A brand new ball game: Bayes net and neural net learning mechanisms in young children.

Alison Gopnik Dept. of Psychology University of California at Berkeley

Clark Glymour Dept. of Philosophy Carnegie-Mellon University and Institute for Human and Machine Cognition, University of West Florida.

gopnik@berkeley.edu

In Processes of change in brain and cognitive development: Attention and performance xxi. 2006 Attention and Performance. 349-372.

Abstract

We outline a new computational account of learning in children using the causal Bayes net formalism. We also present evidence that children as young as two years old use something like causal Bayes net learning mechanisms to infer the causal structure of the world around them. This kind of learning may play an important role in the development of intuitive theories. Finally we contrast causal Bayes net and neural net learning mechanisms.

Over the past thirty years we have discovered an enormous amount about what children know and when they know it. That research has completely transformed the traditional Piagetian view of young children's cognition. Even the youngest infants seem to have abstract representations of the world and are not restricted to "sensori-motor" schemas. Similarly, preschool children are far from being the illogical, egocentric "precausal" creatures Piaget envisioned.

In particular, young children, and even infants, seem to have intuitive theories of the physical, biological and psychological world (for reviews see Gopnik & Meltzoff, 1997; Gelman & Raman, 2002; Flavell, 1999; Wellman & Inagaki, 1997). These theories, like scientific theories, are complex, coherent, abstract representations of the causal structure of the world. Even the youngest preschoolers can use these intuitive theories to make causal predictions, provide causal explanations, and even reason about causation counterfactually (Harris at al . 1996; Hicking & Wellman 2001; Sobel, 2004, Wellman et al, 1997). As with scientific theories, children's theories are much more than just summaries of the evidence—they allow children to draw novel

conclusions and provide novel explanations. Moreover, there is extensive evidence for changes in intuitive theories as children grow older.

Are these changes the result of maturation or are they the result of learning? Several recent findings suggest the latter alternative. First, natural variations in the kinds of evidence children receive can influence the development of intuitive theories. For example, rich city-dwelling children who are relatively deprived of biological experience have a less elaborated view of biology than children growing up on Indian reservations (Ross et al. 2003). Similarly, children with older siblings, who have a wider range of psychological experiences, seem to have an accelerated understanding of intuitive psychology (Ruffman et al. 1998). Moreover, training studies show that providing children with specific kinds of evidence relevant to biological or psychological theories can accelerate the development of those theories (Slaughter, 1996; 1999; 2003). Even providing a child with a pet fish can influence their folk biology (Inagaki & Hitano, 2004).

But the real question for developmental cognitive science is not so much what children know, when they know it or even whether they learn it. The real question is HOW they learn it and WHY they get it right. In the past "theory theorists" have suggested that children's learning mechanisms are

analogous to scientific theory-formation. However, what we would really like is a more precise computational specification of the mechanisms that underlie both types of learning.

The traditional candidates for learning mechanisms in psychology have been variants of associationism, either the mechanisms of classical and operant conditioning in behaviorist theories (e.g, Rescorla & Wagner 1972) or more recently, the mechanisms of dynamical systems theories (eg. Thelen & Smith, 1994) and connectionist theories (e.g. Rumelhart & McLelland 1986, Elman et al 1997; Shultz, 2003, Rogers & McLelland 2004). Of these, only connectionist models really offer computational mechanisms: Behaviorism eschews them altogether, while dynamical systems theories claim that learning is the result of direct physical interactions between the organism and the environment, with no internal computations at all.

Such theories have had great difficulty explaining how apparently rich, complex, abstract, rule-governed representations, the sorts of representations encoded in everyday theories, could be derived from evidence. Typically, associationist theories, both the earlier behaviorist theories and their recent connectionist and dynamic inheritors, denied that such representations really exist. Although children might appear to have rule-governed abstract

representations of the world, these theorists argue that, in fact, they have a collection of much more specific learned associations between inputs and outputs.

Connectionists often qualify this denial by appealing to the notion of "distributed" representations, concepts, and categories, or any psychological process that relates inputs to outputs. The distributed representation is whatever features of the connectionist system produce the relevant outputs from the relevant inputs. On this view however, the representations are not independent accounts of the external world that are responsible for input-output relations. Instead, they are summaries of those input-output relations

Conversely, more nativist theories endorse the existence of abstract rule-governed representations but deny that they are learned. Modularity or "core knowledge" theorists, for example, suggest that there are a few innate causal schemas designed to fit particular domains of knowledge, such as a belief-desire schema for intuitive psychology or a teleological schema for intuitive biology. Development is either a matter of enriching those innate schemas, or else involves quite sophisticated and culture-specific kinds of learning like those of the social institutions of science (see eg. Spelke et al . 1992).

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

To take the computationalist conception seriously, think for a moment of children as very, very complex natural robots. We psychologists want to explain how those robots learn, much as a computer scientist might want to know how a chess playing computer works. The right answer for the computer scientist is a description of an algorithm, perhaps in a high level computer language. Something like that is what is needed for the developmental psychology of learning as well. But that is not what the traditional approaches offer. Behaviorists tell us not to even try to explain--just summarize the input/output regularities. Dynamical systems theorists tell us to describe the physics of the process—which is much like telling a computer scientist who

wants to know how a chess playing computer manages to play chess that she should study the physics of silicon chips. Connectionists tell us to try to model the neural connections and dependencies that produce chilren's remarkable learning capacities—but to do it without biological data about how individual cells act and interact when learning takes place. That is very much like telling the computer scientist that to understand how the computer plays chess, she needs to reconstruct the machine code—not the high level chess playing program.

Computer scientists learned very quickly that computational understanding has levels, and the physical level is useful for building machines but not for understanding the details of their behavior, and the machine code level is typically unintuitive and unrevealing. But there have been few promising theories of learning at the higher computational level that is typically revealing in computer science – that is at the level of representations and algorithms acting on those representations.

Several recent theoretical and empirical developments suggest that this situation may be changing. In particular, recent theoretical advances suggest computational learning procedures which allow abstract, coherent, structured

representations to be derived from patterns of evidence, given certain assumptions. These procedures relate structured representations such as graph structures, grammars or representations of three-dimensional space to patterns of input, particularly patterns of conditional probabilities among events. These computational accounts take the kinds of evidence that have been considered in traditional learning accounts – such as evidence about contingencies among events or evidence about the consequences of actions – and use it to learn structured representations of the kind that have been proposed in traditional nativist accounts. Accounts like these have become increasingly dominant in artificial intelligence and machine learning. In this paper we will focus on one such computational account, the causal directed graphical model or causal Bayes net account of causal knowledge and learning.

