Teaching three-and-a-half-year-olds to reason about ambiguous evidence

Elizabeth Bonawitz\textsuperscript{1}, Adina Fischer\textsuperscript{2}, & Laura Schulz\textsuperscript{3}

In Press, Journal of Cognition and Development

\textsuperscript{1} Department of Psychology, University of California, Berkeley, Berkeley, CA 94720 USA

\textsuperscript{2} Dartmouth Medical School, Hanover, NH 03755 USA

\textsuperscript{3} Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA 02139 USA

\textsuperscript{4} Department of Psychology, Stanford University, Stanford, CA 94305 USA

\textsuperscript{5} Department of Psychology, Harvard University, Cambridge MA 02138 USA
Abstract

Previous research suggests that three-year-olds fail to learn from statistical data when their prior beliefs conflict with the evidence. Are young children’s causal beliefs are entrenched in their folk theories or can young preschoolers rationally update their beliefs with evidence? Motivated by Bayesian accounts of rational inference suggesting that both statistical evidence and children’s prior beliefs should affect learning, we conducted a training study to investigate this question. Children (mean: 45 months) were included in the study if they failed to endorse the statistically more probable cause given ambiguous evidence (evidence in the form AB→E, CA→E, AD→E, etc.) where the recurring cause, A, violated children’s prior beliefs. (A was a psychological cause; the remaining variables were biological.) The children were assigned to a Statistical Reasoning training condition, one of two Prior Belief training conditions (Baserates or Mechanisms), or a Control condition. Relative to the Control condition, children in the test conditions were more likely to endorse the a priori unlikely variable on a free-explanation task. Critically, children in the Statistical Reasoning condition passed the free-explanation task, even though their only information about the belief-violating variable came from the ambiguous evidence. This suggests that teaching children statistical reasoning improves their ability to learn even from data inconsistent with their prior beliefs.

Keywords: Causal learning; Ambiguous evidence; Training study, psychosomatic illness.
The view that children’s causal commitments take the form of naïve theories has been influential in developmental psychology for several decades (Carey, 1985; Harris, German, & Mills, 1996; Hickling & Wellman, 2001; Gopnik, 1988; Gopnik & Meltzoff, 1997; Keil, 1989; Perner, 1991; Sobel, 2004; Wellman, 1990; Wellman, Hickling, & Schult, 1997). This view, the theory theory, maintains that there is a dynamic relationship between children’s naïve theories and evidence such that children’s causal beliefs affect their interpretation of evidence and evidence enables children to revise their beliefs (e.g., Gopnik & Meltzoff, 1997). However, previous studies looking at the relationship between patterns of evidence and children’s folk theories have generated contradictory results, with some studies suggesting that children privilege plausible causal mechanisms over statistical evidence (e.g., Shultz, 1982) and others suggesting that children learn theory-violating causal relations as readily as theory-consistent ones (Kushnir & Gopnik, 2006; Schulz & Gopnik, 2004). To the extent that children’s causal judgments reflect an interaction between their naïve theories and patterns of evidence, such conflicting findings are perhaps not surprising: on any given task, children’s inferences could depend on the strength of their initial beliefs, the strength of the data, and their ability to integrate the two.

Recently, Bayesian analyses have offered a formal account of how learners might integrate statistical evidence with constraints from naïve theories (Tenenbaum, Griffiths, & Kemp, 2006; Kemp, Perfors, & Tenenbaum, 2007). To our knowledge however, only one previous study has directly tested whether, consistent with the predictions of a Bayesian inference model, children’s learning reflects a graded interaction between their domain-specific prior beliefs and the evidence they observe (Schulz, Bonawitz, & Griffiths, 2007). In that study, children were provided with ambiguous data both in contexts where they had strong prior beliefs and ones where they did not. Specifically, preschoolers were read two books in which two candidate causes co-occurred with an effect. Evidence was presented in the form AB→E; CA→E; AD→E, etc. After receiving this evidence, children were asked to identify the cause of the effect on a new trial. While it was possible that B, C, D, etc. were each independent
causes of the effect, it was more probable that the recurring cause, A, was the actual cause. In one book (the Within Domain book), all the causes and the effect were from a single domain; in the other book (the Cross Domains book) the recurring cause, A, came from a different domain (A was a psychological cause of a bodily effect). Thus in the Cross Domains book, there was a conflict between the statistical evidence and children’s prior beliefs.

