Title

Training a Bayesian: Three-and-a-half-year-olds' Reasoning about Ambiguous Evidence

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN

1069-7977

Authors

Bonawitz, Elizabeth Baraff
Fischer, Adina
Schulz, Laura

Publication Date

2008

Peer reviewed

Training a Bayesian: Three-and-a-half-year-olds' Reasoning about Ambiguous Evidence

Elizabeth Baraff Bonawitz, Adina Fischer, & Laura E. Schulz
{liz_b, adinafb, lschulz}@mit.edu
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

Previous work has demonstrated the importance of both naïve theories and statistical evidence to children's causal reasoning. In particular, four-year-olds can use statistical evidence to update their beliefs. However, the story is more complex for three-year-olds. Although three-and-a-half-year-olds perform as well as four-year-olds when statistical evidence is theory-neutral, several studies suggest that they do not learn from statistical evidence when a statistically likely cause is inconsistent with their prior beliefs (e.g., Schulz et al., 2007). There are at least two possible explanations for younger children's failure to use statistical data to update their beliefs: one (the Information Processing account) suggests that younger children have a fragile ability to reason about statistical evidence; the other (a Prior Knowledge account) suggests that in some domains, younger children have stronger prior beliefs and thus require more evidence before belief revision is rational. To distinguish these accounts, we conducted a two-week training study with three-and-a-half-year-olds. Children participated in an *Information Processing Training* condition, a *Prior Belief Training* condition, or a *Control* condition. Relative to the *Control* condition, children in the *Prior Belief Training* condition, but not children in the *Information Processing Training* condition showed an overall improvement in their ability to reason about theory-violating evidence. This suggests that at least some developmental differences in statistical reasoning tasks may be due to younger children's stronger prior beliefs.

Keywords: Causal learning; Ambiguous evidence; Training study.

Theories and Evidence in Preschoolers Causal Reasoning?

One of the hallmarks of a good scientist is her ability to revise her beliefs in the face of counter evidence. However, belief revision should be at once flexible (to permit learning) and conservative (to prevent misleading data from overturning strongly held beliefs). In principle, small amounts of data (e.g., observing the cookies missing from the cookie jar) should suffice to overturn weakly held beliefs (that my partner is dieting) but should leave strong beliefs (in conservation of matter) intact.

Integrating new evidence with prior knowledge may be important in cognitive development as well. Indeed, many researchers have suggested that children's beliefs have some important structural and dynamic similarities to scientific theories. In particular, children seem to have abstract, coherent causal representations of the world that support prediction, explanation, and counterfactual reasoning and – critically -- that are defeasible in the face of counter-evidence (Gopnik, 1996; Carey & Spelke, 1996).

One approach to thinking about how prior beliefs should interact with statistical evidence can be obtained by regarding causal learning as a problem of Bayesian inference. In Bayesian inference, the learner seeks to evaluate a hypothesis about the process that produced some observed data. The learner's a priori beliefs about the plausibility of the hypotheses are expressed in a "prior" probability distribution. The learner seeks to evaluate the "posterior" probability of the hypothesis – their beliefs about the plausibility of the hypothesis after taking into account the evidence provided by the data. The posterior distribution directly combines the evidence obtained, through the likelihood, with the learner's initial beliefs about the plausibility of the hypothesis, expressed in the prior. In the case of causal learning, we can imagine prior probabilities being supplied by a domain-specific theory, stipulating which causal structures are plausible (Tenenbaum, Griffiths, & Niyogi, in press; Tenenbaum & Niyogi, 2003). Thus, Bayesian inference is a rational framework for learning and provides a formal account of how domain-specific theories and domain-general patterns of evidence might interact to affect children's beliefs.

However, the integration of strong prior beliefs with new evidence may not be so natural for young learners -- or even for adults (Chen & Klahr, 1999; Inhelder & Piaget, 1958; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Masnick & Klahr, 2003). For instance, some research suggests that adults interpret identical evidence differently depending on whether the data supports or conflicts with a favored theory. Thus, if two candidate causes are both independent of an effect, learners will cite instances of co-occurrence as evidence for the variable consistent with their theories and instances of non-co-occurrence as evidence against the variable incommensurate with their theories (Kuhn, 1989).

