

# Can Being Scared Cause Tummy Aches? Naive Theories, Ambiguous Evidence, and Preschoolers' Causal Inferences

Laura E. Schulz and Elizabeth Baraff Bonawitz  
Massachusetts Institute of Technology

Thomas L. Griffiths  
University of California, Berkeley

Causal learning requires integrating constraints provided by domain-specific theories with domain-general statistical learning. In order to investigate the interaction between these factors, the authors presented preschoolers with stories pitting their existing theories against statistical evidence. Each child heard 2 stories in which 2 candidate causes co-occurred with an effect. Evidence was presented in the form:  $AB \rightarrow E$ ;  $CA \rightarrow E$ ;  $AD \rightarrow E$ ; and so forth. In 1 story, all variables came from the same domain; in the other, the recurring candidate cause, A, came from a different domain (A was a psychological cause of a biological effect). After receiving this statistical evidence, children were asked to identify the cause of the effect on a new trial. Consistent with the predictions of a Bayesian model, all children were more likely to identify A as the cause within domains than across domains. Whereas 3.5-year-olds learned only from the within-domain evidence, 4- and 5-year-olds learned from the cross-domain evidence and were able to transfer their new expectations about psychosomatic causality to a novel task.

*Keywords:* domain-general and domain-specific causal learning, Bayesian models, naive theories, ambiguous evidence, psychosomatic causes

*Supplemental materials:* <http://dx.doi.org/10.1037/0012-1649.43.5.1124.supp>

By the time children are 5 years old, they understand causal relationships in a variety of domains (Flavell, Green, & Flavell, 1995; Gelman & Wellman, 1991; Gopnik & Meltzoff, 1997; Inagaki & Hatano, 1993; Kalish, 1996; Perner, 1991; Shultz, 1982; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Many researchers have suggested that children's causal knowledge can be best characterized as a set of naive theories: abstract, coherent representations of causal structure that support prediction, intervention, explanation, and counterfactual claims (Carey, 1985; Gopnik, 1988; Gopnik & Meltzoff, 1997; Harris, German, & Mills, 1996; Hickling & Wellman, 2001; Keil, 1989; Perner, 1991; Sobel, 2004; Wellman, 1990; Wellman, Hickling, & Schult, 1997). The view that children's causal representations resemble scientific theories (the *theory theory*) suggests both that patterns of evidence should affect children's causal commitments and that children's causal commitments should affect their interpretation of evidence. Indeed, this dynamic relationship between domain-appropriate

causal beliefs and evidence has been taken as a defining feature of theories (e.g., Gopnik & Meltzoff, 1997).

However, despite the expectation that theory and evidence should interact, developmental psychologists have been largely divided between accounts of causal reasoning emphasizing either domain-specific causal knowledge or domain-general learning from data. Thus, some researchers have suggested that children's naive theories might be generated by domain-specific modules (Leslie, 1994; Scholl & Leslie, 1999) or innate concepts in core domains (Carey & Spelke, 1994; Keil, 1995), whereas other researchers have focused on children's ability to learn causal relations from statistical evidence (Gopnik et al., 2004; Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz & Gopnik, 2004; Shultz & Mendelson, 1975; Siegler & Liebert, 1975; Sobel, Tenenbaum, & Gopnik, 2004; Watson & Ramey, 1972). Although some research on the development of scientific reasoning has emphasized the importance of integrating domain-specific knowledge with domain-general strategies (Barrett, Abdi, Murphy, & Gallagher, 1993; Klahr & Dunbar, 1988; Koslowski, 1996; Koslowski, Okagaki, Lorenz, & Umbach, 1989; Pazzani, 1991; Penner & Klahr, 1996; Schauble, 1990), those studies have focused primarily on adolescents and adults. Surprisingly little research has looked at how prior theories and evidence interact in young children's causal learning.

Moreover, the few studies that have directly compared preschoolers' domain-specific and domain-general causal learning have generated contradictory results. Some studies suggest that children privilege domain-specific mechanism information over domain-general evidence. Work by Shultz (1982), for instance, suggests that preschoolers will override covariation evidence to base causal judgments on the presence or absence of domain-appropriate mechanisms of transmission. For example, in one

---

Laura E. Schulz and Elizabeth Baraff Bonawitz, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; Thomas L. Griffiths, Department of Psychology, University of California, Berkeley.

This research was supported by a James H. Ferry Jr. Fund for Innovation in Research Education Award and a McDonnell Foundation Collaborative Initiative Causal Learning grant to Laura E. Schulz. We thank Darlene Ferranti, Noah Goodman, Kate Hoopell, Adrianna Jenkins, Rebecca Saxe, Josh Tenenbaum, and Wendy Wienerman for helpful comments and suggestions. We are also very grateful to the Discovery Center at the Museum of Science, Boston, and participating parents and children.

Correspondence concerning this article should be addressed to Laura E. Schulz, 46-4011, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: lschulz@mit.edu

study, Shultz showed children a candle with a screen around it. He turned on a fan and then 5 s later turned on a second fan. While turning on the second fan, he moved the screen away from the first fan. The candle extinguished. When children were asked which fan extinguished the candle, children chose the first fan, which was in a position to transmit energy to the candle, rather than the second fan, whose activation was temporally contiguous with the effect. This was taken as evidence that children's causal judgments are more influenced by domain-specific information than domain-general cues, like temporal contiguity. Note, however, that some domain-general information (e.g., the temporal contiguity between removing the screen from the first fan and the candle extinguishing) may have reinforced the domain-specific information about mechanisms of transmission. Thus, it is not clear whether children genuinely privileged the domain-specific information or whether both types of information contributed to children's judgments.

In contrast to the Shultz (1982) studies, other work suggests that preschoolers can use domain-general information to override domain-specific theories. Research suggests, for instance, that 4-year-olds are able to use patterns of conditional dependence and independence to learn that talking to a machine, rather than pushing a button, will make the machine activate or that a block can activate a toy, not through contact, but at a distance (Kushnir & Gopnik, 2007; Schulz & Gopnik, 2004). However, in these studies, the evidence strongly favored the implausible (domain-inappropriate) cause. Target effects never occurred spontaneously and did occur when the domain-inappropriate candidate cause was present by itself. In these contexts, children's prior knowledge seemed to have no effect on their inferences: Children were able to learn causal relationships that violated domain boundaries as easily as within-domain relations. However, judgments made in the face of such unambiguous evidence may not provide either a particularly strong test of domain-general learning mechanisms or a particularly nuanced look at how theories affect the interpretation of statistical data. Thus, although some studies seem to suggest the relative strength of domain-specific knowledge over domain-general learning mechanisms and others suggest the opposite, little research has closely investigated the interaction between the two.

Adding to the complexity, many researchers have suggested that the relationship between theory and evidence may be poorly understood even by older children and naive adults (Chen & Klahr, 1999; Inhelder & Piaget, 1958; Klahr & Dunbar, 1989; 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 support or conflict 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 relationship consistent with their theories and instances of non-co-occurrence as evidence against the relationship incommensurate with their theories (Kuhn, 1989).

Critically, however, such differential treatment of evidence need not be irrational: Small amounts of data (e.g., seeing a vase floating in midair) may suffice to overturn weakly held beliefs (that the magic show was canceled) but should leave strong ones (that unsupported objects fall) intact. To the extent that children's causal judgments reflect normative interactions between naive theories and patterns of evidence, the mixed findings across different studies are perhaps not surprising. On any given task,

children's causal judgments might accord either with their prior knowledge or with the patterns of evidence, depending on the strength of children's initial theories, the strength of the evidence, and children's ability to integrate the two.

A rational answer to the question of how domain-specific theories should interact with statistical evidence can be obtained by approaching causal learning as a problem of Bayesian inference. In Bayesian inference, the learner seeks to evaluate a hypothesis,  $h$ , about the process that produced some observed data,  $D$ . The learner's a priori beliefs about the plausibility of  $h$  are expressed in a prior probability distribution,  $P(h)$ . The learner seeks to evaluate the posterior probability of  $h$ ,  $P(h|D)$ —their beliefs about the plausibility of the hypothesis after taking into account the evidence provided by  $D$ . This can be done by applying Bayes's rule,

$$P(h|D) = \frac{P(D|h)P(h)}{\sum_{h'} P(D|h')P(h')} \quad (1)$$

where  $P(D|h)$  is the likelihood, indicating the probability of generating the data  $D$  if the hypothesis  $h$  were true. (The sum over all hypotheses in the denominator simply ensures that the result is a probability distribution.) The posterior distribution directly combines the evidence obtained from  $D$ , through the likelihood, with the learner's initial beliefs about the plausibility of the hypothesis, expressed in the prior  $P(h)$ . 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, 2007; Tenenbaum & Niyogi, 2003). Thus, Bayesian inference provides a formal account of how domain-specific theories and domain-general patterns of evidence might interact to affect children's beliefs.

