Published in final edited form as:
Psychol Bull. 2012 November ; 138(6): 1085–1108. doi:10.1037/a0028044.

Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory

Alison Gopnik and
University of California at Berkeley

Henry M. Wellman
University of Michigan at Ann Arbor

Abstract

We propose a new version of the "theory theory" grounded in the computational framework of probabilistic causal models and Bayesian learning. Probabilistic models allow a constructivist but rigorous and detailed approach to cognitive development. They also explain the learning of both more specific causal hypotheses and more abstract framework theories. We outline the new theoretical ideas, explain the computational framework in an intuitive and non-technical way, and review an extensive but relatively recent body of empirical results that supports these ideas. These include new studies of the mechanisms of learning. Children infer causal structure from statistical information, through their own actions on the world and through observations of the actions of others. Studies demonstrate these learning mechanisms in children from 16 months to 4 years old and include research on causal statistical learning, informal experimentation through play, and imitation and informal pedagogy. They also include studies of the variability and progressive character of intuitive theory change, particularly theory of mind. These studies investigate both the physical and psychological and social domains. We conclude with suggestions for further collaborative projects between developmental and computational cognitive scientists.

Keywords

Cognitive Development; Bayesian Inference; Theory of Mind; Causal Knowledge; Intuitive Theories

The study of cognitive development suffers from a deep theoretical tension – one with ancient philosophical roots. As adults, we seem to have coherent, abstract and highly structured representations of the world around us. These representations allow us to make predictions about the world, and to design effective plans to change it. We also seem to learn those representations from the fragmented, concrete and particular evidence of our senses. Developmental psychologists actually witness this learning unfold over time. Children develop a succession of different, increasingly accurate, conceptions of the world and it at least appears that they do this as a result of their experience. But how can the concrete particulars of experience become the abstract structures of knowledge?

Address for Correspondence: Alison Gopnik, Dept. of Psychology, University of California at Berkeley, Berkeley, CA, 94704. Gopnik@berkeley.edu.

In the past there have been no satisfying theoretical accounts of how this kind of learning might take place. Instead, traditional empiricist accounts, most recently in the form of connectionist and dynamic systems theories, (Elman, et al., 1996; Thelen & Smith 1994) denied that there actually was the kind of abstract, coherent, structure we seem to see in adult representations. They saw instead a distributed collection of specific associations between particular inputs or a context-dependent assemblage of various functions. Traditional nativist accounts, most recently in the form of modularity and core knowledge theories, (Pinker 1997; Spelke, et al. 1992; Spelke & Kinzler 2007) pointed to the structure, coherence and abstractness of our representations, but denied that they could be learned.

Piaget famously tried to resolve this tension by calling for a “constructivist” theory. But aside from the phrase itself there was little detail about how constructivist learning processes might work. Piaget also made empirical claims; he described developmental evidence that appeared to support constructivism. But in the past thirty years many of those empirical claims have been overturned. The combination of theoretical vagueness and empirical inadequacy doomed the Piagetian account.

Recently, however, a new set of computational ideas promises to reconstruct constructivism. This new “rational constructivism” (Xu, Dewar, & Perfors, 2009) uses the theoretical framework of probabilistic models and Bayesian learning. In tandem, new empirical studies, studies of the mechanisms of learning and studies of the progressive character of development, provide support for these theoretical ideas and suggest new areas of theoretical investigation. In this paper, we weave together this new theoretical and empirical work. The basic computational ideas and experimental techniques we will discuss have been applied to many types of learning–from low-level vision and motor behavior, to phonology and syntax. In this paper, however, we focus on how these new ideas explain the development of our intuitive theories of the world.

Our first aim is to make the computational ideas accessible to mainstream developmentalists (like us). It is certainly rational for psychologists to want to ensure substantial empirical returns before they invest in a new set of formal ideas. So we want to share our own experience of how the formal work can be understood more intuitively and how it can lead to new empirical discoveries. Our second aim is to review and synthesize a large body of empirical work that has been inspired by, and has inspired, the new theoretical ideas. Finally, we will suggest new directions that we hope will lead to yet more empirical and theoretical advances.

The Theory Theory Revisited

20 years ago psychologists began to outline a constructivist set of ideas about cognitive development and conceptual structure sometimes called “the theory theory.” (see e.g. Carey, 1985; Gopnik, 1988; Gopnik & Meltzoff, 1997; Gopnik & Wellman 1992; Keil, 1989; Murphy & Medin 1985; Wellman, 1990; Wellman & Gelman 1992). The theory theory claimed that important conceptual structures were like everyday theories and that cognitive development was like theory revision in science. Children construct intuitive theories of the world and alter and revise those theories as the result of new evidence. Theory theorists pointed to three distinctive aspects of intuitive theories, their structure, function and dynamics. These aspects distinguish the theory theory from other accounts of conceptual structure and development. We will recap those points briefly, and add ideas inspired by the new computational and empirical work.

First, theories have a distinctive structure. They involve coherent, abstract, causal representations of the world. Often these representations include unobservable hidden theoretical entities. Theories also have a hierarchical structure: theories may describe specific causal phenomena in a particular domain but these specific theories may also be embedded in more abstract "framework theories". Framework theories describe, in general terms, the kinds of entities and relations that apply in a domain, rather than specifying those entities and relations in detail.

Second, theories have distinctive cognitive functions. They allow wide-ranging predictions about what will happen in the future. They also influence interpretations of the evidence itself. Moreover, theories allow you to make counterfactual inferences – inferences about what could have happened in the past, or, most significantly, what would happen if you decided to intervene on the world and do something new in the future. These inferences about counterfactuals and interventions go beyond simple predictions about what will happen next, and have been a focus of more recent work.

Finally, theories have distinctive dynamic features. These features reflect a powerful interplay between hypotheses and data, between theory and evidence. In particular, unlike modules or "core knowledge", for example, theories change in the light of new evidence, and they do so in a rational way. Moreover, unlike associationist structures, for example, theories may change quite broadly and generally—in their "higher" principles not just in their local specific details.

Recent work has revealed several new and significant aspects of the dynamics of theory change. First, statistical information, information about the probabilistic contingencies between events, plays a particularly important role in theory-formation both in science and in childhood. In the last fifteen years we've discovered the power of early statistical learning.

Second, we've also discovered the power of informal experimentation. Adults and children themselves act on the world in ways that reveal its causal structure. In science and in childhood, experiments lead to theory change. Children learn about causal structure both through their own interventions on the world, for example, in exploratory play, and through observing the interventions of others, for example in imitation and informal pedagogy

Third, theory change often relies on variability. In the course of theory change, children gradually change the probability of multiple hypotheses rather than simply rejecting or accepting a single hypothesis. Moreover, this process of revision can yield many intermediate steps. Evidence leads children to gradually revise their initial hypotheses and slowly replace them with more probable hypotheses. This results in a characteristic series of related conceptions that forms a bridge from one broad theory to the next.

Developmentalists have charted how children construct and revise intuitive theories. The theory theory has been most extensively applied to intuitive psychological and biological understanding. In infancy and early childhood children begin to construct intuitive theories of their own minds and those of others (e.g., Gopnik & Meltzoff 1997; Gopnik & Wellman, 1994; Wellman 1990). Throughout early childhood and well into the school-age period they construct and change intuitive theories of the biological world (Carey 1985; Gelman, 2003; Inagaki & Hatano 2002). But there is also work on children's understanding of the physical world, starting in infancy and proceeding all the way through adolescence, (e.g., Baillargeon 2008;Smith, Wiser & Carey 1985; Vosniadou & Brewer 1992; Xu 2009) and recently there has been increasing research on children's intuitive theories of the social world (e.g., Dweck 1999; Rhodes & Gelman in press; Seiver, Gopnik, & Goodman, in press).

This work has detailed just what children know when about these crucial domains and has tracked conceptual changes through childhood. For example, in the case of intuitive psychology or "theory of mind" developmentalists have charted a shift from an early understanding of emotion and action, to an understanding of intentions and simple aspects of perception, to an understanding of knowledge vs. ignorance and finally to a representational and then an interpretive theory of mind. Similarly, others have traced successive phases in children's understanding of intuitive biology (Gelman 2003; Inagakai & Hatano 2002).

Of course, there are controversies about just when various conceptual and theoretical changes take place, and what they mean. There is debate about which features of an "intuitive biology" emerge in the preschool period (Gelman 2003) and which only appear later in middle childhood (Carey 1985). There are also debates about infant's abilities to predict the actions of others, and about how these abilities are related to the understanding of the mind that emerges at about age 4 (Leslie 2005; O'Neil 1996; Onishi & Baillargeon 2005; Perner & Ruffman 2005).

There has also been special debate about the right way to think of development in infancy. Some investigators have suggested that there are initial non-theoretical structures, such as perceptual structures or "core knowledge". These structures only become theoretical, and so subject to change and revision, later on as a result of the acquisition of language and the application of analogical reasoning (Carey 2009; Spelke & Kinzler 2007). For others, including us, its "theories all the way down" – we think that even newborn infants may have innate intuitive theories and those theories are subject to revision even in infancy itself (see e.g. Woodward & Needham 2008). By any standard, though, the theory theory has been remarkably fruitful in generating research.

However fruitful, the theory theory has suffered from a central theoretical vagueness. The representations that underpin theories and the learning mechanisms that underpin theory change have both been unclear. The fundamental idea of cognitive science is that the brain is a kind of computer designed by evolution to perform particular cognitive functions. The promise of developmental cognitive science is that we can discover the computational processes that underlie development. The theory theory, like Piagetian constructivism itself, has lacked the precision to fulfill this promise. Crucially, it lacked a convincing computational account of the learning mechanisms that allow theory change to take place. The central analogy of the theory is that children's theories are like scientific theories. But this analogy was only a first step. We need to understand how theory change is possible in principle, either in childhood or in science.

Fortunately, recent advances in the philosophy of science and machine learning have given us a new set of perspectives and tools that allow us to characterize theories and, most significantly, theory change itself. We'll refer to these ideas broadly as the "probabilistic models" approach though they include a number of different types of specific representations and learning mechanisms. We'll give an intuitive and non-technical account of how these models work, and, in particular, how they allow learning and theory change.

Probabilistic models can be applied to many different kinds of knowledge. But one type of knowledge is particularly relevant for intuitive theories – namely causal knowledge. Intuitive theories are representations of the causal structure of the world. This distinguishes them from other types of knowledge, such as knowledge of language, space or number. Probabilistic modeling has led to some more specific and very fruitful ideas about causal knowledge both in the philosophy of science and in computer science.

We also emphasize the hierarchical character of these models. Particularly in recent work, probabilistic models can describe both specific theories and framework theories and learning at both local and more abstract levels.

On the empirical side we will focus on a body of work that has emerged in the past ten years and that differs from earlier types of “theory theory” research as well as cognitive development research more generally. This work goes beyond simply charting what children know and when they know it. One line of research explores particular causal learning mechanisms. Another line of research looks in detail at progressive changes in children’s knowledge and the role that variability plays in those changes. We will interweave reviews of this empirical research with the relevant computational ideas. But we begin by outlining the basic computational ideas themselves.

New Theoretical Advances: Probabilistic Models and Rational Learning

Probabilistic models have come to dominate machine learning and artificial intelligence over the last 15 years, and they are increasingly influential in cognitive science (see e.g., Glymour, 2003; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Oaksford & Chater, 2007). They have also been proposed as a model for theory-like cognitive development (Gopnik, 2000; Gopnik et al. 2004; Gopnik & Tenenbaum 2007). Two features of probabilistic models are particularly important for the theory theory. First, they describe structured models that represent hypotheses about how the world works. Second, they describe the probabilistic relations between these models and patterns of evidence in rigorous ways. As a consequence they both represent conceptual structure and allow learning.

Imagine that there is some real structure in the world – a three-dimensional object, a grammar, or, especially relevant to theories, a network of causal relationships. That structure gives rise to some patterns of observable evidence rather than others – a particular set of retinal images, or spoken sentences, or statistical contingencies between events. That spatial or grammatical or causal structure can be represented mathematically, by a 3-d map or tree structure or a causal graph. You could think of such a representation as a hypothesis about what the actual structure is like. This representation will also allow you to mathematically generate patterns of evidence from that structure. So you can predict the patterns of evidence that follow from the hypothesis, and make new inferences accordingly. For example, a map or a tree or a causal graph will let you predict how an object will look from a different angle, whether a new sentence will be acceptable, or that a new event will be followed by other events. If the hypothesis is correct, then these inferences will turn out to be right.

These generative models then, provide ways of characterizing our everyday representations of the world and explaining how those representations allow us to make a wide range of new inferences. For this reason, a number of cognitive psychologists have used these representations to describe adult knowledge, in particular, causal knowledge (Lu, et al. 2008; Sloman 2005; Waldmann, Hagmayer & Blaisdell 2006), and these representations have been proposed as a way to characterize adult intuitive theories (Rehder & Kim, 2006)

From the developmental point of view, though, the really interesting question is not how we use these representations but how we learn them. Critically, the systematic link between structure and evidence in these models also allows you to reverse the process and to make inferences about the nature of the structure from the evidence it generates. Vision scientists talk about this as “solving the inverse problem”. In vision “the inverse problem” is to infer the nature of 3-d objects from the retinal images they generate. In theory change, the problem is to infer causal structure from the events you observe. Solving the inverse problem lets you learn about the world from evidence. It lets you decide which 3-d map or tree or causal graph best represents the world outside.

The idea that mental models of the structure of the world generate predictions, and that we can invert that process to learn the structure from evidence, is not itself new. It is the basic model underlying both the cognitive science of vision and of language. The big advance has been integrating ideas about probability into that basic framework. If you think of these mental models as logical systems with deterministic relations to evidence the inverse problem becomes extremely difficult, if not impossible, to solve, and that has led to nativist conclusions (e.g., Gold 1967; Pinker 1984). Typically a great many hypotheses are, in principle, compatible with any pattern of evidence, so how can we decide on the best hypothesis? Integrating probability theory makes the learning problem more tractable. Although many hypotheses may be compatible with the evidence, some hypotheses can be more or less likely to have generated the evidence than others.

There are many ways to solve the inverse problem but one of the most powerful and general ways is to use Bayesian inference. Bayesian inference takes off from ideas about probability first formulated by the Rev. Thomas Bayes in the 18th century and instantiated in Bayes' rule. Here is the simplest version of Bayes rule. (This will be the only equation in this paper, but it's a good one.)

$$\text{Bayes Rule:} P(H/E) \propto (P(E/H)\,(P(H)))$$

Bayes' rule is a simple formula for finding the probability that a hypothesized structure (H) generated the pattern of the evidence (E) that you see, that is the probability of H given E, or P(H/E). That probability is proportional to the probability of the pattern of evidence given the hypothesis, P(E/H), and your initial estimate of the probability of the hypothesis, P(H).

Each part of this formula has a conventional name. P(H) is the "prior", the probability of the hypothesis before you looked at the evidence. P(E/H) is the "likelihood", how probable it is that you would see the observed evidence if the hypothesis were true. P(H/E) is the "posterior" – the probability of the hypothesis after you've considered the evidence. Bayes rule thus says that the posterior is a function of the likelihood and the prior.

We can represent a hypothesis as a map, a tree or a causal graph, for example. That map or tree or graph will systematically generate some patterns of evidence rather than others. In other words the representation will establish the likelihood – that is, tell us how likely it is that that hypothesis will generate particular patterns of evidence. If we know the prior probability of the hypothesis, and then observe a new pattern of evidence, we can use Bayes' Law to determine the probability that the hypothesis is true. So we can decide which map, tree or graph is most likely to be correct.

Rather than simply generating a yes or no decision about whether a particular hypothesis is true, the probabilistic Bayesian learning algorithms consider multiple hypotheses and determine their posterior probability. Often, in fact usually, there are many spatial or causal structures or grammars that could, in principle, produce a particular pattern of visual, causal, or linguistic evidence. The structure is "underdetermined" by the evidence. This is the "poverty of the stimulus argument" that led Chomsky and others to argue for innateness. But while many structures may be possible, some of those structures are going to be more likely than others. Bayesian methods give you a way of determining the probability of the possibilities. They tell you whether some hypothesis is more likely than others given the evidence. So we can solve the inverse problem in this probabilistic way.