Of the research that does demonstrate causal learning in children, some suggests that there are uneven contributions from either evidence or prior beliefs. The considerable research suggesting that four and five-year-olds are capable of using statistical data to support belief revision has led many researchers to argue that bottom-up, general statistical learning (e.g., sensitivity to the covariation of events) plays a critical role in children's causal inferences (Gopnik et al., 2004; Keil, 1995; Schulz & Gopnik, 2004). However, other researchers have argued for the importance of domain specific prior beliefs (e.g. Carey & Spelke, 1996; Keil, 1995) and indeed, many studies suggest that both adults and children privilege domain-specific information over domain-general evidence, (Ahn, Kalish, Medin, & Gelman, 1995; Bullock, Gelman & Baillargeon, 1982; Shultz, 1982).

It has perhaps been difficult to evaluate the interaction between domain-specific knowledge and domain-general learning mechanisms because much previous work has focused on unambiguous instances of each. For example, studies arguing for the importance of domain-specific theories have generally involved contexts in which participants have very strong theories and minimal counterevidence; thus, there has been little room for the evidence to play a role in the interpretation of the events. In contrast, studies arguing for the importance of domain-general mechanisms have often provided overwhelming evidence, giving little room for theories to play a role in the interpretation of evidence. Thus, while some research has explored the relative strength of theories and evidence, few studies have demonstrated a graded interaction between the two.

Can being scared give you a tummy ache?

To investigate this interaction, we looked at children's causal judgments in contexts in which we might observe the impact of both naïve theories and patterns of evidence. We provided children with suggestive, but still ambiguous evidence in cases where they did, and cases where they did not, have strong prior beliefs (Bonawitz et al, 2006; Schulz et al, 2007). Children were read two books in which two candidate causes co-occurred with an effect. Evidence was presented in the form: AB → E, AC → E, AD → E, etc. After receiving this statistical evidence, children were asked to identify the cause of the effect on a new trial. While it is possible that B, C, and D were each causes of the effect, it is more probable that A was the single cause. In one book (the *Within Domain* book), all the causes and the effect were from a single domain; in the other condition (the *Cross Domains* book) cause A was from a different domain (creating a conflict between the statistical evidence and children's prior knowledge).

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. Considerable research has shown that children's causal reasoning respects domain boundaries (Carey, 1985; Estes, Wellman, & Woolley, 1989; Hatano & Inagaki, 1994; Wellman & Estes, 1986; Bloom, 2004). In particular, many researchers have suggested that children respect an ontological distinction between mental phenomena and bodily/physical phenomena, (Notaro, Gelman & Zimmerman, 2001). While 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). 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.

Both three-and-a-half- and four-year-old children were read both the *Within* and *Cross Domains* books. Consistent with the predictions of a Bayesian model, four- and five-year-olds inferred that cause A was the relevant cause from both the *Within* and the *Cross Domains* evidence, were more likely to identify A as the cause in the *Within Domain* book than the *Cross Domains* book, and were able to transfer their new expectations about psychosomatic causality to a novel task. However, although three-and-a-half-year-olds readily identified cause A as the target cause in the *Within Domain* book, the younger children failed to learn from the statistical evidence when the evidence violated their prior beliefs; that is, they did not learn at all in the *Cross Domains* book.

What leads to differences in causal learning between younger and older preschoolers?

Why did the three-and-a-half-year-olds respond differently to the *Cross Domains* evidence than the four- and five-year-olds? One explanation consistent with previous research is that three-year-old children might have difficulty making inferences from ambiguous statistical data. If the ability of the three-and-a-half-year-olds to interpret data of this complexity is fragile, any increase in task difficulty (including a conflict with prior knowledge) might compromise children's ability to evaluate the evidence. Research on young children's causal reasoning has suggested that information processing demands can limit 3-year-olds performance even on tasks that only require reasoning within one domain of knowledge, (eg. See Leslie, 1994). A context in which three-year-olds must integrate knowledge across multiple domains may be even more cognitively taxing. Because this account assumes that three-and-a-half-year-olds' failure is due to information processing limitations, we will call this the Information Processing account.