Guided by this account, we looked at how statistical evidence affects children's causal inferences and how children's beliefs about the plausibility of causal hypotheses affect children's interpretation of data. Because we were interested in interactions between naive theories and evidence, we gave children evidence that is formally ambiguous: Children observed events in which two candidate causes simultaneously covary with the effect. We gave children a forced choice between the two causes and manipulated the extent to which each cause was consistent with children's naive theories and with the statistical evidence.

Previous research suggests that preschoolers are able to evaluate evidence of this complexity (i.e., evidence in which candidate causes are never presented in isolation). In one study, for instance, a puppet smelled a bouquet consisting of a tulip and a daisy. The puppet sneezed. The puppet then smelled a bouquet consisting of a tulip and a violet and the puppet sneezed. Children then saw that a bouquet consisting of a daisy and a violet did not make the puppet sneeze. When asked what made the puppet sneeze, children inferred that the tulip, rather than the other flowers, was the cause (Schulz & Gopnik, 2004).

Children can also evaluate ambiguous data with respect to the base rate of candidate causes. Suppose, for instance, children learn that a toy will light up when particular blocks are placed on top of the toy. Children learn either that only 2 of 10 blocks activate the toy (i.e., activating blocks are rare) or they learn that 8 of 10 blocks activate the toy (activating blocks are common). Children in both conditions then see 2 novel blocks, red and blue, placed simulta-

neously on the toy. The toy activates. Children subsequently see a red and yellow block placed simultaneously on the toy. Again, the toy activates. What makes the toy go—just the red block, just the blue and yellow blocks, or all 3 blocks? If activating objects are rare, then it is most likely that only a single block (the red one) is the cause. However, if activating blocks are common, it becomes more plausible that the blue and yellow block or all 3 blocks are causes. Research suggests that children's judgments about activating blocks are sensitive to such base rate information (Tenenbaum, Sobel, Griffiths, & Gopnik, 2007). Similarly, suppose that preschoolers see a red block and a blue block together activate a toy and then see that the red block by itself activates the toy. Children know the red block is a cause, but what about the blue block? If activating blocks are rare, children tend to deny that the blue block is a cause; if activating blocks are common, children are more likely to think the blue block is also causally effective (Sobel et al., 2004).

Our task was formally similar to the tasks used in these previous studies. We read children storybooks in which one cause recurs every day and the other cause is always novel (i.e., the evidence is in the form  $AB \rightarrow E$ ;  $CA \rightarrow E$ ;  $AD \rightarrow E$ ; etc.). One storybook is a *within-domain* story; all variables come from the same domain and thus all causes are a priori equally plausible. If children can engage in domain-general statistical learning from patterns of evidence, we expect that after seeing the evidence, children will infer that A is more probable than any other single cause. The other storybook is a *cross-domains* story: the recurring candidate cause (A) is domain-inappropriate. Thus, A is less plausible than the alternative given the children's naive theories but more plausible given the pattern of evidence. By comparing children's judgments before and after seeing the data, we can evaluate the degree to which children can overcome the biases induced by their naive theories.

Because we wanted to investigate processes that might be applicable to genuine conflicts between theories and evidence, we chose to look at a context in which preschoolers' causal beliefs are robust (and thus might affect children's interpretation of data) but distinct from adult beliefs (and thus might change with evidence). As noted, considerable research suggests 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 (Bloom, 2004; Carey, 1985; Estes, Wellman, & Woolley, 1989; Hatano & Inagaki, 1994; Notaro, Gelman, & Zimmerman, 2001; Wellman & Estes, 1986). Indeed, some researchers have proposed that children may be innate dualists (Bloom, 2004). Thus, although many adults accept the existence of psychosomatic phenomena, preschoolers typically deny that psychosomatic reactions are possible (e.g., they deny that feeling embarrassed can make one blush or that feeling frustrated can cause a headache; Notaro et al., 2001).

Note that this is not to suggest that children deny all relations between mental states and bodily events.<sup>1</sup> In particular, children do understand that volitional mental states (e.g., desires and intentions) can cause intentional action (see, e.g., Bartsch & Wellman, 1989; Meltzoff, 1995; Wellman et al., 1997). Indeed, in the context of voluntary action, bodily events are more typically attributed to psychological causes ("She kicked the ball because she wanted to make a goal") than bodily causes ("She kicked the ball because she lifted her leg, extended her knee, etc."). However, by definition,

involuntary bodily events (e.g., tummy aches and headaches) are not attributed to desires or intentions. Although many adults accept that involuntary bodily states can be caused by psychological states like worrying or fear, children seem to attribute involuntary bodily states exclusively to bodily causes (e.g., illness and injury). Thus, following Notaro et al. (2001), we investigated children's understanding of only a subset of possible relations between psychological and bodily events: cases where a nonvolitional mental state is the cause of an involuntary bodily reaction.<sup>2</sup>

We were interested in how preschool children would interpret formal patterns of evidence suggesting the presence of a psychosomatic cause in light of their strong initial belief that psychosomatic causality is improbable. Thus, in our within-domain task, both the candidate causes and the target effect come from the domain of physiological events. In particular, the candidate causes are bodily contact with different plants and the effect is "itchy spots." In the cross-domains task, all but one of the candidate causes are physiological events (ingestion of different foods), and the effect is also a physiological event (a tummy ache). However, the recurring cause (A) is a psychological event (feeling scared).

Bayesian inference provides us with a rational standard against which we can compare the judgments of children in this task. In this case, the hypotheses concern the relationships between the potential causes and the effect, and the data are the co-occurrences of causes and effects over the days of the story. The prior probabilities of the hypotheses are determined by children's naive theories, which we assume allow a greater probability for relationships within a domain than relationships that cross a domain. A detailed Bayesian model and quantitative predictions of this model for our task are described in Appendix A. This model makes two clear qualitative predictions. First, children's naive theories should affect their interpretation of the evidence: Children should be less likely to choose A in the cross-domains task than in the within-domain task. Second, the evidence should affect children's beliefs: Children should be more likely to choose A after seeing the evidence than at baseline. Earlier research (Sobel et al., 2004) suggests that older (4-year-olds) but not younger (3-year-olds) preschoolers are able to integrate knowledge about base rates and patterns of evidence, so we tested our predictions about the integration of domain-specific prior knowledge and evidence across a range of ages: 3-year-olds, 3.5-year-olds, and 4- and 5-year-olds.

## Experiment 1

In Experiment 1, we read children two storybooks: a within-domain story and a cross-domains story. The evidence (presented in the form  $AB \rightarrow E$ ;  $CA \rightarrow E$ ;  $AD \rightarrow E$ , etc.) is formally identical in the two stories. We predicted that children would be more likely to identify A as a cause when A is domain-appropriate than when it is domain-inappropriate. However, we also predicted that, for both stories, children would be more likely to think A is a cause after seeing the evidence than at baseline.

<sup>1</sup> We are grateful to an anonymous reviewer for stressing the importance of this distinction.

<sup>2</sup> Note that this experiment compares only children's different judgments about the mind and the body. It does not require children to have a fully elaborated naive biology (see, e.g., Carey, 1985).

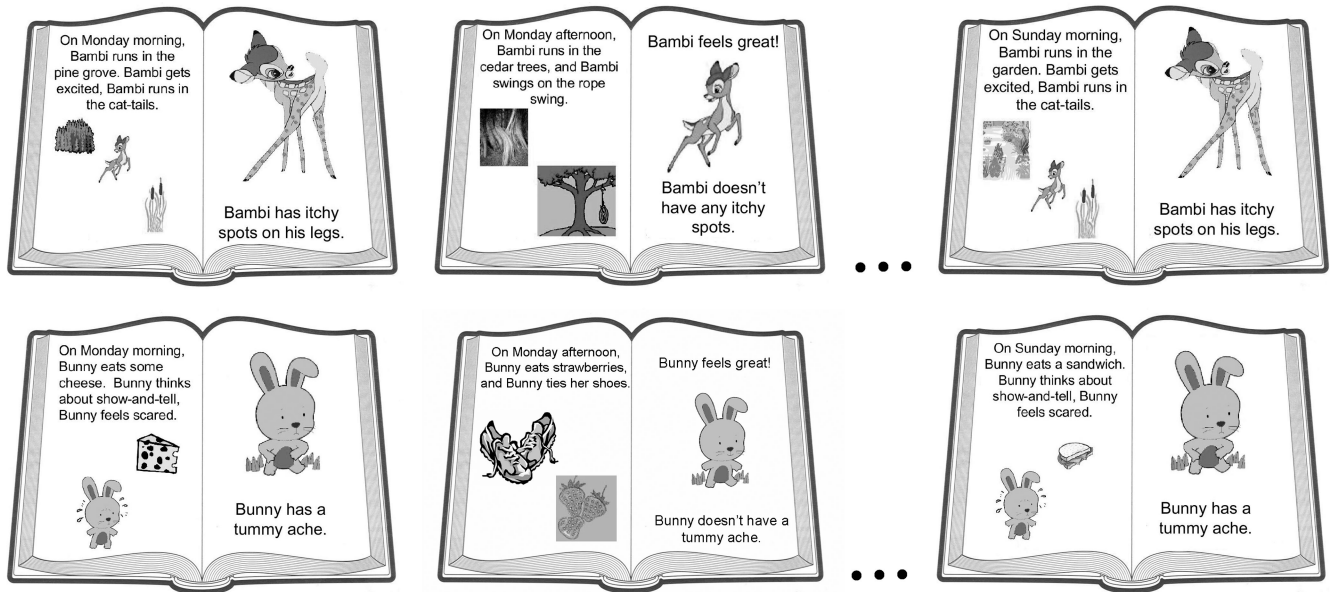


Figure 1. Sample pages from the within-domain storybook (top panel) and the cross-domains storybook (bottom panel) used in Experiments 1 and 2.