Here's an example. Suppose Mary is travelling and she wakes up with a terrible pain in her neck. She considers three possible hypotheses about what caused the pain: perhaps she has a clogged carotid artery, perhaps she slept in an awkward position on that wretched lumpy mattress, perhaps it was that dubious lobster she ate last night. She goes to Web MD and discovers that both a clogged artery and awkward sleeping are much more likely to lead to neckaches than bad shellfish – neckaches have a higher likelihood given a clogged carotid and awkward sleeping than they do given bad shellfish. In fact, she reads that clogged carotids always lead to neckaches -- the likelihood of a neckache given a clogged carotid is particularly high. Should she panic? Not yet. After all, it's much less likely to begin with that she has a clogged carotid artery than that she slept awkwardly or the lobster was bad – awkward sleeping and bad lobsters have a higher prior probability than severely blocked carotids. If you combined these two factors, the likelihood and the prior, you would conclude that a bad night on the lumpy mattress is the most likely hypothesis.

Eventually though, enough evidence could lead you to accept even an initially very unlikely idea. Sufficient additional evidence (the ache persists, an x-ray shows blockage) might indeed lead to the initially unlikely and grim clogged carotid diagnosis. This gives Bayesian reasoning a characteristic combination of stability and flexibility. You won't abandon a very likely hypothesis right away, but only if enough counter-evidence accumulates.

Probabilistic models were originally articulated as ideal rational models of learning. Like ideal observer theory in vision (Geisler, 1989) they tell us how a system could learn best, in principle. In fact, Bayesian inference first emerged in the philosophy of science. Philosophers wanted to determine how scientists ought to react to new evidence, not necessarily how they actually did react. However, ideal observer theory can help us think deeply about how evolution actually did shape the visual system. In the same way probabilistic models can help us think deeply about how evolution shaped human learning capacities. We can compare an ideal learning machine to the learning machine in our skulls.

These ideal rational probabilistic models have both attractions and limitations as theories of the actual representations and learning mechanisms of cognitive development. One attraction is that, at least in principle, this kind of inference would allow children to move from one structured hypothesis to another very different hypothesis based on patterns of evidence. Children need not merely fiddle with the details of an innately determined structure or simply accumulate more and more evidence. They could genuinely learn something new.

Bayesian inference also captures the often gradual and piece-meal way that development proceeds. Empiricists emphasize this aspect of development and it is not easily accommodated by nativism. At the same time, the generative power of structured models can help explain the abstract and general character of children's inferences. Nativists emphasize this aspect of development and it is not easily accommodated by traditional associationist empiricism. And the integration of prior knowledge and new evidence is just what Piaget had in mind when he talked about assimilation and accommodation.

The major drawback of the probabilistic model approach is the vast space of possible hypotheses and possible evidence. Bayesian reasoning gives you a way to evaluate particular hypotheses, given a particular pattern of evidence. However, you still have to decide which hypotheses to evaluate, and equally which evidence to gather. A very large number of hypotheses might be compatible with some particular pattern of evidence, and a child or a scientist (or even a computational learning algorithm) won't be able to enumerate the probability of each one. How do you decide which hypotheses to test in the first place? The evidence you have is also always incomplete. How do you decide when and how to collect new evidence?

In particular, computer scientists talk about the "search problem" – that is the problem of checking all the possible hypotheses against the evidence. There is also a different kind of search problem, namely, how to search for new evidence that is relevant to the hypotheses you want to test. As we will see, there are potential solutions, or at least promising approaches, to both kinds of search problems.

Bayesian reasoning may be applied to everything from olfactory perception in the fly to medical decision-making in a hospital. Bayes rule, by itself, is very general. In fact, it’s too general to explain much without more information about the hypotheses and the likelihoods. Here is where the models come in. The probabilistic model approaches we emphasize specify the structure of the hypotheses in a particular domain. They also specify how these structured hypotheses generate evidence. If we want to characterize the “theory theory” in these terms, we have to find ways to represent both causal relationships and the patterns of evidence they generate.

In what follows we will focus on three recent developments that are particularly relevant to intuitive theories. First, we will describe a subcategory of probabilistic models, called causal Bayes nets, that are particularly relevant to causal knowledge and learning. We will also show, empirically, that children’s causal learning can be understood in terms of Bayes nets.

Second, we will discuss some of the learning mechanisms that are implied by probabilistic causal models, learning mechanisms that could help solve the problem of deciding which hypotheses to test and which evidence to consider. We will outline new empirical evidence which shows that young children learn in a similar way. One way to learn a causal structure, in particular, is to perform planned interventions on the world – experiments. Experiments can not only provide you with more evidence about a causal structure, they can provide you with evidence that is designed to eliminate many possible hypotheses and can help you discriminate between just the most relevant hypotheses. A second way is to watch the outcomes of the interventions other people perform, particularly when those people are knowledgeable teachers. Watching what others do can further narrow the hypotheses and evidence you will consider, and the inferences you will draw. A third, complementary learning technique is to rationally sample just a few hypotheses at a time, testing those hypotheses against one another. This sampling process leads to distinctive kinds of learning with characteristic features of variability and progression.

Finally, we will describe the more recently developed hierarchical causal models. These hierarchical models can characterize broader framework theories along with more specific causal relationships as well as characterize the relations between theories at different levels of abstraction.

Causal Bayes Nets and the Interventionist Theory of Causation

To use Bayes’ law you have to first determine the elements in Bayes’ equation: the hypothesis, the evidence, and the likelihood, that is, the probability of particular patterns of evidence given a particular hypothesis. So you need to have some formal way of describing the hypotheses and systematically relating them to evidence.

Causal hypotheses are particularly important both in science and in ordinary life. As theory theorists noted 20 years ago theories involve coherent and abstract representations of causal relationships. Fortunately, over the last 15 years, computer scientists and philosophers have developed models of causal relations. These models are known as “causal graphical models” or “causal Bayes nets” (Pearl, 2000; Spirtes et al. 1993, 2000). The models also have both inspired and been inspired by a particular philosophical view of causation.

What *is* causation anyway? Theories involve causal relations but what makes those relations distinctively causal? Traditionally philosophers have approached this problem in several different ways. David Hume famously argued that there is no such thing as causation.

Instead, there are simply associations between events. "Mechanists" like Kant in philosophy and Michotte in psychology (Leslie & Keeble 1987; Michotte 1963), argue that causal relations involve particular spatiotemporal patterns, such as contact and launching. Piaget grounded causation in the immediate consequences of our intentional actions

Recently, however, the philosopher James Woodward has articulated and formalized an alternative interventionist account of causation (Woodward 2003). The central idea is that if there is a direct causal relation between A and B, then, other things equal, intervening to change the probability of A will change the probability of B. This view of causality is rather different from the associationist, mechanistic or Piagetian views that underpin earlier work on the development of causal knowledge. But this account dovetails with causal Bayes nets; the models also relate causality to probability and intervention. And, as we will see, the interventionist idea has particularly interesting implications for causal development.

Why is the interventionist idea different from the Humean idea that causation is just correlation? Consider the relation between nicotine-stained yellow fingers and lung cancer. Yellow fingers and lung cancer are correlated and I can predict that if someone has yellow fingers they are more likely to get cancer. But I don't think that yellow fingers cause cancer. This is reflected in the fact that I don't believe intervening to clean someone's fingers will make them any healthier. In contrast, if I believe that smoking causes cancer then I will think that changing the probability of smoking, say in a controlled experiment, will change the probability of lung cancer.

The interventionist account is also different from the mechanistic account. In everyday life we often make causal claims, even when we don't know anything about the detailed mechanisms or spatio-temporal events that underpin those claims (see Keil 2006). The interventionist account explains why – we may not know exactly how a zipper works but we know how to intervene to open or close it.

The interventionist account also suggests why causal relations are so distinctive and so important: Understanding the causal structure of the world allows you to imagine ways that you could do things to change the world, and to envision the consequences of those changes. As we will see, even young children have ideas about causation that fit this picture, although they may, of course, also have more mechanistic conceptions as well.

Causal Bayes nets

Causal Bayes nets were first developed in the philosophy of science, computer science and statistics (Glymour 2001; Pearl 1988, 2000; Spirtes et al. 1993.) Scientists seem to infer theories about the causal structure of the world from patterns of evidence, but philosophers of science found it very difficult to explain how this could be done. Causal Bayes nets provide a kind of logic of inductive causal inference. Scientists infer causal structure by performing statistical analyses and doing experiments. They observe the patterns of conditional probability among variables and "partial out" some of those variables (as in statistical analysis), they examine the consequences of interventions (as in experiments) and they combine the two types of evidence. Causal Bayes nets formalize these kinds of inferences.

In causal Bayes nets, causal hypotheses are represented by directed graphs like the one in Figure 1. The graphs consist of variables, representing types of events or states of the world, and directed edges (arrows) representing the direct causal relations between those variables. Figure 1 is a graph of the causal structure of the woes of academic conferences. The variables can be discrete (like school grade) or continuous (like weight), they can be binary (like "having eyes" or "not having eyes") or take a range of values (like color). Similarly, the direct causal relations can have many forms; they can be deterministic or probabilistic, generative or inhibitory, linear or non-linear. The exact specification of these relations is called the “parameterization” of the graph.

Causal structure and conditional probabilities

The Bayes net formalism specifies systematic connections between the causal hypotheses that are represented by the graphs and particular patterns of evidence. To begin with, the structure of the causal graph itself puts some very general constraints on the patterns of probability among the variables. If we make further assumptions about the parameterization of the graph, that is, about the particular nature of the causal relations, we can constrain the kinds of inferences we make still further. For example, we might assume that each cause independently has a certain power to bring about an effect. This is a common assumption in studies of human causal learning. Or we might even know the exact probability of one event given another, say that there is a 70 percent chance that A will cause B. Or we might know that the evidence we see is a random sample of all the evidence, or instead that it is a sample that is biased in some particular way. Each of these kinds of knowledge can influence our causal inferences.

So, given a particular causal structure and parameterization, only some patterns of probability will occur among the variables. From the Bayesian perspective the graph specifies the likelihood of the evidence given the hypothesis.

To illustrate how this works consider a simple causal problem, partially embedded in the graph of Figure 1. Suppose that I notice that I often can’t sleep when I’ve been to a party and drunk lots of wine. Partying (P) and insomnia (I) covary, and so do wine (W) and insomnia (I). Suppose also that I make some general assumptions about how these variables are likely to be related (the parameterization of the graph). For example, I assume that partying or wine will increase the probability of insomnia, rather than decreasing it, and similarly, that partying will increase the probability of drinking wine. This contrasts with the assumption, say, that wine or partying absolutely determine my insomnia or prevent my insomnia.

There are at least two possibilities about the relations among these variables. Maybe parties cause me to drink wine and that keeps me awake (a causal chain). Maybe parties are so exciting that they keep me awake, and they also independently cause me to drink wine (a common cause). As shown in Figure 2, these possibilities can be represented by two simple causal graphs which include variables like $P_{+/-}$ and $I_{+/-}$ but also specify the nature of the relations between them.

In these graphs $P_{+/-}$, for example, conveys that (to keep things simple) partying can be present (+) or absent (−). $P_{+/-} \rightarrow I_{+/-}$ conveys the fact that partying and insomnia are causally related, and $P_{+} \rightarrow I_{+}$ conveys the more specific hypothesis that more partying leads to more insomnia. So, maybe parties (P_{+}) lead me to drink (W_{+}) and wine keeps me up (I_{+}); or maybe partying (P_{+}) both keeps me up (I_{+}) and lead me to drink (W_{+}). The covariation among the variables by itself is consistent with both these structures.

However, these two graphs lead to different patterns of conditional probability among the three variables, or as statisticians put it, different relations between some variables when other variables are partialled out. Suppose you decide to keep track of all the times you drink and party and examine the effects on your insomnia. If Graph 2a is correct, then you should predict that you will be more likely to have insomnia when you drink wine, whether or not you party. If instead Graph 2b is correct, you will only be more likely to have insomnia when you go to a party, regardless of how much or how little wine you drink.

If I know whether the causal structure of my insomnia is represented by Graph 2a or Graph 2b, and I know the values of some of the variables in the graph (+ or −), I can make consistent and quite general predictions about the probability of other variables. In Bayesian terms, each graph tells us the likelihood of particular patterns of evidence given that particular hypothesis about the causal structure. These predictions can be very wide-ranging -- a simple graph with just a few nodes can generate predictions about a great many possible combinations of events. But causal Bayes nets do more than just allow us to predict the probability of events. They allow us to make more sophisticated causal inferences too.

Bayes nets and interventions

Why think of these graphs as representations of *causal* relations among variables? Here is where the interventionist account of causation comes in. According to the interventionist account, when X directly causes Y, intervening to change the probability of X should change the probability of Y (other things equal). Causal Bayes net algorithms allow us to determine what will happen to Y when we intervene on X.

Predictions about observations may be quite different from predictions about interventions. For example, in a common cause structure like Graph 2b above, we will be able to *predict* something about the value of insomnia from the value of wine. If that structure is the correct one, knowing that someone drank wine will indeed make you predict that they are more likely to have insomnia (since drinking wine is correlated with partying, which leads to insomnia). But intervening on their wine-drinking, forbidding them from drinking, for example, will have no effect on their insomnia. Only intervening on partying will do that.

In causal Bayes nets, interventions systematically alter the nature of the graph they intervene on. In particular, an intervention fixes the value of a variable and in doing so it eliminates the causal influence of other variables on that variable. If I simply decide to stop drinking wine, that means that, no matter what, the wine variable will be set to minus (i.e., W_-); so partying will no longer have any effect. This can be represented by replacing the original graph with an altered graph in which the specific value of some variable is fixed. As a result the arrows directed into the intervened-upon variable will be eliminated (Judea Pearl vividly refers to this process as graph surgery, Pearl 2000). The conditional probabilities among the variables after the intervention can be read off from this altered graph.

Suppose, for example, I want to know the best thing to do to prevent my insomnia. Should I quit partying (intervening to make P_-) or should I quit drinking (intervening to make W_-)? I can calculate the effects of such interventions on the various causal structures, using “graph surgery” to see what the consequences of the intervention will be. The altered graphs in Figure 3, for example, show the same graphs as before but now with an intervention (shown as a firmly grasping fist) on the variable P or the variable W that sets it to a particular value.

If Graph 2a, from Figure 2, is right, and I eliminate partying (set P to P_-) but continue to drink, then when I drink I’ll have insomnia ($W_+ \rightarrow I_+$) but when I don’t I wont ($W_- \rightarrow I_-$). But if I eliminate drinking (set W to W_-) I won’t ever have insomnia (I_-). These are the possibilities shown in Figure 3a. If Graph 2b (from Figure 2) is right, however, if I eliminate drinking but still party, then when I party (P_+) I’ll have insomnia (I_+) and when I don’t (P_-) I wont (I_-). But if I eliminate partying then I’ll eliminate insomnia too ($P_- \rightarrow I_-$). These are the possibilities shown in Figure 3b. So, if Graph 2a is right, I should party sober, but if Graph 2b is right, I should drink at home.

Causal Bayes nets allow us to freely go back and forth from evidence about observed probabilities to inferences about interventions and vice-versa. That’s what makes them causal. They allow us to take a particular causal structure and use it to predict the conditional probabilities of events, and also the consequences of interventions on those events.

We can also use exactly the same formal apparatus to generate counterfactual predictions. Counterfactuals are formally the same as interventions. Instead of saying what should happen when we make the world different, by fixing the value of a variable, we can say what *would have* happened if that variable had been different than it was. The same reasoning that tells me that I should stop drinking to avoid insomnia can tell me that if I had only stopped drinking many years ago, I would have avoided all those sleepless nights.

