Modeling Cross-Domain Causal Learning in Preschoolers as Bayesian Inference

Elizabeth Baraff Bonawitz (liz_b@mit.edu)
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Thomas L. Griffiths (tom_griffiths@brown.edu)
Department of Cognitive and Linguistic Sciences
Brown University
Providence, RI 02912 USA

Laura Schulz (lschulz@mit.edu)
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

This study investigates the interaction between preschoolers' initial theories and their ability to learn causal relations from patterns of data. Children observed ambiguous evidence in which sets of two candidate causes co-occurred with an effect (e.g. *A&B* → *E*, *A&C* → *E*, *A&D* → *E*, etc). In one condition, all candidate causes were from the appropriate domain (a biological cause for a biological effect); in another condition, the recurring candidate cause, *A*, crossed domains (a psychological cause for a biological effect). When all causes were domain-appropriate, children were able to use the data to identify *A* as a cause. When the recurring cause crossed domains, children were less likely to endorse *A*. However, preschoolers were significantly more willing to accept cross-domain causes after seeing the evidence than at baseline. A Bayesian model is proposed to explain this interaction.

Very young children have remarkably sophisticated causal knowledge about the world. Children reason about the causes of mental states (e.g., Meltzoff, 1995), physical systems (e.g., Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982), and biological events (e.g., Gelman & Wellman, 1991; Kalish, 1996). Preschoolers can even make predictions about hidden variables and explain events in terms of unobservable causes (Schulz & Sommerville, in press).

Many researchers have suggested that children's causal knowledge can be characterized as intuitive theories: abstract, coherent, defeasible representations of causal structure (Carey, 1985; Gopnik & Meltzoff, 1997; Wellman, 1990; Keil, 1989). However, relatively little is known about the process of causal learning. While some researchers have suggested that children's naive theories might be instantiated in domain-specific modules (Leslie, 1994; Scholl & Leslie, 1999) or innate concepts in core domains (Carey & Spelke, 1994; Keil, 1995), other researchers have emphasized the role of domain-general learning mechanisms (Gopnik et al., 2004; Schulz & Gopnik, 2004). Very little research (though see Sobel, Tenenbaum, & Gopnik, 2004) has looked at how domain-specific beliefs and domain-general learning mechanisms might interact. In this paper, we provide a formal account of this interaction as rational Bayesian inference and then present two experiments in support of this account suggesting that preschoolers can integrate domain-appropriate prior knowledge with domain-general patterns of evidence.

Theories and Evidence

In the literature on causal learning in children, some studies seem to suggest the relative strength of domain-specific knowledge over domain general learning mechanisms while other findings suggest the opposite. Of the few studies that have directly compared domain-specific and domain-general causal learning, some have suggested that both adults and children privilege domain-specific information over domain-general evidence (e.g., Ahn, Kalish, Medin, & Gelman, 1995; Bullock, Gelman & Baillargeon, 1982; Shultz, 1982). Shultz (1982) for instance, suggests that preschoolers base their causal judgments on knowledge about domain-appropriate mechanisms of transmission rather than evidence of temporal covariation. By contrast, other research suggests that children can use domain general learning mechanisms (such as the conditional probability of events) to override domain boundaries (Schulz & Gopnik, 2004). For example, children can use patterns of evidence to determine that a psychological rather than a physical cause produces a physical effect (Schulz & Gopnik, 2004). Though see Andersson (1986) and Boo and Watson (2001) for examples of over-generalizations of domain general causal notions.

It has been difficult to evaluate the interaction between domain-specific knowledge and domain-general learning mechanisms, because previous work has focused on extreme points. For example, in the Shultz (1982) studies, children were asked to make a judgment after a single instance of temporal co-occurrence, thus there was little room for covariation evidence to affect children's naïve theories. By contrast, in the Schulz and Gopnik

(2004) studies, the covariation data unambiguously supported the domain-inappropriate cause so there was little room for children's naïve theories to affect their evaluation of the evidence. Thus, while some research has explored the relative strength of theories and evidence, few studies have demonstrated a graded interaction between the two.