## Method

**Participants.** Eighty preschoolers were recruited from urban area preschools and the Discovery Center at a large metropolitan science museum. Approximately equal numbers of boys and girls participated (49% girls). Although most children were from White, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented.

Children were tested in three age groups: 3-year-olds (mean age: 39 months; range: 36–41 months), 3.5-year-olds (mean age: 45 months; range: 42–48 months), and 4- to 5-year-olds (mean age: 60 months; range: 50–70 months). The wider age range was used for the oldest age group because pilot work suggested that the performance of 4- and 5-year-olds did not differ on this task. In each age group, 16 children were tested after seeing statistical evidence (the evidence condition). Additionally, 16 4- to 5-year-olds and 16 3-year-olds (8 in the older group; 8 in the younger group) were tested before seeing any statistical evidence (the baseline condition). We tested the 3-year-olds as a single group at baseline because pilot work suggested no difference in baseline performance for the youngest two age groups.

**Materials.** Two storybooks were used in the experiment: a within-domain book and a cross-domains book. Each storybook depicted events occurring over the course of a week. Every morning (Monday–Sunday), two events and an effect occurred. One event (A) and the effect were repeated every morning; the other event varied. Each afternoon, two different events occurred, and the effect failed to occur. (The afternoon events were included to eliminate the possibility that the effect was always present.) Two versions of each storybook were created to counterbalance the order of events.

The within-domain storybook featured a deer (Bambi) who liked to run in different places. Sample text read,

On Monday morning, Bambi runs in the pine grove. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs. On Monday afternoon, Bambi runs in the cedar trees and Bambi plays on the rope swing. Bambi feels great! Bambi doesn't have any itchy spots.

The story continued through the days of the week and ended with, “On Sunday morning, Bambi runs in the garden. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs.”

The cross-domains storybook featured a bunny who was scared of show-and-tell. Sample text read, “On Monday morning, Bunny thinks about show-and-tell. Bunny feels scared. Bunny eats some cheese. Bunny has a tummy ache. On Monday afternoon, Bunny ties her shoes and Bunny eats strawberries. Bunny feels great. Bunny doesn't have a tummy ache.” The story continued through the days of the week and ended with “On Sunday morning, Bunny thinks about show-and-tell, Bunny feels scared. Bunny eats a sandwich. Bunny has a tummy ache.”<sup>3</sup> See Figure 1 for details and Appendix B for the full text of the stories.

**Procedure.** Children were tested individually. The experimenter read both the within-domain story and the cross-domains story to every child (order of stories was counterbalanced among participants). In the baseline condition, children were read only the

<sup>3</sup> In the cross-domains story, the sentence “Bunny thinks about show-and-tell” precedes the target variable (“Bunny feels scared”) in order to delimit the onset of the psychological cause. In order to match the evidence in the two books precisely, we included the sentence “Bambi gets excited” before the target cause, “Bambi runs in the cattails.” It is possible that the inclusion of the lead-in sentence aided children in tracking the evidence in both conditions. However, the inclusion of the lead-in sentence did not affect children's causal inferences at baseline, nor can it account for differences between the within- and cross-domains story.



Sunday page of each story. In the evidence condition, children were read each story in its entirety. At the end of each story, children were asked a test question. In the within-domain story, the test question (in one version's order) was "Why does Bambi have itchy spots? Is it because of running through the garden or because of running through the cattails?" In the cross-domains story, one version of the test question was "Why does Bunny have a tummy ache? Is it because of feeling scared or eating the sandwich?"

## Results

The results are presented in Figure 2. An alpha level of .05 was used throughout this article, and thus all results reported as significant are  $p < .05$  or better. Preliminary analyses revealed no order effects. Because our dependent measure was the number of children making each category choice and we could not be sure the data met the normality assumptions of parametric tests, we used categorical tests (binomial and chi-square tests) throughout.

In the 4- and 5-year-old age group (4 years 0 months to 5 years 6 months) in the within-domain task, there was no significant

difference in the probability with which children chose A or the alternative in the baseline condition ( $N = 16$ ,  $p = ns$ , by binomial test), but children chose A significantly more often than chance (indeed, at ceiling) in the evidence condition ( $N = 16$ , by binomial test). Children were significantly more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 32) = 10.67$ . In the cross-domains task, children had a significant preference for the domain-appropriate cause in the baseline condition ( $N = 16$ , by binomial test) but did not display a statistically significant difference in their choices of A and the alternative in the evidence condition ( $N = 16$ ,  $p = ns$ , by binomial test). Children were significantly more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 32) = 5.24$ . Both at baseline and after seeing the evidence, children were more likely to choose A in the within-domain than cross-domains task: baseline,  $\chi^2(1, N = 32) = 5.24$ ; evidence,  $\chi^2(1, N = 32) = 10.67$ .

In the 3.5-year-old age group (3 years 6 months to 4 years 0 months) in the within-domain task children again did not show a statistically significant difference in their choices between A and the alternative at baseline ( $N = 16$ ,  $p = ns$ , by binomial test) and again significantly preferred A in the evidence condition ( $N = 16$ , by binomial test). Children were more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 32) = 7.57$ . However, in the cross-domains task, children preferred the domain-appropriate cause in the baseline condition ( $N = 16$ , by binomial test) and continued to prefer the domain-appropriate cause in the evidence condition ( $N = 16$ , by binomial test). Children were not significantly more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 31) = 0$ ,  $p = ns$ . Both at baseline and after seeing the evidence, children were more likely to identify A as a cause in the within-domain than cross-domains task: baseline,  $\chi^2(1, N = 32) = 5.24$ ; evidence,  $\chi^2(1, N = 32) = 21.21$ .

Seeing the evidence had no effect for the youngest age group (3 years 0 months to 3 years 6 months). In the within-domain task, children showed no statistically significant preference between A and the alternative even in the evidence condition ( $N = 16$ ,  $p = ns$ , by binomial test). Children were not significantly more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 32) = 0.51$ ,  $p = ns$ . In the cross-domains task, children preferred the domain-appropriate cause in the baseline condition ( $N = 16$ , by binomial test) and were not significantly more likely to choose A after seeing the evidence than at baseline,  $\chi^2(1, N = 32) = .82$ ,  $p = ns$ . Children were more likely to choose A in the within-domain than cross-domains task: baseline,  $\chi^2(1, N = 32) = 5.24$ ; evidence,  $\chi^2(1, N = 32) = 4.57$ .

The different results observed across the different age groups suggest a developmental effect, with age group interacting with the effect of statistical evidence in the cross-domains task. We tested for the possibility of such an interaction using a log-linear model, predicting the frequency with which children chose A over the alternative as function of age group, condition (baseline or evidence), and an Age Group  $\times$  Condition interaction. Removal of the age group factor from the saturated model did not result in a statistically significant increase in lack of fit,  $\chi^2(2, N = 80) = 4.49$ ,  $p = ns$ , whereas removal of condition or the interaction resulted in a statistically significant increase in lack of fit for both the saturated model (condition:  $\chi^2[1, N = 80] = 8.93$ ; Age Group  $\times$  Condition:  $\chi^2[2, N = 80] = 9.48$ ) and the model without