So Bayes nets capture some of the basic structural and functional features of theories. They describe abstract coherent networks of causal relationships in a way that allows predictions, interventions and counterfactual reasoning.

Bayes nets and learning

We just saw that knowing the causal structure lets us make the right predictions about interventions and observations. We can determine the pattern of evidence a particular hypothesis will generate. This lets us calculate the likelihood of a particular pattern of evidence given a particular hypothesis. But we can also use Bayes nets to solve the crucial inverse problem. We can learn the causal structure by observing the outcomes of interventions and the conditional probabilities of events.

Lets go back to the wine-insomnia example. How could you tell which hypothesis about your insomnia is the right one? The Graphs 2a and 2b represent two different causal hypotheses about the world. You could distinguish between the graphs either by intervention or observation. First, you could do an experiment. You could hold partying constant (always partying or never partying) and intervene to vary whether or not you drank wine; or you could hold drinking constant (always drinking or never drinking) and intervene to vary whether or not you partied. This reasoning underlies the logic of experimental design in science.

You could also, however, simply observe the relative frequencies of the three events. If you notice that you are more likely to have insomnia when you drink wine, whether or not you party, you can infer that Graph 2a is correct. If you observe that, regardless of how much or how little wine you drink, you are only more likely to have insomnia when you go to a party, you will opt instead for Graph 2b. These inferences reflect the logic of correlational statistics in science. In effect, as we noted earlier, what you did was to “partial out” the effects of partying on the wine/insomnia correlation, and draw a causal conclusion as a result.

It is not only theoretically possible to infer complex causal structure from patterns of conditional probability and intervention (Glymour & Cooper, 1999; Spirtes et al., 1993). It can actually be done. Computationally tractable learning algorithms have been designed to accomplish this task and have been extensively applied in a range of disciplines (e.g., Ramsey et al. 2002; Shipley 2000). In some cases, it is also possible to accurately infer the existence of new previously unobserved variables (Richardson & Spirtes, 2003; Silva et al., 2003; Spirtes, et al. 1997).

Causal Bayes nets are particularly well suited to Bayesian learning techniques (Griffiths & Tenenbaum 2007; Heckerman et al., 1999). Bayesian graphical networks allow us to easily determine the likelihood of patterns of evidence given a causal hypothesis. Then we can use Bayesian learning methods to combine this likelihood with the evidence and the prior probability of the hypothesis. We can infer the probability of particular graphs from a particular pattern of contingencies among variables, or from the outcome of some set of controlled experiments.

We will say more later about probabilistic Bayesian models, and in particular hierarchical models. But let's begin by showing how the ideas we've described so far apply to empirical research with children.

Empirical Work on Bayes Nets and Bayesian Reasoning in Children

Over the past ten years, a number of researchers have explored whether children might have Bayes net-like representations of causal structure, and whether they can learn causal structure from evidence in the way that the formalism suggests. We know that even infants can detect complex statistical patterns. In fact, statistical learning has been one of the most important recent areas of developmental research on linguistic and perceptual learning (e.g., Gomez, 2002; Kirkham & Johnson 2002; Saffran et al. 1996; Wu, Gopnik, Richardson, & Kirkham, 2011). This research shows that even young infants are sensitive to some of the statistical regularities in the data that would be necessary to engage in Bayesian causal learning at all.

But more recent research goes further. It demonstrates that very young children, even infants, can actually use those statistics to make inferences about causal structure. Researchers have also explored whether children use that knowledge in ways that go beyond simple association. And they have explored whether children can make similar causal inferences from the outcomes of interventions – their own or those of others. Finally, they have also asked whether children will integrate their prior knowledge with new evidence in a Bayesian way, and whether they will go beyond learning about observable variables to posit unobservable ones. The quick answer to all these questions is yes.

The methodology of all these experiments has been similar. Obviously, it is not possible to explicitly ask very young children about conditional probabilities or interventions. Indeed, the judgment and decision-making literature has demonstrated that even adults have a great deal of difficulty with explicit and conscious probabilistic reasoning (see e.g., Kahneman & Tversky, 1996). On the other hand, there is evidence that human minds unconsciously use Bayesian inference extensively in areas like vision and motor control (Kersten, Mamassian, & Yuille, 2004; Wolpert, 2007). We can ask whether children might also implicitly use these inference techniques to develop intuitive theories.

Researchers studying intuitive theories have usually tried to discover a typical child's knowledge of familiar causal generalizations, and to track changes in that knowledge as children grow older. We can ask whether children of a particular age understand important causal relationships within domains such as psychology, biology and physics. But if we want to understand the fundamental mechanisms of causal learning we also need to give children causal problems that they haven't already solved. So researchers have given children controlled evidence about new causal systems to see what kinds of causal conclusions they will draw.

Causal Learning in Young Children

Learning causality from probability—The Bayes net approach to causation suggests that children might be able to go beyond learning the immediate consequences of their actions, as Piaget suggested, associating correlated events, as the classical associationist or connectionist accounts suggest (Rogers & McLelland 2004), or understanding specific physical events that involve contact and movement (Leslie & Keeble 1987; Michotte 1963). Instead, children might be able to learn new causal structure from patterns of probability.

Moreover, according to the interventionist account of causation children should be able to use that structure to design new interventions on the world. In a first set of experiments Gopnik et al. (2001) showed just that. Children saw a “blicket detector” -- a machine that lit up and played music when some combinations of objects but not others were placed on it, as depicted in Figure 4. For example, children might see that the machine did not activate when you put B alone on it, but did activate when you placed A on it and continued to do so when B was added to A (as in Figure 4). Given your prior knowledge about the machine, it could have any one of the causal structures represented by the Bayes nets in Figure 2. However, according to the formalism, the pattern of evidence is only compatible with the first structure where A is a blicket and B is not.

Then children were asked to design an intervention to make the machine go or turn off. If the causal structure is that illustrated in the top left of the possibilities in Figure 4, you should intervene on A and not B to make the machine stop. 2-, 3- and 4-year-olds could use the pattern of covariation between the blocks and the machine’s activation to infer the causal structure of the machine. Then they could use that causal knowledge to figure out how to make the machine go or stop. They would add only A, and not B or A and B, to make the detector activate. They would remove only A, and not B or A and B, to make the machine stop. Gweon and Schulz (2011) found similar abilities to infer causation from covariation in infants as young as 16 months.

Sobel et al. (2004) found that preschool children would also make correct causal inferences from more complex statistical patterns, particularly “backward blocking”. Backward blocking is a kind of causal inference that requires children to learn about the causal efficacy of an object using information from trials in which that object never appeared. For example, children saw A activate the machine by itself, and then saw A and B together activate the machine. The fact that A alone was sufficient to activate the machine made the children think that B was less likely to be a blicket.

There are two interesting points about this inference. First, unlike the inference in our first example, it is probabilistic. B could still be a blicket, but this hypothesis is less likely if A activated the machine. Second, this inference is particularly difficult to explain with standard associationist theories. Sobel and Kirkham (2007) found that children as young as 18 months old also showed similar capacities for backward blocking. They also found that, in an anticipatory looking task, even 9-month-olds seemed to infer causation from covariation.

Children can also infer more complicated kinds of causal structure. Gopnik et al. (2004) showed that children could use a combination of interventions and statistics to infer the direction of a causal relation, that is, whether A caused B or vice-versa. Still further, Schulz et al. (2007) showed that 4-year-old children could use this kind of evidence to infer more complex causal structures involving three variables. In these experiments, they distinguished between a causal chain, where A causes B causes C (as in Graph 2a earlier) and a common cause structure where A causes B and C (as in Graph 2b earlier).

In these examples, the causal relations were complex but deterministic. Kushnir et al. (2005) showed that 4-year-old children could also make inferences about probabilistic relationships. Children could use probabilistic strength to infer causal strength – they thought that a block that set off the machine 2 of 3 times was more effective than one that worked 2 of 6 times (although both set off the machine two times).

Integrating prior knowledge and new evidence—Bayesian inference combines evidence, likelihoods, and the prior probability of hypotheses. Do children take prior knowledge into account in a Bayesian way when they are making causal inferences? Several recent studies show that they do, but that, also in a Bayesian way, new evidence can lead them to overturn an initially likely hypothesis. Thus, Sobel, Tenenbaum, and Gopnik (2004) and Griffiths, Sobel, Tenenbaum, & Gopnik (2011) showed that children would take the baseline frequency of blickets into account when they made new inferences in a backwards blocking task. They made different inferences when they were told beforehand that blickets were rare or common. If blickets were rare, for example, children were less likely to conclude that a block was a blicket than if blickets were common.

Kushnir and Gopnik (2007) explored how children integrated prior knowledge about spatio-temporal causal relationships with new evidence. To begin with children clearly preferred a hypothesis about a blicket machine that involved contact, as we might expect based on perceptual or mechanism accounts of causality (e.g., Leslie & Keeble 1987; Muentener & Carey 2010). They assumed that a block would have to touch the blicket detector to make it go. However, children overcame that prior assumption when they were presented with statistical evidence that blickets could act remotely, without contact. When they saw that the machine was most likely to activate when an object was waved above it, rather than touching it, they concluded that contact was unnecessary.

Other studies show the influence of prior knowledge on causal learning. In particular, Schulz and Gopnik (2004) and Schulz et al. (2007), explored whether children believe that causal relations can cross domains – that, for example, a physical cause could lead to a psychological effect or vice-versa. Many studies suggest that children are initially reluctant to consider such hypotheses – they have a very low prior probability (e.g., Notaro, Gelman & Zimmerman 2001). However, Schulz and Gopnik (2004) showed that 4-year olds would use statistical information to learn about cross-domain causal relations. For example, children initially judged that talking to a machine would not make it go. But if they saw the appropriate conditional probabilities between talking and activation, they became more willing to consider the cross-domain cause. Schulz et al. (2007), then gradually gave children more and more statistical evidence supporting a cross-domain hypothesis. This systematically shifted children's inferences in precisely the way a Bayesian model would predict. As children got more and more evidence in favor of the hypothesis, they were more and more likely to accept it. These cross–domain inferences are a good example of an initially low probability hypothesis that may be confirmed by the right pattern of evidence.

Unobserved causes—Children don't just use statistical patterns to infer observed causal relations, like the fact that the blicket lights up the detector. They also use conditional probabilities to infer the existence of unobserved causes--hidden "theoretical entities." Gopnik et al. (2004) found that when the observed variables couldn't explain the evidence, children would look for unobserved variables instead. Children saw two simple stick puppets, which we'll call W and I, that moved and stopped together – they covaried, like Wine and Insomnia in Figure 2. This pattern of covariation indicated that there was some causal link between the two events, but didn't specify exactly what causal structure led to that link. Then children saw the experimenter intervene to move W, with no effect on I and vice-versa intervene to move I with no effect on W. These two interventions ruled out the two obvious causal hypotheses W → I and I → W. Then children were asked if W made I move, I made W move, or something else made them both move. They chose "something else" as the right answer in this condition, but not in a similar control condition. Moreover, many of the children searched for the unobserved cause, looking behind the puppet apparatus. So 4-year-olds had concluded that some unobserved common cause, U, influenced W and I, and therefore W←U→I. Similarly, Schulz and Somerville (2006) found that when children saw an indeterministic machine – that is a machine that went off only 2 of 6 times--they inferred that some hidden variable was responsible for the failures.

Inferring psychological causation—Most of these initial experiments involved physical causation. However, children also extend these causal learning techniques to other domains. Schulz and Gopnik (2004) found that 4-year-old children used covariation to infer psychological and biological causes as well as physical ones. In a particularly striking experiment, Kushnir, Xu & Wellman (2010), found that children as young as 20 months old would use statistical Bayesian reasoning to infer the desires of another person.

To set the stage, recall that a causal model doesn't just specify causal structure – it can also specify the relations between the causal structure (including the parameterization of that structure) and the evidence. The default assumption for many causal models, including the models we typically use in science, is that the evidence we see is a random sample from an underlying distribution. When the evidence doesn't fit that pattern we either have to revise our assumptions about the causal structure, or revise our assumptions about the sampling process. This is the logic behind significance tests. When there is less than a five percent chance that the pattern of evidence we see was the result of a random sampling process, we infer that there is some additional causal factor at work.

To begin with, Xu and Garcia (2009) demonstrated that 9-month-olds were sensitive to sampling patterns. The experimenter showed the infants a box full of white and red ping-pong balls, in an 80-20 proportion. Then she took some balls from the box. A natural causal model would be that this sample was randomly generated. In that case, the distribution of balls in the sample should match the distribution of the balls in the box. Indeed, infants looked longer when a sample of mostly red balls was taken from a box of mostly white balls, than when a sample of mostly white balls was extracted. These infants initially seemed to assume that the balls were a random sample from the distribution in the box.

This result is interesting for several reasons. For one thing notice that the violations of expectancy were not impossible – you could, after all, pull mostly white balls from a mostly red box--but merely improbable. Infants appeared to be sensitive to the probability of different outcomes. It's as if the infants said to themselves "Aha! less than .05 probability that this occurred by chance!" But would the surprising evidence drive the children to another causal model?

Going one step further, Kushnir et al. (2010) found that, in fact, 20-month-olds interpreted this non–random sampling causally and psychologically. An experimenter took frogs from a box of almost all ducks or she took frogs from a box of almost all frogs. Then she left the room and another experimenter gave the child a small bowl of frogs and a separate bowl of ducks. When the original experimenter returned she extended her hand ambiguously between the bowls. The children could give her either a frog or a duck. When she had taken frogs out of the box that was almost all ducks, children gave her a frog. In this case, the infants concluded that she wanted frogs. In contrast, when she had taken frogs from a box of almost all frogs children were equally likely to give her a frog or a duck. In this case, the infants concluded that she had merely drawn a random sample from the box, rather than displaying a preference for frogs. So these 20-month-old infants had inferred an underlying mental state—a desire--from a statistical pattern.

In a still later study, Xu and Ma (2011) showed that both 2–year-olds and 16-month-olds would use non-random sampling to learn that an adults desires might differ from their own. This is especially interesting because in an earlier "theory of mind" study Gopnik & Repacholi (1997) demonstrated that 18 month olds could spontaneously appreciate the fact that their own desires differed from the desires of others, but 15 month olds could not.

(c et al. (in press) showed that children could make particularly complex inferences about covariation with probabilistic data in a social setting. Four-year-olds saw that an action probabilistically covaried either with a person or with a situation (e.g., Sally plays on a trampoline rather than a bicycle three out of four times, while Josie only plays on it one out of four times, or in a contrast condition Sally and Josie both play on the trampoline ¾ of the time but only play on the bicycle ¼ of the time.) Four-year-olds correctly inferred that the action was caused by a feature of the person in the first case, but by a feature of the toy in the second. Moreover, in this study, 4-year-olds explained the “person-caused” patterns of probabilistic covariation by inferring consistent and long-lasting personal traits, and used those traits to predict future patterns of behavior. This finding is striking because these kinds of attributions reflect an intuitive theory of traits that usually emerges later in middle childhood (cf. Dweck, 1999). When children receive the appropriate evidence, however, they are able to make such inferences even at a much earlier age.

Both this study and the Xu and Ma study of desires are also interesting because they show that these learning mechanisms can help explain the naturally occurring changes in children’s theory of mind that were the focus of earlier work. Children not only make correct inferences in an artificial setting like the blicket detector, they do so in more everyday settings. In such situations the data can drive children towards a theory change that occurs in normal development.

Dynamic Features of Theories

To sum up so far, probabilistic models can provide a formal account of both the structure and function of specific intuitive theories. They can represent complex causal hypotheses mathematically. They can also explain mathematically how those hypotheses can generate new predictions, including probabilistic inferences, counterfactual claims, and prescriptions for interventions. The Bayesian interaction between prior knowledge and current evidence can also help explain the interpretive function of theories, the way they lead us to interpret and not just record new data. On the Bayesian view, our prior knowledge shapes the inferences we draw from new data, as Piaget pointed out long ago. Moreover, as we’ve just seen, children as young as 16 months old can actually make these kinds of inferences, and by four these inferences are both ubiquitous and sophisticated.

