Why Learning Can Be Hard: Preschooler's Causal Inferences

Elizabeth Baraff Bonawitz & Laura Schulz

Massachusetts Institute of Technology
Bldg. 46-4005, 77 Massachusetts Avenue
Cambridge, MA 02139
{liz_b, lschulz}@mit.edu

Abstract

Human intelligence has long inspired new benchmarks for research in artificial intelligence. However, recently, research in machine learning and AI has influenced research on children's learning. In particular, Bayesian frameworks capture hallmarks of children's causal reasoning: given causally ambiguous evidence, prior beliefs and data interact. However, we suggest that the rational frameworks that support rapid, accurate causal learning can actually lead children to generate and maintain incorrect beliefs. In this paper we present three studies demonstrating these surprising misunderstandings in children and show how these errors in fact reflect sophisticated inferences.

Ambiguity and Rational Models

Even if the causal structure of the world were deterministic, learning would not be easy. One trial may be sufficient to learn whether a causal relationship did or did not exist, and just one trial would be enough to overturn an incorrect belief, but only if the space of possible correct models was small. Unfortunately, data in the actual environment is not clear-cut. The space of possible models is often vast. And, accurate inferences of the actual causal models of the world are plagued by ambiguity in our observations.

Ambiguity arises for a number of reasons. Firstly, our perceptual machinery is not optimal, and observations may be noisy: we may not see a relevant event; we may think we see an event that did not actually happen; or, we may forget what we have perceived. Secondly, events frequently occur simultaneously. If the observer doesn't know which variables are a priori the relevant variables, then it becomes exceptionally difficult to resolve the numerous confounds. Related to this point is the fact that different causal events happen on different time scales; without knowing how much time should transpire between a cause and its outcome makes the confounding problem exponentially more complex. Finally, not all variables are observable. Thus, even given perfect perception without confounds and known time-scales, it may be difficult to be aware of the actual hidden causes or effects.

Despite these numerous limitations, advances in artificial intelligence have led to models that capture some of the inherent structure of the world. To do so, these models go beyond statistical covariation data and incorporate prior knowledge with evidence. One approach to thinking about how prior beliefs should interact with statistical evidence can be obtained by regarding causal learning as a problem of Bayesian inference. In Bayesian inference, the learner seeks to evaluate a hypothesis about the process that produced some observed data. The learner's a priori beliefs about the plausibility of the hypotheses are expressed in a "prior" probability distribution. The learner seeks to evaluate the "posterior" probability of the hypothesis – their beliefs about the plausibility of the hypothesis after taking into account the evidence provided by the data. The posterior distribution directly combines the evidence obtained, through the likelihood, with the learner's initial beliefs about the plausibility of the hypothesis, expressed in the prior. In the case of causal learning, we can imagine prior probabilities being supplied by a domain-specific theory, stipulating which causal structures are plausible (Tenenbaum, Griffiths, & Niyogi, 2007; Tenenbaum & Niyogi, 2003).

Powerful Causal Learners

A Bayesian framework captures both the robustness of structured representations and also the flexibility of bottom-up, statistical learning and offers a natural framework with which to consider how causal learning operates for human children. Indeed, many cognitive developmentalists have suggested that children's beliefs have important structural and dynamic similarities to these rational models: children have abstract coherent, causal representation of the world, but evidence plays a critical role in how those theories are formed and revised. (Gopnik, 1996; Carey & Spelke, 1996). In just the first

Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

few years of life, children acquire richly structured beliefs, can use those beliefs to make optimal choices for exploring the environment (e.g. Schulz & Bonawitz, 2007); can explain surprising events by appealing to unobserved variables (Schulz & Sommerville, 2006); and can make sophisticated inferences about causal events, even when the observed data is ambiguous (e.g. Gopnik et al., 2004).

However, sometimes children do not seem to be so rational. Indeed, sometimes children just get things wrong. For example, even though adults endorse psychosomatic illness, children believe that mental states are not causally connected to biological events; thus, they believe that things like being worried can't lead to tummy aches. Or children will believe that their toy blocks will always balance in the geometric center of the block, even when the block is unevenly weighted and will actually balance at the center of mass. Even more strikingly, children get it wrong in the face of evidence. They observe that "Daddy gets headaches whenever he worries, and Mommy's face turns red when she's mad," but nonetheless argue that mental states can't cause biological events, or they repeatedly attempt to balance an unevenly weighed block in the center, even though they observe that it continues to fall (Karmilof-Smith & Inhelder, 1974).