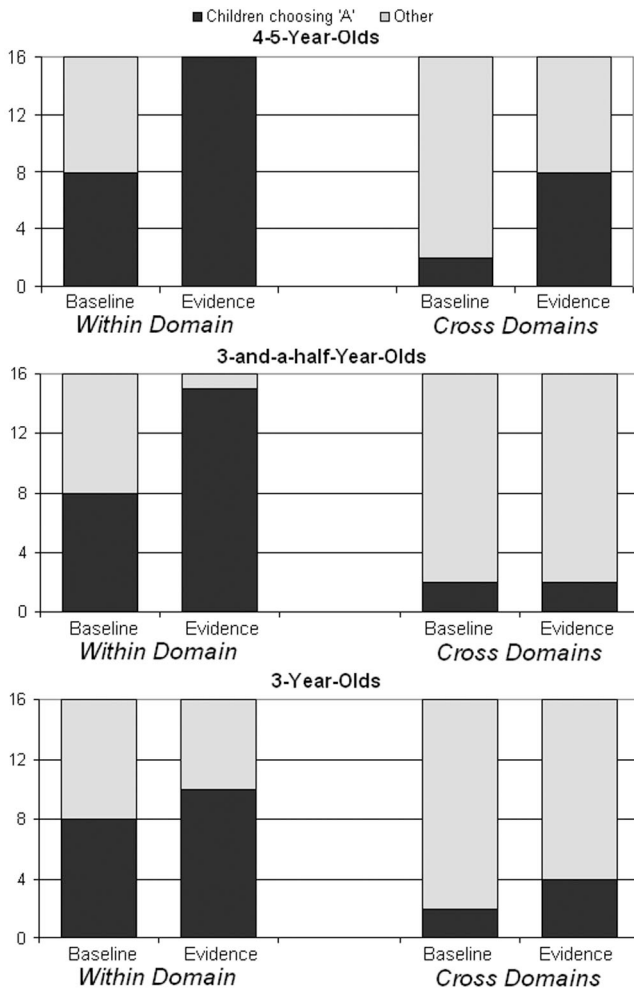


Figure 2. Children's responses to the storybook task in Experiment 1. The vertical axis shows the number of children selecting the different responses.

the effect of age group (condition:  $\chi^2[1, N = 80] = 8.93$ ; Age Group  $\times$  Condition:  $\chi^2[2, N = 80] = 7.74$ ). These results indicate that the best model includes condition and the Age Group  $\times$  Condition interaction as predictors, supporting the hypothesis that the 4- and 5-year-olds responded differently to the cross-domain evidence than did the 3-year-olds.

### Discussion

The results of Experiment 1 suggest both that children's domain-specific theories interact with their interpretation of evidence and that the nature of this interaction changes over the course of development. Overall, we found a graded interaction between prior knowledge and evidence of the kind predicted by our Bayesian model: All but the youngest children learned that A was a cause when A was consistent with their theories, and all of the children were less likely to identify A as a cause when A violated their theories. (Note that this finding rules out a simple associative explanation of children's inferences; the association between Variable A and the effect was identical within and across domains.) Critically, the oldest preschoolers seemed to learn from the evidence even when the evidence conflicted with their prior beliefs. After seeing the data, 4- and 5-year-olds were able to entertain a causal possibility (that being scared might cause tummy aches) that they did not endorse without seeing those data.

By contrast, the younger 3-year-olds (3 years 0 months to 3 years 6 months) had a strong preference for domain-appropriate causes and apparently failed to learn from the evidence throughout. It is not clear whether this failure is due to competence or performance deficits. Our task was quite complex and younger children might have been able to learn from the data in a simpler or more supported task. Alternatively, the youngest children might have understood the evidence, but the evidence might not have overcome the children's initial inductive biases, even within domains. (That is, the youngest children might have found it relatively more difficult to believe that one type of plant might make one itch and others might not.) Further research might disambiguate these accounts.

However, for the 3.5-year-olds (3 years 6 months to 4 years 0 months), the discrepancy between within- and cross-domain reasoning was particularly striking. Although the evidence in the tasks was formally identical, 94% of the children in this group inferred that A was the cause in the within-domain task (no different than the 4- and 5-year-olds), whereas only 12% inferred that A was the cause in the cross-domains task (no different than at baseline).

Why did the 3.5-year-olds respond differently to the cross-domains evidence than the 4- and 5-year-olds? There are at least three possible explanations. One possibility is that 3-year-old children might have difficulty making inferences from ambiguous statistical data. If (as suggested by the failure of the younger 3-year-olds to use the evidence at all) the ability of the 3.5-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. Alternatively, the younger children might have a stronger belief in domain boundaries than older children. The data might have been insufficient to overcome 3-year-olds' initial inductive bias that psychological causes are unlikely to generate bodily effects. Finally, the older 3-year-olds and the 4- and 5-year-olds might not differ either

with respect to their ability to evaluate evidence or their initial domain-specific theories. However, younger children might be less able than older children to update their beliefs on the basis of surprising evidence. Again, further research might disambiguate these accounts.

Overall, the results of Experiment 1 suggest that domain-general and domain-specific information interact to affect children's causal learning, consistent with the prescriptions of Bayesian inference. Moreover, older preschoolers can use statistical evidence to make inferences against their domain-specific theories even when the data are ambiguous. That is, children can infer a domain-inappropriate causal relationship even when the evidence does not formally rule out the causal relationship consistent with their initial theories.

However, Experiment 1 also suggests that children's learning is relatively conservative; children were less likely to learn from statistical evidence that conflicted with their theories than from evidence consistent with their theories. If children's learning is conservative, then the children might not generalize much beyond the task itself. That is, even those children who endorse psychosomatic causes in the cross-domains story might be reluctant to endorse psychosomatic causality in general. Alternatively, the children might be more willing to accept the possibility of other psychosomatic events. In Experiment 2, we look at how exposure to the evidence in the cross-domains story affects children's inferences about other psychosomatic events.

### Experiment 2

In earlier research on children's understanding of the limits of psychological explanations, Schult and Wellman (1997) showed preschoolers actions that were physically and biologically possible (e.g., jumping up and down; drinking orange juice) or impossible (walking through a wall; staying awake forever). They found that preschoolers distinguished possible and impossible events in both domains; that is, children understood that one could do possible events if one wanted to but one could not do impossible ones. In Experiment 2, we used a similar method to look at preschoolers' judgments about the possibility of physical, psychological, and psychosomatic events. We asked children to make these possibility judgments either at baseline or after reading the cross-domains story used in Experiment 1. If the children interpret the storybook evidence conservatively, then children who hear the cross-domains story should be no more likely than children at baseline to say that other types of psychosomatic events are possible. However, if children generalize broadly, then children who hear the cross-domains story should be more likely than children at baseline to endorse psychosomatic causality.

### Method

*Participants.* Thirty children (mean age: 58 months; range: 49–71 months) were recruited from urban area preschools. Fifty percent of the participants were girls. Although most children were from White, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented. Children were randomly assigned to a baseline possibility judgments condition or an evidence and possibility judgments condition.

*Materials.* Six pictures were used (see Appendix C). The pictures showed 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). In the evidence and possibility judgments condition, the cross-domains storybook from Experiment 1 was also used.

**Procedure.** Children were tested individually. In the baseline possibility judgments condition, children were shown each of the six pictures in one of two fixed semirandom orders: (Order 1) physically impossible; psychogenic headache; biologically impossible; biologically possible; physically possible; psychogenic sickness; (Order 2) biologically possible; psychogenic sickness; physically possible; physically impossible; psychogenic headaches; biologically possible. The experimenter read the children a brief passage about the events in the picture (see Appendix C). At the end of each passage, children were 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?" The evidence and possibility judgments condition was identical to the baseline possibility judgments condition except that children were first tested on the cross-domains storybook as in the evidence condition of Experiment 1.

### Results and Discussion

In the evidence and possibility judgments condition, children's responses to the cross-domains storybook replicated the results for this age group in Experiment 1. Sixty percent of the children chose "being scared," not significantly different from the 50% who chose being scared in the evidence condition of Experiment 1,  $\chi^2(1, N = 31) = 0.125, p = ns$ , and significantly more than the 12% of children who chose being scared in the baseline condition of Experiment 1,  $\chi^2(1, N = 31) = 7.63$ .

In the possible/impossible picture task, 1 child in the baseline possibility judgments condition and 1 child in the evidence and possibility judgments condition answered yes to all six questions, and 1 child in the baseline possibility judgments condition answered no to all six questions. To ensure that children could properly distinguish possible and impossible events, we eliminated these children from further analysis, leaving 13 children in the baseline possibility judgments condition and 14 children in the evidence and possibility judgments condition.

The critical question was whether children would be more likely to say that psychosomatic events were possible in the evidence and possibility judgments condition than in the baseline possibility judgments condition. In fact, there was no difference in children's possibility judgments between the conditions,  $\chi^2(1, N = 27) = .07, p = ns$  (see Table 1). In both conditions, children denied the possibility of both psychosomatic events significantly more often than expected by chance (by binomial test), and no other patterns of responses occurred more often than chance (by binomial test). Within the evidence and possibility judgments condition, children who chose being scared in the cross-domains story were no more

Table 1  
*Children's Possible/Impossible Judgments in Experiment 2*

Condition	No. of children who said that the event could happen (all others said the event could not happen)	
	Baseline possibility judgments ( $n = 13$ )	Evidence and possibility judgments ( $n = 14$ )
Biologically possible	12 (92)	11 (79)
Biologically impossible	2 (15)	2 (21)
Physically possible	13 (100)	14 (100)
Physically impossible	0 (0)	0 (0)
Psychogenic (headaches)	4 (31)	5 (36)
Psychogenic (sickness)	3 (23)	5 (36)

*Note.* Percentages are in parentheses.

likely than children who chose food to say that the other psychosomatic events were possible,  $\chi^2(1, N = 14) = 2.14, p = ns$ .