In this paper we look at children's causal judgments in contexts in which we might observe the impact of both naïve theories and patterns of evidence. Specifically, we look at whether children's domain-specific theories affect their interpretation of evidence and whether patterns of evidence affect children's domain-appropriate beliefs. Intuitively, a within-domain cause will always be favored over a cross-domain cause in the absence of evidence to the contrary. However, as evidence accumulates in favor of the unlikely cause, domain-general learning may override domain-specific knowledge and a priori unlikely causes may be favored. First, we will present a rational account of this interaction, which is formalized in a theory-based Bayesian model. Second, we will use this model to predict children's responses to complex patterns of evidence.

Reasoning with Ambiguous Evidence Within and Across Domains

In the current study, we show preschoolers storybooks in which two candidate causes covary with an effect; one cause recurs and the other causes are always novel (i.e., the evidence is in the form $A\&B\rightarrow E$; $A\&C\rightarrow E$; $A\&D\rightarrow E$ … etc.) In the within-domain story, all the causes are domain-appropriate. If children are able to learn from the data, they should infer that 'A' is the cause. However, in the Cross-Domain story, the recurring cause (A) is domain-inappropriate. Thus A is less plausible than the alternative cause given the children's naïve theories but more plausible given the pattern of evidence. By comparing children's judgments before and after seeing the evidence, we can evaluate the degree to which children can overcome the initial biases induced by their causal theories.

Because we wanted to investigate processes that might underlie genuine instances of theory change, we chose a context in which children's theories are both robust and distinct from adult theories. As noted, considerable research has shown that children's causal reasoning respects domain boundaries. In particular, many researchers have suggested that children respect an ontological distinction between mental phenomena and bodily/physical phenomena (Carey, 1985; Estes, Wellman, & Woolley, 1989; Hatano & Inagaki, 1994; Notaro, Gelman, & Zimmerman, 2001; Wellman & Estes, 1986). Although adults accept that some events (e.g., psychosomatic phenomena) can cross the mental/physical divide, preschoolers typically deny that psychosomatic reactions are possible (e.g., they deny that feeling frustrated can cause a headache or that feeling embarrassed can make you blush; Notaro, Gelman & Zimmerman, 2001). We were interested in how preschool children would interpret formal patterns of evidence suggesting the presence of a psychosomatic cause in light of a strong initial belief in domain boundaries.

Theory-based Bayesian Inference

Bayesian inference provides a natural framework in which to consider how prior knowledge and data interact. We propose to model children's causal inferences in a framework with two critical components. First, we assume that children's judgments are the result of a Bayesian inference, comparing a set of hypotheses as to the causal structure that underlies the observed data. Second, we assume that these hypotheses are generated by a causal theory. This Bayesian model captures the two critical components of children's reasoning: their ability to update their beliefs given new evidence, and the soft constraints imposed by their prior knowledge.

To capture children's reasoning on the storybook task, we model their inferences as weighing the probability of one explanation over another. That is, children are explicitly asked in the task, "Why does {character} have {symptom}? Is it because of {Explanation 1} or is it because of {Explanation 2}?" We model the probability that the child chooses Explanation 1 as

$$\frac{P(\text{Explanation 1} \mid D)}{P(\text{Explanation 1} \mid D) + P(\text{Explanation 2} \mid D)} \quad (1)$$

which directly contrasts the two potential explanations given the data, D, observed. The probability of each possible explanation given the data is computed by summing over all causal models that are consistent with the explanation. This is formalized as:

$$P(\text{Explanation 1} \mid D) = \sum_{h \in H} P(\text{Explanation 1} \mid h) P(h \mid D) \quad (2)$$

where h is a hypothesis as to the underlying causal structure, and H is the space of all hypotheses. We represent hypothetical causal structures as causal graphical models (Pearl, 2000; Spirtes, Glymour, & Schienes, 1993), consisting of a graphical structure indicating the causal relationships among a set of variables, where nodes are variables and relationships are indicated by arrows from cause to effect, and a set of conditional probability distributions giving the probability that each variable takes on a particular value given the values of its causes. We assume that the probability of the explanation given a particular causal structure h is $1/k$, where k is the set of candidate causes that are present and possess a causal relationship with the effect in h.