Some researchers have suggested that limitations in learning from evidence can be accounted for generally as information processing limitations. For example, children may have problems encoding the evidence: they may be easily distractible; they may not be sensitive to pragmatics in the way adults are presenting the data, etc. Additionally, some researchers have suggested that children have more impoverished strategies for storing information in long-term memory and later recalling it (e.g. Chi, 1976). Some researchers have even suggested that integrating new evidence with their prior beliefs may not be so natural for young learners -- or even for adults (Chen & Klahr, 1999; Inhelder & Piaget, 1958; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Masnick & Klahr, 2003). For instance, some research suggests that adults interpret identical evidence differently depending on whether the data supports or conflicts with a favored theory. Thus, if two candidate causes are both independent of an effect, learners will cite instances of co-occurrence as evidence for the variable consistent with their theories and instances of non-co-occurrence as evidence against the variable incommensurate with their theories (Kuhn, 1989).

Addressing children's surprising cognitive mistakes in these terms, however, neglects the explanatory power of computational accounts. It is important to note that we do not make quantitative predictions from a specific model; instead we suggest that Bayesian inference offers a framework with which to think about more qualitative explanations for children's quirky misunderstandings. A rational approach suggests that the mechanisms that help children rapidly learn and conservatively maintain beliefs could be the same mechanisms that lead to errors. In what follows, we present three research projects that support the claim that the surprising errors we see children making may actually be the result of a more optimal approach to learning.

Strong prior beliefs & cross domain reasoning

Considerable research has shown that children's causal reasoning respects domain boundaries (Carey, 1985; Estes, Wellman, & Woolley, 1989; Hatano & Inagaki, 1994; Wellman & Estes, 1986; Bloom, 2004). In particular, many researchers have suggested that children respect an ontological distinction between mental phenomena and bodily/physical phenomena, (Notaro, Gelman & Zimmerman, 2001). While adults accept that some events (e.g., psychosomatic phenomena) can cross the mental/physical divide, preschoolers typically deny that psychosomatic reactions are possible (e.g. they deny that feeling frustrated can cause a headache or that feeling embarrassed can make you blush). We were interested in how preschool children would interpret formal patterns of evidence suggesting the presence of a psychosomatic cause in light of a strong initial belief in domain boundaries.

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

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

Why did the three-and-a-half-year-olds respond differently to the *Cross Domains* evidence than the four- and five-year-olds? The differential treatment of evidence

by younger and older preschoolers may be interpreted in terms consistent with the Bayesian framework. Younger children might be modeled as having a stronger prior belief in domain boundaries than the older children. This may arise because children have less experience with psychosomatic events and thus believe them to be unlikely (Prior Belief Baserate account); or, younger children may have difficulty positing a mechanism by which domain-violating events (such as psychosomatic illness) are possible (Prior Belief Mechanism account). Thus, although a small amount of evidence might be needed to convince an older child of psychosomatic causality (given that older children have other experiences supporting such claims), more evidence may be needed to overturn younger children's resistance to believing in psychosomatic illness.

A second possibility is that three-year-old children might have difficulty making inferences from ambiguous statistical data. If the ability of the three-and-a-half-year-olds to interpret data of this complexity is fragile, any increase in task difficulty (including a conflict with prior knowledge) might compromise children's ability to evaluate the evidence. We'll call this the Information Processing account. Indeed, in Schulz, Bonawitz, & Griffiths (2007), a third group of younger 3-year-olds failed to learn from the evidence even in the *Within Domain* book, supporting the claim that younger children may have more difficultly with the task in general.

Finally, there's good reason to believe that both the prior belief and information processing accounts could be true. Consistent with the prior belief accounts, Notoro et al (2001) have demonstrated that with age, children begin to report psychogenic phenomenon as both physical and psychologically caused, supporting the claim that a belief in this domain boundary may lessen with age and experience. On the other hand, research on young children's causal reasoning has suggested that information processing demands can limit 3-year-olds performance even on tasks that only require reasoning within one domain of knowledge, (e.g. See Leslie, 1994) and therefore a context in which three-year-olds must integrate knowledge across multiple domains may be even more cognitively taxing. Thus, previous research supports the proposal that 3-and-a-half-year-old children may have both a stronger prior belief in domain boundaries and a more fragile ability to reason from ambiguous evidence.