Causal Bayes Nets. Causal directed graphical models, or causal Bayes nets, have been developed in the philosophy of science and statistical literature over the last fifteen years (Glymour 2001; Pearl 2000; Spirtes et al. 1993). "Theory theorists" in cognitive development point to an analogy between learning in children and learning in science. Causal Bayes nets

provide a formal account of a kind of inductive inference that is particularly important in scientific theory-formation. Scientists infer causal structure by observing the patterns of conditional probability among events (as in statistical analysis) by examining the consequences of interventions (as in experiments) or, usually, by combining the two types of evidence.

In causal Bayes nets, causal relations are represented by directed acyclic graphs. 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 (see figure 1). The variables can be discrete (like school grade) or continuous (like weight), they can be binary (like "having eyes" or "not having eyes") or many valued (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 the nature of these relations is called the "parameterization" of the graph.

Insert Figure 1 about here

The structure of a causal graph constrains the conditional probabilities among the variables in that graph, no matter what the variables are or what the parameterization of the graph is. In particular, it constrains the conditional independencies among those variables. Given a particular causal structure, only some patterns of conditional independence will occur among the variables.

Conditional and unconditional dependence and independence can be precisely defined mathematically. Two variables X and Y are unconditionally independent in probability if and only if for every value x of X and y of Y the probability of x and y occurring together equals the unconditional probability of x multiplied by the unconditional probability of y. That is p (x & y) = p (x) * p (y.*)*.Two variables are independent in probability conditional on some third variable Z if and only if p (x, y | z) = p (x | z) * p (y | z). That is for every value x,y, and z of X, Y and Z the probability of x and y given z equals the probability of x given z multiplied by the probability of y given z.

The structure of the causal graph puts constraints on these patterns of probability among the variables. These constraints can be captured by a single formal assumption, the Causal Markov Assumption as follows:

The Causal Markov Assumption: For any variable X in a causal graph, X is independent of all other variables in the graph (except for its own direct and indirect effects) conditional on its own direct causes.

Causal Bayes nets also allow us to determine what will happen when we intervene from outside to change the value of a particular variable. When two variables are genuinely related in a causal way then, holding other variables constant, intervening to change one variable should change the value of the other. Indeed, philosophers have recently argued that this is just what it means for two variables to be causally related (Woodward, 2003). Given a causal graph, particular interventions will only have effects on some variables and not others. The Bayes net formalism captures these relations through a second assumption, an assumption about how interventions should be represented in the graph.

The Intervention Assumption: A variable I is an intervention on a variable X in a causal graph if and only if 1) I is exogenous (that is, is not caused by any other variables in the graph) 2) directly fixes the value of X to x and 3)

does not affect the values of any other variables in the graph except through its influence on X.

Given this assumption we can accurately predict the effects of interventions on particular variables in a graph on other variables. In causal Bayes nets interventions systematically alter the nature of the graph they intervene on, and these systematic alterations follow directly from the formalism itself. In particular, when an external intervention fixes the value of a variable it also eliminates the causal influence of other variables on that variable. This can be represented by replacing the original graph with an altered graph in which arrows directed into the intervened upon variable are eliminated (Judea Pearl vividly refers to this process as graph surgery (Pearl 2000)). The conditional dependencies among the variables after the intervention can be read off from this altered graph. This same inferential apparatus can be used to generate counterfactual predictions.

A central aspect of causal Bayes nets, indeed the thing that makes them causal, is that they allow us to freely go back and forth from evidence about observed probabilities to inferences about interventions and vice-versa.

These two assumptions, then, allow us to take a particular causal structure and accurately predict the conditional probabilities of events, and also the consequences of interventions on those events, from that structure.

To illustrate this consider three simple causal graphs ,1) a chain X → Y→ Z, 2) a common effect structure Z ← X → Y. and 3) a common cause structure X→Z←Y. Suppose, for example, that I notice that I often can't sleep when I've been to a party and drunk lots of wine, partying (X) and insomnia (Z) covary, and so do wine (Y) and insomnia (Z). This covariation by itself is consistent with all three of the structures above. Maybe parties lead me to drink and wine keeps me up, maybe parties both keep me up and lead me to drink, maybe parties don't affect my wine drinking but parties keep me up and wine independently keeps me up.

However, each structure will lead to a different pattern when I intervene on wine and partying, if, for example, I experiment by intentionally sitting in my room alone and drinking one night, and then partying dead sober the next. I can calculate the effects of such interventions on each of the three causal structures, using "graph surgery" and predict the results. I will get different results from these experiments depending on the true causal

structure (solitary drinking will lead to insomnia, and sober partying won't for graph 1, sober partying will lead to insomnia and solitary drinking won't for graph 2, and both experiments will lead to insomnia for graph 3).

Even if I can't experiment, however, I can still discriminate these three graphs by looking at the patterns of conditional probability among the three variables. If graph #1 is right, and there is a causal chain that goes from parties to wine to insomnia, then $Y \perp Z \mid X$ – the probability of insomnia occurring is independent (in probability) of the probability of party-going occurring conditional on the occurrence of wine-drinking. If graph #2 is right, and parties are a common cause of wine and insomnia, then $X \perp Y \mid Z$ – the probability of wine-drinking occurring is independent (in probability) of the probability of insomnia occurring conditional on the occurrence of party-going.

Insomnia might also be a common effect of both wine-drinking and parties ($X \rightarrow Y \leftarrow Z$). In this case, X is <u>not</u> independent of Z conditional on Y. The intuitions here are less obvious, but they reflect the fact that, in this case, knowing about the effect and about one possible cause gives us information about the other possible cause. We can illustrate this best with a different example. Suppose X is a burglar, Y is the burglar alarm sounding, and Z is the

neighboring cat tripping the alarm wire, so that Y is a common effect of X and Z. If we hear the alarm sound and see the cat tripping the wire, we are less likely to conclude that there was a burglar than if we simply hear the alarm sound by itself. Similarly, if partying and wine-drinking really were completely independent – if we were just as likely to drink wine at home, then if someone else just knew that we had insomnia and that we had been at a party that would actually lessen the probability that we had also been drinking wine. This effect is called "explaining away" (see Pearl 2000, Spirtes et al. 1993 for discussion).

These systematic relations between causation, intervention and conditional probability allow a range of systematic predictions. We can also use the formalism to work backwards and learn the causal graph from patterns of conditional probability and intervention. This type of learning requires an additional assumption – the faithfulness assumption.

The Faithfulness Assumption: In the joint distribution on the variables in the graph, all conditional independencies are consequences of the Markov assumption applied to the graph.

Given the faithfulness assumption, it is possible to infer complex causal structure from patterns of conditional probability and intervention (Glymour & Cooper, 1999; Spirtes et al., 1993). Computationally tractable learning algorithms have been designed to accomplish this task and have been extensively applied in a range of disciplines (eg Ramsey et al. 2002; Shipley 2000). In some cases, it is also possible to accurately infer the existence of new unobserved variables that are common causes of the observed variables (Silva et al. 2003 , Spirtes et al. 2003).