The probability of a particular causal structure given the data is expanded via Bayes rule as

$$P(h \mid D) \propto P(D \mid h) P(h) \quad (3)$$

where $P(h)$ is the prior probability of a particular causal structure, implementing the constraints imposed by the prior knowledge of the learner, and $P(D|h)$ is the "likelihood", indicating the probability of the data D under the causal model h. The precise values of these two probabilities are determined by the causal theory entertained by the observer.

Generating Causal Models from a Causal Theory

An important notion in developmental psychology is the idea that children have rich causal theories of the world. As proposed by Tenenbaum and Niyogi (2003) and Tenenbaum, Griffiths, and Niyogi (in press), we model the theory that guides the inferences made by children in our task as a simple scheme for generating causal graphical models. In this scheme, we allow for several types of domains. These domains can include biological causes, psychological causes, physical effects, biological effects, etc., as illustrated schematically in Figure 1. Causal relationships can only exist between nodes on the top line (causes) and nodes on the bottom line (effects). Causes are likely to have relationships with their domain-related effects, as given by the thick, solid arrows. However, we also allow a small probability that a cause from one domain can lead to an effect in another domain. This assumption is illustrated by the thin arrows connecting elements across domains.

This framework theory provides a simple recipe for generating the space of causal graphical models that could describe a particular situation. The prior probability associated with each model is simply its probability of being generated by the theory. The process of generating a model breaks down into four steps:

1. Represent all possible causes and all possible effects as a set of nodes in a causal graphical model.
2. For each cause and effect in the same domain, generate a causal relationship (an arrow) between the corresponding nodes with probability *p.*
3. For each cause and effect in different domains, generate a causal relationship (an arrow) between the corresponding nodes with probability *q*.
4. Specify the conditional probability distribution for the effects given their causes.

We will now describe these steps in detail.

Causal nodes In our model, the number of nodes are given by the number of different variables observed. In the current study, we only learn about the presence of a single effect over seven days, following the $A\&B\rightarrow E$, $A\&C\rightarrow E$, etc, pattern discussed above. This produces eight candidate causes, so there are 2^8 possible causal models (each candidate cause either does or does not influence the effect).

Causal arrows Causal arrows between nodes are generated according to the theory. As expressed above, if

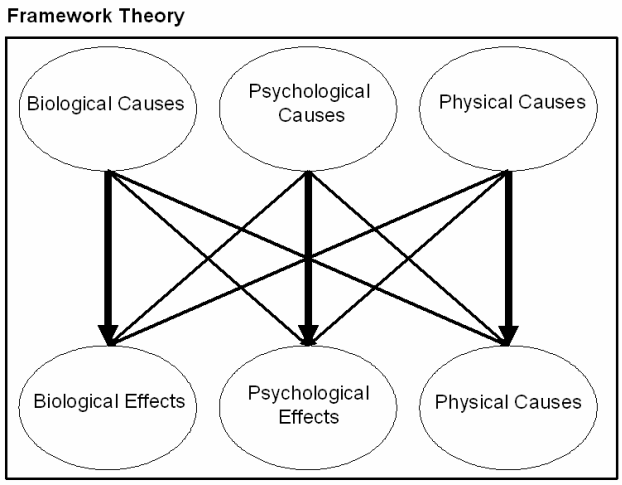

Figure 1: Schematic of framework theory that includes causal connections *within-domain* and *cross-domain*.

a cause and effect are both within the same domain, then the probability a relationship exists is relatively high and given by p. In contrast, if the link between two nodes is cross domain, then a relationship is very unlikely, and is given a lower probability, q. Assuming that each relationship is generated independently, we can evaluate the prior probability of each of the 256 possible models by multiplying the probabilities of the existence or non-existence of the causal relationships involved.