Psychosomatic training study

In order to investigate these different accounts, we designed a two-week intervention for 3-and-a-half-year-old children who initially failed a *Cross Domains* pretest book (identical to the book used in Schulz et al., 2007). That is, we included children who rejected the statistically likely domain-violating candidate cause in favor of the statistically unlikely within-domain cause. Children were exposed to one of four different sets of training books. One set of books was designed to give children evidence that the baserate of psychosomatic events was likely, testing the Prior Belief Baserates account. The prior belief mechanism books taught children about possible mechanisms of psychosomatic illness, without giving them information about the baserate of psychosomatic events. The third set of books provided information processing training by giving children practice reasoning from ambiguous statistical evidence (AB→E; AC→E, AD→ E, etc.); however, these books always presented children with within-domain evidence, and the evidence never dealt with psychosomatic causality. Finally, a fourth set of non-training books were designed for a control group of children in which psychological variables were repeatedly mentioned but were not causally connected to biological outcomes.

We ended the training by reading children a final *Cross Domains* book (formally identical to the initial forced-choice book but involving different specific stimuli); however, there are important reasons to be cautious interpreting a shift in responses from food to worrying as a sign that children were learning. Research has shown that preschool children may vary responses when asked the same question twice (e.g. Poole & White, 1991), probably because children interpret the act of asking again as the interviewer's desire for alternative information (Memon, Cronin, Eaves, & Bull, 1993). Thus, after reading the *Cross Domains* book, we gave children a Free Explanation task as our final test book. The responses of 4-year-olds on this task in previous studies suggested that the explanation task is a reliable test of whether or not the 3-year-olds in our training study have genuinely learned about the plausibility of psychosomatic events.

If the Prior Belief Baserates account is correct, and three-and-a-half-year-olds' failure to update their beliefs with evidence is due to a strong prior belief that psychosomatic events are unlikely, then presenting children with evidence that psychogenic events are common should improve children's performance on the final explanation test book. However, if children fail to endorse psychosomatic events because they do not understand the causal mechanism, then training on the causal mechanisms should improve their performance on the free explanation book. If the Information Processing account is correct and children fail the belief revision task because of a fragile ability to reason from ambiguous evidence, then multiple exposures to this form of evidence could help children use the ambiguous data to revise their beliefs in a final *Cross Domains* book about psychosomatic events and extend that knowledge to the free explanation test book. Of course, it is possible that three-and-a-half-year-olds have *both* a fragile ability to interpret ambiguous evidence *and* a strong prior belief in domain boundaries. In that case, all three training studies should impact the children's performance as compared to the children in the *Control* training.

Results and Discussion of training study

Critically, the pattern of responses on the free explanation test book demonstrated a significant effect of the training conditions. Children's responses were coded as either appealing to the target psychological cause in the story (e.g. feeling nervous; thinking about school) or to external biological causes not mentioned in the story (e.g., eating too much food, "bumping his belly", or even, as one child notably put it, "has too much poo poo in him"). 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 and unambiguously into one of the two categories. As compared to the *Control* condition all three training conditions seemed to help children: significantly more children appealed to psychological explanations in the *Prior Belief Mechanisms* condition ($\chi2$ (1, N = 38) = 7.03, p < .01) and the *Information Processing* condition, ($\chi2$ (1, N = 38) = 5.58, p = .02), and marginally more children in the *Prior Belief Baserates* condition, ($\chi2$ (1, N = 38) = 3.14, p = .08). There were no significant differences between training conditions ($\chi2$ (2, N = 57) = .93, p = NS).