These results also tell us something about the dynamics of theory change – about learning. They show that children are learning about causal structure in a normatively correct way – given the right evidence they draw appropriate causal conclusions. But we can also ask more deeply about the specific learning processes that are involved in theory change. In some ways this is the most interesting question for developmentalists. How do children resolve the search problems we described earlier? How do they decide which hypotheses to test, and which evidence to use to test them? Here also new empirical work and new computational insights dovetail. We will describe three different ways children could home in on the correct hypotheses and the best evidence. They can act themselves, performing informative experiments. They can watch and learn from the actions of others, particularly actions that have a pedagogical purpose. And they can use sampling techniques.

Learning from interventions: Exploration, experimentation and play—One of the insights of the causal models approach is that deliberately intervening on the world, and observing the outcomes of those interventions, is a particularly good way to figure out the causal structure of the world. The framework formally explains our scientific intuition that experiments tell you more about causal relationships than simple observations do. In particular, the philosopher of science Frederick Eberhardt has mathematically explored how interventions allow you to infer causal structure from data (Eberhardt 2007; see also Cook, Goodman, & Schulz, 2011).

It turns out that by intervening yourself, you can rapidly get the right evidence to eliminate many possible hypotheses, and to narrow your search through the remaining hypotheses. A less obvious, but even more intriguing, result is that these interventions need not be the systematic, carefully controlled experiments of science. The formal work shows that even less controlled interventions on the world can be extremely informative about causal structure. Multiple simultaneous interventions can be as effective as intervening on just one variable at a time. “Soft’ interventions, where the experimenter simply alters the value of a variable can be as effective as more controlled interventions, where the experimenter completely fixes that value. What we scientists disparagingly call a “fishing expedition” can still tell us a great deal about causal structure – you don’t necessarily need the full apparatus of a randomized controlled trial.

These ideas have led to a renewed and reshaped investigation of children’s play. Anyone who watches young children has seen how they ceaselessly fiddle with things and observe the results. Children’s play can look like informal experimentation. Indeed, historically, Piaget, Montessori, Bruner, and most preschool teachers have agreed that children learn through play (see Hirsh-Pasek & Golinkoff 2003; Lillard, 2005). But, how could this be, given the equally convincing observation that children’s play is often just that--playful – that is, undirected and unsystematic? In fact, other research demonstrates that even older children and naïve adults are bad at explicitly designing causally informative experiments (Chen & Klahr 1999; Kuhn 1962). If children’s playful explorations are so unconstrained how could they actually lead to rational causal learning?

Recent research by Schulz (Bonawitz et al. 2011; Cook et al. 2011; Schulz et al., 2007, 2008; see also Legare, 2012) has begun to address this issue. Schulz and her colleagues have shown that children’s exploratory play involves a kind of intuitive experimentation. Children’s play is not as structured as the ideal experiments of institutional science. Nevertheless, play is sufficiently systematic so that, like scientific fishing expeditions, it can help children discover causal structure. This research also shows that children don’t just draw the correct conclusions from the evidence they are given—they actively seek out such evidence.

In an illustrative series of experiments Schulz and Bonawitz (2007) assessed how preschool children explored a new “jack-in-the-box” type of toy. The toy had two levers that produced two effects (a duck and/or a puppet could pop up). Crucially, Schulz and Bonawitz compared two conditions, one where the causal structure of the toy was ambiguous and one where it was clear. In the *confounded* condition, an adult and the child pushed both levers simultaneously and both effects appeared. With this demonstration it was completely unclear how the toy worked. Maybe one lever produced the duck, and the other produced the puppet, maybe one lever produced both effects, maybe both levers produced both effects, etc. In the *unconfounded* condition, on the other hand, the adult pushed one lever and it systematically produced a single effect, and then the child pushed the other lever which systematically produced the other effect. In this unconfounded condition the causal structure of the toy was clear.

The experimenter placed this “old” toy and a new, simpler, single-lever toy in front of the child. Then she left the child alone, free to play with either toy. If children’s play is driven by a desire to understand causal structure, then they should behave differently in the two conditions. In the confounded condition, they should be especially likely to explore the “old” toy. In that condition the old toy’s causal structure is unclear, and further intervention could help reveal it. In the unconfounded condition, however, interventions will have no further benefit, and so children should play with the new toy instead. Indeed, 3- and 4-year-old children systematically explored the old toy rather than the new one in the confounded condition but not the unconfounded one. Moreover, after they had finished exploring the toy, children in the confounded condition showed that they had figured out how the toy worked. In a second study (Cook, Goodman & Schulz, 2011), children showed an even more sophisticated implicit ability to determine which experiments would be most informative, given their background knowledge of the causal context.

So, when young children were given a causally puzzling toy to play with, they spontaneously produced interventions on that toy and they did this in a rational way. Of course, children's play is not rational in the sense that it is explicitly designed to be an optimally effective experiment. But children's actions ensure that they receive causally relevant and informative evidence. Once that evidence is generated through play, children can use it to make the right causal inferences.

This research is not only intriguing in itself. It also shows how research inspired by probabilistic models can shed light on classic developmental questions.

Learning from interventions: Imitation, observation and pedagogy—So we can learn about causation by experimenting ourselves. But we can also learn about causation by watching what other people do, and what happens as a result. At times other people even try to demonstrate causal relations and to teach children about what causes what. This kind of observational causal learning goes beyond simple imitation. It isn't just a matter of mimicking the actions of others, instead children can learn something new about how those actions lead to effects. In fact, a number of experiments suggest that, at least by age 4, children can use information about the interventions of others in sophisticated ways to learn new causal relationships (e.g., Buchsbaum et al. 2011; Gopnik et al. 2004; Schulz et al. 2007). For example, by age 4 and perhaps earlier, children can distinguish confounded and unconfounded interventions, and recognize that confounded interventions may not be causally informative (Kushnir & Gopnik, 2005). In more recent experiments, Buchsbaum et al. (2011) showed that 4-year-olds would use statistical information to infer meaningful, causally effective goal-directed actions from a stream of movements. In an imitation task, children saw an experimenter perform five different sequences of three actions on a toy, which activated or did not activate on each trial. A statistical analysis of the data would suggest that only the last two actions of the three were necessary to activate the toy. When children got the toy they often produced just the two relevant actions, rather than imitating everything that the experimenter did.

Further studies show that while children often observe correlational information, they apparently privilege some of those correlations over others. In particular, recent findings suggest that very young children act as if correlations that result from the direct actions of others are especially causally informative.

Bonawitz et al. (2010) showed 4-year-olds and 2-year-olds simple correlations between two events that were not the outcome of human action. One box would spontaneously move and collide with a second box several times. Each time the second box would light up and then a toy plane a few inches away would spin. No human action was involved. Then they asked the children to make the plane spin themselves. The obvious course is to push the first box against the second. Four-year-olds would do this spontaneously and they would also look towards the plane as soon as they had done so. Interestingly, however, 2-year-olds were strikingly unlikely to spontaneously move the box in order to make the plane go. Although they would happily move the box if they were specifically asked to do so, even then they did not look towards the plane and anticipate the result. However, these younger children were much more likely to act themselves and to anticipate the result when they observed a human agent bring about exactly the same events. That is, when they saw an experimenter push the first block against the second and then saw the plane spin, 2-year-olds would both push the block themselves and anticipate the result. Meltzoff, et al. (in press) compared 24-month-olds, 3-year olds and 4-year-olds with different stimuli, and additional controls, with similar results—younger children were more likely to make causal inferences from correlations when the correlations were the result of human actions

Intriguingly, children can learn even more effectively from other people by making implicit assumptions about the intentions of those people. In particular, children appear to be sensitive to the fact that evidence may be the result of pedagogy – the intention to teach. Recently, Csibra and Gergeley (e.g., 2006) have suggested that even infants are sensitive to pedagogy and make different inferences when evidence comes from a teacher. For Csibra and Gergeley, this is the result of an innate set of cues pointing to pedagogical intent, such as the use of motherese and eye contact, which automatically lead children to make particular kinds of inferences.

Alternatively, however, pedagogy might have an effect by leading children to assume different kinds of probabilistic models. Shafto and Goodman (2008) have modeled these inferences in Bayesian terms, and have made quite precise predictions about how learners should make rational causal inferences from pedagogical and non-pedagogical evidence. Four-year-olds act in accord with those predictions (Buchsbaum et al. 2011; Bonawitz et al 2011).

The central idea behind the Bayesian pedagogical models is that children not only model the causal structure of the world, they also model the mind of the person teaching them about the world. Remember that causal models can specify how evidence is sampled. When the 20-month-olds in Kushnir et al.'s frog and duck study saw a non-random sample, they inferred that the agent deliberately intended to pick the frogs. This can also work in reverse – you can use what you know about someone's intentions to make inferences about how the evidence was sampled. In particular, if one person is trying to teach another they should provide an informative sample, rather than a random one. So if a learner knows that they are being taught, they can assume that the sample is informative.

For example, suppose a person shakes up novel toys in a bag, blindly extracts a few and labels each with a novel name, "dax". Contrast this with the case where instead the person looks inside and deliberately extracts exactly the same toys, shows you each one, and labels it a "dax". This second case provides pedagogical evidence. You can assume the teacher drew the sample nonrandomly to instruct you about these toys, in particular. As a result you can make different inferences about the word and the objects. Specifically, in the second case you can assume that the word is more likely to apply only to the sampled toys than to all the toys in the bag, or that all the sampled toys will behave in the same way, while the toys that were not sampled will behave differently. Even infants make these inferences (Gweon et al. 2010; Xu & Tenenbaum 2007).

In general, implicit pedagogy is an enormous asset for learning. It allows children to focus on just the hypotheses and evidence that are most relevant and significant for their culture and community. On the other hand, implicit pedagogy also has disadvantages. It may lead children to ignore some causal hypotheses. Bonawitz et al. (2011), showed children a toy that could behave in many different and non-obvious ways (pressing a button could make it beep, squeezing a bulb lit it up, etc.). When a demonstrator said that she was showing the child how the toy worked, children would simply imitate the action she performed. When the demonstrator activated the toy accidentally, children would explore the toy and discover its other causal properties.

Research on “over-imitation” also illustrates this effect. In these studies, children see another person act on the world in a complicated way to bring about an effect. Sometimes in these circumstances children act rationally, reproducing the most causally effective action (Gergeley, Bekkering & Király 2002; Southgate, et al. 2009; Williamson, Meltzoff & Markman 2008). Sometimes, though, they simply reproduce exactly the sequence of actions they see the experimenter perform--they over-imitate (Horner & Whiten 2005; Lyons et al. 2006; Tomasello, 1993). These conflicting results may seem puzzling. From a Bayesian perspective, though, this variability could easily reflect an attempt to balance two sources of information about the causal structure of the event. The statistics themselves are one source. The other is the assumption that the adult is trying to be informative.

To illustrate, earlier we described the Buchsbaum et al. (2011) study in which children rationally picked out and imitated only the causally relevant actions from a longer string. Buchsbaum et al. also did exactly the same experiment but now included pedagogical information – the experimenter said “Here’s my toy, I’m going to show you how it works”. In this case children were much more likely to “over-imitate”, that is, they assumed that everything the adult did was causally effective and imitated all her actions. Moreover, a Bayesian model predicted exactly how much children would over-imitate. Again, probabilistic models can illuminate a classical developmental problem – how, when and why children imitate.

The interventionist causal Bayes net framework suggests that children might learn causal structure especially effectively from their own interventions and from the interventions of others, particularly when those others are trying to teach them. Experimenting yourself can provide especially rich information about causal structure. Attending to the interventions of others can point children even more narrowly to just the statistical relationships that are most likely to support causal inferences. Understanding that those interventions are pedagogical adds still more information. Empirical work shows that preschoolers do indeed learn particularly effectively in these ways. Probabilistic models can predict these learning patterns quite precisely.

Sampling and variability—Earlier we described the search problems, both the problem of choosing which hypotheses to test and the problem of finding evidence to test them. Since an extremely large number of alternative hypotheses might be compatible with the evidence, we can’t learn by simply enumerating all the alternatives and testing each one. Performing interventions and observing the interventions of others can help to solve this problem. These interventions give the child additional evidence that is particularly well-designed to eliminate some relevant alternatives and discriminate among others. This kind of “active learning” has been explored in machine learning in reinforcement learning paradigms, as well as in the causal Bayes net literature, but has only just begun to be applied to Bayesian learning more generally.

Instead, the most common solution to the search problem in Bayesian machine learning is based on hypothesis sampling (see e.g., Robert & Casella, 1999). (This is different from the evidence sampling we talked about earlier in the ping-pong ball study) This solution focuses on searching among hypotheses given the evidence, rather than searching for evidence given a hypothesis.

The probabilistic Bayesian view suggests that, at least abstractly, the learner has a distribution of many possible hypotheses, some more and some less probable, rather than having just a single hypothesis. Since it is impossible to test all the hypotheses at once, the system randomly but systematically selects some hypotheses from this distribution and tests just those hypotheses. In some versions of sampling, hypotheses that are more probable to begin with are more likely to be sampled than less probable ones, but in all versions the system will try even low-probability hypotheses some of the time. Then the system can test the sampled hypotheses against the evidence. As a result of this test, the system will, in a Bayesian way, change the probability of these hypotheses, and it will adjust the distribution of all possible hypotheses accordingly. Then it samples again from this new distribution of hypotheses and so on.

Theoretically, we can think of these algorithms as procedures that search through an extremely wide space of hypotheses, and take samples from that space in order to find the correct one. In fact, however, these procedures work in a way that will be more plausible and familiar to developmental psychologists. They may, for example, take a likely hypothesis, and then “mutate’ that hypothesis to generate a number of new, slightly different hypotheses with different probabilities. Or they may take a likely hypothesis and then consider that hypothesis along with a few other similar hypotheses. Then this set of hypotheses can be tested against the data, the probability of the hypotheses can be updated and the process can be repeated. In many cases, perhaps rather surprisingly, statisticians can prove that, in the long run, this kind of step-by-step constructivist process will give the same answer you would get by searching through all the hypotheses one by one.

There are many varieties of sampling but all of them have an interesting feature: Variability among hypotheses becomes a crucial hallmark of the learning process. The probabilistic Bayesian learner entertains a variety of hypotheses, and learning proceeds by updating the probabilities of these varied hypotheses.

Developmental researchers have increasingly recognized that children also entertain multiple hypotheses and strategies at the same time. Children are typically variable. Individual children often perform correctly and incorrectly on the same task in the same session, or employ two or three different strategies on the same task on adjacent trials. As Robert Siegler (1995; 2007) has cogently emphasized this variability may actually help to explain development rather than being just noise to be ignored. We don’t just want to know that children behave differently at 4 than at 3, but why they behave differently. Variability can help.

Siegler’s examples typically come from number development. In his studies children use variable strategies for exactly the same addition problems. But the same pattern applies to intuitive theories like theory of mind or naïve biology. Consider standard change-of-location false belief tasks. A child sees Judy put her toy in the closet and go away. Judy doesn’t see that her Mom then shifts the toy to the dresser drawer. Judy returns and the child is asked “Where will Judy look for her toy?; In the closet or in the drawer?” In one intensive study (Liu, et al. 2007) almost 50 preschoolers were given 20 to 30 false-belief tasks. At one level of analysis, children were quite consistent: 65% of them passed more than 75% of these tasks, they were consistently correct, and an additional 30% passed fewer than 25% of these tasks, they were consistently incorrect – they said that Judy would look first in the drawer, the “realist” answer. Only three children were in the middle showing a fully mixed pattern. But when you examine the data in more detail, it becomes clear that there is enormous variability: *All* the children produced a mix of incorrect realism answers and correct false-belief answers.