Conditional probability distribution The conditional probability distribution allows us to evaluate the probability of a specific model, h, generating the data observed over m trials, $P(d_m|h)$. These data consist of the values taken on by all variables on that trial – the presence or absence of the causes and effects. We assume that the probability of each cause being present or absent is constant across all of the causal models, and the only difference is in the probability they assign to the occurrence of the effect on that trial. We assume that the conditional probability of the effect given the set of causes is 1 if any cause which influences the effect is present, and ε otherwise (this corresponds to a noisy-OR parameterization where each cause has a strength of 1 and the background has a strength of ε). The probability of the full set of data, D, accumulated over the course of the storybook is given by

$$P(D|h) = \prod_m P(d_m|h) \quad (4)$$

where the data observed on each trial in the story are assumed to be generated independently.

Model Predictions

The predictions of the model given this pattern of evidence are represented in Figure 3. We implemented our intuition of relatively low *cross-domain* probability

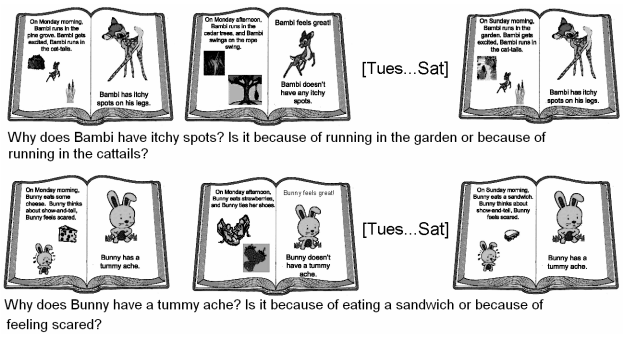

Figure 2: W*ithin* and *cross-domain* storybooks used in Experiment 1.

by setting q = .1 and set a higher *within-domain* probability of p = .4. As described above, we also assumed a small ε = .001. Importantly, the model demonstrates the shift between favoring the within-domain candidate cause at baseline to favoring the cross-domain candidate cause after evidence. We conducted an experiment to test the predictions of this model.

Experiment 1

The goal of experiment 1 was to look at whether or not children would also be able to integrate domain-general learning with their strong domain-specific priors.

Methods and Design

Participants Thirty-two four and five-year-olds (range = 4;0 to 5;11, M = 5;0) participated. Children were randomly assigned to either a Baseline Condition or an Evidence Condition.

Materials Two picture storybooks were used as the stimuli (see Figure 2). Both books featured events occurring over a week, starting on Monday and ending on Sunday so children received 7 'days' of evidence. The Within Domain storybook featured a deer who liked to run in different places. The deer got itchy spots on his legs every morning. Evidence was presented as described above: *A&B→E*; *A&C→E*; *A&D→E*, etc. The recurring candidate cause (A) was running through cattails, the other cause varied (e.g., running through a meadow, a garden, etc.) (To show that the effect was not always present, the deer ran through different places in the afternoons and never got itchy spots). The Cross Domain book was identical except that it featured a bunny rabbit who got a tummy ache in the mornings (but not the afternoons). Feeling scared was the recurring cause; the other candidate cause varied among types of food Bunny ate (e.g., cheese, a sandwich, etc.) Two sets of each book were created to counterbalance the order of events.

Procedure Each child was read both the *within-* and *cross-domains* storybook (order was counterbalanced) in a quiet location. In the Evidence Condition, children were asked at the end of the story, "Why does [Bambi/Bunny] have [itchy spots/tummy ache]? Is it because of [running in the garden/eating a sandwich] or because of [running in the cattails/feeling scared]?" Children in the Baseline Condition saw the same storybooks, only the Monday-Saturday events were not included, and the story went straight to the final, Sunday page.