After only four twenty-minute training sessions over two weeks, three-year-olds showed an improvement in their ability to recognize causal variables that conflicted with their prior knowledge. This improvement was manifest across three quite different tasks, and yet children showed improvement relative to the *Control* condition in all three conditions. Perhaps the most striking finding is that three-year-olds, after receiving Training, were able to extend their inferences about psychosomatic illness to a free explanation transfer task. Children were, on average, three times more likely than children in the *Control* condition to endorse the possibility of psychosomatic illness, suggesting that children had genuinely revised their beliefs from the limited training. Given that children in the *Prior Belief Baserates* training were only marginally more likely than children in the *Control* condition to transfer their knowledge, this conclusion should be drawn with caution for this condition. Nonetheless, this study provides suggestive evidence that developmental changes in children's difficulty in learning from counterintuitive evidence may be both due to a fragile ability to integrate ambiguous evidence with prior knowledge in this task and due to initially stronger beliefs in domain boundaries.

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

It also raises some important questions for consideration, however. What leads to such a strong, initial belief in domain boundaries? Is this perhaps a rational bias to have because it helps the learner identify likely potential causes in ambiguous contexts? Indeed, if in the natural environment entities within a domain are more likely to be causally connected than entities across domains, this prior could bootstrap correct causal inferences for young learners.

Simplicity as a means for causal inference

Despite the growing literature demonstrating children's expertise in appealing to and generating accurate causal explanations, there is relatively little research exploring how children choose between multiple explanations. Previous work by Lombrozo (2007) has demonstrated that adults are sensitive to the simplicity of competing causal explanations, where simplicity is quantified as the number of causes invoked in an explanation. As suggested above, this principle (known as Ockham's Razor) may provide a basis by which explanations can be evaluated. Additionally, data from Lombrozo's work suggests that simpler explanations are assigned a higher prior probability, with the consequence that disproportionate probabilistic evidence is required before a complex explanation will be favored over a simpler alternative. While these findings suggested that adults use simplicity as a basis for evaluating explanations, no work had investigated whether children are sensitive to these same principles of simplicity and probability.

In a follow-up experiment, Bonawitz and Lombrozo (*in review*) directly examined whether and how children employ simplicity in a task formally similar to Lombrozo's original work, but adapted to create a more interactive procedure that would capture the interest of and be understandable to preschool children. To accomplish this, we designed a toy in which we could independently vary the simplicity and probability of candidate explanations. Children saw a device with colored chips that generated one or two effects: the activation of a light and of a fan.

If anything, children have a greater need than adults for sophisticated domain-general strategies for evaluating explanations under uncertainty. This would suggest that children should mirror adults in using simplicity as a constraint on inductive judgments. One possibility is that children will actually rely on simplicity more than adults do, and prefer simpler explanations regardless of their relative probabilities. This outcome would also be predicted if the basis for adults' reliance on simplicity comes from cognitive constraints, such as limited working memory. But if Ockam's razor is to help rather than hinder learning, one might predict that children will mirror adults in integrating a preference for simplicity with probabilistic evidence. Under this reasoning, children should show a preference for simpler explanations, but should also demonstrate more nuanced responses, favoring the more complex explanation as the probability evidence becomes more compelling.

A final possibility is that presenting young children with this kind of intuitively conflicting information will be overwhelming, and children will be unable to successfully integrate a preference for simplicity with probabilistic evidence. If this is the case, children may respond on the basis of a single dimension (choosing either probability or simplicity at chance), or select among the response options at random.

Procedure: Children were introduced to the toy box, the red, green, and blue chips, and the activator bucket. Children were shown that when a red chip was placed in the activator bucket, the bottom globe lit-up and spun around. Children were then shown that when the green chip was placed in the activator bucket, the top fan lit-up and spun around. Finally children were shown that the blue chip made both the bottom and top parts of the toy spin-around and light up. In all cases, the toy was actually activated surreptitiously by the experimenter. Pilot testing confirmed that the illusion of the chips causing the activation was so strong that even adults believed the chips were somehow causing the toy to work and were not aware that the experimenter was actually controlling the toy.

Next the experimenter took out the clear container and asked the children to help count out chips into the container. In all conditions only 1 blue chip was added to the container. To manipulate the probability of the more complex explanation (red & green), we varied the number of red and green chips such that there were 3 of each (red & green) in the 1:1 condition, 6 of each in the 2:1 Condition, 12 of each in the 4:1 Condition, and 18 of each in the 6:1 Condition.