Related data come from false belief explanation tasks. In these tasks, rather than asking for a prediction--“Where will Judy search for her toy?”--the experimenter shows the child that

Judy actually goes to the wrong place and asks for an explanation--"Why is Judy looking in the closet?" Young children offer cogent explanations; in fact, their explanations are often better than their parallel predictions (Wellman 2011). But they produce a mix of very different explanations. On successive tasks a typical child might answer "she doesn't want her toy anymore" (desire explanation), "it's empty" (reality explanation), "she doesn't know where it is" (knowledge-ignorance explanation) and "she thinks her toy's there" (belief-explanation). Amsterlaw and Wellman (2006) tested 3- and 4-year-old children on 24 such false belief explanation tasks over 6 weeks. Reality explanations were more prevalent early on and knowledge-ignorance plus belief explanations were more prevalent later. But all the children were variable, often producing two or three different explanations on the same day.

This kind of variability has sometimes led developmentalists to claim that there are no general shifts in children's understanding. Instead, children's performance is always intrinsically variable and context-dependent (see e.g., Greeno, 1998; Lave & Wenger, 1991; Thelen & Smith, 1994). However, such claims stand in tension with what appear to be genuine broad, general changes in children's knowledge. On the other hand, researchers who are interested in charting these broader changes often treat this variability as if it is simply noise to be ignored. But this in turn does not jibe with evidence that this variability actually helps children learn (Amsterlaw & Wellman 2006; Goldin-Meadow 1997; Siegler 1995).

The probabilistic approach helps us reconcile variability with broad conceptual change. In fact, variability may actually tell us something important about how broader changes take place. If children are sampling from a range of hypotheses then variability makes sense. The gradually increasing prevalence of belief explanations, for example, might reflect the fact that those hypotheses become more likely as the evidence accumulates. As a result they are more likely to be sampled and confirmed.

But does the variability in children' answers, both within and across children, actually reflect the probability of different hypotheses, as the Bayesian view would suggest? For example, will children produce many examples of high-probability hypotheses and just a few examples of low-probability hypotheses? In a first attempt to answer that question Denison et al. (2010) designed a simple experiment where the probability of different hypotheses could be precisely defined. They told 4-year-old children that either red or blue chips placed into a machine could make it go, showed them a bag of mixed red and blue chips, shook the bag and invisibly tipped out one of the chips into the machine, which activated. Then they asked the children several times whether they thought that the chip in the machine was red or blue.

In this case the probability of different hypotheses directly reflects the distribution of chips in the bag. If there are 80 red chips and 20 blue ones then there will be an 80% chance that the "red chip" hypothesis is right. If children were simply randomly responding they should guess red and blue equally often. If they were simply trying to maximize their successful answers, they should always say red. But if their responses are the result of a sampling process, they should choose red 80% of the time and blue 20% of the time, that is they should "probability match". If the distribution is 60/40 instead of 80/20 they should adjust their responses so that they guess "red chip' less often, and 'blue chip" more often. In fact, this is just what the children do.

Moreover, children did this in a way that went beyond the simple probability matching we see in reinforcement learning (Estes, 1950). In an additional experiment, children saw two bags, one with two blue chips and one with a mix of 14 red and 6 blue chips. The experimenter picked one of the closed bags at random and tipped the chip into the bag. The probability that a blue chip was in the machine equaled the probability that it would fall out of the bag times the 50% chance that that bag was chosen. So blue chips were actually more likely to end up in the machine than red ones, even though there were more red than blue chips overall. In this condition, children's responses did not simply match the frequency of the chips overall, greater for red than blue, but rather matched the probability that chips of each color would end up in the machine, greater for blue than red. So children were not simply updating their behavior based on reinforcement, or matching their responses to the perceptual distribution of the chips. Instead they seemed to be genuinely generating hypotheses based on their probability.

More Theoretical Advances: Hierarchical Bayes Nets

Bayes nets are good representations of particular causal structures, even complex causal structures. However, according to the theory theory often children are not just learning particular causal structures but are also learning abstract generalizations about causal structure. For example, in addition to learning that my desire for frogs causes me to take them out of the box, children may develop a broader generalization that I always prefer frogs to other toys. Or more generally still, they may conclude that desires are likely to differ in different people.

In fact, classsic empirical "theory theory" research showed that children develop more abstract, framework knowledge over and above their specific causal knowledge. For example, when they make judgments about objects, children often seem to understand broad causal principles before they understand specific details. 3- and 4-year-olds, like adults, know that biological objects, like an egg or a pig, have different insides than artifacts, like a watch or a piggy bank. They also know that those insides are important to identity and function. At the same time, however, they are notably inaccurate and vague about just what those insides actually are (Gelman & Wellman 1991). They say that biological objects have blood and guts (even an egg) and artifacts have gears and stuffing inside (even a piggy bank). Similarly, in causal tasks, children assume that objects with similar causal powers will have similar insides, even before they know exactly what those insides are actually like (Sobel et al. 2007).

These broader generalizations are important in both scientific and intuitive theories. Philosophers of science refer to "overhypotheses" (Goodman 1955), or "research programs" (Laudan 1977), or "paradigms" (Kuhn 1962) to capture these higher-order generalizations. Cognitive developmentalists have used the term "framework theories" (Carey 2009; Wellman 1990; Wellman & Gelman 1992). In their framework theories, children assume there are different kinds of variables and causal structure in psychology versus biology versus physics. In fact, they often understand these abstract regularities before they understand specific causal relationships (see e.g., Simons & Keil 1995).

Some nativists argue that this must mean that the more abstract causal knowledge is innate. In contrast, constructivists, including Piaget and theory theorists, insist that this more abstract causal knowledge could be learned. But how could this be? Bayes nets tell us how it is possible to learn specific causal structure. How is it possible, computationally, to learn these more abstract over-arching causal principals?

Griffiths and Tenenbaum (2007; 2009; Tenenbaum, et al. 2011) inspired by both philosophy of science and cognitive development, have formulated computational ways of representing and learning higher-order generalizations about causal structure. Following Gelman, et al. (2003) they call their approach hierarchical Bayesian modeling (HBM) or, sometimes, theory-based Bayesian modeling. The idea is to have meta-representations, that is, representations of the structure of particular Bayes nets and of the nature of the variables and relationships involved in those causal networks. These higher-level beliefs can constrain the more particular hypotheses represented by particular Bayes nets. Moreover, these higher-level generalizations can themselves be learned by Bayesian methods.

In standard Bayesian modeling a particular Bayes net represents a specific hypothesis about the causal relations among particular variables. Hierarchical Bayesian models stack up hypotheses at different levels. The higher levels contain general principles that specify which hypotheses to entertain at the lower level.

Here is an example, from biology. Consider a family of causal graphs representing causal relations between sleeping, drinking, and exposure to cold etc. on the one hand and metabolizing energy, being active, being strong, growing, having a fever, and healing etc. on the other, as in Figure 5. Perhaps the relations might be captured by the graph of nodes and arrows on the left hand side of Figure 5, graph A. Perhaps instead, the correct causal structure is graph B or graph C. In fact, something very like graph B is what Inagaki and Hatano (2002, 2004) describe as the naïve “vitalistic biology” apparent in the cognition of 4-, and 5-year-olds (see their 2004 p. 42, Figure 2.3), while something like graph C would be more like the theory of scientific medicine.

Although graphs B and C themselves differ they are actually both versions of the same more abstract graph schema. In this schema, all the nodes fall into three general categories: Input situations, that is, external forces that affect an organism (sleep, exposure, eating), Outcome occurrences, that is characteristics of the organism itself (getting sick, healing, growing), and internal biological Processes (metabolizing energy). Both graphs B and C (in contrast to graph A) have this general form, but differ in specifics. Both could be generated from a simple higher order framework theory that goes something like this:

1. There are three types of nodes: Input, Process, and Outcome
2. Possible causal relationships *only take the form of* Input→Process and Process→Outcome, or in total Input→Process→Outcome.

Note that Input→Process→Outcome is more general and abstract than any of the more specific graphs in Figure 5. Input→Process→Outcome is at a higher level in several ways. Input, Process, and Outcome are not themselves any of the nodes in any of the graphs (which are instead “eat food”, “heal”, etc.). Further, Input→Process→Outcome does not itself directly “contact” the evidence, which would include dependencies between “eats food” and “gets sick”, or “eats food” and “metabolizes energy”.

Recall that Bayesian reasoning means that we can solve the inverse problem and determine the posterior, the probability of the hypothesis given the evidence, by using what we know about the likelihood and the prior. P (H|E) is a function of P (H) and P(E|H).

The idea behind *hierarchical* Bayesian learning is to use the same reasoning but now relate a particular hypothesis to a higher level framework theory instead of relating the hypothesis to evidence. We can think of the specific hypothesis H, as evidence for a higher-level theory, T. Then we can consider the probability of T given H, P(H|T). That is, we can specify the likelihood of lower-level hypotheses given higher-level theories--P($H_{| |}$T)--just as we can specify the likelihood of the evidence given a lower-level hypothesis--P(E|H). High-level theories can act as constraints on inference at the lower level. Lower-level hypotheses provide evidence for the higher-level theories. Moreover, T could also be related to a yet more general framework theory T_1, and so on to produce a hierarchy of theories at different levels.

Going back to Figure 5, we can think of graphs A, B, and C as specific theories of biology. Input→Process→Outcome is a higher-level framework theory, not a specific theory. This framework theory generates some specific theories (e.g., graphs B and C) but *not* others (not graph A, for example). The higher-level framework theory Input→Process→Outcome does not directly contact the data. But because it generates some specific theories and not others, the higher-level framework theory will indirectly confront the data via the specific theories it generates.

There are a variety of possible framework theories just as there are a variety of specific theories and even more specific hypotheses. Bayesian inference lets us specify the probability of different framework theories, just as it lets us specify the probability of different specific theories and different hypotheses. We can infer the probability of a framework theory from the probability of the specific theories it generates, just as we can infer the probability of a specific hypothesis from the evidence.

Computational work on HBMs has shown that, at least normatively, hierarchical Bayesian learning can actually work. Higher level framework theories can indeed be updated in a Bayesian way via evidence that contacts only lower level hypotheses. Griffiths & Tenenbaum (2007) provide several simple demonstrations, Kemp, et al. (2007) and Goodman, et al. (2011) provide more comprehensive and complex ones. These demonstrations show that it is possible, in principle, for learning to proceed at several levels at once—not just at the level of specific hypotheses but also at the level of specific theories and, even more abstractly, at the framework theory level.

At the least, these demonstrations provide intriguing thought experiments. They suggest that data-driven learning can not only change specific hypotheses but can also lead to more profound conceptual changes, such as the creation of more abstract theories and framework theories. These computational thought experiments underwrite the feasibility of constructivist accounts.

HBMs also help address the hypothesis search problem. If a learner initially assumes a particular framework theory, she might only test the specific theories that are generated by that framework theory, rather than other hypotheses. If a 4-year-old believes the “vitalist” framework theory of biology she may consider hypotheses about whether sleeping well or eating well makes you grow, but she won’t initially consider the hypothesis that growing makes you sleep well or that exposure to cold makes you heal poorly.

Constructivists insist that the dynamic interplay between structure and data can yield both specific kinds of learning and more profound development as well. Hierarchical Bayesian models provide a more detailed computational account of how this can happen. On the hierarchical Bayesian picture local causal learning can, and will, lead to broader, progressive, theory revision and conceptual change.

Abstract Learning in Childhood

While these hierarchical Bayesian models have been inspired by developmental data, they are new. So only a few very recent experimental developmental studies have specifically tested hierarchical Bayesian ideas.

Learning abstract schemas

Dewar and Xu (2010) have shown that even infants can infer abstract regularities from patterns of data in their category learning. In the causal case, more focally, Schulz et al. (2008) designed an experiment in which blocks from different underlying and nonobvious categories would interact according to different general causal principles. When some blocks banged together they made a noise but when others banged together they did not. In fact, you could explain the pattern of evidence by assuming that the blocks fell into three categories, X, Y and Z. X activates Y and Y activates Z but X does not activate Z. Four-year-old children used a few samples of evidence to infer the existence of these three distinct categories and then generalized this high-level rule to new cases. They assumed that new blocks would fall into one of these three general categories.

Lucas et al. (in press) investigated whether children could infer generalizations about the logic of a causal system. First they showed children a "blicket detector" that followed either a disjunctive rule (all red blocks make it go) or a conjunctive one (you need both a red and blue block to make it go). Then they placed a new set of differently colored blocks on the detector. The children saw an ambiguous pattern of evidence that was compatible in principle with either the disjunctive or conjunctive principle. 4-year-olds generalized the prior higher-order rule to the new blocks – they assumed that the new blocks acted disjunctively or conjunctively based on the earlier evidence.

These two recent studies show that given appropriate patterns of evidence 4-year-old children can go beyond inferring specific causal relationships. They can also infer more abstract generalizations in the way that hierarchical Bayes nets would propose. Even before the children knew exactly which blocks caused exactly which effects they had inferred some general principles – the blocks would fit into one of three categories, or they would act conjunctively.

Progressive learning in childhood

Hierarchical Bayesian models also rely crucially on variability among hypotheses. They test multiple specific theories and framework theories, and update the probability of those theories in the light of new evidence, just as probabilistic Bayesian models do in general. From this hierarchical perspective variability can be thought of not only "synchronically" (children adopt multiple approaches at one time) but also "diachronically" (different approaches emerge over time). This means that as hierarchical Bayesian learning proceeds over multiple iterations, intermediate transitional hypotheses emerge. In particular, in the learning process some abstract hypotheses progressively come to dominate others but then themselves become dominated by still others. HBMs, as they dynamically operate on evidence over time, result in characteristic progressions of intermediate hypotheses—progressions of specific hypotheses and more abstract theories and framework theories (Ullman, Goodman, & Tennenbaum 2010).

If children are hierarchical probabilistic Bayesian learners, then they should also produce intermediate hypotheses, and those hypotheses should improve progressively. In fact, children's conceptual development does progress in this way. Indirect evidence of such progressions is available, for example, in studies of children's naïve astronomy (Vosniadou & Brewer 1992) and naïve biology (Inagaki & Hatano 2002). But more direct evidence is provided by "microgenetic" studies that track conceptual change longitudinally over days or weeks (see e.g., Siegler 2007). Recent research on preschoolers' theories of mind illustrates this approach..

Some recent research claims that 12- to 15-month-old infants are already aware that actors act on the basis of their beliefs and false beliefs (e.g., Onishi & Baillargeon 2005; Surian, et al. 2007). It is not yet clear how to best interpret these infant "false belief" findings nor how to integrate them with the preschool research. Some insights from the probabilistic models framework may be relevant to this problem, however. First, we can ask whether children represent a coherent network of causal beliefs, like a complex causal Bayes net, or instead simply have isolated representations of particular causal links. The claim that 3- and 4-year-olds have different theories of mind is not, and never was, based on their performance on false-belief tasks alone. Instead, theory theorists argued for a theoretical change based on the simultaneous and highly correlated emergence of many conceptually related behaviors. These behaviors involve explanation as well as prediction, and involve the understanding of sources of information, appearance and reality, and representational change as well as predictions about actions (Wellman 2002; Gopnik & Wellman 1992). For example, passing false-belief tasks is highly and specifically correlated with children's ability to understand that their own beliefs have changed, and that appearances differ from reality (Gopnik & Astington 1988). Similarly, 3-year-old children who are trained on understanding belief show an improved understanding of the sources of their knowledge, but do not show a similarly improved understanding of number conservation (Slaughter & Gopnik 1997). Infants may have early pieces of this network but only integrate them into a coherent whole between 3 and 5.

A second related issue concerns the distinction between prediction and causal representation that we outlined earlier. The infant findings largely come from "looking-time methods" which reflect whether or not infants predict specific outcomes. It may be that these results reflect the fact that infants have collected correlational evidence that will later be used to construct causal representations. Infants on this view, might be like Tycho Brahe – the astronomer who collected the predictive data that Copernicus and Galileo used to construct the heliocentric causal theory. Brahe could predict with some accuracy what the stars would do but he could not explain why they acted this way. Later causal representations would then allow children to use this information in more sophisticated ways, for example, to design novel interventions on the world, to make counterfactual inferences or to provide explanations. An infant might notice that people's actions are consistently related to their perceptions, without treating those patterns as evidence for a causal model of the mind.