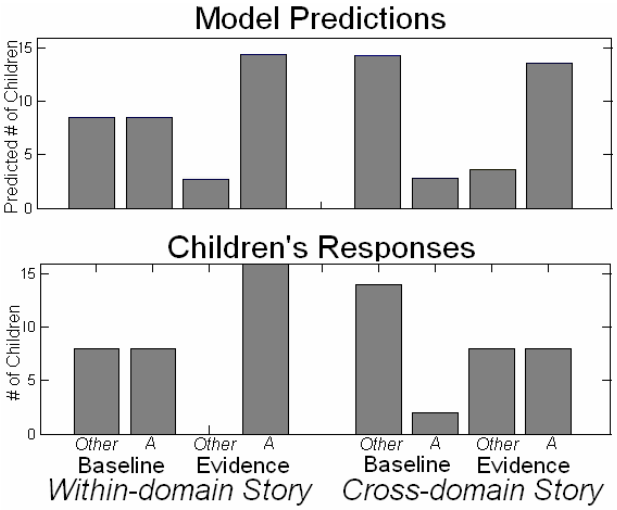

Figure 3: Model predictions and children's responses in Experiment 1.

Results

Preliminary analysis revealed no order effects. In the Baseline Condition, children chose at chance between the candidate causes in the *within-domain* storybook and almost always chose the domain-appropriate variable (food) in the *cross-domains* storybook. Children were significantly more likely to identify A as the cause in the Evidence Condition than at Baseline in both the *within-domain* and *cross-domains* storybooks (*within-domain*: χ^2 (1, N = 32) = 10.67, p < .01 ; *cross-domains*: χ^2 (1, N = 32) = 5.23, p < .05). However, children were less likely to choose A in the *cross-domains* storybook than in the *within-domain* storybook, (χ^2 (1, N = 32) = 10.67, p < .01). See Figure 1 for details.

As shown in Figure 3, our model accurately predicted children's responses. The model gives correct relative weights to the variables at baseline in both the *within-domain* and *cross-domains* conditions. The model also favored the posterior probability of 'cattails' over 'garden'. It was slightly less successful at capturing the degree to which children would choose 'feeling scared' as the cause; the model predicted that the posterior probability of 'feeling scared' as the candidate cause should have been significantly greater than 'sandwich'. Children showed slightly greater resistance to parting with their initial inductive biases. Importantly however, the model captured the overall pattern of children's learning; children were significantly more willing to select 'feeling scared' after seeing the evidence then at baseline.

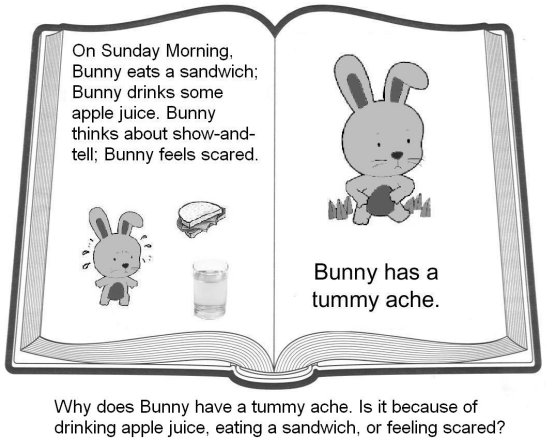

Figure 4: Example page from *cross-domains* storybook used in Experiment 2.

Discussion

As predicted by our Bayesian model, the results of Experiment 1 suggest that domain-specific theories and domain-general learning mechanisms interact. Children were more likely to use the evidence to identify A as a cause when A was consistent with their theories than when A violated their theories. Critically however, children also seemed to learn from the evidence. After seeing the data, preschoolers were able to entertain a causal possibility (that being scared might cause tummy aches) that they did not endorse at baseline.

Although children reading the *cross-domains* storybook identified A as a cause more often after seeing the evidence than at baseline, only 50% of the children chose A as a cause in the Evidence condition. Because children were given a forced choice between two causes, it is unclear whether these children were actually learning from the evidence or if they were merely confused by the cross-domain storybook and guessing at chance.