These learning algorithms typically take two forms. In Bayesian learning, a prior probability for various causal graphs is calculated first (Heckerman, 1999). Then these prior probabilities are updated given the evidence about the actual conditional probabilities in the data, using Bayes rule. In constraint based learning the graphs are constructed from the probabilities directly, step by step in a more bottom-up way (Spirtes et al. 1993). These two types of learning can also be combined.

The Bayesian methods have the advantage that they can easily integrate prior knowledge about the plausibility of various causal structures into the learning process. The disadvantage is that the number of possible causal structures rapidly becomes intractable, so that in practice various kinds of

greedy search heuristics have to be used. The constraint-based methods have the disadvantage that they are less able to constrain search by using prior knowledge, but have the complementary advantage that they do not require such knowledge. They can also be shown to be asymptotically correct –given enough data these algorithms will always yield the right answer about which graphs are compatible with that data.

Causal Bayes net representations and learning algorithms allow learners to accurately predict patterns of evidence from causal structure and to accurately learn causal structure from patterns of evidence. They constitute a kind of inductive causal logic. It is possible to prove that only certain patterns of evidence will follow from particular causal structures, given the Markov, Intervention and Faithfulness assumptions, just as only certain conclusions follow from particular logical premises, given the axioms of logic.

Causal Bayes nets are also analogous to the representations and algorithms that allow the visual system to accurately infer spatial structure from retinal patterns—the kinds of representations and deductions captured in "ideal observer" theories in vision (Gopnik et al. 2004). The visual system implicitly assumes that there is a world of three-dimensional moving objects and then makes assumptions about how those objects lead to particular

patterns on the retina. By making the further assumption that the retinal patterns were, in fact, produced by the objects in this way, the system can work backwards and infer the structure of objects from those patterns (see eg. Palmer, 1999). Causal Bayes net inferences involve similar assumptions and allow similar deductions. Just as the visual system assumes that the patterns at the retina were produced by three dimensional objects in a particular way and then uses those assumptions to infer the objects from the retinal patterns, causal Bayes net assume that causal structure produced patterns of evidence and use those assumptions to learn the structure from the evidence.

Causal Bayes nets, then, provide a way of formally specifying accurate inductive causal inferences just as logic provides a way of formally specifying accurate deductive inferences and "ideal observer" theories in vision provide a way of formally specifying accurate visual inferences.

Causal Bayes nets as a model of children's causal learning: Empirical results.

This leads to a further question. Do human beings use these kinds of ideal rational computations to learn about the world? In vision there is extensive evidence that human visual systems are close to ideal observers, and

although people sometimes make logical errors, there is also extensive evidence that they use logic to draw everyday conclusions. Recently, several investigators have suggested that adults' causal knowledge might involve implicit forms of Bayes nets representations (Gopnik & Glymour 2002; Rehder & Hastie, 2001; Steyvers et al, 2003; Waldmann; 2001). In particular, it turns out that Patricia Cheng's causal power model of human causal learning (Cheng, 1997) is a special case of Bayes net learning, with a particular constrained set of graph structures and a particular parameterization (Glymour, 2001; Glymour & Cheng 1999). However, Steyvers et al (2003) have shown that adults can learn more complex structures, which go well beyond the causal power model, particularly, if they are allowed to perform interventions.

Of course, adults have extensive experience and often explicit tuition in causal inference. If young children could use versions of Bayes nets assumptions and computations they would have a powerful tool for making causal inferences. They might, at least in principle, use such methods to uncover the kind of causal structure involved in everyday intuitive theories.

In recent work we have been exploring this possibility by seeing how young children use conditional probability and intervention to make

judgments about causation (Gopnik, 2000; Gopnik & Sobel, 2000; Gopnik & Schulz, 2004; Gopnik et al, 2001; 2004; Schulz & Gopnik, 2004; Sobel et al. 2004). Our basic technique is to present children with novel causal relationships, relationships they have never experienced or heard about before. Then we present children with carefully controlled information about conditional probabilities and interventions and see what causal conclusions they draw.

Two and a half year olds can discriminate conditional independence and dependence even with controls for frequency, and can use that information to make judgments about causation, at least when the causal relations are generative, deterministic and non-interactive. In these experiments we showed children various combinations of objects placed on a new machine, "the blicket detector". The blicket detector is a square box which lights up and plays music when some combinations or objects, but not others are placed on it. The children were told that "blickets make the machine go" and were asked to identify which objects were blickets.

For example, in Gopnik et al, 2001, children saw the sequence of events depicted in Figure 2a, and the control sequence depicted in Figure 2b. In Figure 2a the effect E (the detector lighting up) is correlated with both A

and B. However, E is independent in probability of B conditional on A, but E remains dependent on A conditional on B. In Figure 2b each block activates the detector the same number of times as in Figure 2a but the conditional independence patterns are the same for A and B. Children consistently choose A rather than B as the blicket, in the first condition, and choose equally between the two blocks in the second condition. Assuming that the causal relations are deterministic, generative and non-interactive, a Bayes net account would generate a similar prediction.

Insert Figure 2 about here

The experiments with two year olds, however, only required that children discriminate between conditional probabilities of 1 and <1. Moreover, the paradigms could also be explained by the use of a causal form of the Rescorla-Wagner associative learning procedure. In fact, you could think of the Rescorla-Wagner rule precisely as a procedure that approximates the conditional independencies among events in the world that indicate causal structure in certain very simple cases.

In similar experiments, however, four-year-old children also used principles of Bayesian inference to combine prior probability information with information about the conditional probability of events. Moreover, they did this in a backwards blocking paradigm that is not easily explained by the Rescorla Wagner rule (Sobel et al. 2004). For example, suppose children see the sequence of events in Figures 2c and 2d. On a Bayes net account, the causal structure of 2c is plain, A does not cause the effect and B does, and the children also say this. However, the causal structure of 2d is ambiguous, it could be that A and B both make the detector go, but it is also possible that only A does. Note that, in contrast, on a Rescorla Wagner account the associations between B and the detector should be the same in both these cases – the independent association or lack of association of A and the detector should have no influence. However, the backward blocking result is predicted by Cheng's model. (Note that this is a case of "explaining away", knowing that the effect was caused by A makes it less likely that it was also caused by B.)

Indeed, children say that B is a blicket much less often in the 2d condition than the 2c condition. However, we can increase the prior probability of the "A only" structure by telling the children beforehand that

almost none of the blocks are blickets. Children who are told that blickets are rare are more likely to choose the "A only" structure – that is to say that A is a blicket but B is not. This is not predicted by either Rescorla-Wagner or Cheng models.