Regardless of whether these distinctions explain the changes from infancy to early childhood, there are definitely changes in how children think about the mind during the preschool years. Preschoolers don't just go from failing to passing false-belief tasks between 2 and 5. Instead they develop a series of insights about the mind. Importantly, individual differences in how rapidly or in what sequence children achieve these insights predict other key childhood competences. These include how children talk about people in everyday conversation, their engagement in pretense, their social interactional skills, and their interactions with peers (Astington & Jenkins 1999; Lalonde & Chandler 1995; Watson et al. 1999). These individual differences are related to IQ and executive function, but also have a distinct developmental profile (e.g., Carlson & Moses 2001; Wellman et al. 2008). None of this is consistent with the idea that all the concepts of intuitive psychology are in place innately and they are only masked by performance problems.

Let's look at the progressive changes in preschooler's theory of mind more closely. It has been clear for a long time that children can understand peoples' desires and intentions before they understand their false beliefs (Wellman & Liu 2004 provide a meta-analytic review). But this transition actually involves a more revealing, extended set of conceptual progressions. This becomes apparent when children are assessed using a recently established Theory of Mind Scale (Wellman & Liu 2004). The scale includes carefully constructed tasks that assess children's understanding of (1) Diverse Desires (people can have different desires directed at the same thing), (2) Diverse Beliefs (people can have different beliefs about the same situation), (3) Knowledge-Ignorance (something can be true, but someone doesn't know it), (4) False Belief (something can be true, but someone might believe something different), and (5) Hidden Emotion (someone can feel one way but display a different emotion). Preschoolers solve these tasks in this order and the same 5-step progression characterizes American (Wellman & Liu 2004; Wellman, et al. in press), Australian (Peterson, et al. 2005), and German (Kristen, et al. 2006) preschoolers.

If these progressions reflect hierarchical Bayesian learning, then they should vary depending on the learners' experiences. Recent studies demonstrate two kinds of variation. One is variation in timetables. Deaf children of hearing parents go through the same 5-step sequence as their hearing peers but they are very delayed. They take 12 to 15 years to proceed through the same steps that take hearing children 5 or 6 years (Peterson et al. 2005; Wellman et al. 2011). In contrast, deaf children of deaf parents, who learn sign as a native language, do not show these delays. The deaf-of-hearing children have much less conversational experience than hearing or deaf-of-deaf children and this probably leads to the very delayed appearance of each step, even though the sequence of those steps is the same.

The other variation involves differences in sequencing which reflect different childhood experiences. For example, American and Chinese children are immersed in different languages and cultures, which emphasize quite different aspects of intuitive psychology. These differences lead to related differences in their early theory-of-mind progression – the sequence of insights is different rather than delayed (Wellman, et al. 2006).

To elaborate, a theory of mind is the product of social and conversational experiences that may vary from one community to another. Western and Chinese childhood experiences could be crucially different. Cultural psychologists have suggested that Asian cultures emphasize the fact that people are interdependent and function in groups, while Western cultures emphasize independence and individuality (Nisbet, 2003). These differences lead to different emphases. Asian cultures may focus on common perspectives while American cultures focus on the diversity of beliefs. Moreover, Western and Chinese adults seem to have very different everyday epistemologies. Everyday Western epistemology is focused on truth, subjectivity, and belief; Confucian-Chinese epistemology focuses more on pragmatic knowledge acquisition and the consensual knowledge that all right-minded persons should learn (Li 2001; Nisbet 2003). Indeed, in conversation with young children, Chinese parents comment predominantly on "knowing", whereas U.S. parents comment more on "thinking" (Tardif & Wellman 2000).

In accord with these different cultural emphases Chinese preschoolers develop theory of mind insights in a different sequence than Western children, Both groups of children understand the diversity of desires first. But Chinese children, unlike Western children consistently understand knowledge acquisition before they understand the diversity of beliefs. (Wellman, et al. 2006; Wellman, et al. 2011).

This extended series of developmental achievements fits the hierarchical Bayesian learning, theory-construction perspective. On the Bayesian view children should develop a characteristic sequence of theories as their initial hypotheses become progressively revised in the face of new evidence. In fact, Goodman et al. (2007) have proposed a Bayesian model of theory-of-mind learning that captures the characteristic changes in children's understanding. Moreover, the probabilistic Bayesian perspective would predict that the sequence of theories can differ depending on the learner's exact "diet" of evidence. The comparison of deaf and hearing children and American and Chinese children supports this prediction.

The Blessings of Abstraction

So far we have shown that hierarchical Bayesian models can provide computational accounts that explain how children might learn abstract framework theories from specific theories, which are learned from evidence. But the developmental findings also suggest that children sometimes develop abstract framework theories before they develop detailed specific theories.

The conventional wisdom in psychology has been that learning at a lower level of generalization and abstraction – more concrete learning – must precede higher-order, more abstract, and more general learning. This idea has been presupposed both by empiricists and by nativists. So when infants or very young children understand abstract structure, this is often taken to mean that this structure is innate (see e.g., Spelke 1992). Although some developmentalists have stressed instead that young children often seem to *learn* abstract regularities before specific ones (Simons & Keil 1995; Wellman & Gelman, 1998), this is a distinctively unconventional proposal—it might appear that this kind of learning just couldn't work.

In fact, recent work on hierarchical Bayesian modeling has shown that sometimes abstract generalizations can actually precede specific ones (e.g., Goodman, et al. 2011). In principle, then, children may be learning both specific facts about my desires, and generalizations about all desires, from the same evidence and at the same time.

A simple example can illustrate this. Suppose I show you a pile of many bags of colored marbles and your job is to learn the color of the marbles in each bag. Now I draw some marbles out the bags. I start with Bag 1. I take out 1 then 2 then 3 then 4 red marbles in a row. Each time I ask you what you think will happen next. After a while you'll predict that yet another red marble will appear from Bag 1. Then I repeat this with Bag 2. This time I take out a succession of blue marbles. By Bag 3 you might well conclude: "I don't know the color, but whatever the first one is I bet all the rest in that bag will be the same color." You've learned an abstract regularity, and your knowledge of that regularity *precedes* your learning of the specific color in Bag 3 or, for that matter, any other remaining bag. Note that this abstract structure was certainly not innate, it was learned. Under the right circumstances abstract regularities (e.g., T = all marbles in Bag 3 will be the same color) can be learned in advance of the specifics (H = all marbles in Bag 3 are purple). The philosopher Nelson Goodman (1955) called these more abstract principles "overhypotheses"--hypotheses about which hypotheses are likely.

Kemp, et al. (2007) demonstrate how overhypotheses can be learned in hierarchical Bayesian models, and building on these ideas, Goodman, et al. (2011) have used hierarchical Bayesian modeling to provide a striking set of computational results they call "the blessing of abstraction." They have shown that, in hierarchical Bayesian models, it is often as easy to learn causal structure at several levels at once as it is to simply infer particular causal structure. Moreover, learning *both* higher-level and specific structures from the data can be no slower (that is, requires no more data samples) than learning only the specific structures and having the abstract ones "innately" specified at the start.

Probabilistic hierarchical Bayesian learners thus learn abstract structures alongside and even before the specifics that those regularities subsume. Arguably children do the same thing.

Conclusions

The interchange between cognitive developmental scientists and probabilistic Bayesian modelers has already informed us about theories, learning, and development, and it promises further insights. The empirical results we have described are not easily compatible with either traditional empiricism or nativism. The children in these studies are clearly not simply associating particular inputs. Instead they infer more abstract causal structures, including the very abstract structures involved in overhypotheses or framework theories. At the same time, it is plain that these structures are not simply present innately and then triggered, or masked by performance constraints. In the learning experiments, children receive evidence for new causal structures that they do not already know – even very young children learn these new structures rapidly and accurately. Similarly, the gradual sequence of progressively more accurate theories, and the fact that this sequence unfolds differently with different evidence patterns, are both difficult to explain in nativist terms.

Thinking about computation provides new empirical projects for developmentalists and thinking about development also provides new projects for the modelers. We conclude by briefly listing some of these future projects.

Several advances in the computational world could inspire new developmental investigations. Developmentalists have only just begun to explore how the computational work on hierarachical Bayesian models might be reflected empirically in the development of framework theories. There is also computational work which suggests that what appear to be causal primitives, like the very idea of intervention or causation itself, could, in principle, be constructed from simpler patterns of evidence (Goodman et al. 2011). There is important empirical work to be done exploring which representations are in place initially and which are learned. Models can help direct such empirical endeavors.

Perhaps the most significant area where new computational work can inform development involves the algorithmic instantiations of computational principles. Most of the formal work so far has been at what David Marr (1982) called the computational level of description. That is, the models show how it is possible, normatively, and in principle, to learn structure from patterns of evidence. However, we can also ask how it is possible to actually do this in real time in a system with limited memory and information-processing capacity. There are many different specific algorithms being explored in machine learning, and we don't yet know which algorithms children might use or approximate. Similarly, formal work specifies how active intervention can inform learning, and these ideas can also be found in work on "active learning" in computation. Although we know that children actively explore the world, we don't know in detail how their exploration shapes and is shaped by learning.

Ultimately, of course, we would also like to know how these algorithms are actually instantiated in the brain. A few recent studies have explored the idea that neural circuits instantiate Bayesian inferences (Knill & Pouget 2004) but this work is only just beginning.

Equally, paying attention to development raises new questions for computationalists. The conceptual changes that children go through are still more profound than any the computational models can currently explain. Even hierarchical Bayes nets are still primarily concerned with testing hypotheses against evidence, and searching through a space of hypotheses. It is still not clear exactly how children generate what appear to be radically new hypotheses from the data.

Some learning mechanisms have been proposed in cognitive development to tackle this issue, including the use of language and analogy. In particular, Carey (2009) has compellingly argued that specific linguistic structures and analogies play an important role in conceptual changes in number understanding, through a process she calls "Quinean bootstrapping". There is empirical evidence that the acquisition of particular linguistic structures can indeed reshape conceptual understanding in several other domains, closer to intuitive theories (see e.g. Casasola, 2005; Gopnik, Choi & Baumberger, 1996; Gopnik & Meltzoff, 1997; Pyers & Senghas 2009; Shusterman & Spelke, 2005).

But it is difficult to see how language or analogy alone could lead to these transformations. In order to recognize that a linguistic structure encodes some new, relevant conceptual insight it seems that you must already have the conceptual resources that the structure is supposed to induce. Similarly, philosophers have long pointed out that the problem with analogical reasoning is the proliferation of possible analogies. Because an essentially infinite number of analogies are possible in any one case, how do you pick analogies that reshape your conceptual understanding in relevant ways and not get lost among those that will simply be dead ends or worse? In the case of mathematical knowledge, these problems may be more tractable because such knowledge is intrinsically deductive. But in the case of inductively inferring theories there are a very wide range of possible answers. When many linguistic structures could encode the right hypothesis, or many analogies could be relevant, the problem becomes exponentially difficult. These proposals thus suffer from the same constructivist problem we have been addressing all along. And so, again, characterizing the influence of language and analogy in more precise computational terms might be very helpful. If probabilistic and hierarchical Bayesian models can help solve the riddle of induction, then perhaps they can shed light on these other learning processes as well.

Moreover, empirically, developmentalists have discovered several other phenomena that seem to be involved in theory change but that have yet to be characterized in computational terms. Much recent work in developmental psychology has explored how children can learn from the testimony of others (Koenig, Clément, & Harris, 2004). As we noted earlier, computationalists are just starting to provide accounts of the sort of social learning that is involved in intuitive pedagogy: there is still much to be done in understanding how children learn from other people.

Similarly, there is a great deal of work suggesting that explanations play an important role in children's learning (Wellman 2011). Even very young children ask for and provide explanations themselves and respond to requests for explanations from others (e.g., Callanan & Oakes 1992), and these explanations actually seem to help children learn (Amsterlaw & Wellman 2006; Siegler 1995; Legare 2012). But there is no account of explanation in computational terms.

Schulz and colleagues have shown that exploratory play has an important role in causal learning. But other kinds of play, particularly pretend and imaginary play, are equally ubiquitous in early childhood and seems to have an important role in early learning. However, that role is still mysterious computationally.

The relation between learning in infancy and in the preschool period is also unresolved. There is extensive evidence that very young infants detect statistical patterns. There is also evidence that 16 to 20-month-olds can infer causal structure from those patterns (Gweon & Schulz, 2012; Kushnir et al. 2010; Sobel & Kirkham, 2007, Xu and Ma, 2011). We still don't know what happens in between.

More generally, there are questions about the relationship between learning and broader development. For the hierarchical probabilistic models framework, and for that matter, for the theory theory itself, there is no principled difference between inference, learning and development. Accumulated experience can lead to profound and far-reaching developmental change – the equivalent of "paradigm shifts" in science. Nevertheless, we can still ask whether there is something about children, in particular, that makes them different kinds of learners than adults.

Both evolutionary and neurological considerations suggest that this might be true. Childhood is a period of protected immaturity in which children are free to learn and explore without the practical constraints of adult life. Children's brains appear to be more generally flexible and plastic than adult brains, and children seem to be particularly flexible learners. From a computational perspective, some of this flexibility may reflect the fact that children simply have less experience and so have lower "priors" than adults do. But there may also be more qualitative differences between adult and child learners. In the reinforcement learning literature, for example, there is a distinction between "exploring" and "exploiting".

Different computational mechanisms may be most effective when an organism is trying to learn novel information and when it is trying to make the best use of the information it already has. Children may be designed to start out exploring and only gradually come to exploit. Computational models that reflect and explain these broad developmental changes would be particularly interesting.

We have shown that new computational ideas coupled with new cognitive development research promise to reconstruct constructivism. The new computational research relies on probabilistic Bayesian learning and hierarchical Bayesian modeling. The new cognitive development research studies the mechanisms of childhood causal learning. The new studies show how exploration and experimentation, observation and pedagogy, and sampling and variability lead to progressively more accurate intuitive theories. These advances provide a more empirically and theoretically rich version of the theory theory. Collaboration between cognitive development and probabilistic modeling holds great promise. It can help produce more precise developmental theories and more realistic computational ones. It may even explain, at last, how children learn so much about the world around them.

Acknowledgments

We are grateful to Noah Goodman, Laura Schulz and two anonymous reviewers for very helpful comments and to Sheba Shakir and Sophie Bridges for assistance. The paper was supported by NICHD grant HD022149 to HW and NSF BCS-1023875 to AG. Most significantly. we acknowledge and deeply thank the members of the McDonnell Causal Learning Collaborative and the James S. McDonnell Foundation itself. The ideas and research reported here are a direct consequence of the vision of both our colleagues in the collaborative and the foundation that supported them.

References

Amsterlaw J, Wellman HM. Theories of mind in transition: A microgenetic study of the development of false belief understanding. Journal of Cognition and Development. 2006; 7(2):139–172.10.1207/s15327647jcd0702_1

Astington JW, Jenkins JM. A longitudinal study of the relation between language and theory-of-mind development. Developmental Psychology. 1999; 35(5):1311–1320.10.1037/0012-1649.35.5.1311 [PubMed: 10493656]

Baillargeon R. Innate ideas revisited: For a principle of persistence in infants' physical reasoning. Perspectives on Psychological Science. 2008; 3(1):2–13.10.1111/j.1745-6916.2008.00056.x [PubMed: 22623946]

Bonawitz E, Shafto P, Gweon H, Goodman ND, Spelke ES, Schulz L. The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. Cognition. 2011; 102(3):322–330.10.1016/j.cognition.2010.10.001 [PubMed: 21216395]

Bonawitz EB, Ferranti D, Saxe R, Gopnik A, Meltzoff AN, Woodward J, Schulz LE. Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. Cognition. 2010; 115(1):104–117.10.1016/j.cognition.2009.12.001 [PubMed: 20097329]

Buchsbaum D, Gopnik A, Griffiths T. Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. Cognition. 2011; 120(3):331–340.10.1016/j.cognition.2010.12.001 [PubMed: 21338983]

Callanan MA, Oakes LM. Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. Cognitive Development. 1992; 7:213–233.