After mixing the chips in the container in front of the child, the experimenter got out the opaque rigid bag and said: "Now I'm going to put all my chips into this bag." After pouring the chips into the bag, the experimenter sat the bag on top of the container on top of the toy and then 'accidentally' knocked the bag on its side, so that the opening fell in and towards the 'activator bucket' and away from the child. The bag fell such that the child could not see or hear what chips actually fell into the activator. As soon as the bag fell, the top and bottom portions of the toy activated. During the 'accidental' fall, the experimenter exclaimed "Oops! I knocked my bag over! I think one or two chips may have fallen into my toy! What do you think fell into the toy?" Children's explanations were recorded. [1]

Results and Discussion of Simplicity Experiment

Children's responses fell unambiguously and uniquely into one of three categories: simple (blue chip only), complex (red & green chips only), or other (red only, green only, both blue and red, both blue and green, or one of each color). Collapsing across all four conditions, children were no more likely to choose the simple explanation over the complex explanation ($\chi^2(1) = 1.17$, *n.s.*). Critically, the distribution of explanations differed significantly as a function of condition ($\chi^2(6) = 20.25$, $p < .01$). As the complex explanation became more likely, children became increasingly likely to select it over the simpler alternative (see Figure 1). Before interpreting this trend in relation to simplicity and probability, it's worth rejecting an alternative explanation for this finding, namely that as the total number of chips involved in the task increased, children simply became more inclined to cite explanations involving multiple chips. Were this the case, the total number of explanations citing more than one chip should differ across conditions. This was not the case ($\chi^2(3) = 5.69$, *n.s*).

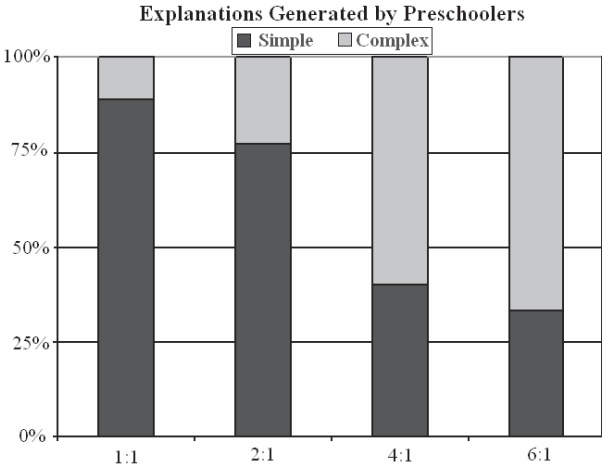

Figure 1: Percent simple and complex explanations given by preschoolers as probability of complex explanations increases over condition.

The fact that a greater number of children selected the complex explanation as it became more likely suggests that children are sensitive to probability in evaluating competing causal explanations. Yet when these explanations were equally likely (in the 1:1 condition), significantly more children selected the simple explanation ($\chi^2(1) = 7.58$, $p < .01$), suggesting that all else being equal, children prefer simple explanations. Moreover, these explanatory virtues competed, with children continuing to select the simpler explanation even when it was less likely to be true. Even in the 6:1 condition, over 30% of children chose the simple explanation.

In sum, Bonawitz & Lombrozo's findings demonstrate that children are not only sensitive to simplicity and probability in evaluating competing causal explanations, but that as a group they can integrate these bases for evaluation. The data provide two reasons to believe that children treat simplicity as a probabilistic cue that constrains inductive inferences. First, as a group children continued to value simplicity even when the complex explanation was more probable. This suggests that simplicity is not merely a basis for evaluating explanations in the absence of probabilistic evidence; rather, simplicity

[1] Children were given a memory check which made sure that children were aware that both the blue chip and the red and green chips combined could cause the toy to activate both the top and bottom. The probability ratios for each condition were computed by assuming that accidentally tipping the bag was equally likely to result in one or two chips and that the probability of a second chip being of a given color was affected only by the fact that the first chip decreased the pool of available chips of that color by one (i.e. the chips were otherwise conditionally independent).

is treated as commensurate with baserate information. Second, children were asked which chips they thought fell into the toy. Unlike Lombrozo (in press), which asked adults to rate the quality of explanations, this prompt is explicitly about the true state of affairs. That children used simplicity as a basis for inferring a property of the world suggests that simplicity is not regarded as a desired but undiagnostic property, but rather as a legitimate basis for inference.