Experiment 2: Cross Domain Learning with Multiple Variables

To address the concerns raised in the previous experiment, we replicated the *cross-domains* Evidence Condition of Experiment 1 but provided children with three potential candidate causes (two within-domain candidate causes and one cross-domain candidate cause). If children are learning from the evidence, they should be significantly more likely to pick 'feeling scared' than either of the other variables; if children are confused by the evidence, they should pick 'feeling scared' at chance (33% of the time).

Methods and Design

Participants Sixteen 5-year-olds (range = 4;2 to 6;0, M = 4;10) participated.

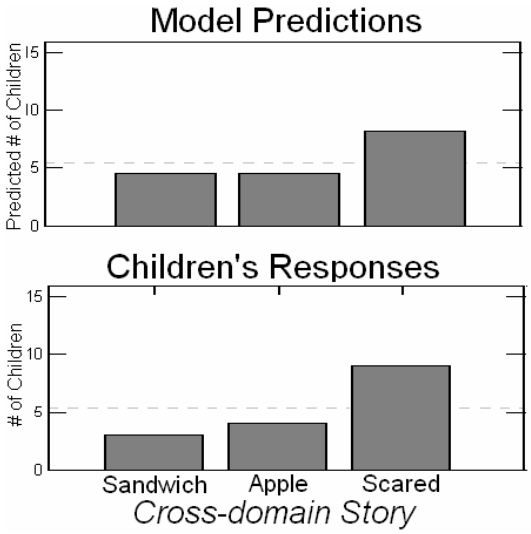

Figure 5: Model predictions and children's responses in Experiment 2. Dashed line represents chance.

Materials A *cross-domains* book similar to the Cross-Domain book in Experiment 1 was created. However, instead of only eating one food per day, Bunny ate one food and drank one liquid and felt scared, (See Figure 4). Evidence about the cause of tummyaches followed the pattern: *ABC* → *E, ADF* → *E, AGH* → *E,* etc.

Procedure The procedure was identical to the procedure in Experiment 1; however, children were only tested in the Evidence Condition on the Cross-Domain book.

Results and Discussion

After seeing the evidence, children were significantly more likely to identify *C* as the cause then at chance, (binomial test, test proportion: 0.33, $p < .05$) indicating that the children were not confused by the data, but rather that they inferred that being scared was a possible cause for Bunny's tummy ache, (see Figure 5). Children did not choose either of the other two variables above chance (binomial test, test proportion: 0.33, p = ns). Using the same parameter values for p and q as in Experiment 1, our model was also able to predict the children's response, (see Figure 5). Importantly, the model predicted a strong posterior probability of the cross-domain cause, but relatively weak posteriors on other two within-domain candidate causes, sandwiches and apple juice. The results from Experiment 2 corroborate the findings in Experiment 1 and suggest that children learn from the evidence and are able to overcome their initial theories.

Discussion

This research demonstrates the important contributions that domain-specific theories make to children's interpretation of evidence, as well as the role that evidence can play in affecting domain-specific beliefs. We have also offered a formal account of children's theory-based learning in terms of Bayesian inference. By providing a formal account, we hope to make clear the

interaction between domain-specific prior knowledge and domain-general learning mechanisms.

In our framework domain-specific knowledge is captured by the priors specified by the framework theory, and domain-general learning is represented in terms of Bayesian inference. The framework theories represent the set of constraints on possible causal relations and Bayesian modeling provides a framework for learning these constraints at multiple levels. From the studies presented here, it is unclear whether children in our experiments underwent theory change (at the framework level), or if children instead simply learned something specific about Bunny's unfortunate condition, without updating their beliefs about psychosomatic illness in general. While the broader question of learning framework theories is beyond the scope of this paper, in principle, theory-based Bayesian inference could capture this more general learning. As children accumulate evidence about instances of psychosomatic illness, the prior for cross-domain causal events in general (i..e., psychological causes generating biological effects) increases. However, future work might look at the extent to which patterns of evidence can effect genuine theory change.