Carey, S. Conceptual change in childhood. Cambridge, Mass: MIT Press; 1985.

Carey, S. The origin of concepts. New York: Oxford University Press; 2009.

Carlson SM, Moses LJ. Individual differences in inhibitory control and children's theory of mind. Child Development. 2001; 72(4):1032–1053.10.1111/1467-8624.00333 [PubMed: 11480933]

Casasola M. Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. Developmental Psychology. 2005; 41(1):183–192.10.1037/0012-1649.41.1.183 [PubMed: 15656748]

Chen Z, Klahr D. All other things being equal: Acquisition and transfer of the control of variables strategy. Child Development. 1999; 70(5):1098–1120.10.1111/1467-8624.00081 [PubMed: 10546337]

Cook C, Goodman N, Schulz LE. Where science starts: Spontaneous experiments in preschoolers' exploratory play. Cognition. 2011; 120(3):341–349.10.1016/j.cognition.2011.03.003 [PubMed: 21561605]

Csibra, G.; Gergely, G. Social learning and social cognition: The case for pedagogy. Processes of Change in Brain and Cognitive Development. In: Munakata, Y.; Johnson, MH., editors. Attention and Performance. Vol. XXI. Oxford: Oxford University Press; 2006. p. 249-274.

Denison, S.; Bonawitz, L.; Gopnik, A.; Griffiths, T. Preschoolers sample from probability distributions. Poster presented at the Cognitive Science Society; Portland, OR. August 2010; 2010.

Dewar KM, Xu F. Induction, overhypothesis, and the origin of abstract knowledge. Psychological Science. 2010; 21(12):1871–1877.10.1177/0956797610388810 [PubMed: 21078899]

Dweck, CS. Self-theories: their role in motivation, personality, and development. Philadelphia, PA: Psychology Press; 1999.

Elman, JL.; Bates, EA.; Johnson, MH.; Karmiloff-Smith, A.; Parisi, D.; Plunkett, K. Rethinking innateness: A connectionist perspective on development. Cambridge, Mass: MIT Press; 1996.

Eberhardt F, Scheines R. Interventions and causal inference. Philosophy of Science. 2007; 74(5):981–995.10.1086/525638

Estes WK. Toward a statistical theory of learning. Psychological Review. 1950; 57(2):94–107.10.1037/h0058559

Geisler WS. Sequential ideal-observer analysis of visual discriminations. Psychological Review. 1989; 96(2):267–314.10.1037/0033-295x.96.2.267 [PubMed: 2652171]

Gelman, SA. The essential child: origins of essentialism in everyday thought. Oxford ; New York: Oxford University Press; 2003.

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian analysis. 2. New York: Chapman & Hall; 2003.

Gelman SA, Wellman HM. Insides and essences: Early understandings of the non-obvious. Cognition. 1991; 38(3):213–244.10.1016/0010-0277(91)90007-Q [PubMed: 2060270]

Gergely G, Bekkering H, Kiraly I. Rational imitation in preverbal infants. Nature. 2002; 415(6873): 755.10.1038/415755a [PubMed: 11845198]

Glymour, C.; Cooper, GF. Computation, causation, and discovery. Menlo Park, CA: AAAI Press; 1999.

Glymour, CN. The mind's arrows: Bayes nets and graphical causal models in psychology. Cambridge, MA: MIT Press; 2003.

Gold EM. Language identification in the limit. Information and Control. 1967; 10(5):447–474.10.1016/S0019-9958(67)91165-5

Goldin-Meadow S. When gestures and words speak differently. Current directions in psychological scienc. 1997; 6(5):138–143.10.1111/1467-8721.ep10772905

Gómez RL. Variability and detection of invariant structure. Psychological Science. 2002; 13(5):431–436.10.1111/1467-9280.00476 [PubMed: 12219809]

Goodman, N. Fact, fiction, and forecast. Cambridge, MA: Harvard University Press; 1955.

Goodman ND, Ullman TD, Tenenbaum JB. Learning a theory of causalitiy. Psychological Review. 2011; 118(1):110–119.10.1037/a0021336 [PubMed: 21244189]

Gopnik A. Conceptual and semantic development as theory change: The case of object permanence. Mind & Language. 1988; 3(3):197–216.10.1111/j.1468-0017.1988.tb00143.x

Gopnik A, Choi S, Baumberger T. Cross-linguistic differences in early semantic and cognitive development. Cognitive Development. 1996; 11(2):197–225.10.1016/s0885-2014(96)90003-9

Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, Danks D. A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review. 2004; 111(1):3–32.10.1037/0033-295X.111.1.3 [PubMed: 14756583]

Gopnik, A.; Meltzoff, AN. Words, thoughts, and theories. Cambridge: MIT Press; 1997.

Gopnik A, Sobel DM, Schulz LE, Glymour C. Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental Psychology. 2001; 37(5):620–629.10.1037/0012-1649.37.5.620 [PubMed: 11552758]

Gopnik A, Wellman HM. Why the child's theory of mind really is a theory. Mind and Language. 1992; 7:145–171.10.1111/j.1468-0017.1992.tb00202.x

Gopnik, A.; Wellman, HM. The theory theory. In: Hirschfeld, L.; Gelman, S., editors. Domain specificity in cognition and culture. New York: Cambridge University Press; 1994. p. 257-293.

Greeno JG. The situativity of knowing, learning, and research. American Psychologist. 1998; 53(1):5–26.10.1037/0003-066x.53.1.5

Griffiths TL, Chater N, Kemp C, Perfors A, Tenenbaum JB. Probabilistic models of cognition: Exploring representations and inductive biases. Trends in Cognitive Sciences. 2010; 14(8):357–364.10.1016/j.tics.2010.05.004 [PubMed: 20576465]

Griffiths TL, Sobel DM, Tenenbaum JB, Gopnik A. Bayes and blickets: Effects of knowledge on causal induction in children and adults. Cognitive Science. Oct 4.2011 [Epub ahead of print]. 10.1111/j.1551-6709.2011.01203.x

Griffiths, T.; Tenenbaum, JB. Two proposals for causal grammars. In: Gopnik, A.; Schulz, L., editors. Causal learning: Psychology, philosophy, and computation. New York: Oxford University Press, Inc; 2007. p. 323-345.

Griffiths TL, Tenenbaum JB. Theory-based causal induction. Psychological Review. 2009; 116(4): 661–716.10.1037/a0017201 [PubMed: 19839681]

Gweon H, Schulz L. 16-month-olds rationally infer causes of failed actions. Science. 2011; 332(6037): 1524–1524.10.1126/science.1204493 [PubMed: 21700866]

Gweon H, Tenenbaum JB, Schulz LE. Infants consider both the sample and the sampling process in inductive generalization. Proceedings of the National Academy of Sciences. 2010; 107(20):9066–9071.10.1073/pnas.1003095107

Hirsh-Pasek, K.; Golinkoff, RM. Einstein never used flash cards: How our children really learn–and why they need to play more and memorize less. Emmaus, PA: Rodale Inc; 2003.

Heckerman, D.; Meek, C.; Cooper, GF. A Bayesian approach to causal discovery. In: Glymour, CN.; Cooper, GF., editors. Computation, causation, and discovery. Menlo Park, CA: Cambridge, MA: AAAI Press; MIT Press; 1999.

Horner V, Whiten A. Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). Animal Cognition. 2005; 8(3):164–181.10.1007/s10071-004-0239-6 [PubMed: 15549502]

Inagaki, K.; Hatano, G. Young children's naive thinking about the biological world. New York: Psychology Press; 2002.

Inagaki K, Hatano G. Vitalistic causality in young children's naive biology. Trends in Cognitive Sciences. 2004; 8(8):356–362.10.1016/j.tics.2004.06.004 [PubMed: 15335462]

Kahneman D, Tversky A. On the reality of cognitive illusions. Psychological Review. 1996; 103(3): 582–591.10.1037/0033-295x.103.3.582 [PubMed: 8759048]

Keil, FC. Concepts, kinds, and cognitive development. Cambridge, Mass: MIT Press; 1989.

Kemp C, Goodman ND, Tenenbaum JB. Learning to learn causal models. Cognitive Science. 2010; 34(7):1185–1243.10.1111/j.1551-6709.2010.01128.x [PubMed: 21564248]

Kemp C, Perfors A, Tenenbaum JB. Learning overhypotheses with hierarchical Bayesian models. Developmental Science. 2007; 10(3):307–321.10.1111/j.1467-7687.2007.00585.x [PubMed: 17444972]

Kersten D, Mamassian P, Yuille A. Object perception as Bayesian inference. Annual Review of Psychology. 2004; 55(1):271–304.10.1146/annurev.psych.55.090902.142005

Kirkham NZ, Slemmer JA, Johnson SP. Vital statistical learning in infancy: Evidence of a domain general learning mechanism. Cognition. 2002; 83(2):B35–B42.10.1016/S0010-0277(02)00004-5 [PubMed: 11869728]

Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences. 2004; 27(12):712–719.10.1016/j.tins.2004.10.007 [PubMed: 15541511]

Koenig MA, Clément F, Harris PL. Trust in testimony: Children's use of true and false statements. Psychological Science. 2004; 15(10):694–698.10.1111/j.0956-7976.2004.00742.x [PubMed: 15447641]

Kristen S, Thoermer C, Hofer T, Aschersleben G, Sodian B. Skalierung von "theory of mind" aufgaben (Scaling of theory of mind tasks). Zeitschrift fur Entwicklungspsychologic und Padagogische Psychologie. 2006; 38(4):186–195.10.1026/0049-8637.38.4.186

Kuhn, TS. The structure of scientific revolutions. Chicago: University of Chicago Press; 1962.

Kushnir T, Gopnik A. Young children infer causal strength from probabilities and interventions. Psychological Science. 2005; 16(9):678–683.10.1111/j.1467-9280.2005.01595.x [PubMed: 16137252]

Kushnir T, Gopnik A. Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. Developmental Psychology. 2007; 43(1):186–196.10.1037/0012-1649.43.1.186 [PubMed: 17201518]

Kushnir T, Xu F, Wellman HM. Young children use statistical sampling to infer the preferences of other people. Psychological Science. 2010; 21(8):1134–1140.10.1177/0956797610376652 [PubMed: 20622142]

Lalonde CE, Chandler MJ. False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. Cognition and Emotion. 1995; 9(2–3):167–185.10.1080/02699939508409007

Laudan, L. Progress and its problems: toward a theory of scientific growth. Berkeley: University of California Press; 1977.

Lave, J.; Wenger, E. Situated learning: Legitimate peripheral participation. Cambridge: Cambridge University Press; 1991.

Legare CH. Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. Child Development. 2012; 83:173–185. [PubMed: 22172010]

Leslie AM. Developmental parallels in understanding minds and bodies. Trends in Cognitive Sciences. 2005; 9(10):459–462.10.1016/j.tics.2005.08.002 [PubMed: 16125434]

Leslie AM, Keeble S. Do six-month-old infants perceive causality? Cognition. 1987; 25(3):265–288.10.1016/s0010-0277(87)80006-9 [PubMed: 3581732]

Lillard, AS. Montessori: The science behind the genius. Oxford, UK ; New York: Oxford University Press; 2005.

Lu H, Yuille A, Liljeholm M, Cheng PW, Holyoak KJ. Bayesian generic priors for causal learning. Psychological Review. 2008; 115:955–984.10.1037/a0013256 [PubMed: 18954210]

Lucas C, Gopnik A, Griffiths T. Developmental differences in learning the forms of causal relationships. Proceedings of the Cognitive Science Society. (in press).

Lyons DE, Santos LR, Keil FC. Reflections of other minds: How primate social cognition can inform the function of mirror neurons. Current Opinion in Neurobiology. 2006; 16(2):230–234.10.1016/j.conb.2006.03.015 [PubMed: 16564687]

Marr, D. Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman; 1982.

Meltzoff AN, Waismeyer A, Gopnik A. Learning about causes from people: Observational causal learning in 24-month-old infants. Developmental Psychology. (in press).

Michotte, A. The perception of causality. New York: Basic Books; 1963.

Muentener P, Carey S. Infants' causal representations of state change events. Cognitive Psychology. 2010; 61(2):63–86.10.1016/j.cogpsych.2010.02.001 [PubMed: 20553762]

Murphy GL, Medin DL. The role of theories in conceptual coherence. Psychological Review. 1985; 92(3):289–316.10.1037/0033-295x.92.3.289 [PubMed: 4023146]

Notaro PC, Gelman SA, Zimmerman MA. Children's understanding of psychogenic bodily reactions. Child Development. 2001; 72(2):444–459.10.1111/1467-8624.00289 [PubMed: 11333077]

Oaksford, M.; Chater, N. Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press; 2007.

O'Neil DK. Two-year-old children's sensitivity to a parent's knowledge state when making requests. Child Development. 1996; 67(2):659–677.10.1111/j.1467-8624.1996.tb01758.x

Onishi KH, Baillargeon R. Do 15-month-old infants understand false beliefs? Science. 2005; 308(5719):255–258.10.1126/science.1107621 [PubMed: 15821091]

Pearl, J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo, CA: Morgan Kaufman; 1988.

Pearl, J. Causality: Models, reasoning, and inference. New York: Cambridge University Press; 2000.

Perner J, Ruffman T. Infants' insight into the mind: How deep? Science. 2005; 308(5719):214–216.10.1126/science.1111656 [PubMed: 15821079]

Peterson CC, Wellman HM, Liu D. Steps in theory-of-mind development for children with deafness or autism. Child Development. 2005; 76(2):502–517.10.1111/j.1467-8624.2005.00859.x [PubMed: 15784096]

Pinker, S. Language learnability and language development. Cambridge, Mass: Harvard University Press; 1984.

Pinker, S. How the mind works. New York: Norton; 1997.

Pinker, S. Learnability and cognition: The acquisition of argument structure. Cambridge, Mass: MIT Press; 1991.

Pyers JE, Senghas A. Language promotes false-belief understanding: Evidence from learners of a new sign language. Psychological Science. 2009; 20(7):805–812.10.1111/j.1467-9280.2009.02377.x [PubMed: 19515119]

Ramsey J, Gazis P, Roush T, Spirtes P, Glymour C. Automated remote sensing with near infrared reflectance spectra: Carbonate recognition. Data Mining and Knowledge Discovery. 2002; 6(3): 277–293.10.1023/a:1015421711749

Rehder B, Kim S. How causal knowledge affects classification: A generative theory of categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32(4):659–683.10.1037/0278-7393.32.4.659 ; 10.1037/0278-7393.32.4.659.supp (Supplemental)

Rhodes M, Gelman SA. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. Cognitive Psychology. (in press).

Robert, CP.; Casella, G. Monte Carlo statistical methods. New York: Springer-Verlag; 1999.

Rogers, TT.; McClelland, JL. Semantic cognition: A parallel distributed processing approach. Cambridge, Mass: MIT Press; 2004.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274(5294):1926–1928.10.1126/science.274.5294.1926 [PubMed: 8943209]

Schulz LE, Bonawitz EB. Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. Developmental Psychology. 2007; 43(4):1045–1050.10.1037/0012-1649.43.4.1045 [PubMed: 17605535]

Schulz LE, Gopnik A. Causal learning across domains. Developmental Psychology. 2004; 40(2):162–176.10.1037/0012-1649.40.2.162 [PubMed: 14979758]

Schulz LE, Gopnik A, Glymour C. Preschool children learn about causal structure from conditional interventions. Developmental Science. 2007; 10(3):322–332.10.1111/j.1467-7687.2007.00587.x [PubMed: 17444973]

Schulz LE, Sommerville J. God does not play dice: Causal determinism and preschoolers' causal inferences. Child Development. 2006; 77(2):427–442.10.1111/j.1467-8624.2006.00880.x [PubMed: 16611182]

Schulz LE, Standing HR, Bonawitz EB. Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. Developmental Psychology. 2008; 44(5): 1266–1276.10.1037/0012-1649.44.5.1266 [PubMed: 18793061]

Seiver E, Gopnik A, Goodman N. "Did she jump because she was not afraid or because the trampoline was safe?" Causal inference and the development of social cognition. Child Development. (in press).