There were no significant differences in children's tendency to judge the biologically impossible, physically impossible, and psychogenic events as impossible; similarly, children were equally likely to judge the physically possible and biologically possible events as possible (by McNemar's test,  $p = ns$  throughout). In both conditions, children indicated that biologically and physically possible events were possible and biologically and physically impossible events were impossible significantly more often than expected by chance (by binomial test).

The results of Experiment 2 suggest that children interpret the evidence in the cross-domains story quite conservatively. Observing the evidence in the cross-domains story did not affect children's willingness to accept other causal relationships (e.g., between worrying and headaches or between being nervous and feeling sick). Why is children's learning so constrained? One possibility is that children's causal generalizations are affected by their understanding of the domains involved. If preschoolers think that tummy aches, headaches, and feeling sick are distinct forms of illness, they might not readily generalize causes of tummy aches to other ailments. (Anecdotally, for instance, the children seemed to identify feeling sick primarily with vestibular upset and throwing-up; several children volunteered reminiscences on the topic in that context but never otherwise). Alternatively, children's generalizations might have been affected by the extent to which they treated being scared, worrying, and being nervous as the same type of causal event; children might not have appreciated the commonality among the psychological variables. Indeed, different beliefs about the commonality among mental states or bodily states may affect even adults' generalizations of psychosomatic causality. Adults may accept, for instance, that worrying can cause tummy aches but deny that worrying can cause cancer; similarly, they may accept that anxiety can cause headaches but deny that excitement causes headaches.

Alternatively, children might have failed to generalize psychosomatic causality from the cross-domains story to the possible/impossible judgment task simply because the evidence for psychosomatic causality provided by the cross-domains story was relatively weak. Other plausible candidate causes (e.g., food) were

always present and children saw only a total of seven trials. Given the conflict between the statistical evidence and children's prior beliefs, such conservative learning from minimal data is rational; if children have strong prior beliefs and the evidence against these beliefs is relatively limited, children's naive theories should be robust to the anomalous data. The results of Experiment 2 suggest that children might engage in just this sort of authentic but conservative learning.

### Experiment 3

However, given children's failure to generalize their inferences, one might be skeptical that children genuinely learn from the statistical evidence in the first place. Although in both Experiments 1 and 2, children were significantly more likely to identify being scared as a cause in the cross-domains task after seeing the evidence than at baseline, children did not choose being scared significantly more often than chance in either experiment (50% of children chose being scared in Experiment 1; 60% chose being scared in Experiment 2). Because children had a forced choice of two variables, it is not clear whether the children genuinely learned to infer psychosomatic causes from the evidence or whether the surprising evidence confused the 4- and 5-year-olds and led them to choose at chance.

In Experiment 3, we introduced two measures to distinguish authentic learning from chance performance. First, we modified the cross-domains story so that three candidate causes (one domain inappropriate, two domain appropriate), rather than two, covary with the effect every day. If the children are confused by the evidence and choosing at chance, they should choose being scared 33% of the time. However, if children genuinely learn from the data, we would expect to replicate the results of Experiments 1 and 2: Children should choose being scared more often than chance and more often than either of the other variables.

Second, we asked children to extend their inferences from the forced-choice task to a free-explanation task. If children genuinely learn the target psychosomatic causal relation (between being scared and tummy aches), they should be able to apply this knowledge to explain a new instance of the same target relation, even if they are unwilling to endorse psychosomatic causes in general. To assess children's ability to engage in this near transfer of their learning, we read children a passage about a puppy that is worried about the first day of school and has a stomachache. We chose to use the free-explanation task because we believed that following up the storybook task with a possible/impossible judgment question about the same target relation ("Can that happen? Can worrying cause tummy aches?") might lead the children to believe that we were questioning their original responses. Because preschoolers are vulnerable to changing their answers on repeated questioning, we believed the free-explanation task would be a more sensitive measure of children's understanding.

Children were given the explanation task both at baseline (the baseline explanation condition) and after having read the revised cross-domains story (the evidence and explanation condition). If the cross-domains story does not affect children's learning, then there should be no difference between the two conditions. However, if children do learn from the evidence in the cross-domains story, they should be more likely to attribute the puppy's stom-

achache to worrying in the evidence and explanation condition than in the baseline explanation condition.

### Method

*Participants.* Forty children (mean age: 58 months; range: 49–71 months) were recruited from urban area preschools. Fifty percent of the participants were girls. Although most children were from White, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented. Children were randomly assigned to a baseline explanation or an evidence and explanation condition.

*Materials.* The cross-domains storybook of Experiment 1 was modified so that every morning of the week, three events (two domain appropriate and one domain inappropriate) and the effect occurred. The domain-inappropriate event (being scared) was repeated each day; the other two events always varied. Thus, a sample test page read, "Bunny thinks about show-and-tell. Bunny feels scared. Bunny eats a sandwich. Bunny drinks apple juice. Bunny has a tummy ache." Two different versions of each book were created so that for half the children being scared was the last of the three events and for half the children being scared was the first of the three events. Additionally, a novel puppy storybook was used. The text preceding the test question read in its entirety: "This is Puppy. Puppy is worried because next week he starts school. The first day of school makes Puppy worried. Puppy's stomach hurts."

*Procedure.* Children were tested individually. In the baseline explanation condition, children were read only the puppy book. The test question was open ended: "Why do you think Puppy's stomach hurts?" Children were asked to offer an explanation, and no prompts were given. The evidence and explanation condition was identical except that children were first tested on the revised cross-domains book as in Experiments 1 and 2.

### Results and Discussion

Children's responses to the cross-domains story in this experiment replicated the results in Experiments 1 and 2. Fifty-five percent of the children in the evidence and explanation condition chose being scared compared with the 50% of children who chose being scared in the evidence condition of Experiment 1 and the 60% of children who chose being scared on the storybook task in the evidence and possibility judgments condition of Experiment 2. Although in this experiment children were faced with a choice of three variables rather than two, there were no significant differences between children's tendency to choose being scared in this experiment and the evidence conditions of Experiments 1,  $\chi^2(1, N = 36) = 0.09, p = ns$ , and 2,  $\chi^2(1, N = 36) = 0.31, p = ns$ . Children were significantly more likely to choose being scared in this experiment than in the baseline condition of Experiment 1,  $\chi^2(1, N = 36) = 6.96$ . Within this experiment, children chose being scared significantly above chance ( $N = 20$ , by binomial test) and did not choose either of the other two variables above chance ( $N = 20, p = ns$ , by binomial test). There was a trend for children to choose being scared over any other variable,  $\chi^2(1, N = 20) = 4.0, p = .08$ .

On the free-explanation task, we coded children's explanations for reference to physical/bodily variables (e.g., sickness, hunger, or



injury) and psychological variables (e.g., worrying about the first day of school). Explanations fell uniquely and unambiguously into a bodily, psychological, or “I don’t know” category. In the baseline explanation condition, 55% of the children gave only bodily/physical explanations, 30% of the children gave only psychological explanations, and 15% of the children said, “I don’t know.” By contrast, in the evidence and explanation condition, 20% of the children referred only to bodily/physical causes, 70% of the children referred only to psychological causes, and 10% said “I don’t know.” The bodily/physical explanations all referred to food or hunger, with the exception of a single child in the baseline explanation condition who said “itchy stomach.” The psychological explanations all referred to being the first day of school and/or being worried, sad, scared, or nervous. Children were significantly more likely to reference psychological causes in the evidence and explanation condition than in the baseline explanation condition,  $\chi^2(1, N = 40) = 6.67$ .

We also analyzed the data to see whether the children who chose being scared in the cross-domains story were more likely to offer psychological explanations on the puppy book than those who did not. Of the 11 children who chose being scared in the cross-domains story, 8 (73%) offered psychological explanations in the transfer task; of the 9 children who did not choose being scared in the cross-domains story, 6 (67%) offered psychological explanations in the transfer task. There was no significant difference between these groups (Fisher’s Exact Test,  $p = ns$ ).<sup>4</sup> This suggests that even those children who did not identify being scared as the causal variable in the cross-domains story may have learned enough from the evidence to treat worrying a relevant causal variable in the free-explanation task.

The results of Experiment 3 suggest that children are not merely confused by statistical evidence that violates their prior beliefs; rather, children draw accurate inferences from such evidence. Preschoolers were able to use the statistical evidence to identify a psychological variable as a likely candidate cause of a bodily effect in both a forced-choice and free-explanation task. These experiments suggest that 4- and 5-year-olds can genuinely learn novel causal relations from limited amounts of data, even when the evidence conflicts with the children’s prior beliefs.