Four-year-olds can also perform even more complex kinds of reasoning about conditional dependencies, and they do so in many domains, biological and psychological as well as physical. In one experiment children were shown a monkey puppet and various combinations of flowers in a vase (see Figure 3) (Schulz & Gopnik, 2004). They were told that some flowers made the monkey sneeze and others didn't. Then they were shown the following sequence of events: Flowers A and B together made monkey sneeze. Flowers A and C together made monkey sneeze. Flowers B and C together did not make monkey sneeze. Children correctly concluded that A would make the monkey sneeze by itself, but B and C would not. In a control condition they saw each flower make the monkey sneeze with the same frequency, and they chose between the flowers at chance. This result can be explained by Rescorla-Wagner as well as Bayes net models with suitable assumptions, but cannot be explained by Cheng's causal power model.

Insert Fig. 3 about here

Learning from Interventions

Conditional probability is one basic type of evidence for causation. The other basic type of evidence involves understanding interventions and their consequences. Look again at the intervention assumption.

The Intervention Assumption: A variable I is an intervention on a variable X in a causal graph if and only if 1) I is exogenous (that is, is not caused by any other variables in the graph) 2) directly fixes the value of X to x and 3) does not affect the values of any other variables in the graph except through its influence on X.

The technical definition of intervention in this assumption may look formidable but it actually maps well onto our everyday intuitions about intentional goal-directed human actions. We assume that such actions are the result of our freely willed mental intentions, and so unaffected by the

variables they act on (Clause 1). Clause 2 is basic to understanding goal-directed action. When actions are genuinely goal-directed we can tell whether our actions are effective: that is whether they determine the state of the variables we act upon, and we modify the actions if they are not. Clause 3 is essential to understanding means-ends relations. When we act on means to gain an end we assume that our actions influenced other variables (our ends) through, and only through, the influence on the acted-upon variable (the means).

Moreover, we assume that these features of our own interventions are shared by the interventions of others. This is an important assumption because it greatly increases our opportunities for learning about causal structure – we learn not only from our own actions but also from the actions of others.

There is evidence that this conception of intervention, as freely willed goal-directed actions that may be performed by oneself or others, is in place in children at least by the time they are 18 months old (Meltzoff, 1996). In addition, children can clearly learn from interventions in simple cases. For example, they can learn which effects directly follow from their actions in trial and error learning.

Learning from Combinations of Conditional Probabilities and Interventions

The crucial aspect of causal Bayes nets, however, is that intervention and conditional probability information can be coherently combined and inferences can go in both directions. Animals have at least some forms of the ability to infer conditional probabilities, and even conditional independencies, among events -- as in the phenomenon of blocking in classical conditioning (Rescorla & Wagner 1972; Shanks, 1985: Shanks & Dickinson; 1987). They also have at least some ability to infer causal relations between their interventions and the events that follow them, as in operant conditioning and trial- and error learning. However, there is, at best, only very limited and fragile evidence of non-human animals' ability to combine these two types of learning in a genuinely causal way. Why is it that when Pavlov's dogs associate the bell with food, they don't just spontaneously ring the bell when they are hungry? The animals seem able to associate the bell ringing with food, and if they are given an opportunity to act on the bell and that action leads to food, they can replicate that action. Moreover, there may be some transfer from operant to classical conditioning. However, there is no evidence that animals can go directly from learning novel conditional

independencies to designing a correct novel intervention. Moreover, surprisingly primates show only a very limited and fragile ability to learn by directly imitating the interventions of others, an ability that is robustly present in one-year-old humans (Povinelli, 2000: Tomasello & Call, 1997).

In contrast, we've shown that very young children solve causal problems in a way that suggests just this coordination of observation and action. Preschool children, for instance, can use contingencies, including patterns of conditional independence, to design novel interventions to solve causal problems. Three-year-olds in the blicket detector experiments use information about conditional independence to produce appropriate interventions (such as taking a particular object off the detector to make it turn off) that they have never seen or produced before (Gopnik et al., 2001).

Even more dramatically, four-year-olds used patterns of conditional dependence to craft new interventions that required them to cross domain boundaries, and overturn earlier knowledge (Schulz & Gopnik, 2004). For example, children were asked beforehand whether you could make a machine light up by flicking a switch or by saying "Machine, please go". All of the children said that flicking the switch would work but talking to the machine would not. Then children saw that the effect was unconditionally dependent

on saying “Machine, please go”, but was independent of the switch conditional on the spoken request. When children were then asked to make the machine stop 75% said “Machine, please stop”. Moreover, these children were more likely to predict that a new machine could be activated by talking to it than a control group of children.

Note that this cross-domain result is particularly difficult to explain using the apparatus of either associationist or nativist theories. Children clearly had built up very strong associations and had extensive experience with the within-domain causal relations. On nativist accounts one of the most characteristic aspects of core knowledge is that inferential principles are restricted to particular domains. Yet children overrode these associations and principles after only a few presentations of the relevant causal evidence – evidence that definitively pointed to cross-domain causal relations.

Most crucially, however, four-year-olds can also combine patterns of conditional dependence and intervention to infer causal structure and do so in a way that recognizes the special character of intervention (Gopnik et al., 2004). This kind of inference is naturally done by Bayes nets and is not a feature of either associationist or causal power accounts of causal learning. Children can even do this when the relations between the events are

probabilistic rather than deterministic. And children can use such combinations of information to infer the existence of unobserved variables.

For example, we showed four-year-olds a novel "puppet machine" in which two stylized puppets moved simultaneously. They were told that some puppets almost always, but not always, made others go. In one condition they saw the experimenter intervene to move puppet X, and puppet Y also moved simultaneously on five of six trials. On one trial the experimenter moved X and Y did not move. In the other condition they simply observed the puppets move together simultaneously five times, while on one trial the experimenter intervened to move X and Y did not move. The covariation between X and Y was the same in both cases. However, children accurately concluded that X made Y move in the first case, while Y made X move in the second. Again these results would not be predicted either on an associationist or causal power account, but follow directly from Bayes net assumptions.

Similarly, children could use the pattern of interventions and covariation to normatively infer an unobserved variable that was a common cause of two observed variables. Children were again shown the puppet machine but now they saw the two objects move together several times, and then saw the experimenter intervene to move the puppets. In one condition,

like the condition described above, the experimenter moved X and Y did not move. When asked to explain why the puppets moved together children said they did so because Y moved X. In the unobserved condition the experimenter intervened on both puppets. When she intervened on X, Y did not move, when she intervened on Y, X did not move. When children were asked to explain why the objects moved together they said that they did so because of some hidden factor. Moreover, adults referred to a hidden structure even when the relations between the puppets were probabilistic, and they discriminated among different hidden structures (Kushnir et al. 2003). We have some preliminary evidence that children may behave similarly.

