under contract; Weisberg & Sobel, 2012). However, the demands of counterfactual reasoning would not seem to require the wanton disengagement with reality manifest across these diverse cognitive phenomena. As Jerry Fodor noted sardonically in response to Steven Pinker's suggestion that we appreciate fiction because it offers us insight into situations we might encounter in real life:

> *… what if it turns out that, having just used the ring that I got by kidnapping a dwarf to payoff the giants who built me my new castle, I should discover that it is the very ring that I need in order to continue to be immortal and rule the world? It is important to think out the options betimes, because a thing like that could happen to anyone and you can never have too much insurance (Fodor, 1998).*

Here is a different proposal. What matters about our fictions is not that they tell us the content of possible worlds or that they exercise our ability to reason through the consequences of false premises. What matters is our ability to create the false premises in the first place. Being able to disengage from data may be requisite to being the kind of creature that can go beyond data-driven learning. Indeed, it may be that thinking of new ideas requires precisely the ability to impose a kind of cognitive firewall between the criteria for constructing an idea and the criteria for verifying it. This is not to say that there are not constraints, even on our fictions (Harris, 2000; Shtulman, 2009; Weisberg & Bloom, 2009; Weisberg & Goodstein, 2009; Weisberg & Sobel, 2012; Wyman, Rakoczy, & Tomasello, 2009). However, the constraints most important for learning may be "narrative" constraints. Good narratives do not have to be true, but they do have to do all of the following: provide an abstract representation of a problem and its solution, satisfy criteria consistent with the narrative goal (e.g. perhaps by fulfilling causal and subsumptive explanatory demands), and fulfill a functional role for entertainment, persuasion, illustration, provocation, explanation, soothing, or stimulation. In short, the constraints on a good narrative are plausibly the rational constraints for hypothesis generation generally.

On this account, we spend our time telling stories for the same reason that monkeys spend their time climbing trees, not because it usually solves a problem but because it is hard to know when a problem will appear. If we engage in the activity continually, then when a problem does come along, we may find ourselves making the right leap at the right time. In engaging, from early childhood onward, in acts of fictional narratives, in telling stories about what we experience even in our sleep, in retaining this ability even in the face of devastating insults to our bodies and brains, and in finding this engagement sufficiently pleasurable that we seek it out in our cinemas, theaters, novels, and fire circles, we may be manifesting the most distinctively human aspect of our ability to learn: the ability to step away from the real world in order to better see the world as it really is.

ACKNOWLEDGMENTS

Thanks to Fei Xu and Tamar Kushnir for (respectively) talking me into starting and finishing this chapter and for many helpful ideas along the way. Thanks to Rebecca Saxe and the members of the MIT Imagination Reading Group—Eyal Dechter, Daniel Friel, Julian Jara-Ettinger, John McCoy, Max Siegel, Tomer Ullman, and Nathan Winkler-Rhoades—for hours of interesting conversation that helped constrain and generate these ideas. Finally, warm thanks to Susan Kaufman, Margot and Isaac Schulz, and Josh Tenenbaum for improving my fit to both the data and the grammar —and for helping me tell a better story.

REFERENCES

Adolph, K. E., Eppler, M. A., & Gibson, E. J. (1993). Development of perception of affordances. *Advances in Infancy Research, 8*, 51–98.

Amsterlaw, J., & Wellman, H. (2006). Theories of mind in transition: a microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7*, 139–172.

Baillargeon, R., & Luo, Y. (2002). Development of the object concept. In *Encyclopedia of cognitive science*, *3*, (pp 387–391). London: Nature Publishing Group.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*(5304), 1293–1295.

Bell, W. J. (1991). *Searching behaviour: The behavioural ecology of finding resources*. New York: Chapman/Hall.

Berger, S. E., Adolph, K. E., & Lobo, S. A. (2005). Out of the toolbox: toddlers differentiate wobbly and wooden handrails. *Child Development, 76*(6), 1294–1307.

Berry, D. A., & Fristedt, B. (1985). *Bandit problems: Sequential allocation of experiments*. London: Chapman and Hall.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French, et al. (Eds.), *Midwest studies in philosophy*, Vol. X, (pp. 615–678). Minneapolis: University of Minnesota Press.

Bonawitz, E. B., & Griffiths, T. L. (2010). Deconfounding hypothesis generation and evaluation in Bayesian models. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Bonawitz, E. B. & Lombrozo, T. (in press). Occam's rattle: children's use of simplicity and probability to constrain inference.