Consistent with the prescriptions of Bayesian inference, kindergartners and older preschoolers (50-70 months; mean: 60 months) correctly inferred that A was the cause in both cases but were more likely to identify A as the cause in the Within Domain book than the Cross Domains book. Critically however, younger preschoolers (42-48 months; mean: 45 months) did not. Although the three-and-a-half-year-olds readily identified cause A as the target cause in the Within Domain book (indeed, they were indistinguishable from the older children), they failed to learn at all in the Cross Domains book: they consistently chose the within-domain cause.

The dramatic contrast between the three-and-a-half-year-olds’ impressive reasoning about statistical evidence in neutral contexts (near ceiling) and their poor reasoning (near floor) in belief-violating contexts leaves open the possibility that very young children might be able to use the relative probabilities of events to distinguish candidate causes, but that children’s prior beliefs about plausible causal relations restrict the kinds of evidence they are willing to consider in the first place.

Note that the majority of previous work showing that children can use statistical data to revise their prior beliefs has looked at older children (four to six-year-olds) and used unambiguous deterministic evidence (e.g., Legare, Wellman, & Gelman, in press; Schulz & Gopnik, 2004; though see Kushnir & Gopnik, 2006 for other work suggesting that four to six-year-olds can use probabilistic, ambiguous evidence to revise their beliefs). The few studies that have demonstrated belief-revision in younger children (three-year-olds) have used only unambiguous evidence (Kushnir & Gopnik, 2006; Sobel, 2006). Here we look at whether evidence-based teaching is effective in helping young
preschoolers revise their causal beliefs given ambiguous data. Inspired by rational Bayesian inference accounts, we introduce three training conditions intended to independently affect children’s statistical reasoning and their prior beliefs.

In one training condition (the *Statistical Reasoning* condition) we teach children to draw accurate inference when multiple causal relations are possible but one is more probable than the others (i.e., we read children stories in which evidence is presented in the form $AB \rightarrow E; CA \rightarrow E; AD \rightarrow E$, etc. and teach children that $A$ is the most likely cause of the effect, $E$). This condition was motivated in part by the finding (Schulz et al., 2007) that young three-year-olds (3;0 - 3;5) fail to learn from such evidence even in a theory-neutral (Within Domain) condition. Given that even the neutral task was challenging for three-year-olds, three-and-a-half-year-olds might not have been able to handle the increased difficulty posed by a conflict with prior beliefs. If giving three-and-a-half-year-olds practice reasoning from evidence in *theory-neutral* contexts improves their ability to draw accurate inferences from *theory-violating* evidence, this would suggest that even young preschoolers can use formal properties of evidence, not merely to distinguish plausible causes, but also to draw inferences at a more abstract level: inferring the causal status of previously unrecognized causal variables.

As noted however, the Bayesian analysis suggests that children’s causal judgments depend not only on the strength of the evidence but also on their prior beliefs. Thus in the other two training conditions, we intervene on children’s causal beliefs. While adults accept that psychosomatic phenomena can cross domain boundaries, preschoolers typically deny that psychological events can cause physical/bodily events (e.g., that feeling embarrassed can make your face to turn red; Notaro, Gelman, & Zimmerman, 2001). How might evidence-based teaching help children revise their prior beliefs about psychosomatic causality? Arguably, young children have relatively limited exposure to psychosomatic events. If children believe the base rate of psychosomatic causality is low, then they might (rationally) resist accepting a psychological cause as the most probable explanation of a bodily
effect. Thus, one approach to intervening on children’s prior beliefs is to manipulate their perception of the frequency of psychosomatic events. We will call this the Prior Beliefs Base Rate condition.