Critically however, the differential treatment of evidence by younger and older preschoolers may not be irrational from a Bayesian perspective: rather, it is possible that the younger children might have a stronger prior belief in domain boundaries than the older children. Notoro et al (2001) demonstrated that with age children begin to report psychogenic phenomenon as both physical and psychologically caused, supporting the claim that a belief in this domain boundary may lessen with age and experience. Thus, although a small amount of evidence might be needed to convince an older child of psychosomatic causality (given that older children have other experiences supporting such claims), more evidence may be needed to overturn younger children's resistance to believing in psychosomatic illness. Thus the three-and-a-half-year-olds might have failed to learn from the statistical data in the previous study, not because of information processing limitations but because the data presented might have been insufficient to overcome three-year-olds' initial inductive bias that psychological causes are unlikely to generate bodily effects. We will call this the Prior Belief account.

Psychosomatic training study

In order to evaluate these accounts, we designed a two-week training study for three-and-a-half-year-olds. In one condition (*Information Processing Training*) children were repeatedly exposed to ambiguous evidence (AB→E; AC→E, AD→ E, etc.); however, these books always presented children with within-domains evidence and the evidence never dealt with psychosomatic causality. If the Information Processing account is correct and children fail the belief revision task simply because of a fragile ability to reason from ambiguous evidence, then multiple exposures to this form of evidence could help children use the ambiguous data to revise their beliefs in a final test book about psychosomatic events.

In the other condition (*Prior Belief Training*), children were not given additional exposure to ambiguous evidence, but they were taught that psychosomatic events were relatively common. If the Prior Belief account is correct and three-and-a-half-year-olds failure to update their beliefs with evidence is due, not to their fragile ability to interpret evidence, but instead to strong prior beliefs in a psychological/biological domain boundary, then presenting children with evidence that psychogenic events are common should improve children's performance on both a final test book and a transfer task.

Of course it is possible that three-and-a-half-year-olds have *both* a fragile ability to interpret ambiguous evidence *and* a strong prior belief in domain boundaries. In that case, both training studies should impact the children's performance. An additional group of children were tested in a *Control* training condition in which psychological variables were repeatedly mentioned but were not causally connected to biological outcomes.

In all three conditions, children were first tested on a *Cross Domains* test book (identical to the book used in Schulz et al., 2007). Children were included in the study only if they failed the initial *Cross Domains* book (that is, they rejected the statistically likely domain-violating candidate cause in favor of the statistically unlikely within-domain cause). All children then received 5 training books over a two-week period. At the end of the training, children's performance was assessed on three measures: 1) a final *Cross Domains* test book (formally identical to the initial book but involving different specific stimuli) 2) a Free Explanation book (requiring a free explanation rather than a forced choice response) and 3) a Transfer book (assessing children's willingness to generalize from one psychosomatic event to others).

Methods and Design

Participants Forty eight children (mean age: 45 months; range: 39-48 months) were recruited from Boston area preschools. Fifty four percent of the participants were girls and a range of ethnicities resembling the diversity of the population was represented. Children were randomly assigned to an *Information Processing Training* condition, a *Prior Belief Training* condition, or a *Control* condition. The experimenter met with the children for four 20-minute sessions over a period of two weeks.

Materials Two Cross Domains test books were created, as well as a Free Explanation test book, and a set of Transfer cards. Additionally, five training books for each condition (*Prior Belief, Information Processing, Control)* were created for a total of 15 training books.

Test books: (1) *Cross Domains Books:* These stories presented ambiguous but statistically compelling evidence supporting a psychological cause of an illness. In both books, the repeating psychological cause was always paired with the character eating different types of food (the biological candidate causes). At the end of the book, children were asked a forced choice question: "What is the cause of (Bunny, Beaver's) (Belly Ache, Tummy Hurting)? Is it because of (feeling worried, feeling scared) or because of eating (cornbread, sandwich)?" The order of events (psychological or food) was counter-balanced across books. (2) *Free Explanation Book:* This story book presented children with an open explanation task, reading: "This is Puppy. Puppy is nervous because it's his first day of school. {Experimenter turns page.} Oh, oh! Puppy's stomach hurts! Why do you think puppy's stomach hurts?" (3) *Transfer Cards:* Six picture cards were created showing a physically possible event (throwing a ball in a lake and making a splash); a physically impossible event (brushing a window with a feather and breaking it); a biologically possible event (skipping rope and getting tired); a biologically impossible event (stomping on the ground and making a tomato grow); and two psychosomatic events (worrying and getting a headache; being nervous and feeling sick). Children were shown each of the six pictures in one of two fixed semi-random orders and asked yes or no questions about the possibility of the event. For example, for the physically impossible event, children were asked: "Can that happen? Can Tony break the window with a feather?" The two psychogenic questions were: "Can that happen? Can Leslie get a headache from worrying too much?" and "Can that happen? Can Jordan start to feel sick from being nervous and upset?"