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. E. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*(4), 215–234.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition, 113*, 262–280.

Brauer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). Making inferences about the location of hidden food: social dog, causal ape. *Journal of Comparative Psychology, 120*(1), 38–47.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science, 14*(1), 107–133.

Buchsbaum, D., Bridgers, S., Weisberg, D., & Gopnik, A. (2012). The power of possibility: causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions, 367*(1599), 2202-2212.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carruthers, P. (2009). How we know our own minds: the relationship between mind-reading and metacognition. *Behavioral and Brain Sciences, 32*, 121–138.

Chang, L. (2006). *Wisdom for the soul: Five millennia of prescriptions for spiritual healing*. Washington: Snosophia.

Charnov, E. L. (1976). Optimal foraging: the marginal value theorem. *Theoretical Population Biology, 9*, 129–136.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356–373.

Cook, C., Goodman, N., & Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition, 120*(3), 341–349.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*, 1704–1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*, 876–879.

Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping (SLAM): part I the essential algorithms. *Robotics and Automation Magazine, 13*(2), 99–110.

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science, 306*(5703), 1903–1907.

Fodor, J. (1998). *In critical condition: Polemical essays on cognitive science and the philosophy of mind. Representation and mind*. Cambridge, Mass: MIT Press.

Gazzaniga, M. S. (1998). The split brain revisited. *Scientific American, 279*(1), 35–39.

Gentner, D., Brem, S., Ferguson, R. W., Wolff, P., Markman, A. B., & Forbus, K. D. (1997). Analogy and creativity in the works of Johannes Kepler. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403–459). Washington, DC: American Psychological Association.

Gentner, D. (2002). *Analogical reasoning, psychology of encyclopedia of cognitive science*. London: Nature Publishing Group.

Gentner, D., Holyoak, K. J., & Kokinov, B. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.

Gentner, D., & Smith, L. (2012). Analogical reasoning. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (2nd ed.) (pp. 130–136). Oxford, UK: Elsevier.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (Ed.), *Testing scientific theories, minnesota studies in the philosophy of science*, Vol. X. Minneapolis: University of Minnesota Press.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Erlbaum.

Griffiths, T. L., & Tenenbuam, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition, 103*(2), 180–226.

Gittins, J. (1989). *Multi-armed bandit allocation indices*. Wiley.

Gittins, J. C., & Jones, D. M. (1979). A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika, 66*(3), 561–565.

Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review, 118*(1), 110–119.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal understanding: the evolution, function and phenomenology of the theory-formation system. In F. Keil, & R. Wilson (Eds.), *Cognition and explanation*. Cambridge, Mass: MIT Press.

Gopnik, A. and Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms and the theory. *Psychological Bulletin*. doi: 10.1037/a0028044.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition, 103*(2), 180–226.

Harris, P. L. (2000). *The work of the imagination*. Blackwell.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*(3), 233–259.

Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135–175.

Holyoak, K., & Thagard, P. (1996). *Analogy in creative thought*. Cambridge, MA: MIT Press.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcome of physical events: two-year-olds fail to reveal knowledge of solidity and support. *Child Development, 71*(6), 1540–1554.

Howson, C. (1985). Some recent objection to the Bayesian theory of support. *The British Journal for the Philosophy of Science, 36*, 305–309.

Kaelbling, L. P. (1993). *Learning in embedded systems*. Cambridge, Mass: MIT Press.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligence Research, 4*, 237–285.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*(1–2), 99–134.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*(2), 263–292.

Keil, F. C., & Wilson, R. A. (2000). In *Explanation and cognition*. Cambridge, MA: MIT Press.

Kelemen, D. (1999). Functions, goals and intentions: children's teleological reasoning about objects. *Trends in Cognitive Sciences, 12*, 461–468.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, 105*(31), 10687–10692.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher, & W. Salmon (Eds.), *Scientific explanation*. University of Minnesota Press.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer, & C. R. B. Joyce (Eds.), *Human judgment: the social judgment theory approach*. Amsterdam: North-Holland.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211–228.

Kopelman, M. D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery, and Psychiatry, 50*, 1482–1487.

Kopelman, M. D., Ng, N., & Van den Brooke, O. (1997). Confabulation extending across episodic, personal and general semantic memory. *Cognitive Neuropsychology, 14*, 683–712.