Another reason children might resist psychosomatic causality is because they do not understand how psychological states affect bodily states. Research suggests that both adults and children are more willing to accept causal relations for which they can imagine plausible mechanisms (e.g. Ahn, Kalish, Medin, & Gelman, 1995; Shultz, 1982). Thus, we might increase children’s acceptance of psychosomatic causality by offering an explanation of how emotional states might cause bodily outcomes. We will call this the Prior Beliefs Mechanism condition.

If children who initially fail to endorse psychosomatic causality succeed after any or all of the three training conditions, this would suggest that well before children begin formal education, rational inductive inference mechanisms support children’s ability, not only to reason about theory-neutral causal relations, but also their ability to reason about a priori unlikely events. Note that children might succeed in some of the training conditions and not others (e.g., intervening on children’s prior beliefs might be helpful but intervening on their statistical reasoning ability might not). However, given that the Bayesian analysis suggests that both factors play a role, we predict that all three manipulations should independently improve children’s performance. By contrast, if three-year-olds’ causal beliefs are relatively entrenched (e.g., because they are constrained by core knowledge in domains like naïve physics and naïve psychology) then we would expect children to be resistant to belief revision and perform no better than children in a control condition.

Training Study

We designed a two-week training study to investigate these accounts. Because we were interested in children’s ability to reason about theory-violating evidence, children were included in the study only if they initially endorsed the plausible within-domain cause rather than the statistically likely but theory-violating cause in a pretest book (identical to the Cross Domains book used in Schulz et al., 2007).
Eighty children were assigned to one of four conditions: a *Statistical Reasoning* training, a *Prior Belief Baserates* training, a *Prior Belief Mechanisms* training, and a *Control* condition. At the final session, children were again read a Cross Domains storybook (formally identical to the initial book but with different specific stimuli). However, given that previous research suggests that preschoolers tend to vary their responses when asked versions of the same question twice (e.g. Memon, Cronin, Eaves, & Bull, 1993; Poole & White, 1991), we believed that this measure might not be a very sensitive index of children’s learning. Thus our dependent measure of interest was a free explanation task, adapted from Schulz et al., 2007 (originally given to four and five-year-olds). The free explanation measure provides a strong test of children’s learning in that children have to transfer their knowledge to a novel task.

In the free explanation task children were told about a puppy dog who was scared about the first day of school and had a tummy ache; children were asked to explain why the puppy had a tummy ache. Researchers in the earlier study found that at baseline, four and five-year-olds ignored the only variable mentioned in the story (being scared) and strikingly, invented their own domain-appropriate explanations instead (e.g., “because he fell on his stomach”; “because he ate too much food”); children who had first been exposed to the evidence in the Cross Domains book adopted the psychosomatic explanation. If the training conditions support three-and-a-half-year-olds’ ability to learn the target causal relation, they should be able to transfer their learning and explain the bodily event using the psychological explanation.

*Methods and Design*

*Participants.*

Eighty children (mean age: 45 months; range: 39-48 months; 54% girls) were recruited from preschools in a metropolitan area. An experimenter met individually with each child for four 20-minute sessions over a period of two weeks. No two sessions were on consecutive days. Most of the children
were white and middle class but a range of ethnicities resembling the diversity of the population was represented.

To ensure that participants entering the training did not already endorse psychosomatic events, children were given an initial Cross Domains test book; those who passed were dropped from the study and replaced. Replicating Schulz et al., 2007, 82% of the three-and-a-half-year-olds tested on the initial Cross Domains storybook failed the task (i.e., chose the theory-consistent rather than the statistically probable cause) and thus met the inclusion criteria for the study. Children were randomly assigned to a Statistical Reasoning Training condition, Prior Belief Baserates Training condition, a Prior Belief Mechanisms Training condition, or a Control condition (20 children per condition). There were no age differences among the four conditions ($F(3, 76) = 1.48, p = ns$).