Why were children able to transfer their inferences about psychosomatic causes from the cross-domains story to the free-explanation task in this experiment but not from the cross-domains story to the possible/impossible judgment tasks in Experiment 2? We believe the difference between the experiments can be explained in a number of ways. First, the change in stimuli across the tasks might have impaired children’s transfer of information in Experiment 2; that is, children might have found it more difficult to transfer their inferences from the storybook to the picture tasks than from one storybook to another storybook. Second, the possible/impossible judgment task might have been more difficult than the free explanation task. If so, the greater difficulty of the task might have made the transfer of knowledge less likely. Finally, as hypothesized, children might have been less willing to generalize their knowledge about one psychosomatic causal relation to psychosomatic causes in general than to extend their inferences about a single psychosomatic causal relation. As noted, given the minimal evidence for psychosomatic causality provided by the cross-domains story, it is rational that children might have interpreted the

data conservatively and transferred their inferences more readily in Experiment 3 than in Experiment 2.

## General Discussion

Collectively, these three experiments suggest that children learn about causal relationships by taking into account both statistical evidence and constraints from their naive theories, consistent with the predictions of Bayesian inference models. Given identical evidence, preschoolers were more likely to identify a variable as a cause when the variable was consistent with their theories than when it violated their theories. Older preschoolers were able to use ambiguous, domain-violating evidence to make inferences about psychosomatic causality that they did not make at baseline. Moreover, children were able to learn from the cross-domain evidence even though their initial domain-specific theories were not ruled out but merely rendered less probable by the data. Finally, children’s learning was sufficiently robust that children who observed evidence for psychosomatic causality were more likely than children at baseline to offer psychosomatic explanations in a novel task.

The role of Bayesian inference in our analysis of children’s ability to combine statistical evidence and constraints from naive theories was intended to be similar to that of ideal observer analysis in vision (e.g., Yuille & Kersten, 2006) and rational analysis in the study of adult cognition (e.g., Anderson, 1990; Marr, 1982; Shepard, 1987). Bayesian inference provides a rational solution to the problem of updating one’s beliefs in the light of new evidence and can thus guide researchers in exploring how well children solve this problem. In particular, a Bayesian model can allow researchers to make both qualitative predictions about how evidence and theories interact and quantitative predictions about the conclusions that are warranted from a particular combination of observed data and constraints derived from a theory. The Bayesian model presented in Appendix A provides one such set of predictions, showing that the judgments of the 4- and 5-year-olds in our experiments are close to the probabilities entertained by an ideal Bayesian learner using a particular causal theory.

We do not claim that Bayesian inference is the only way to define a model that could reproduce our results. Although the effect of domain on children’s judgments is inconsistent with accounts of causal learning based purely on the strength of association or patterns of covariation between events (Cheng, 1997, 2000; Shanks, 1985; Shanks & Dickinson, 1987; Spellman, 1996), such accounts could predict the data reported here if augmented with initial assumptions about the strength of causal relationships within and across domains. Our intent was not to explore the mechanism by which children make these judgments but rather whether children solve the abstract computational problem of combining theory and evidence in a way that is consistent with the prescriptions of Bayesian inference. In our studies, an appropriately augmented associative mechanism remains a possible explanation for how children could approximate the rational Bayesian solution to this problem, as it would essentially build in the two critical components of Bayesian inference: initial beliefs regarding

<sup>4</sup> A Fisher’s Exact Test was used in this case because the expected counts in the cells of the contingency table were less than four.

possible causal relationships and revision of these beliefs in light of evidence. Further experiments would be necessary to explore the adequacy of such an account, although we note that the results of several previous studies of causal reasoning in children would seem to provide evidence against simple associative models as a general explanation for performance on this kind of task (e.g., Sobel et al., 2004; Tenenbaum et al., 2007).

The results of our experiments also raise several other questions. As noted, it is not clear whether the developmental differences among 3-year-olds, 3.5-year-olds, and 4- and 5-year-olds are due to changes in children's ability to evaluate evidence, changes in children's initial inductive biases, or both. The relative influence of domain-specific and domain-general information might change dramatically over the course of development, depending on age-related commonalities and differences in children's commitment to domain-specific theories, sensitivity to evidence, and ability to integrate the two. Further research might disambiguate the role of each of these factors.

Of note, however, the finding that 3.5-year-olds did not use the theory-violating evidence to change their judgments is consistent with a wide variety of research suggesting that 3-year-olds have particular difficulty changing their minds in the face of evidence. On paradigms as diverse as theory of mind tasks, ambiguous figure tasks, and dimensional change card sort tasks, 3-year-olds' initial inferences seem to be remarkably impervious to feedback (Gopnik & Rosati, 2001; Kirkham, Cruess, & Diamond, 2003; Munakata & Yerys, 2001; Russell, Jarrold & Potel, 1994; Zelazo, Frye, & Rapus, 1996). Researchers have suggested process-level theories, such as the theory of attentional inertia (e.g., Diamond & Kirkham, 2005; Kirkham et al., 2003), to account for many of these phenomena. In future work, it would be interesting to investigate the relationship between such process-level theories and computational level accounts of the biasing effects of prior knowledge, like the Bayesian model proposed here.

Additionally, the studies here do not tell us precisely what children learned. We have suggested that children can use domain-general evidence to learn at least one particular psychosomatic relationship: that being scared causes tummy aches. We also know that children did not learn to accept that psychosomatic causes were possible in general. However, children's learning might have been either more narrow or more broad than this summary suggests. Children might have learned that being scared can cause tummy aches only in the context of storybooks—or they might have begun, but not completed, a process of fundamentally altering their understanding of domain boundaries. Additional research might establish the extent to which patterns of evidence can influence children's naive theories.

Further research might also establish the extent to which children's learning is affected by varying the amount and type of evidence children observe. In our studies, children saw candidate causes covary deterministically with an effect seven times. Differences in the quantity, quality, and presentation of the data (more trials, negative evidence, evidence from interventions, probabilistic evidence, etc.) might influence both children's willingness to override domain-specific beliefs and their willingness to generalize from the data to novel events.

Note also that in this study children saw a single consistent pattern of evidence: Candidate causes were paired together, and one variable was always held constant while the other variable

always changed. In the real world, data are unlikely to be packaged in such a consistent manner. Causes often act stochastically, multiple variables can change simultaneously, and unobserved causes may be present. In such contexts, it may be far more difficult both to detect and to draw inferences from recurring variables. Furthermore, children might be better able to track evidence in the pedagogical context of a story than in the world at large; conversely, the inferences children make about events in a story might be particularly unlikely to generalize beyond the story itself. Further research might investigate children's ability to draw inferences from ambiguous evidence in a broader range of contexts and in cases where the presentation of the data is less tightly controlled.

Finally, in our studies, children observed the evidence in the absence of any explanation of how psychological events might cause bodily events. Many researchers have proposed that an understanding of causal mechanisms is fundamental to an understanding of causal relationships (Ahn, Kalish, Medin, & Gelman, 1995; Bullock, Gelman, & Baillargeon, 1982; Koslowski, 1996; Shultz, 1982) and it would be interesting to know how offering children explicit information about causal mechanisms might affect their learning. It seems probable that children might be more willing to learn from a combination of evidence and information about plausible processes of causal transmission than from evidence in isolation. Conversely, researchers have suggested that evidence about the covariation of interventions and outcomes can support inferences about causal mechanisms (Schulz, Kushnir, & Gopnik, 2007; Schulz & Sommerville, 2006). It would be interesting to know whether evidence for novel observed causal relations might prompt children to posit novel mechanistic explanations.

More generally, it seems probable that children begin to learn about psychosomatic causality, not merely because they observe or are told about covariations between psychological and bodily events but because adults explicitly assert the existence (or non-existence) of such causal relationships. We do not know how merely telling children about a causal relationship affects children's interpretation of evidence. Nor do we know how different cultural beliefs about psychosomatic causality might influence children's learning. Further research might investigate the interaction between information conveyed through cultural transmission and children's learning from evidence.

What these studies do suggest is that Bayesian inference captures a hallmark of causal learning in early childhood: conservatism with respect to prior knowledge but flexibility in the face of new evidence. Although learners might lack a metacognitive understanding of the relationship between theories and evidence, rational computations integrating new data and prior knowledge could form part of an implicit human learning mechanism, allowing the process of theory formation to be both adaptive and stable. The results of these studies suggest that even very young children can integrate prior knowledge and evidence to make normative causal inferences, giving children a powerful mechanism for developing and revising their naive theories about the world.