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Tests of optimal sampling by foraging great tits. *Nature, 275*, 27–31.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*(3), 929–944.

Leonard, J. J. & Durrant-whyte, H. F. (1991). Simultaneous map building and localization for an autonomous mobile robot. *Intelligent robots and systems' 91.'Intelligence for mechanical systems, proceedings IROS'91. IEEE/RSJ International Workshop*, 1442–1447.

Lockman, J. J. (2000). A perception-action perspective on tool use development. *Child Development, 71*(1), 137–144.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*(10), 464–470.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*(3), 232–257.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*(2), 167–204.

Lombrozo, T., Williams, J., & McClelland, J. L. (1988). Connectionist models and psychological evidence. *Journal of Memory and Language, 27*(2), 107–123.

McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences, 26*(8), 423–428.

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: orangutans use water as tool. *Biology Letters, 3*(5), 453–455.

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science, 6*(4), 413–429.

Nietzsche, F. (1882). *The gay science (Walter Kaufmann, Trans.)*. New York: Vintage. (Original work published 1974).

Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.

Norman, D. A. (1999). *The invisible computer*. MIT Press.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608–631.

Phelps, E. A., & Gazzaniga, M. S. (1992). Hemispheric differences in mnemonic processing: the effects of left hemisphere interpretation. *Neuropsychologia, 30*, 293–297.

Piaget, J. (1937). *NonEnLa construction du reel chez l'enfant*. Oxford: Delachaux & Niestle.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition, 123*, 199–217.

Roscoe, R., & Chi, M. (2007). Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research,* 77(4), 534–574.

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science, 36*(4), 321–350.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review, 117*(4), 1144–1167.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382-389.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050.

Shi, L., Feldman, N. H., & Griffiths, T. L. (2008). Performing Bayesian inference with exemplar models. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott, & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). New York: Cambridge University Press.

Shtulman, A. (2009). The development of possibility judgment within and across domains. *Cognitive Development, 24*, 293–309.

Sloman, S. (1994). When explanations compete: the role of explanation on judgements of likelihood. *Cognition, 51*(1), 1–21.

Sobel, D., & Kushnir, T. (2006). The importance of decision making in causal learning form interventions. *Memory & Cognition, 34*(2), 411–419.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*(4), 753–766.

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

Spelke, E. S. (1999). Innateness, learning, and the development of object representation. *Developmental Science, 2*, 145–148.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489.

Strehl, A., Mesterharm, C., Littman, M., & Hirsh, H. (2006). Experience-efficient learning in associative bandit problems. *Proceedings of the 23rd international conference on machine learning (ICML-06).*

Strevens, M. (2004). The causal and unification accounts of explanation unified – causally. *Nous, 38*, 154–179.

Stulp, G., Emery, N. J., Verhulst, S., & Clayton, N. S. (2009). Western scrub-jays conceal auditory information when competitors can hear but cannot see. *Biology Letters, 5*(5), 583–585.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge: MIT Press.

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

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). *Science, 331*(6022), 1279–1285.

Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics*. Cambridge, Mass: MIT Press.

Tong, S., & Koller, D. (2001). Active learning for structure in Bayesian networks. *Seventeenth International Join Conference on Artificial Intelligence*. 863–869.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory acquisition as stochastic search. *Cognitive Development*. http://dx.doi.org/10.1016/j.bbr.2011.03.031.

Walker, C., & Gopnik, A. (under contract). Causality and imagination. In M. Taylor (Ed.), *The development of imagination*. New York: Oxford University Press.

Weisberg, D. S., & Bloom, P. (2009). Young children separate multiple pretend worlds. *Developmental Science, 12*(5), 699–705.

Weisberg, D. S., & Goodstein, J. (2009). What belongs in a fictional world? *Journal of Cognition and Culture, 9*, 69–78.

Weisberg, D. S., & Sobel, D. M. (2012). Young children discriminate improbable from impossible events in fiction. *Cognitive Development, 27*, 90–98.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science, 34*, 776–806.

Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford University Press.

Woodward, J. (2009). Scientific explanation. Available at:. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (spring 2009 ed.). http://plato.stanford.edu/archives/spr2009/entries/scientific-explanation/.

Wyman, E., Rakoczy, H., & Tomasello, M. (2009). Normativity and context in young children's pretend play. *Cognitive Development, 24*(2), 146–155.