Although the content of children's framework theories and the priors over those theories may differ from adult theories, Bayesian inference suggests a universal system for integrating theories and evidence. Most importantly, this computational account captures a hallmark of children's causal theories: that children's inferences are conservative with respect to their prior knowledge and yet flexible in the face of new evidence.

Acknowledgments

Special thanks to the participating daycares in Cambridge, MA & Portland, OR, and to the Boston Museum of Science, to Wendy Weinerman and Elanna Levine for data collection, to Noah Goodman, Fei Xu, Kate Hooppell, and Anna Jenkins for thoughtful discussion, and to the Singleton Fellowship. This research was additionally supported by a McDonnell Foundation and James H. Ferry Fund grant to L.S.

References

Ahn, W., Kalish, C.W., Medin, D.L., Gelman, S.A., 1995. The role of covariation versus mechanism information in causal attribution. Cognition 54, 299–352.

Andersson, B. (1986). The experiential gestalt of causation: A common core to pupils' preconceptions in science. European Journal of Science Education, 8, 155–171.

Boo, H.K, & Watson, J.R. (2001) Progression in High School Students' (Aged 16--18) Conceptualizations about Chemical Reactions in Solution, *Science Education*, **85,** 568.

Bullock, M., Gelman, R., & Baillargeon, R. (1982) The development of causal reasoning. In W.J. Freidman (Ed.), *The developmental psychology of time* (pp.209-254). New York: Academic Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press/Bradford Books.

Estes, D., Wellman, H.M., & Woolley, J.D. (1989) Children's understanding of mental phenomena. In H. Reese (Ed.), *Advances in child development and behavior*. New York: Academic Press.

Gelman, S.A., & Wellman, H.M. (1991) Insides and Essence: Early Understandings of the Nonobvious. *Cognition, 38*(3), 213-244.

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

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

Hatano, G., & Inagaki, K. (1994) Young children's naïve theory of biology. *Cognition*, *50*, 171-188.

Kalish, C. (1996) Causes and symptoms in preschoolers' conceptions of illness. *Child Development*, *67*(4), 1647-1670.

Keil, F.C. (1989) *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Keil, F.C. (1995). The growth of causal understandings of natural kinds. In D. Sperber & D. Premack (Eds.), *Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation.* (pp. 234-267). New York: Clarendon Press/Oxford University Press.

Leslie, A.M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* (pp. 119-148). New York: Cambridge University Press.

Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838-850.

Notaro, P.C., Gelman, S., & Zimmerman, M.A. (2001) Children's Understanding of Psychgenic Bodily Reactions, *Child Development 72(2),* 444-459.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press.

Scholl, B.J., and Leslie, A.M. (1999) Modularity, development and "theory of mind." *Mind Lang.* 14:131–153.

Schulz, L.E., & Gopnik, A. (2004) Causal Learning Across Domains, *Developmental Pscyhology, 40(2),* 162-176.

Schulz, L.E., Sommerville, J. (in press). God does not play dice: Causal determinism and children's inferences about unobserved causes. *Child Development.*

Shultz, T.R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1), 1-51.

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

Spelke, E.S, Breinlinger, K., Macomber, J., & Jacobson, K. (1992) Origins of Knowledge. *Psychological Review, 99*(4), 605-632.

Spirtes, P., Glymour, C., and Scheines, R. (1993) *Causation, Prediction, and Search*, Springer-Verlag, New York.

Tenenbaum, J.B., Griffiths, T. L., and Niyogi, S. (in press). Intuitive theories as grammars for causal inferences. To appear in Gopnik, A., & Schulz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford University Press.

Tenenbaum, J.B. and Niyogi, S. (2003). Learning Causal Laws. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.*

Wellman, H. (1990). *The Child's theory of mind.* Cambridge, MA: MIT Press.