In even more recent work we have shown that four-year-old children could also use a pattern of interventions and covariation to normatively infer more complex structures, in particular, to distinguish a causal chain from a common effects structure from a causal conjunction (Schulz & Gopnik, 2003). We showed children another new machine – the gear-toy. The gear-toy consisted of two gears with a switch on the side. When you flicked the switch the gears moved together. This simple machine could involve (at least) four different causal relations between the gears, the switch could make A go which could make B go, the switch could make B go which could make A go,

the switch could independently make A go and make B go, or the switch and A together could make B go and vice-versa. We showed children different patterns of intervention on the gears and the switch, and pictures representing the different causal relationships (basically cuter versions of causal graphs with smiling gears pushing each other instead of letters and arrows). Children appropriately picked the right picture from the evidence, and, given the pictures, they could correctly infer what pattern of evidence would result.

Our empirical work so far has looked at a relatively limited range of causal inferences. Our causal structures involved at most three variables and the variables were simple discrete two-valued variables. Children only had to discriminate between conditional probabilities of 1 vs < 1. The causal relations were generative and for the most part (with the exception of the gear toy conjunctions) they were not interactive.

In other experiments we have shown that four-year-old children will infer inhibitory relations between two variables, and they will discriminate more finely among different degrees of probabilistic strength between two variables. In particular children say that an a object that sets off the blicket detector two out of three times has stronger causal powers than one that sets

off the detector only one out of three times (Kushnir & Gopnik, in press). We have not yet shown whether children can infer more complex causal structure when the causal relations have these sorts of parameterizations.

Causal Bayes nets do allow such inferences, and allow inferences about much more complex structures involving multiple variables with multiple parameterizations. Moreover, they allow inferences from pure covariation as well as from interventions. Most of our experiments involved a combination of observation and intervention, rather than just observations per se. Children's causal learning mechanisms almost certainly involve a more restricted subset of the general causal Bayes nets methods, though we don't yet know the limit of that subset.

For example, it may be that children assume that causal relations are fundamentally deterministic with a single "error term", an assumption that requires interventions to infer complex structure. They may be more closely analogous to, say, experimental chemists who can infer causal structure from a single experiment, than to epidemiologists who infer causal structure from big statistical databases.

Moreover, the youngest children in our experiments were 2 1/2 and most of the children were three or four. We don't know how far back these learning capacities go, and whether, for example, they are present in infancy.

Nevertheless the learning capacities we have demonstrated in children extend well beyond those predicted by any other theory of causal learning, In our experiments we carefully controlled the apparatus so that were no spatio-temporal or mechanical cues that discriminated causal structure. And, as we note above, although individual experiments might be explained by associationist or causal power theories many of the experiments (the prior knowledge backwards blocking experiment and the puppet machine and gear toy experiments, in particular) are not easily explicable by either of these theories.

Causal Bayes nets and connectionism.

How are causal Bayes nets related to other theories of children's learning, particularly connectionist or neural net theories?

Neural nets and Bayes nets are very similar as formal mathematical structures, but they have been developed and are typically applied to different purposes in conjunction with quite different algorithms. (see Table 1)

Formally, Bayes nets are akin to a kind of feed-forward neural net, but there are differences as well. Both Bayes nets and neural nets have the same graphical structure, they consist of nodes (units) that are connected by arrows (connections). The characteristic Markov property (or the extended version of that property called d-separation) is shared by both Bayes nets and neural nets. However, not all Bayes nets assign weights to individual node-to-node connections (arrows), and, in Bayes nets , unlike neural nets, different weights may be assigned for different values of the parent nodes. Some recurrent neural nets that satisfy the d-separation property (Pearl, 1988) are likewise cyclic Bayes nets, and vice-versa (Spirtes, et al., 2000).

The most important differences involve the semantics of the models, differences that also influence the accompanying algorithms. Nodes in Bayes nets almost always have an external reference; representation is not distributed. Instead the individual nodes in Bayes nets represent individual variables in the external world, and the individual connections represent individual causal links among those variables. Even the hidden nodes that represent unobserved variables in Bayes nets denote external properties, features or relations that could be observed in principle.

This contrasts with typical neural net models. In these models input and output variables refer to entities outside the network, but individual hidden nodes and connection weights do not. A group of hidden nodes, their connections and weights, may indirectly represent some aspect of the external world, but these representations are distributed. Nodes and arrows (that is, neural units and connections) do not directly map to individual variables and causal relations in the world, as they do in causal Bayes nets.

Also unlike typical neural net models, Bayes nets do not distinguish input from output nodes—they are all on the same footing. This enables them to flexibly account for inferences in many directions, including inferences from outputs to inputs as well as the other way around. In this way Bayes net algorithms can not only generate predictions, they can also use the same apparatus to generate counterfactuals and explanations.

There are several different kinds of algorithms for inference and learning associated with Bayes nets. One set of algorithms presuppose a fixed, parameterized Bayes net. The algorithms can be used to update the probability of all nodes given values of any subset of nodes, and thus to predict the values of some nodes from the values of others. Feed-forward neural nets typically

provide a more restricted form of such updating, confined to predicting output from input values.

A second set of algorithms updates probabilities for all nodes given external interventions. As noted above interventions in causal Bayes Nets force variables to take a particular value or force a probability distribution on one or more nodes. The same set of algorithms can be used to generate certain kinds of counterfactual claims. At least as currently formulated, neural net models do not distinguish interventions and counterfactuals from predictions in this way.

Similarly, unlike neural net models of the kind developed so far, the links in causal Bayes nets have an intervention interpretation. Assume we know the network and the joint probability distribution in the objective system the network models. Then we can compute the probability distribution that will result from an external manipulation of a variable represented in the network – we can compute what will happen if we intervene on that system in a particular way. In fact, this is precisely what makes causal Bayes nets *causal*. Neural net models do not interpret the connections between nodes in this way.

In a nutshell: in a neural net psychological model, the hidden nodes, linkages and weights are about psychology or, were the model biologically serious, about specific nerve cells or complexes and their connections. But they are not directly about the world outside the organism – they do not individually represent variables in the world or causal relations between them. The functional input/output relations and the "distributed representations" are about the world and about psychology. In a causal Bayes net, everything is about the world (right or wrong). As a psychological model, Bayes nets represent what the subject thinks about the world.

Given this basic semantic difference there are several ways in which Bayes nets modeling and classical neural net modeling might be considered as complementary rather than competitive. Perhaps the most obvious is that, as we are applying them here at least, the two formalisms involve different levels of description. We think of Bayes nets as what Marr (1982) called a "computational" level of description, in the same way, that in vision science the geometric relations between two-dimensional projections and three-dimensional objects involve a computational levels of representation. In

principle, these sorts of computations might be implemented at a lower level by a neural net.

In our work Bayes net models are intended to describe what and how children think as they learn about causal relations and use that knowledge. We have no doubt that ultimately children do this by means of neural synaptic connections, though we have little idea how the brain performs this or other high-level computations. What is needed is a kind of neural compiler—in the computer science sense—which would show how these representations and algorithms could be carried out by specific actual nerve cells or complexes through their synaptic connections. That kind of problem seems fundamental for a unified account of brain and mind, and will depend on interdisciplinary work by computer scientists, neuroscientists and psychologists (see eg. O'Reilly & Munakata, 2000).