Shafto, P.; Goodman, N. Teaching games: Statistical sampling assumptions for pedagogical situations. Proceedings of the 30th Annual Conference of the Cognitive Science Society; 2008.

Shipley, B. Cause and correlation in biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge, UK: Cambridge University Press; 2000.

Shusterman, A.; Spelke, ES. Language and the development of spatial reasoning. In: Carruthers, P.; Laurence, S.; Stich, S., editors. The innate mind: Structure and contents. New York: Oxford University Press; 2005. p. 89-106.

Siegler, RS. Children's thinking: How does change occur?. In: Schneider, W.; Weinert, F., editors. Memory performance and competencies. Hillsdale, NJ: Erlbaum; 1995.

Siegler RS. Cognitive variability. Developmental Science. 2007; 10(1):104–109. [PubMed: 17181707]

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

Sloman, SA. Causal models: How people think about the world and its alternatives. New York: Oxford University Press; 2005.

Smith C, Carey S, Wiser M. On differentiation: A case study of the development of the concepts of size, weight, and density. Cognition. 1985; 21(3):177–237.10.1016/0010-0277(85)90025-3 [PubMed: 3830547]

Sobel DM, Kirkham NZ. Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. Developmental Science. 2007; 10(3):298–306.10.1111/j.1467-7687.2007.00589.x [PubMed: 17444971]

Sobel DM, Tenenbaum JB, Gopnik A. Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive Science. 2004; 28(3):303–333.10.1207/s15516709cog2803_1

Sobel DM, Yoachim CM, Gopnik A, Meltzoff AN, Blumenthal EJ. The Blicket within: Preschoolers' inferences about insides and causes. Journal of Cognition and Development. 2007; 8(2):159–182.10.1080/15248370701202356 [PubMed: 18458796]

Southgate V, Chevallier C, Csibra G. Sensitivity to communicative relevance tells young children what to imitate. Developmental Science. 2009; 12(6):1013–1019.10.1111/j.1467-7687.2009.00861.x [PubMed: 19840055]

Spelke ES, Breinlinger K, Macomber J, Jacobson K. Origins of knowledge. Psychological Review. 1992; 99(4):605–632.10.1037/0033-295x.99.4.605 [PubMed: 1454901]

Spelke ES, Kinzler KD. Core knowledge. Developmental Science. 2007; 10(1):89–96.10.1111/j.1467-7687.2007.00569.x [PubMed: 17181705]

Spirtes, P.; Glymour, C.; Scheines, R. Causation, prediction and search, Springer Lecture Notes in Statistics. 2. MIT Press; 1993. 2000

Spirtes, P.; Christopher, M.; Richardson, T. Causal inference in the presence of latent variables and selection bias. Uncertainty in artificial intelligence: Proceedings of the eleventh conference; San Francisco, CA: Morgan Kaufmann; 1997. p. 499-506.

Surian L, Caldi S, Sperber D. Attribution of beliefs by 13-month-old infants. Psychological Science. 2007; 18(7):580–586.10.1111/j.1467-9280.2007.01943.x [PubMed: 17614865]

Tardif T, Wellman HM. Acquisition of mental state language in Mandarin- and Cantonese-speaking children. Developmental Psychology. 2000; 36(1):25–43.10.1037/0012-1649.36.1.25 [PubMed: 10645742]

Tenenbaum, JB.; Griffiths, T.; Niyogi, S. Intuitive theories as grammars for causal inference. In: Gopnik, A.; Schulz, L., editors. Causal learning: Psychology, philosophy, and computation. New York: Oxford University Press, Inc; 2007. p. 301-322.

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: Statistics, structure, and abstraction. Science. 2011; 331(6022):1279–1285.10.1126/science.1192788 [PubMed: 21393536]

Thelen, E.; Smith, LB. A dynamic systems approach to the development of cognition and action. Cambridge, Mass: MIT Press; 1994.

Tomasello M. It's imitation, not mimesis. Behavioral and Brain Sciences. 1993; 16(4):771–772.10.1017/S0140525X00032921

Ullman, TD.; Goodman, ND.; Tennenbaum, JB. Theory acquisition as stochastic search. Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society; 2010.

Vosniadou S, Brewer WF. Mental models of the earth: A study of conceptual change in childhood. Cognitive Psychology. 1992; 24(4):535–585.10.1016/0010-0285(92)90018-w

Waldmann MR, Hagmayer Y, Blaisdell AP. Beyond the information given: Causal models in learning and reasoning. Current Directions in Psychological Science. 2006; 15(6):307–311.10.1111/j. 1467-8721.2006.00458.x

Watson AC, Nixon CL, Wilson A, Capage L. Social interaction skills and theory of mind in young children. Developmental Psychology. 1999; 35:386–391.10.1037/0012-1649.35.2.386 [PubMed: 10082009]

Wellman, HM. The child's theory of mind. Cambridge, Mass: MIT Press; 1990.

Wellman HM. Reinvigorating explanations for the study of early cognitive development. Child Development Perspectives. 2011; 5:33–38.

Wellman HM, Fang F, Liu D, Zhu L, Liu G. Scaling of theory-of-mind understandings in Chinese children. Psychological Science. 2006; 17(12):1075–1081.10.1111/j.1467-9280.2006.01830.x [PubMed: 17201790]

Wellman HM, Fang F, Peterson CC. Sequential progressions in a theory of mind scale: Longitudinal perspectives. Child Development. 2011; 82(3):780–792.10.1111/j.1467-8624.2011.01583.x [PubMed: 21428982]

Wellman HM, Gelman SA. Cognitive development: Foundational theories of core domains. Annu Rev Psychol. 1992; 43:337–375.10.1146/annurev.ps.43.020192.002005 [PubMed: 1539946]

Wellman HM, Liu D. Scaling of theory-of-mind tasks. Child Development. 2004; 75(2):523–541.10.1111/j.1467-8624.2004.00691.x [PubMed: 15056204]

Wellman HM, Lopez-Duran S, LaBounty J, Hamilton B. Infant attention to intentional action predicts preschool theory of mind. Developmental Psychology. 2008; 44(2):618–623.10.1037/0012-1649.44.2.618 [PubMed: 18331149]

Williamson RA, Meltzoff AN, Markman EM. Prior experiences and perceived efficacy influence 3-year-olds' imitation. Developmental Psychology. 2008; 44(1):275–285.10.1037/0012-1649.44.1.275 [PubMed: 18194026]

Wolpert DM. Probabilistic models in human sensorimotor control. Human Movement Science. 2007; 26(4):511–524.10.1016/j.humov.2007.05.005 [PubMed: 17628731]

Woodward, J. Making things happen: A theory of causal explanation. New York: Oxford University Press; 2003.

Wu R, Gopnik A, Richardson DC, Kirkham NZ. Infants learn about objects from statistics and people. Developmental Psychology. 2011; 47(5):1220–1229.10.1037/a0024023 [PubMed: 21668098]

Xu F, Denison S. Statistical inference and sensitivity to sampling in 11-month-old infants. Cognition. 2009; 112(1):97–104.10.1016/j.cognition.2009.04.006 [PubMed: 19435629]

Xu, F.; Dewar, K.; Perfors, A. Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. In: Hood, BM.; Santos, L., editors. The origins of object knowledge. Oxford, UK: Oxford University Press; 2009. p. 263-284.

Xu F, Garcia V. Intuitive statistics by 8-month-old infants. Proceedings of the National Academy of Sciences. 2008; 105(13):5012–5015.10.1073/pnas.0704450105

Xu F, Ma L. Young children's use of statistical sampling evidence to infer the subjectivity of preferences. Cognition. 2011; 120:403–411. dx.doi.org/10.1016/j.cognition.2011.02.003. [PubMed: 21353215]

Xu F, Tenenbaum JB. Word learning as Bayesian inference. Psychol Rev. 2007; 114(2):245–272.10.1037/0033-295X.114.2.245 [PubMed: 17500627]

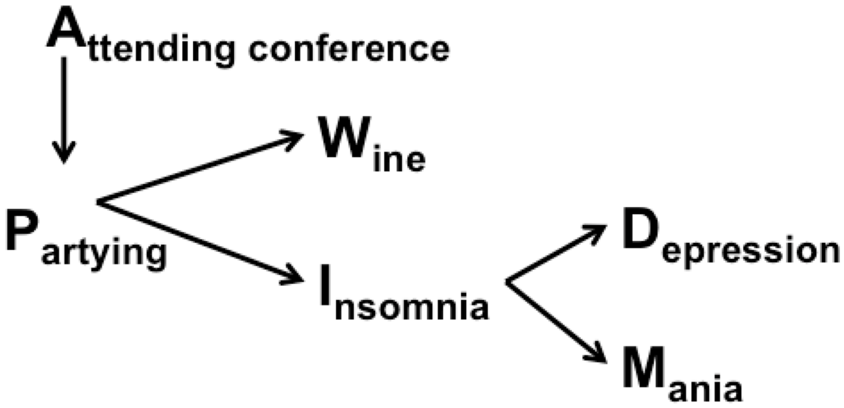

Figure 1.
Causal Bayes net of academic conferences (and their consequences). Causal Bayes nets can connect any variables with connected edges. In this example, to keep things concrete, A = attending a conference; P = partying; W = drinking wine; I = insomnia; D = depression; and M = mania.

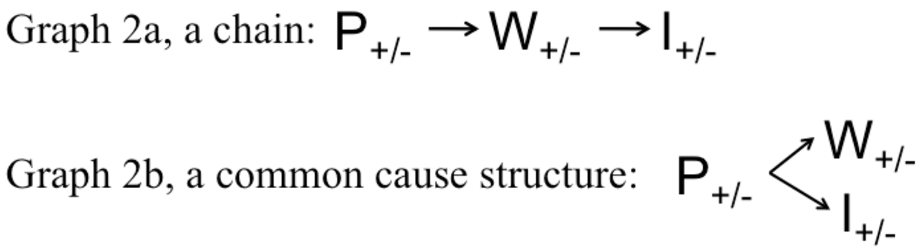
Graph 2a, a chain: P+/- → W+/- → I+/-
Graph 2b, a common cause structure: P+/-
W+/-
I+/-

Figure 2.
Simple causal graphs of two alternative causal relations between partying (P), drinking wine (W), and insomnia (I).

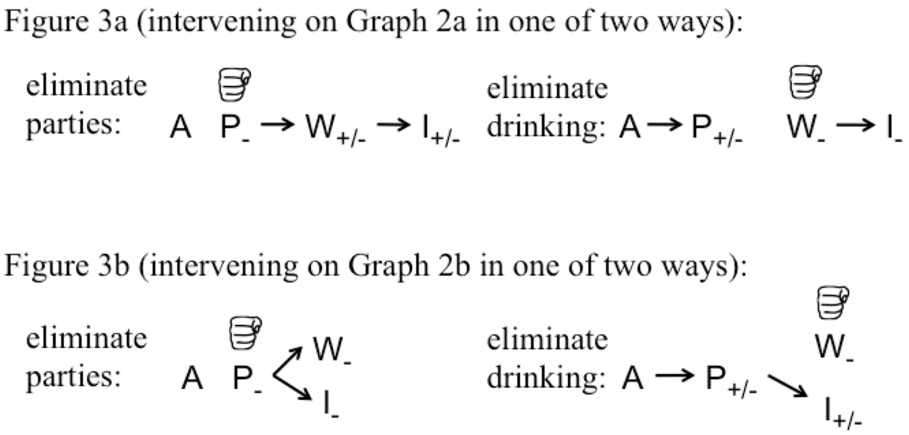

Figure 3.

Altered graphs showing the results of interventions on Graphs 2a and 2b (from Figure 2) under two different interventions: eliminating partying or eliminating wine.

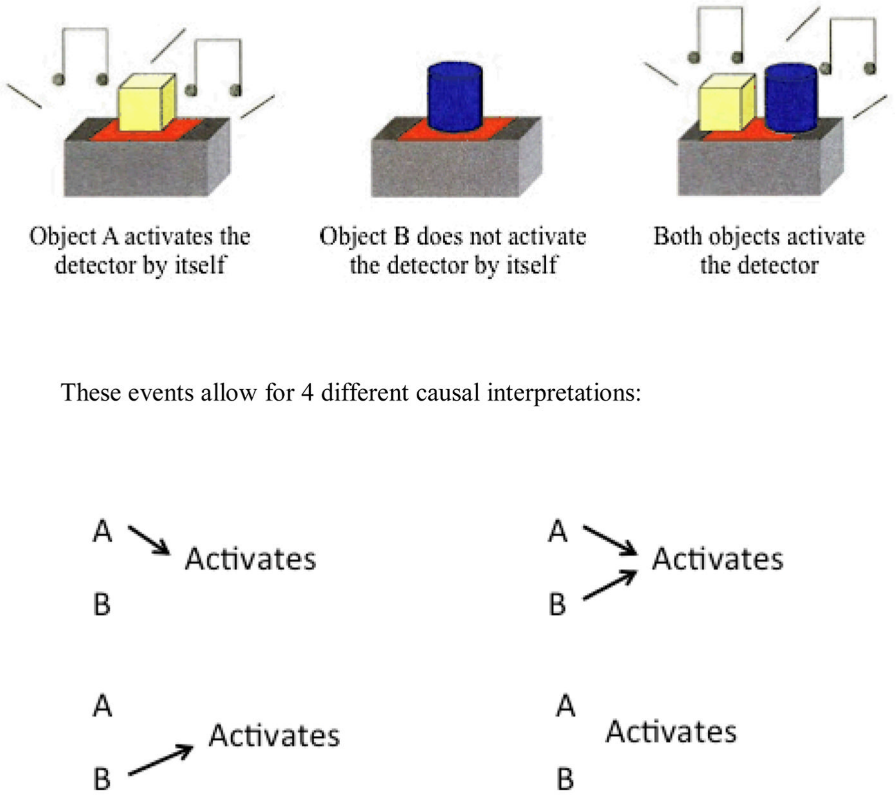
Object A activates the detector by itself
Object B does not activate the detector by itself
Both objects activate the detector
These events allow for 4 different causal interpretations:
A
Activates
B
A
Activates
B
A
Activates
B
A
Activates
B

Figure 4.
Example blicket detector and a sequence of events that do and do not activate the detector. These events allow for four different causal interpretations, presented in abbreviated Bayes net form at the bottom of the figure.

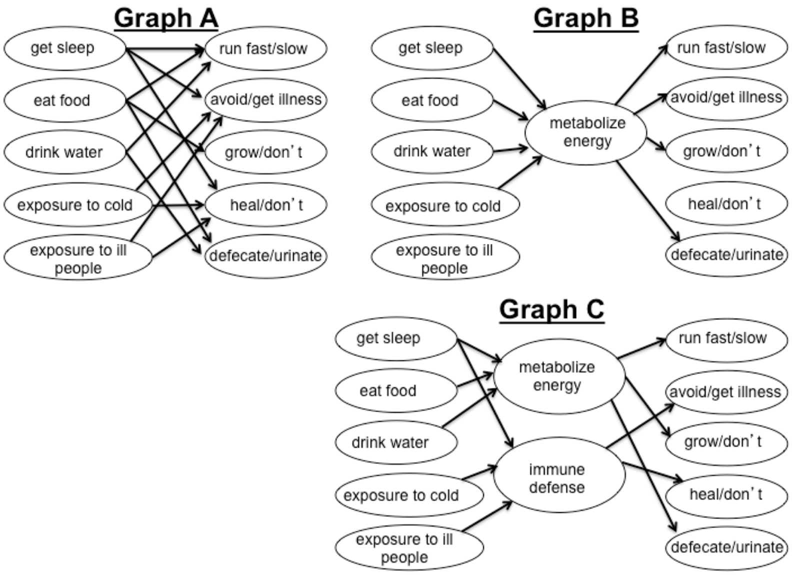

Figure 5.
Three different causal Bayes nets of commonplace biological events.