Materials.

Two Cross Domains books and a Free Explanation test book were used. Additionally, five different training books were used in each of the four conditions (Statistical Reasoning, Prior Belief Baserate, Prior Belief Mechanisms, and Control), for a total of 20 training books. The training books were each approximately 20 pages long and had approximately 9 words per page. (See Figure 1.)

Cross Domains books: One book was used to see if children met the initial inclusion criteria and one was used on the final day, preceding the Free Explanation Test book. The books were identical except for details of the stimuli. In each book, a character (Bunny or Beaver) ate a different food, experienced a recurring psychological cause (feeling worried; feeling scared), and a recurring biological effect (belly ache; tummy hurting) each morning of a seven-day week. Each afternoon, the character ate two different foods and felt fine. At the end of the story, children were asked a forced choice question about the events of that morning: “Why does (Bunny’s, Beaver’s) (belly ache? tummy hurt)? Is it because of (feeling worried, feeling scared) or because of eating (the cornbread, the sandwich)?” The order of events (psychological or food) was counterbalanced throughout.
Training books: Five books, each involving unique characters and candidate relations, were used in each training condition.

Statistical Reasoning Training: In each book, a character experienced a pair of candidate causes (one recurring and one varying each day) and a consistent effect in a format identical to the Cross Domains books (AB→E; CA→E; AD→E, etc.). In each book all the variables were drawn from a single domain; no domains were psychological. At the end of each story, children were given a forced choice between causal variables (e.g.: “Why does Bambi have itchy spots? Is it because of running in the cattails or running in the garden?”)

Prior Belief Baserate Training: Each book showed ten characters in a classroom. All ten characters experienced the same emotion (e.g. boredom waiting for a hamster to do a trick). Eight of the ten characters had a bodily 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 bodily reaction to the psychological emotion happened to very many or very few characters in the story (e.g., “Can you remind me: did very many students get sleepy or did very few students get sleepy?”).

Prior Belief Mechanisms Training: Each book explained that a particular psychological state could generate bodily effects and offered a brief account of how this might happen (e.g. “When Peter feels embarrassed, his brain makes different things happen to his body…his cheeks turn pink and he starts to blush. That’s because Peter’s brain changes the way energy moves through his body and can send energy to his cheeks.”) At the end of each book, children were asked to repeat the explanation for the bodily outcomes in the books (e.g., “Can you explain to me: what made Peter blush?”)

Control: The control books told a story about a character who experienced a recurring psychological state throughout the day (e.g. “Tom is excited because today is his birthday. In the morning, Tom’s mom gives him a present. Tom is very excited to open his first present.”). To match the
level of engagement in the other training conditions, children were asked memory questions at the end of each story.

*Free Explanation test book:* This book read in its entirety: “This is Puppy. Puppy is nervous because it’s his first day of school. Oh, oh! Puppy’s stomach hurts!” Children were asked: “Why does Puppy’s stomach hurt?”

*Procedure.*

Children were tested individually in a quiet room at their daycare. Participants were first tested on one of the two Cross Domains books (particular book counterbalanced between children; the other book was then used on the final day).

Children who met the criteria for the training study were read the first book from their assigned condition (see Figure 1). The experimenter then met with the child three more times over the course of two weeks. 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). The experimenter gave feedback if the child answered incorrectly during the training sessions (i.e., in the Statistical Reasoning training, the experimenter pointed to the recurring variable and showed the child how it occurred each day along with the effect; in the Prior Beliefs Base Rates training, the experimenter pointed to the number of children with the bodily response and observed that it was “very many” rather than “very few”; in the Prior Beliefs Mechanism training, the experimenter repeated the explanation for the bodily effect; in the Control condition, the children were reminded of the correct information). On the final day (Day 4) the children were read the Cross Domains storybook and then tested on the Free Explanation book (the order was fixed so that if children learned from the Cross Domains book, they could use the evidence and transfer their knowledge for the Free Explanation question). No feedback was given on the final day.
Preliminary analyses.