Why might children privilege simpler explanations? One possibility is that simpler explanations are less demanding. They involve fewer cognitive resources, and are thus easier to generate and reason about. While an explanation of this kind is possible, it seems unlikely on the basis of the data. Children had no problem generating complex explanations in the 6:1 condition, and a comparison of these data to Lombrozo (2007) suggests that children are not more simplicity-prone than adults, as one might expect were cognitive limitations the basis for the preference.

Might a strong bias for simpler explanations be adaptively useful? There are a few reasons to think such a bias could be beneficial, even if it results in 'errors' when over-extended. First, simpler explanations are in many cases more likely than alternatives. When the base rates of all potential causes are equal (and less than 50%), the joint probability of causes happening together will be smaller than the probability of only one cause. In fact, the causes appearing in a complex explanation could be considerably more common and still be less likely in conjunction than the single cause in a simple explanation. Given this regularity, children may have learned that simplicity is a reliable heuristic for probability. And because probability information is often unavailable, such a heuristic could be widely applicable and often correct (or at least unfalsified). Second, a preference for simpler explanations has been defended on methodological grounds. Simpler explanations tend to rule out a greater number of possible observations, which means that they are easier to falsify. Third, arguments from statistics and computer science suggest that complex explanations run the risk of fitting "noise" or idiosyncratic properties of the possible observations one has sampled. To the extent that simpler explanations avoid these dangers, they will generalize more effectively to future situations.

Conserving beliefs facing surprising evidence

Given that a learner may sometimes make mistakes, ideally, belief revision should be at once flexible (to permit learning) and conservative (to prevent misleading data from overturning strongly held beliefs). Identifying and exploring surprising evidence may not be for the sake of overturning beliefs but rather to discover the potentially present hidden variables that can account for the surprising observations. Thus, belief revision is rationally not necessary when alternative variables can be used to 'explain away' data that conflicts with one's beliefs.

Previous research has demonstrated that children are remarkably good at generating explanations that appeal to alternative variables in order to maintain their beliefs. For example, Schulz and Sommerville (2006) showed that when given the choice, children appealed to an alternative variable that may have accounted for the stochastic way a switch caused a box to light rather than overturn their strong prior beliefs in determinism. Goodman et al (2006) tested children with a false belief task and found that when children were presented with outcomes that were surprising with respect to their theories, they were more likely to appeal to alternative causes than in cases when the evidence confirmed their theories. And Legare et al (in press) found that children were more likely to generate explanations about surprising events rather than unsurprising events.

However, no previous studies have explored whether children's own spontaneous play can lead to the discovery of alternative variables given surprising evidence. Rational frameworks predict that if children's exploration leads to the discovery of a hidden variable, then children should explain the events in terms of that variable in order to maintain beliefs; by contrast, in cases where there are no hidden variables, children should revise their beliefs.

Getting ahead with a theory of balance

In order to test these predictions, we needed to find a domain where children had strong, yet flexible beliefs. Because the development of children's theories has been well established in the domain of balancing blocks, this domain is particularly conducive for investigating the relationship between children's folk theories and their exploratory play and learning. In a seminal study, Karmiloff-Smith & Inhelder (1974) looked at children's understanding of balance between the ages of 4 and 9 years of age. They demonstrated that between 6 and 8 years, children first entertain a "Center Theory", believing that regardless of the center of mass, an object should be balanced at its geometric center. Center Theorists repeatedly attempt to balance unevenly weighted blocks at their geometric center. Gradually, children develop the correct, adult theory of balance: "Mass Theory". Mass Theorists, who are slightly older on average than Center Theorists, understand that in order for a block to be stable, it must be balanced over its center of mass. Children's understanding of balance has subsequently been investigated by many researchers (e.g. Halford, 2002; Janson, 2002; Normandeau, 1989; Siegler, 1976). However, much of this literature focuses on the transition between incorrect and correct rules and strategies and not on the processes, like exploratory play, that might generate the evidence that could support such discoveries.