Part of the appeal of neural net models is the sense that the "ultimate" theory must be such a model--but such a theory will need to be accompanied by a compilation theory for higher order cognitive functions. One way of viewing contemporary neural net models of psychological functions is that they are surmises about features of such a compiler. So understood, explaining how the causal inferences revealed in our experiments and

captured by the causal Bayes net formalism could be implemented in the brain represents a challenge to neural net modeling.

Very recently, Rogers and McLelland (Rogers & McLelland, 2004) have taken up this challenge to model several features of our data with particular types of neural net simulations. One thing that children (and adults) do in causal inference is to make predictions about novel patterns of contingencies among events, based on their experience of past contingencies. In our first studies, for example, children predict that the yellow but not the blue block will now make the detector go. Rogers and McLelland show that these predictions can be modelled by a sequential neural net, and they suggest ways in which such a net could be derived from a particular set of learning experiences. One might interpret these simulations as a hypothetical implementation of at least some aspects of Bayes net computations.

On the other hand, these models do not, as yet, capture several other important features of Bayes net inferences and of our data. Our experiments show that children distinguish passive associations between variables from associations produced by external interventions on one or more variables, and in simple cases use that different information in a way that accords with causal Bayes net principles. They show that children will, from such data,

correctly infer the existence of unobserved common causes acting on two variables. They show that children will, from learned causal relations, correctly (both objectively correctly and correctly according to causal Bayes net representations) infer the effects of interventions they have never before seen. Further, they do all of these things from a quite small collection of samples.

It would be a step forward if neural net theorists were able to further extend models of the kind described in Rogers and McLelland smoothly to accommodate these phenomena. Such models could be thought of in both Bayes net and neural net terms and the explanatory power of such models could come from both sources.

A second way in which the two types of theories may be complementary concerns the developmental trajectory and origins of Bayes net learning mechanisms. As we noted above the youngest subjects in our studies so far were 2 1/2 years old. We suggest that the learning mechanisms we have discovered are responsible, at least in part, for the sort of developmental changes in intuitive theories that we see in the preschool years. But this leaves open the question of where the learning mechanisms themselves come from.

It is possible that the learning mechanisms we see in these children, with their characteristic links between contingency, causation, intervention

and explanation, are in place innately. On the other hand it is also certainly possible that these mechanisms are themselves derived from an earlier history of experience. A neural net that could be extended to implement the full set of Bayes net predictions would presumably be derived from some particular set of experiences – some learning history. Again, Rogers and McLelland suggest some ways that a net might be constructed from a certain learning history that could simulate some, though not all, of the causal inferences a Bayes net system would make and our children do make. The further empirical question becomes whether children, in fact, experience that kind of learning history.

A third way that the two systems may be complementary is simply that they model different types of knowledge and learning. The kinds of causal learning that are well-modeled by causal Bayes nets are powerful and general but they are still specific to the domain of causality. There are many other kinds of learning, and there may be many kinds of learning mechanisms, just as the algorithms involved in vision are very different from those in audition.

Consider, for example, the contrast between the baseball knowledge of Billy Beane, the general manager of the Oakland A's and that of Barry Bonds, the star hitter of the rival San Francisco Giants. Even Noam Chomsky and

Steven Pinker would agree that baseball knowledge is not likely to be the result of a dedicated innate module, so it must have been learned. Both Billy and Barry have complex motor representations that allow them to take the input of a ball pitched in a certain location with a certain velocity, and translate that information into muscle output that will have a particular effect on the ball. Both also have basic causal knowledge of how baseball works.

Barry, over literally hundreds of thousands of trials, has honed and perfected that motor knowledge to the point where he can hit (or decide not to hit) virtually any kind of pitch that is thrown at him. The computations that allow him to make these decisions in a split second must be of staggering complexity and subtlety. Billy, in contrast, never made it to the majors.

Billy, however, understands the full complex causal structure that relates the various skills and actions of a roster of forty or so players to a particular outcome variable – winning pennants. Billy's causal knowledge allows him to predict the outcomes of particular decisions and interventions (encouraging walks, avoiding steals, hiring college rather than high-school players, trading players as they become stars) on winning. It also allows him to provide explanations of those outcomes (over a long season on-base percentage causes the maximum number of wins), and to consider the

outcomes of counterfactuals (if only Jeremy Giambi had slid they would have won the playoff against the Yankees, if they had retained Miguel Tejada they would have done no better this year).

Barry can modulate his motor output to fit his perceptual input better than anyone who has ever lived. But when it comes to explanations and counterfactuals, about the best he can come up with is the classic baseball restatement of phenomenology "I was seeing the ball well" or "I'd have hit better if I'd been seeing the ball well". Similarly, while Barry can adjust his swing to changing circumstances he is unlikely to know beforehand what the consequences of interventions on that swing are likely to be. He may be able to experience a change of stance, say, and adjust to its effects, but he won't be able to say beforehand what that effect will be. In so far as Billy's representation of the causal structure of baseball is indeed accurate, he will be able to plan effective interventions a priori. His most important decisions, in fact, are made before each season starts.

Barry, aside from what must be staggering natural neural connectivity, has learned his skill the Carnegie Hall way, practice, practice, practice, by taking on feedback from thousands of trials. Note that unlike Barry, however, Billy has only had a few trials to predict the outcome, and he can't afford to

make adjustments only after the play-off feedback has been provided. Instead he infers causal structure by collecting massive amounts of statistical data and considering the results of experiments (such as the failed five closer rotation in Boston).

Similarly, there is no matter of fact, about whether Barry Bonds neural representations are wrong or right, Barry's sensori-motor knowledge doesn't seem accurately described as being true or false and it "represents" the causal structure of baseball only in a very indirect way. Billy's causal claims about baseball do have that character (and many have been hotly debated).

It should be obvious by now that Barry's skills seem well suited to a process of connectionist modeling, while Billy's seem more suited to a causal Bayes net representation. This is consistent with the applications of connectionism, and dynamic modeling more generally. These ideas have been most effectively applied to the learning of such classification skills as handwriting recognition or indeed to skilled motor learning in adults or in children.

This contrasts with learning the kind of causal knowledge that underwrites intuitive theories. That kind of learning allows explanation, intervention and counterfactuals as well as prediction. It can be learned in a

few trials rather than hundreds or thousands. It seems better modeled by the new formalisms. At least when they are constructing intuitive theories children seem more like Billy than Barry.

Until recently, it has seemed that connectionist learning was the only computational game in town – the alternative was some form of innate triggering, a faintly mystical faith in constructivism, or a vague analogy to scientific induction. Causal Bayes nets and related learning mechanisms may supplement rather than replacing connectionist ones. But they still make it a whole new ball game.