Information Processing Training Books: Ambiguous evidence was presented in each of the 5 storybooks used in this condition in the format of AB → E; AC → E; …; AG → E. In these books, events were always within-domain and non-psychological so that children would not also be getting

Day 4 Test Results % of children's correct responses				
	Cross Domains	Free Explanation	Transfer Cards	Overall Averages
Control	**25***	13	21	17
Information Processing	**27***	**57**[1]	25	33
Prior Belief	**50****	**40**[2]	**50**[1]	**45**[3]

* $p < .05$; ** $p < .01$ as compared to day 1 book
[1] $p < .05$; [2] $p = .08$; [3] $p < .01$ as compared to *Control* Condition

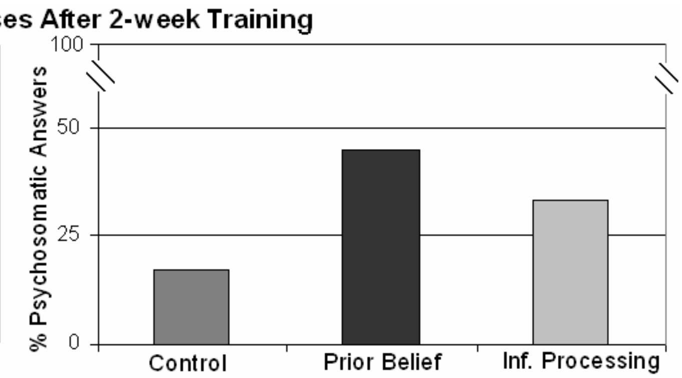

Figure 1: Table to the left shows the percent psychosomatic answers given on the Cross Domains test book, the Free Explanation test book, the transfer cards, as well as the overall average number of psychosomatic responses given on the final test day. The figure on the right reflects the average percent correct psychosomatic responses given by children in each of the three training conditions.

evidence that might affect their beliefs about psychosomatic illness. At the end of each story, children were given a forced choice between two causal variables, for example, "Why does Bambi have itchy spots? Is it because of running in the cattails or running in the garden?" If children choose the statistically unlikely variable, they were corrected by being show that the statistically likely variable was always present with the effect and was therefore a better explanation.

Prior Belief Training Books: Five different storybooks presented evidence that psychosomatic reactions happen fairly often. Books showed ten characters in a classroom. All the characters experienced a similar emotion (e.g. boredom waiting for a hamster to do a trick). Eight of the ten characters were then shown to have a physical reaction (e.g. Sue gets sleepy; Charles gets sleepy; Josh does not get sleepy). At the end of the book children were given a forced choice question asking whether the physical reaction to the psychological emotion happened to very many or very few characters in the story (e.g., "Can you help remind me: did very many students get sleepy or did very few students get sleepy?") If children chose incorrectly, the experimenter went back through the book with the child pointing out how many characters got sleepy and corrected the child.

Control Books: Control books were created to both impose a time delay between the initial and final test sessions, and to make sure that the repetition of mental events in the *Prior Belief Training* would not explain children's choice of the psychological causes in the test books. These books told a story about a character who had a recurring psychological state (i.e., happy, angry, or sleepy). To remain consistent with the level of engagement children had in the other training conditions, children were asked a question at the end of each story that was irrelevant to psychosomatic causal reasoning but that verified that the children were engaged with the story. No corrective feedback was given for children's responses to these questions.

Procedure Children were read storybooks in a quiet room at their daycare. Children were randomly assigned to each condition. Participants were first tested with one of the two Cross Domains test books, particular books counterbalanced between children (the other book was then used for the test book at the end of training). Only children who chose the non-psychological (statistically unlikely) cause were included in the training and were then immediately read book 1 for whatever condition the child was assigned. Children were then visited at their daycare three more times over the course of two weeks, with no two consecutive training days. On each of the second and third visits, children were read the two books appropriate to their training condition (Books 2 & 3 on Day 2; and Books 4 & 5 on Day 3). On the final day (Day 4) the second cross domains test story book, the free explanation book, and the far transfer cards were read.