Training Books: Across the training period, children’s performance on the training books improved. In all three training conditions, children were more likely to answer the prompts at the end of the training books correctly on the last day’s training book than on the first day’s (first book 45%, last book 75% in Statistical Reasoning: McNemar (n = 20), p < .05; first book 50%, last book 75% in Prior Belief Baserates: (n = 20), p = .06; first book 35%, last book 85% in Prior Belief Mechanism: (n = 20), p < .01). This suggests that the training itself was implemented successfully.

Cross Domains Book (Final Day): Responses on the final Cross Domains book were coded as appealing to the recurring psychosomatic cause or to the alternative domain-appropriate cause (i.e., the particular food). Compared to their responses on the original test book (at floor due to the inclusion criteria), children were more likely to appeal to psychosomatic causes in all conditions (Statistical Reasoning: McNemar (n = 20) = 35%, p < .01; Prior Belief Baserates: (n = 20) = 45%, p < .01; Prior Belief Mechanism: (n = 20) = 45%, p < .01; Control: (n = 20) = 30%, p < .05). There were no significant differences between conditions. As noted however, performance on the second Cross Domains book is both a relatively insensitive and relatively weak, measure of learning: insensitive because preschoolers are vulnerable to varying their responses in response to repeated questioning (Memon et al., 1993; Poole & White, 1991); weak because learning from one Cross Domains book to another does not require any degree of generalization. Children’s ability to transfer their learning to the open-ended explanation task is thus both a more reliable and more stringent assessment of whether the training affected children’s beliefs about the plausibility of psychological variables as potential causal explanations of bodily effects.

Results:

Children’s responses on the Free Explanation book were coded as appealing to the target psychological cause in the story (e.g. feeling nervous; thinking about school), to external domain-
appropriate bodily causes not mentioned in the story (e.g., “eating too much food”, “bumping his belly”) or other. Two children (one in the Prior Belief Baserates Training and one in the Control condition) responded, “I don’t know”. Otherwise, children’s responses fell uniquely into the psychogenic or bodily category.

There were no significant differences between the three training conditions ($\chi^2 (2, n = 60) = .93, p = ns$). As predicted however, children were significantly more likely to appeal to psychological explanations across all three training conditions than the Control condition ($\chi^2 (2, n = 80) = 6.94, p < .01$): 50% of the children appealed to the psychological cause in the Statistical Reasoning condition; 40% in the Prior Belief Base Rate condition, and 55% in the Prior Belief Mechanism training, but only 15% of children did so in the Control condition (see Figure 2). Note that few children in the Control condition invoked the psychological variable on the free explanation task despite the fact that they read stories that repeatedly referred to a character experiencing psychological states (Tom is excited, Lisa is angry, Erin is sleepy, etc.).

Comparing individual training conditions with the Control condition, significantly more children appealed to psychological explanations in the Statistical Reasoning condition, ($\chi^2 (1, n = 40) = 5.58, p = .02$) and the Prior Belief Mechanism condition ($\chi^2 (1, n = 40) = 7.03, p < .01$) than the Control condition and marginally more children did so in the Prior Belief Baserates condition, ($\chi^2 (1, n = 40) = 3.14, p = .08$). Interestingly, the results in the Mechanism condition were more robust than those in the Base Rates condition, consistent with the claim that children’s prior beliefs may be particularly sensitive to information about plausible causal mechanisms (Ahn et al. 1995; Shultz, 1982). Note however, that the fact that children also robustly succeeded in the Statistical Reasoning training suggests that, consistent with the Bayesian analysis, both prior beliefs and the strength of the evidence independently affect children’s causal inferences.

**Discussion**
These results suggest that children as young as three-and-a-half are not unduly entrenched in their prior causal commitments; very young preschoolers rationally update their beliefs from evidence. Following a brief training, involving only five storybooks, children were able to invoke previously unrecognized causal variables as explanations for events.