## References

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

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Barrett, S. E., Abdi, H., Murphy, G. L., & Gallagher, J. M. (1993). Theory-based correlations and their role in children's concepts. *Child Development, 64*, 1595–1616.
- Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946–964.
- Bloom, P. (2004). *Descartes' baby*. New York: Basic Books.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (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. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). New York: Cambridge University Press.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.
- Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive power. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227–253). Cambridge, MA: MIT Press.
- Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think. *Psychological Science, 16*, 291–297.
- Estes, D., Wellman, H. M., & Woolley, J. (1989). Children's understanding of mental phenomena. In H. Reese (Ed.), *Advances in child development and behavior* (pp. 41–86). New York: Academic Press.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60*(1, Serial No. 24).
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition, 38*, 213–244.
- Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind and Language, 3*, 197–217.
- Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1–31.
- Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.
- Gopnik, A., & Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science, 4*, 175–183.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*, 620–629.
- Griffiths, T. L. (2005). *Causes, coincidences, and theories*. Unpublished doctoral dissertation, Stanford University.
- Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*, 233–259.
- Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50*, 171–188.
- Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology, 37*, 668–683.
- Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind–body distinction. *Child Development, 64*, 1534–1549.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence* (A. P. S. Milgram, Trans.). New York: Basic Books.
- Kalish, C. (1996). Causes and symptoms in preschoolers' conceptions of illness. *Child Development, 67*, 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. P. D. Sperber (Ed.), *Causal cognition: A multidisciplinary debate*. New York: Clarendon Press/Oxford University Press.
- Kirkham, N., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–476.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.
- Klahr, D., & Dunbar, K. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109–143). Hillsdale, NJ: Erlbaum.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation isn't enough: The role of causal mechanism, sampling method and sample size in causal reasoning. *Child Development, 60*, 1316–1327.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674–689.
- Kuhn, D., Amsel, E., & O'Laughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology, 43*, 186–196.
- 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.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- 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*, 67–98.
- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.
- Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335–337.
- Notaro, P. C., Gelman, S. A., & Zimmerman, M. A. (2001). Biases in reasoning about the consequences of psychogenic bodily reactions: Domain boundaries in cognitive development. *Merrill-Palmer Quarterly, 48*, 427–449.
- Pazzani, M. J. (1991). A computational theory of learning causal relationships. *Cognitive Science, 15*, 401–424.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.
- Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development, 67*, 2709–2727.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Russell, J., Jarrold, C., & Potel, D. (1994). What makes strategic deception difficult—the deception or the strategy? *British Journal of Developmental Psychology, 12*, 301–314.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31–57.

- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and “theory of mind.” *Mind & Language*, *14*, 131–153.
- Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children’s understanding of the limits of psychological explanation. *Cognition*, *62*, 291–324.
- Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*, 162–176.
- Schulz, L. E., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Interventions and causal inference. In A. Gopnik & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation* (pp. 67–85). New York: Oxford University Press.
- Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers’ causal inferences. *Child Development*, *77*, 427–442.
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *37*(B), 1–21.
- Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 229–261). San Diego, CA: Academic Press.
- Shepard, R. N. (1987, September). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1, Serial No. 194).
- Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, *46*, 394–399.
- Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, *10*, 401–402.
- Sobel, D. M. (2004). Exploring the coherence of young children’s explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology*, *22*, 37–58.
- Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children’s causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303–333.
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.
- Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, *7*, 337–342.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction and search*. Cambridge, MA: MIT Press.
- Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 301–322). New York: Oxford University Press.
- Tenenbaum, J. B., & Niyogi, S. (2003). Learning causal laws. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Tenenbaum, J. B., Sobel, D. M., Griffiths, T. L., & Gopnik, A. (2007). *Bayesian inference in causal learning from ambiguous data: Evidence from adults and children*. Manuscript submitted for publication.
- Watson, J. S., & Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly*, *18*, 219–227.
- Wellman, H. M. (1990). *The child’s theory of mind*. Cambridge, MA: MIT Press.
- Wellman, H. M., & Estes, D. (1986). Early understanding of mental entities: A reexamination of childhood realism. *Child Development*, *57*, 910–923.
- Wellman, H. M., Hickling, A. K., & Schult, C. A. (1997). Young children’s psychological, physical, and biological explanations. In H. M. Wellman & K. Inagaki (Eds.), *The emergence of core domains of thought: Children’s reasoning about physical, psychological, and biological phenomena* (pp. 7–25). San Francisco: Jossey-Bass/Pfeiffer.
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*, 301–308.
- Zelazo, P. D., Frye, X., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

(Appendixes follow)



## Appendix A

## Bayesian Model

Bayesian inference provides a rational account of how children should go about combining theory and evidence. It can also be used to make precise quantitative predictions as to whether conclusions are justified by the observation of data, given a set of assumptions about the constraints provided by naive theories. In Experiment 1, children are asked, “Why does [character] have [symptom]? Is it because of [A] or because of [B]?” We model the probability that children choose explanation A as

$$\frac{P(\text{Explanation A}|D)}{P(\text{Explanation A}|D) + P(\text{Explanation B}|D)} \quad (2)$$

This directly contrasts the two possible explanations given the data observed (a similar expression applies in Experiment 2, contrasting the three possible explanations). The probability of each candidate explanation being selected given the data is computed by summing over all possible causal models that are consistent with the explanation. This is formalized as,

$$P(\text{Explanation A}|D) = \sum_{h \in H} P(\text{Explanation A}|h)P(h|D) \quad (3)$$

where  $h$  is a hypothesis as to the underlying causal structure, and  $H$  is the space of all hypotheses.

We represent hypotheses using causal graphical models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993), where nodes correspond to variables, arrows from cause to effect represent relationships between variables, and a set of conditional probability distributions captures the probability that each variable takes on a particular value given the values of its causes. We assume that the probability of a cause being selected as an explanation given a particular causal structure  $h$  is  $1/k$ , where  $k$  is the size of the set of candidate causes that are present and possess a causal relationship with the effect in  $h$ , and where the proposed explanation is a member of this set. The probability of a particular causal structure given the data is obtained via Bayes’s rule (Equation 1), using a prior  $P(h)$  and likelihood  $P(D|h)$  derived from a causal theory.

As proposed by Tenenbaum and Niyogi (2003), Griffiths (2005), and Tenenbaum et al. (2007), we model the framework theory that guides children’s inferences as a simple scheme for generating causal graphical models. In this scheme, we allow for different domains. Causal variables have relationships with effect variables; causes are likely to have relationships to effects within their domain, however, there is also a small probability that a cause from one domain can lead to an effect in another domain.

The prior probability associated with each model is simply its probability of being generated by the theory. The process of generating a causal graphical model from this theory breaks down into three steps. First, we identify the nodes (causes and effects) in

the model. In our case, the nodes simply correspond to the set of causes and effects that appear in the story. Second, we generate the causal relationships between these nodes. If cause and effects are within domain, then the probability a relationship exists is relatively high and given by  $p$ . If the link between two variables crosses domains, then a relationship is unlikely and is given a lower probability,  $q$ . With  $n$  causes, there are  $2^n$  possible causal models. Assuming that each relationship is generated independently, we can evaluate the prior probability of each of these models by multiplying the probabilities of the existence or non-existence of the causal relationships involved. The particular values of the probabilities  $p$  and  $q$  depend on the child’s theory. Such theories might change with age and experience; that is, younger children might think cross-domain events are more or less probable than older children. We assume that children think the probability of cross-domain events is low (but not extremely low) by setting  $q = .1$ , and by setting a higher within-domain probability  $p = .4$ .

Finally, we specify the conditional probability of the effect given the causes present in the causal model. This allows us to evaluate the probability of a specific model,  $h$ , generating the data observed on the  $m$ th day,  $P(d_m|h)$ . These data consist of the values taken on by all variables on that day—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 day. We then take the conditional probability of the effect given the set of causes to be 1 if any cause that influences the effect is present, and  $\epsilon$  otherwise, corresponding to a noisy-OR parameterization (Pearl, 2000), where each cause has a strength of 1 and the background has a strength of  $\epsilon$ . We assumed that the probability of an effect in the absence of any causes was low, with  $\epsilon = .001$ . The probability of the full set of data,  $D$ , accumulated over the course of the story is given by

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

where the data observed on each day are assumed to be generated independently.

As can be seen comparing the results predicted by the Bayesian model in Figure A1 with the 4-year-olds’ responses in Experiments 1 and 2, our model accurately predicted the responses of the oldest children, with a Pearson product–moment correlation coefficient of  $r(9) = .85$ . The model gives correct relative weights to the variables at baseline in both the within-domain and cross-domains conditions. Critically, the model predicted the increased A responses after evidence in all conditions, while still capturing the more subtle graded interaction between theory and evidence.