Results

One child was dropped from the *Information Processing Training* condition for failure to give an interpretable answer on the initial test book. There was no other attrition during the training, leaving 16 children in the *Prior Belief* and *Control* condition, and 15 children in the *Information Processing* condition. There were no age differences between groups and no difference in the numbers of males and females in each condition. There were no main effects of age.

There were no differences across conditions for children's correct answers on the training books. Children's responses to questions in the *Information Processing, Prior Belief,* and *Control* training books were not significantly different from chance on Day 1 ($p = NS$, by binomial test). However, by Days 2 and 3, children were significantly more likely to choose the correct answer in all conditions{(*Control*, Day 2, $p < .05$; Day 3, $p < .01$); (*Information Processing*, Day 2, $p < .05$; Day 3, $p < .01$); (*Prior Belief*, Day 2, $p < .05$; Day 3, $p < .01$) by binomial test}. These results suggest that the training books in all conditions were difficult enough initially, and that children were genuinely learning from the training.

Overall Results We first analyzed how often children gave the correct, psychosomatic response across all three final measures: the Cross Domain Test book, the Free Explanation Book, and the Transfer Test cards. Overall, children in the *Prior Belief Training* condition were significantly more likely to correctly give psychosomatic responses across the board (M = 44.8%) than were children in the *Control* condition (M = 17.2%), (t(32) = 3.06, p < .01), (See Figure 1). By contrast, children in the *Information Processing* condition (M = 33.3%) were not more likely to endorse psychosomatic causality than children in the *Control* condition (t(32) = 1.60, p = *NS*). There were no significant differences between the *Prior Belief* and *Information Processing* conditions (t(32) = 1.02, p = *NS*). These results suggest that at least the *Prior Belief Training* significantly improved children's ability to recognize the relevant causal variable in the final test tasks. We next analyzed the effect of the training on each test task separately.

Cross Domains Test Books Responses of children were coded as either appealing to the recurring psychosomatic cause or to the alternative cause (i.e., the particular food) for the test book in each condition. Compared to their responses on the original test book, children were significantly more likely to appeal to psychosomatic causes in all conditions (*Control*: $\chi 2$ (1, N = 32) = 4.57, p < .05; *Information Processing*: $\chi 2$ (1, N = 30) = 6.00, p < .05; *Prior Belief*: $\chi 2$ (1, N = 32) = 10.67, p < .01), (See Figure 1). There were no significant differences between conditions; however, in terms of raw numbers, children in the *Prior Belief Training* condition were most likely to appeal to being worried or scared on the last days book (50% of the children, compared to only 25% of children in the *Control* Condition and 27% in the *Information Processing Training* condition).

Free Explanation Test Book Children's responses were coded as either appealing to the target psychological cause in the story (e.g. feeling nervous; thinking about school) or to external biological causes not mentioned in the story (e.g., eating too much food, "bumping his belly", or even, as one child notably put it, "has too much poo poo in him"). Children's responses fell uniquely and unambiguously into one of these two categories. Two children (one in the *Information Processing Training* and one in the *Prior Belief* Training) refused to provide an explanation and were removed from analyses for just the Explanation Test books, resulting in a total of 16 children in the *Control* condition, 15 children in the *Prior Belief* condition, and 14 children in the *Information Processing* condition compared for this task. Because children had no previous exposure to these books, success on the training was measured by comparing the number of children's psychological explanations in the two training conditions to the *Control* condition.

Training seemed to significantly help children in the *Information Processing Training* condition and marginally help children in the *Prior Belief Training*, with more children appealing to psychological explanations in these conditions than in the *Control* condition (*Information Processing*: $\chi 2$ (1, N = 30) = 7.63, p < .01; *Prior Belief*: $\chi 2$ (1, N = 32) = 3.06, p = .08), (See Figure 1). There were no significant differences between the two training conditions ($\chi 2$ (1, N = 29) = 1.29, p < *NS*).

Given that in the explanation book, the information processing demands were minimal, it may seem odd that the *Information Processing Training* improved children's responses on this. We believe that exposure to the recurring variables in the *Information Processing Training* books might have encouraged children to attend to elements within the story (as opposed to other relevant information), and thus supported children's success on this test task.