What can our training study tell us about the role of statistical inference and prior beliefs in children’s reasoning about theory-violating evidence? First, consider the implications of the Statistical Reasoning training. The children in this condition were given no more information about psychosomatic events than children in the Control condition -- and indeed, were given less exposure than children in the Control condition to psychological variables in general (psychological variables were never mentioned during their training). Nonetheless, children in the Statistical Reasoning training were more likely than children in the Control condition to adopt psychosomatic explanations. Thus critically, although all children saw identical data in the two Cross Domains books, the children taught to reason about statistical evidence were better able than children in the Control condition to bring this evidence to bear on the explanation task. This suggests that children as young as three-and-a-half can use ambiguous statistical data, not only to draw accurate inferences about plausible causal variables, but to attribute causal power to theory-violating causal variables.

The success of the Prior Belief conditions suggests that either increasing children’s perception of the base rate of a target causal relation or increasing their understanding of the target causal mechanism may increase children’s willingness to appeal to causal relations that they previously dismissed. Although the explanations given to children in the Mechanism training were very shallow, akin to what researchers have called “coarse sketches” (Keil, 2003; p. 371), they dramatically improved children’s willingness to invoke psychological variables as causes of bodily events, whereas simple repeated exposure to psychological variables (as in the Control condition) was ineffective. Consistent with other
research, this suggests that even lean representations of causal mechanisms can support accurate causal judgments (see Keil, 2003; Rosenblit & Keil, 2002).

However, these results fall short of demonstrating that manipulating children’s prior beliefs improves their ability to learn from (erstwhile) implausible data. Children in the Prior Belief conditions may have been better able to learn from the statistical evidence in the Cross Domains books and bring this evidence to bear on the free explanation task, but it is also possible that these training conditions directly increased the children’s willingness to appeal to psychosomatic causes. Further research is needed to isolate the effect of rational constraints from the effects of other task demands to see whether changing prior knowledge can change the statistical inferences of even very young children.

Finally, although this study suggests that three-and-a-half-year-olds can use rational inductive inference to revise their beliefs, we do not know at what level of abstraction such revision occurred. Previous work (Schulz et al., 2007) suggested that four and five-year-olds changed their inferences only at a quite specific level (e.g., entertaining the possibility that worrying could cause tummy aches) rather than at the level of more abstract theories (revising their beliefs about psychogenic causality generally). In light of that, it is noteworthy that three-and-a-half-year-olds in the Prior Beliefs training conditions in the current study were able to generalize from other instances of psychosomatic causation to the test exemplar. This suggests that at least some abstract inferences were enabled by those training conditions. Further research might establish the degree to which different interventions transform children’s ability to reason about theory-violating evidence at different levels of abstraction.

Collectively however, these results suggest that the importance of rational constraints on causal learning in early childhood. More importantly, these results suggest the malleability of such constraints. If we can effectively teach three-and-a-half-year-olds to revise their beliefs with evidence, we may all have something to learn.
References


Teaching Ambiguous Evidence


Figure 1: Study design and sample pages from the training books

**Day 1**
- Cross-Domains Book
- Training Book 1

**Day 2**
- Training Book 2
- Training Book 3

**Day 3**
- Training Book 4
- Training Book 5

**Day 4**
- Cross-Domains Book
- Test Book

**2 Weeks**

**Statistical Reasoning Training Books**
(Example Book)

**Prior Belief Mechanism Training Books**
(Example Book)

**Prior Belief Baserate Training Books**
(Example Book)

**Control Condition Books**
(Example Book)

- "Why does Bambi have itchy spots? Is it because of running in the garden or running in the cattails?"
- "Can you explain, what made Peter blush when ... he was embarrassed?"
- "How many children got sleepy? Very ... many or very few?"
- "What made all of the checkers fall off of the ... table?"
Figure 2: Percentage of psychosomatic explanations generated by three-and-a-half-year-olds in each of the four conditions.