To a Center Theorist, a block with a conspicuously heavy side balancing on its geometric center may not be surprising; however, this evidence should surprise a Mass Theorist. Conversely, to a Center Theorist, a block with

one heavy side balancing under its center of mass might be surprising, but that evidence should not surprise a Mass Theorist. Rational frameworks generate the following predictions. Firstly, children observing surprising evidence should be more likely to play with the block than children observing unsurprising evidence. Secondly, children in the surprising evidence condition should be more likely to appeal to an alternative variable to explain away the surprising data than children in the unsurprising condition. Thirdly, children who cannot explain conflicting data should be more likely to revise their beliefs.

To investigate how children's theories affect their exploratory play, we used a method similar to the free play paradigm of Schulz and Bonawitz (2007), (Bonawitz, Lim, & Schulz, 2007). We first presented children with a balance and a set of blocks. Then we showed children that the block balanced either according to their theoretical beliefs (unsurprising evidence condition), or in conflict with their beliefs (surprising evidence condition) and then let them choose to play freely with either the balancing block (the familiar toy) or a peg and ring toy (the novel toy). After a minute of free play, we returned the block to its initial balancing state and asked children what made it balance. Lastly, we gave children a new block to make a final balance attempt with. We predicted that children who observe identical evidence but have different theories will show different patterns of exploratory play, will appeal to different explanatory variables, and will differentially update their beliefs by the final balance.

Results and Discussion of Balance Study

Overall, children were more likely to explore the familiar toy (the block) when the evidence conflicted with their theories than when it confirmed their theories. To compare the amount of time playing with the blocks, we ran a two-way-between subjects ANOVA with theory and type of evidence as the between subjects variables and time spent playing with the blocks as the dependent measure. Comparisons between conditions revealed no main effect of theory (averaging across the two conditions, Center Theorists and Mass Theorists played for equal amounts of time) and no main effect of evidence type (averaging across the two conditions by theory, children who saw the block balancing at the geometric center played as long as children who saw the block balancing at the center of mass). However, comparisons revealed a significant interaction: children spent more time playing with the block when the evidence conflicted with their theories than when the evidence confirmed their theories.

Recall that if children observe evidence that goes against their initial beliefs they may respond by trying to 'explain away' the evidence (e.g., by positing the presence of a hidden variable) or by revising their beliefs. As it happens, a hidden variable was present – in all conditions, a magnet was always under the block at the point of contact. All the children discovered the magnet during the course of free play. And of course the magnet is, in all cases, a sufficient explanation for the block sticking to the platform. Nonetheless, children might be more likely to appeal to the magnet as an explanatory variable when evidence conflicts with rather than confirms their theories.

In our original study, overall, children in the conflicting condition were significantly more likely than children in the confirming condition to appeal to the magnet as an explanatory variable (χ^2 (1, N = 62) = 4.13, $p < .05$). However, this effect was driven by the older Mass Theorists. We tested a second group of Center Theorists, removing some of the cognitive load of the experiment by removing the free play and adding a familiarization trial at the beginning to help them think about generating explanations. Comparing this group of Center Theorists to our Mass Theorists, we found a significant interaction of explanation by theory by condition. While 67% of Mass Theorists appealed to the magnet in the condition that conflicted with their predictions, only 8% of the Center Theorists appealed the magnet, observing the same data (that is consistent with the center theory). In contrast, when the block was balanced over the center of mass (conflicting with the Center Theorist's predictions, but consistent with the Mass Theorists), the pattern of results reversed. Only 20% of Mass Theorists appealed to the magnet as an explanatory variable, where-as more than half the Center Theorists did.

As our final measure, we looked at the degree to which Center Theorists learned during the course of the experiment and revised their predictions for the final test trials. Note that there are two ways Center Theorist children could get conflicting evidence: the block failing to balance in the expected location (over the geometric center), or the block successfully balancing in the unexpected location (over the center of mass). Of course, in this case, the successful balance could be explained away by the presence of the magnet.

We compared the learning of children who only had evidence they could explain away with that of children who only had evidence they cannot explain away. To do so, we looked at the final balance attempts of our initial group of Center Theorists who saw the block balance in the unexpected location but who never tried to balance the block in the middle (8/16). All of these children could explain unexpected balances by appealing to magnet. We also tested a new group of sixteen Center Theorists who only received evidence they could not explain away; that is, they saw the same evidence as the first group, but critically the magnet was not present in the play block and thus could not be used as an alternative variable to explain away the surprising evidence.