Far Transfer Test Cards Children had some difficultly with this task, with several children in each condition showing a bias, responding all 'yes' or all 'no' to all six questions (3 children in the *Control* condition; 3 children in the *Prior Belief* condition; and 4 children in the *Information Processing* condition). For the analyses of the Transfer cards, these children were removed, resulting in a total of 13 children in the *Control* condition, 13 in the *Prior Belief* condition, and 11 in the *Information Processing* condition. Children in the *Prior Belief Training* were significantly more likely to endorse the possibility of psychosomatic causality than were children in the *Control* condition (*Prior Belief*: $\chi 2$ (1, N = 32) = 4.28, p < .05); however, children in the *Information Processing* condition (*Information Processing*: $\chi 2$ (1, N = 30) = .08, p < *NS*) were no more likely to endorse psychosomatic causality than were children in the *Control* condition, (see Figure 1).

Discussion

After only four twenty-minute training sessions over two weeks, three-year-olds showed an improvement in their ability to recognize causal variables that conflicted with their prior knowledge. This improvement was manifest across three quite different tasks and these findings were strongest for the *Prior Belief Training*, in which children showed improvement relative to the *Control* condition on all three measures. Perhaps the most striking finding is that three-year-olds, after receiving *Prior Belief Training*, were able to extend their inferences about psychosomatic illness to a transfer task. Children in the *Prior Belief Training* were twice as likely as children in the *Control* condition to endorse the possibility of psychosomatic illness, suggesting that children had genuinely revised their beliefs from the limited training.

Children who received the *Information Processing Training* showed more improvement than children in the

Control condition only on the Free Explanation task. As noted, such improvement might be due largely to the children's increased attention to elements within the story as opposed to relevant external factors. However, children in the *Information Processing* condition did not show improvement relative to children in the *Control* condition on the other two measures. This suggests that improving children's ability to process ambiguous evidence has a relatively limited influence on children's ability to learn from counterintuitive statistical evidence. Given that there were no significant differences between the *Prior Belief* and *Information Processing Training* conditions, this conclusion can only be drawn with great caution. Nonetheless, this study provides suggestive evidence that developmental changes in children's difficulty in learning from counterintuitive evidence may be due primarily, not to younger children's greater difficulties in processing statistical information, but to their initially stronger beliefs in domain boundaries.

What did the *Prior Belief Training* do for the children? From a Bayesian perspective, it seems likely that the training changed children's perception of the prior probability of psychosomatic events, thus making the posterior probability of the psychosomatic hypothesis stronger. This raises the question of how children's estimates of the probability of events are generated in the first place. One possibility is that children might believe domain-violating events are rare because they have trouble imagining a mechanism by which psychological events could influence biological events. Further research might explore the extent to which training children on causal mechanisms improves their ability to reason about causal variables previously perceived as theory-violating.

Overall, given that all the three-year-olds in this experiment initially failed the cross-domains task, the success of the children following such a short training is promising. In particular, it suggests that even very young children may be rational learners, conservatively but flexibly integrating new evidence with prior beliefs.

Acknowledgments

Thanks to Irene Headen, Michael Obilade, Noah Goodman, Wendy Weinerman, Darlene Ferranti, and members of the Early Childhood Cognition lab, and participating daycares. Research supported from the 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.

Bloom, P. (2004). Descartes' Baby. New York: Basic Books

Bonawitz, E.B., Griffiths, T.L., & Schulz, L. (2006) Modeling Cross-Domain Causal Learning in Preschoolers as Bayesian Inference. *Proceedings of the 28th Annual conference of the Cognitive Science Society.* Vancouver, Canada.

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.

Carey, S., & Spelke, E. S. (1996). Science and core knowledge. *Philosophy of Science*, 63(4), 515-533.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. Child Development, 70(5), 1098-1120.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.

Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, 63, 4, 485-514.

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.

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

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. P. D. Sperber (Ed.), *Causal cognition: A multidisciplinary debate.* New York, NY: Clarendon Press/Oxford University Press.

Kuhn, D., Amsel, E., & O'Laughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674-689.

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.

Masnick, A. M., & Klahr, D. (2003). Error Matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, 4(1), 67-98.

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

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

Schulz, L., Bonawitz, E.B., & Giffiths, T. (2007) Can being scared make your tummy ache? Naïve theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology.*

Tenenbaum, J.B., Griffiths, T. L., and Niyogi, S. (2007). 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 & Estes, 1986; Carey, S., & Spelke, E.S. (1996) Science and core knowledge. *Philosophy of Science, 63*(4), 515-533.