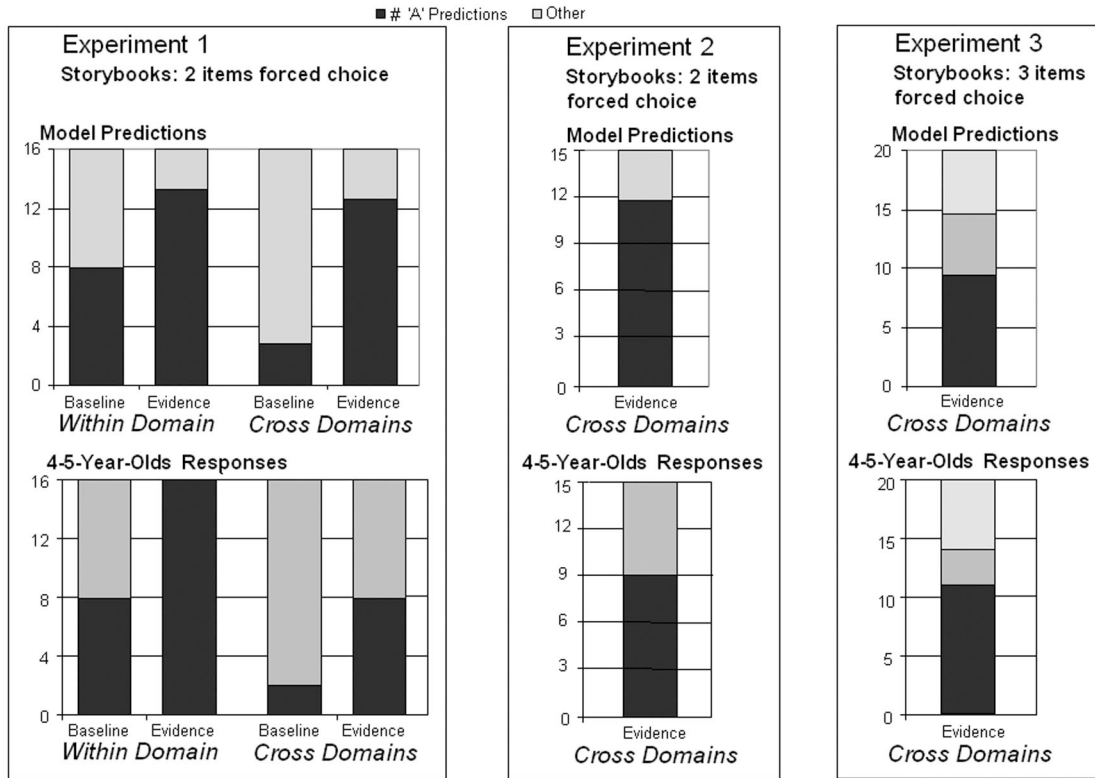


Figure A1. Predictions of the Bayesian inference model compared with 4- and 5-year-olds' responses to the storybook task in Experiments 1, 2, and 3, when  $q$  (cross-domain probability) = .1 and  $p$  (within-domain probability) = .4. The vertical axis shows the number of children selecting the different responses.

Appendix B

Text of Storybooks Used in Experiment 1

Text of Within-Domain Storybook Used in Experiment 1

Title: *Bambi's Adventures*

This is Bambi. Bambi likes to prance and run in lots of different places. Running is fun for Bambi. On Monday morning, Bambi runs in the pine grove. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs. On Monday afternoon, Bambi runs in the cedar trees and Bambi plays on the rope swing. Bambi feels great! Bambi doesn't have any itchy spots. On Tuesday morning, Bambi gets excited. Bambi runs in the cattails. Bambi runs in the grass. Bambi has itchy spots on his legs. On Tuesday afternoon, Bambi reads a book and Bambi runs through the rock bed. Bambi feels great! Bambi doesn't have any itchy spots. On Wednesday morning, Bambi runs in the marsh. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs. On Wednesday afternoon, Bambi runs through the apple orchard and Bambi plays with his toy truck. Bambi feels great!

Bambi doesn't have any itchy spots. On Thursday morning, Bambi gets excited. Bambi runs in the cattails. Bambi runs in the leaves. Bambi has itchy spots on his legs. On Thursday afternoon, Bambi plays jump rope and Bambi runs in the sand. Bambi feels great! Bambi doesn't have any itchy spots. On Friday morning, Bambi runs in the bushes. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs. On Friday afternoon, Bambi runs through the playground and Bambi roller skates. Bambi feels great! Bambi doesn't have any itchy spots. On Saturday morning, Bambi gets excited. Bambi runs in the cattails. Bambi runs in the grass. Bambi has itchy spots on his legs. On Saturday afternoon, Bambi gets his hair brushed and Bambi runs through the blueberry patch. Bambi feels great! Bambi doesn't have any itchy spots. On Sunday morning, Bambi runs through the garden. Bambi gets excited. Bambi runs in the cattails. Bambi has itchy spots on his legs. The next day Bambi's spots were all gone. Have fun Bambi! The End.

(Appendixes continue)

Text of Cross-Domains Storybook Used in Experiment 1

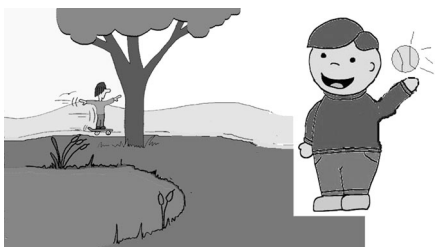
*Title: Bunny's Big Week*

This is Bunny. Bunny is scared because next week she has to give show-and-tell. Show-and-tell makes Bunny scared. On Monday morning, Bunny thinks about show-and-tell. Bunny feels scared. Bunny eats some cheese. Bunny has a tummy ache. On Monday afternoon, Bunny ties her shoes and Bunny eats strawberries. Bunny feels great! Bunny doesn't have a tummy ache. On Tuesday morning, Bunny eats a popsicle. Bunny thinks about show-and-tell. Bunny feels scared. Bunny has a tummy ache. On Tuesday afternoon, Bunny eats some toast and Bunny takes a bath. Bunny feels great! Bunny doesn't have a tummy ache. On Wednesday morning, Bunny thinks about show-and-tell. Bunny feels scared. Bunny eats French fries. Bunny has a tummy ache. On Wednesday afternoon, Bunny plays bingo and Bunny eats pasta. Bunny feels great! Bunny doesn't have a tummy ache. On

Thursday morning, Bunny eats a muffin. Bunny thinks about show-and-tell. Bunny feels scared. Bunny has a tummy ache. On Thursday afternoon, Bunny eats some yogurt and Bunny brushes her teeth. Bunny feels great! Bunny doesn't have a tummy ache. On Friday morning, Bunny thinks about show-and-tell. Bunny feels scared. Bunny eats some soup. Bunny has a tummy ache. On Friday afternoon, Bunny plays on the monkey bars and Bunny eats a banana. Bunny feels great! Bunny doesn't have a tummy ache. On Saturday morning, Bunny eats a carrot. Bunny thinks about show-and-tell. Bunny feels scared. Bunny has a tummy ache. On Saturday afternoon, Bunny eats some tofu and Bunny builds a snowman. Bunny feels great! Bunny doesn't have a tummy ache. On Sunday morning, Bunny thinks about show-and-tell. Bunny feels scared. Bunny has a tummy ache. The next day Bunny gave show-and-tell. She did very well and everyone clapped! Hurray for Bunny! The End.

Appendix C

Text and Pictures Used in the Far Transfer Study of Experiment 2



Physically possible: This is Alex. Alex is at the playground. Alex is throwing a ball. Alex is throwing the ball near a lake. Alex's friend tells Alex that if he throws the ball into the water it will make a big splash. Can that happen? Can Alex make a splash by throwing the ball into the water?



Physically impossible: This is Tony. Tony is at the park. Tony picks up a feather. Tony brushes the feather against a car window. Tony's friend tells Tony that if he keeps brushing the feather on the window, the window will break. Can that happen? Can Tony break the window with a feather?



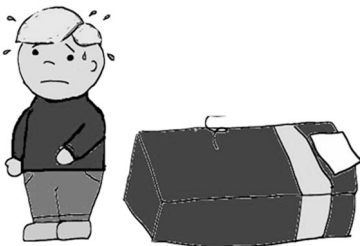
Biologically possible: This is Erin. Erin is in her backyard. Erin is jumping rope. Erin has been jumping rope for a long time. Erin's friend tells Erin that if she keeps jumping rope she will get very tired. Can that happen? Can Erin get tired from jumping rope for a long time?



Biologically impossible: This is Mel. Mel is in the garden. Mel is playing with the soil. Mel pats the soil with her hand. Mel's friend tells her that if she pats the soil a lot, the soil will sprout a tomato. Can that happen? Can Mel make a tomato grow by patting the soil?



Psychogenic headache: This is Leslie. Leslie is on the school bus. It's Leslie's first day of school today. Leslie is worried about the first day of school. Leslie worries and worries. Leslie's friend tells her that if she keeps worrying she'll get a headache. Can that happen? Can Leslie get a headache from worrying too much?



Psychogenic sickness: This is Jordan. Jordan is in his bedroom. Jordan is upset and nervous because he has to stay with a babysitter. Jordan feels very upset about the babysitter. Jordan's friend tells Jordan that if he keeps being upset and nervous he will start to feel sick. Can that happen? Can Jordan start to feel sick from being nervous and upset?

Received April 7, 2006  
 Revision received February 9, 2007  
 Accepted March 5, 2007 ■