Children in both groups were given a novel block at the end of the play period and were asked to balance it. When the magnet was present during the play period (and thus the counter-evidence could be explained away), only 1/8 Center Theorist 'switched' and tried to balance the final block over the center of mass. However, when the magnet was absent, 10/16 Center Theorists 'switched' and tried to balance the final block over the center of mass. In fact, the one child in the explaining away condition who did change

his theory successfully balanced the block upside-down over the center of mass, thus generating conflicting evidence that could not be accounted for by the magnet.

Children were significantly more likely to revise their balance attempts when the evidence generated during the play period could not be explained away than when it could. This suggests that children genuinely learn from exploratory play, but also that children are sensitive to alternative variables that create noise in observations. Thus, the sometimes slow transition from one theory to another might be partially accounted for by children's ability to explain surprising data by spontaneously appealing to auxiliary hypotheses.

Discussion

In this paper we have tried to argue that approaches in artificial intelligence have helped inspire developmentalists in thinking about the problems human children encounter in learning. Bayesian frameworks in particular suggest that the problem being solved is inferring the causal structure that generates the evidence; to solve this problem, statistical likelihoods and prior beliefs are integrated. We have demonstrated that this framework naturally captures how human learners balance conservativism while flexibly integrating evidence, and rapidly, yet accurately, generate causal knowledge of the world.

In a sense, these experiments also exhibit the adage that 'there is no such thing as a free lunch'. The same mechanisms that are in place to help children rapidly develop and conservatively maintain beliefs (given ambiguous data) may be the same mechanisms that lead to errors. Therefore, we hope that by exploring the cases when children have erroneous beliefs has informed us about mechanisms that in fact facilitate learning.

References

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

Bonawitz, E.B., & Lombrozo, T. (in review) The role of simplicity in preschooler's and adult's causal explanations.

Bonawitz, E.B., Fischer, A., & Schulz, L.E. (2008) Training a Bayesian: Three-and-a-half-year-olds' Reasoning about ambiguous evidence. Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society. DC, USA.

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

Bonawitz, E.B., Lim, S., & Schulz, L.E. (2007) Weighing the Evidence: Children's theories of Balance affect play. Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society. Nashville.

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

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

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? Memory & Cognition, 4: 559-572.

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

Estes, D., Wellman, H. M., & Woolley, J. (1989). Children's understanding of mental phenomena. In H. Reese (Ed.), Advances in child development and behavior (pp. 41-86). NY: Academic Press.

Goodman, N.D. et al. (2006) Intuitive Theories of Mind: A Rational Approach to False Belief. Proc. 28th Annual Conference of the Cognitive Science Society. Vancouver, Canada.

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

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

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

Halford, G.S.,et al. (2002) Young children's performance on the balance scale: The influence of relational complexity. Journal of Experimental Child Pscychology, 81, 417-445.

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

Jansen, B.R.J., & van der Maas, H.L.J. (2002) the development of children's rule use on the balance scale task. Journal of Experimental Child Psychology, 81, 383-416.

Karmiloff-Smith, A., & Inhelder, B. 1974, If you want to get ahead, get a theory. Cognition, 3(3), p. 195-212.

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

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

Lombrozo, T. (2007). Simplicity and probability in causal explanation. Cognitive Psychology, 55(3), 232-257

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

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

Schulz, L., & Bonawitz, E.B. (2007) Serious fun: Preschoolers play more when evidence is confounded. Developmental Psychology, Jul Vol 43(4) 1045-1050.

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

Schulz, L. E. & Sommerville, J. (2006). God does not play dice: Causal determinism and children's inferences about unobserved causes. Child Development, 77 (8), 427-442.

Siegler, R.S. (1976) Three aspects of cognitive development. Cognitive Psychology, 8, 481-520.

Tenenbaum, J.B., Griffiths, T. L., and Niyogi, S. (2007). Intuitive theories as grammars for causal inferences. Gopnik, A., & Schulz, L. (Eds.), Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Tenenbaum, J.B. and Niyogi, S. (2003). Learning Causal Laws. 25th Annual Conference of the Cognitive Science Society..

Wellman & Estes, 1986; Carey, S., & Spelke, E.S. (1996) Science and core knowledge. Philosophy of Science, 63(4), 515-533.