Acknowledgments

The research reported in this paper was supported by National Science Foundation Grant DLS0132487 and a Hewlett Foundation Grant to the Center for Advanced Studies in the Behavioral Science to the first author. This chapter was written during AG's fellowship year at the Center and she thanks the staff and administration of that wonderful institution. The paper also benefited greatly from discussion with Jay McLelland, Thomas Shultz, Denis Mareschal Michael Thomas, Thomas Richardson and John Campbell and comments from two anonymous reviewers.

Figure Titles

Figure 1: A causal graph

Figure 2: Screening-off and backwards blocking

Figure 3: Biological screening-off

Figure 1: A causal Bayes net

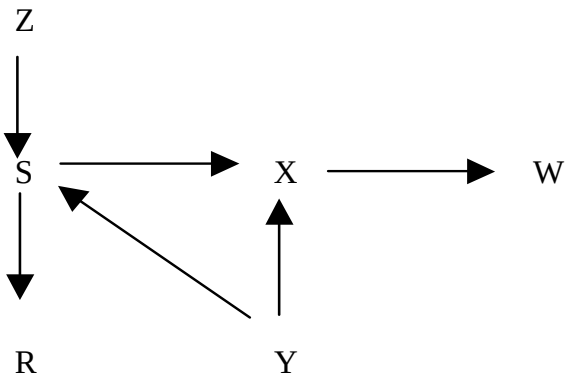
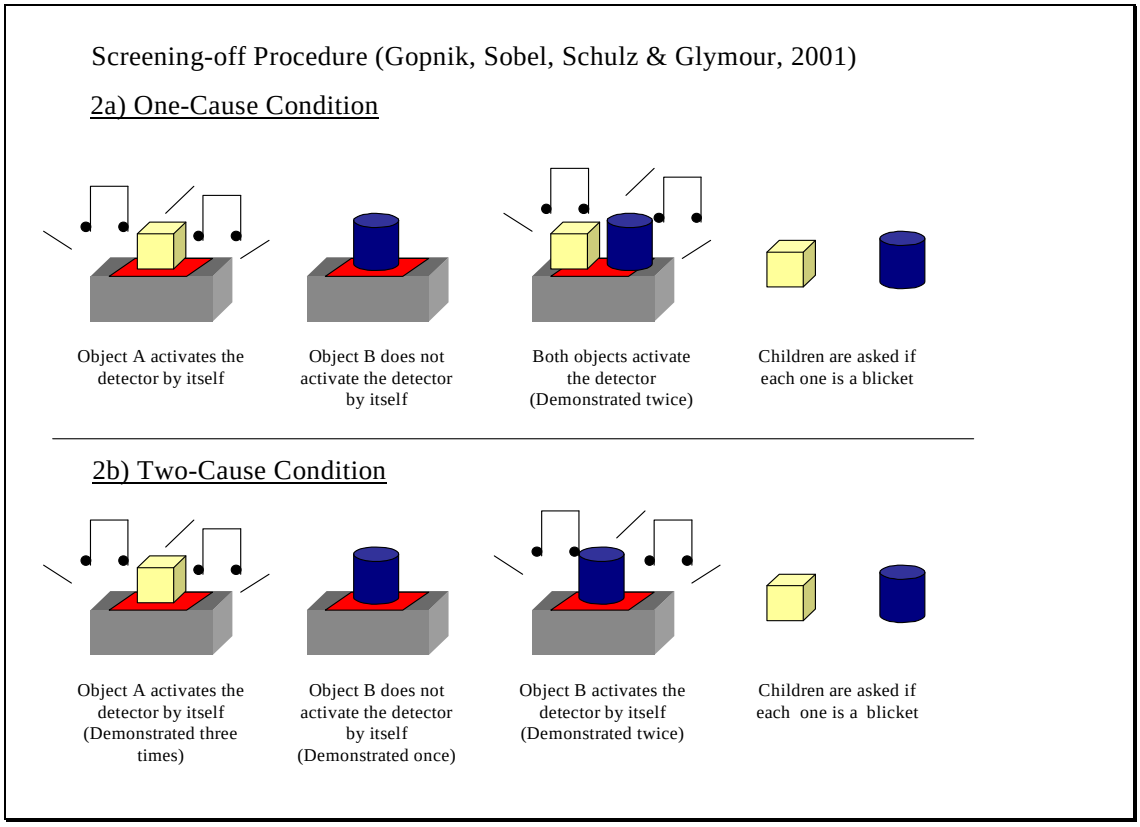
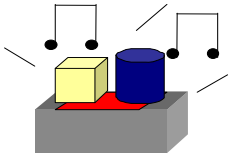


Figure 2: Screening-off and backwards blocking

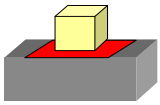


Backwards Blocking Procedure (Sobel, Tenenbaum, & Gopnik, 2004)

2c) Inference



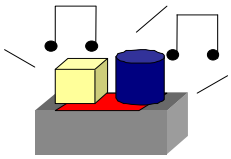
Both objects activate the detector



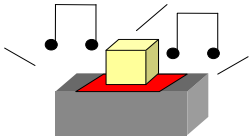
Object A does not activate the detector by itself

Children are asked if each is a blicket they are asked to the machine go

2d) Backward Blocking



Both objects activate the detector

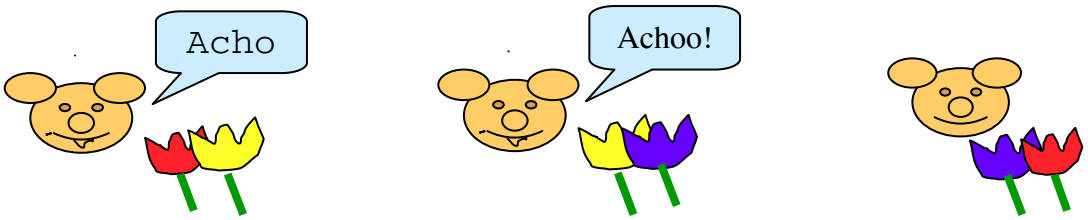


Object Aactivates the detector by itself

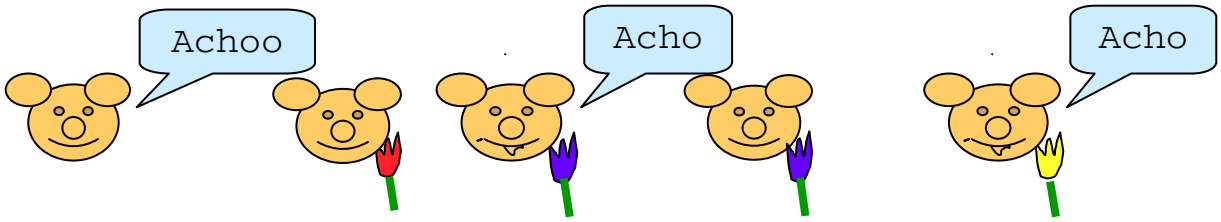
Children are asked if each is a blicket they are asked to the machine go

Figure 3: Screening-off in a biological task. (Schulz & Gopnik, 2004)

Test Condition: Children see that the red and yellow flowers together make Monkey sneeze and that the blue and yellow flowers together make Monkey sneeze, but that the red and blue flowers together do not make Monkey sneeze.



Control Condition: Children see identical frequency information but each flower is presented singly; the red and blue flower each make Monkey sneeze half time; the yellow flower makes Monkey sneeze all the time.



In each condition, children are asked which flower makes the Monkey sneeze. Children choose the yellow flower in the test condition but choose at chance in the frequency control

Table 1

Feature	**Neural Network**	**Bayes Network**
Graph	Directed Graph	Directed Graph
Vertices	Random Variables	Random Variables
Vertex function	Of Parent Variables in the Graph	Of Parent Variables in the Graph
Constraints on the Joint Probability Distribution on the Vertices	None	D-separation
Cyclic Graphs Allowed	Yes	Yes
Distinguished Input and Output Variables	Yes	No
Associated Theory of Interventions	No	Yes
Graph Structure Estimation by	Heuristic Pruning	Asymptotically Correct Search (Bayesian or Constraint Based)

Parameter Estimation by	Variety of iterative methods, e.g., back-propagation	Maximum likelihood or Bayesian posteriors
Interpretation of Vertices	None, or as representations of "internal" Objects (e.g., nerve cells)	Representations of Variable Properties of "external" Systems
Usual Data Application	Classification	Causal and Statistical Modeling

References

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367-405.

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

Flavell, J. H (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*, 21-45.

Gelman, S. A, & Raman, L. (2002). Folk biology as a window onto cognitive development. *Human Development*, Vol 45(1), 61-68.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Glymour, C. & Cheng, P. (1999). Causal mechanism and probability: a normative approach. In K. Oaksford & N. Chater (eds.) *Rational models of cognition.* Oxford: Oxford University Press.

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

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

Gopnik, A. & Glymour, C. (2002). Causal maps and Bayes nets: A cognitive and computational account of theory-formation. In P. Carruthers, S. Stich, M. Siegal, (Eds.) *The cognitive basis of science.* Cambridge: Cambridge University Press.

Gopnik, A., Glymour. C., Sobel, D., Schulz, L. , Kushnir, T. & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 1, 1-31.

Gopnik & Meltzoff, A.N. (1997). *Words, thoughts and theories.* Cambridge, MA: MIT Press.

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

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about causal properties in categorization and induction. *Child Development, 71,* 1205-1222.

Gopnik, A. & L. Schulz (2004). Mechanisms of theory-formation in young children. *Trends in Cognitive Science, 8.* 8.

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

Heckerman, D., Meek, C. and Cooper, G. (1999). A Bayesian approach to causal discovery. In C. Glymour and G. Cooper (Eds*). Computation, Causation, and Discovery,* pp. 143-67. Cambridge, MA: MIT Press.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology, 5,* 668–683.

Inagaki, K., & Hatano, G. (2004). Vitalistic causality in young children's naive biology. *Trends in Cognitive Sciences.* Vol 8(8), 356-362

Kushnir, T. & Gopnik, A. (in press). Young children infer causal strength from probabilities and interventions. Psychological Science

Kushnir, T. Gopnik. A., Schulz, L, & Danks, D. (2003). Inferring hidden causes. In R. Alterman & D. Kirsch (eds*). Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society, Cognitive Science Society*: Boston MA.

Marr, D. (1982*). Vision: a computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended

acts by 18-month-old children. *Developmental Psychology, 31*, 838-850

O' Reilly, R. & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience.* Cambridge, MA: MIT Press.

Palmer, S. (1999). *Vision science: From photons to phenomenology.* Cambridge, MA: MIT Press.

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

Povinelli, D. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works.* New York: Oxford University Press.

Ramsey, J., Roush, T., Gazis, P., & Glymour, C. (2002). Automated remote sensing with near-infra-red reflectance spectra: Carbonate recognition. *Data Mining and Knowledge Discovery, 6*, 277–293.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 3*, 323–360.

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

Richardson, T., & Spirtes, P. (2003). Causal inference via ancestral graph models. In Green, P. Hjort N., & Richardson, S. (eds.) *Highly structured stochastic systems.* Oxford, Oxford University Press.

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

Ross, N., Medin, D., Coley, J. D., Atran, S. (2003) Cultural and experimental differences in the development of folkbiological induction. *Cognitive Development 18* (1), 25-4

Ruffman, T.; Perner, J.; Naito, M.; Parkin, L.; & Clements, W. A. (1998). Older (but Not Younger) Siblings Facilitate False Belief. *Developmental Psychology. 34 (1),* 161-174.

Rumelhart, D. & McLelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Schulz, L. & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40, 2,* 162–176.

Schulz, L. & Gopnik, A. (2003). Causal learning in children. Paper presented at the Society for Research in Child Development, Tampa, Florida, April.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 21* (pp. 229-261). San Diego, CA, US: Academic Press, Inc.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology, 37*(1), 1-21.

Shipley, B. (2000). *Cause and correlation in biology.* Oxford, England: Oxford University Press.

Shultz, T.R (2003). *Computational developmental psychology.* Cambridge, MA: MIT Press.

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

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*, 2967–2988.

Slaughter, V., Jaakkola, R., & Carey, S. (1999). Constructing a coherent theory: Children's biological understanding of life and death. In M. Siegal & C. Peterson (Eds.), *Children's understanding of biology and health* (pp. 71–96). Cambridge MA: Cambridge University Press.

Slaughter, V., Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology* 46 (1) 1-30

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

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology, 22,* 37-58.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science. 28,3.*

Spirtes, P., Glymour, C., & Scheines, R. (1993*). Causation, prediction, and search (Springer Lecture Notes in Statistics).* New York: Springer-Verlag.

Steyvers, M., Tenenbaum, J.B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*,1.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* The MIT Press, Cambridge, MA.

Tomasello, M. & Call, J. (1997). *Primate cognition.* London : Oxford University Press.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 1,* 27–58.

Wellman, H. M., Hickling, A. K., & Schult, C. A. (1997). Young children's psychological, physical, and biological explanations. In *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena. New directions for child*

development, No. 75 (Wellman, H. M. and Inagaki, K., eds), pp. 7-25, San Francisco, CA, US: Jossey-Bass/Pfeiffer.

Wellman, H. M. and Inagaki, K., (Eds.) (1997). *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena. New directions for child development, No. 75* San Francisco, CA, US: Jossey-Bass/Pfeiffer.

Woodward, J., (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford Univ. Press.