The Rational Child:

Theories and Evidence in Prediction, Exploration, and Explanation

By

Elizabeth R. Bonawitz

B.S., Northeastern University, 2002

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2009

Author:_____
Department of Brain and Cognitive Sciences
April 21, 2009

Certified by:_____
Laura Schulz
Assistant Professor of Cognitive Science
Thesis Supervisor

Accepted by:_____
Earl Miller
Chairman, Department Committee on Graduate Students

<div align="center">

The Rational Child:

Theories and evidence in Prediction, Exploration, and Explanation

by

Elizabeth R. Bonawitz

Submitted to the Department of Brain and Cognitive Sciences
On April 21, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

</div>

Abstract

In this thesis, rational Bayesian models and the Theory-theory are bridged to explore ways in which children can be described as Bayesian scientists. I investigate what it means for children to take a rational approach to processes that support learning. In particular, I present empirical studies that show children making rational predictions, exploration, and explanations. I test the claim that differences in prior beliefs or changes in the observed evidence should affect these behaviors. The studies presented in this thesis encompass two manipulations: in some conditions, children's prior beliefs are equal, but the patterns of evidence are varied; in other conditions, children observe identical evidence but children's prior beliefs are varied. I incorporate an additional approach in this thesis, testing children within a variety of domains, tapping into their intuitive theories of biological kinds, psychosomatic illness, balance, and physical systems. Chapter One introduces the problem. Chapter Two explores how evidence and children's strong beliefs about biological events and psychosomatic illness influence their forced-choice explanations in a story-book task. Chapter Three presents a training study to further investigate the developmental differences discussed in Chapter Two. Chapter Four looks at how children's strong differential beliefs of balance interact with evidence to affect their predictions, play, explanations, and learning. Chapter Five looks at children's exploratory play with a jack-in-the-box, (where children don't have strong, differential beliefs), given different patterns of evidence. Chapter Six investigates children's explanations following theory-neutral evidence about a mechanical toy. Chapter Seven concludes the thesis. The following chapters will suggest that frameworks combining evidence and theories capture children's causal learning about the world.

Thesis Supervisor: Laura Schulz
Title: Assistant Professor of Cognitive Science

One day, some 25 years ago, my mother-in-law found her four-year-old son distraught over a fish tank that was filled with shaving cream. A naturally curious and helpful child, Keith had been observing his mother taking care of the fish for sometime and wanted to know how to assist. Despite being sketchy on the details, when Keith discovered that the fish tank looked a bit murky, he was determined to help clean it. He had remembered his mother explaining that charcoal was important for cleaning the tank (though he couldn't quite recall the details of the process, such as there being a specific kind of *activated* charcoal that was added to the *filter*.) He ran to the basement (where dad kept his *grilling* charcoal), grabbed the bag, returned to the fish tank, and dumped a hefty amount directly into the water. (It was quite dirty; the more charcoal the cleaner it would be!) He was rather surprised when the tank water immediately turned black. *Hmm,* he thought, *I'm certain that mom said she added charcoal, but maybe there was something else I need to add as well. Maybe if I add something white, like shaving cream, it will counter-act the black water!* Shaving cream also seemed like an especially good hypothesis for cleaning the tank because it was kept in the bathroom cabinet which contained all sorts of other people-safe cleaning supplies. When the first dollop of cream didn't seem to neutralize the charcoal, Keith decided to empty the majority of the can in (he had used a lot of charcoal and besides, it was fun to spray.) As it became apparent that the shaving cream wasn't working and the fish were no longer moving, panic set-in; it was only moments later that mom arrived on the scene.

Preface

Children are peculiar creatures, aliens that have stumbled into our otherwise rational existence. Despite their tiny stature, they have unnatural powers over us (note the humiliating acts a parent will perform in public simply to elicit a small giggle or smile). Children seem to predict the impossible, act without purpose, and generate explanations that seem like nonsense. They dump charcoal and shaving cream into a fish tank in order to clean it and are surprised and upset when the result is less than effective.

Yet, as we peer at them through our scientist lenses, these extraterrestrials begin to look familiar. We see glimpses of ourselves in them and their actions begin to take purpose. Are these aliens are in fact young scientists, making interventions and testing their beliefs, delighting in their exploration as scientists delight in new data? They generate explanations about the events that happen in our universe, forming causal theories (as unworldly as they may be). They balance their theories with evidence as a scientist must, with a skeptical eye towards data that conflicts with strongly held beliefs but a willingness to revise those beliefs when evidence is compelling. As we focus more closely we see that perhaps it is not just that children are scientists, but rather that science is possible because we were once children.

Of course, the comparison illuminates some differences between child and scientist. Children are not formally taught the rigors of science as adults in academic settings are. They cannot articulate why careful controls need to be taken to test a hypothesis, nor can they invent the correct controls to take. They do not write-up their findings and present them at conferences to other child-scientists. (Though, watching children on a playground sometimes generates this illusion; especially when a particularly interesting insect, toy, or new game has been discovered.) Overall, children do not seem to be meta-cognitively aware of their process of hypothesis testing and evidence collecting as good scientists must. Nonetheless, I will suggest that we are studying many of the same processes of scientific discovery when we study children's causal learning.

Any psychologist would agree that testing children poses significantly greater challenges than testing adult populations. As participants, children are hard to come by: parents must be solicited, more stringent IRB protocols must be in place, and daycares and classrooms must be interrupted for testing. Children are impatient: studies must be short, interesting, and involve unique forms of compensation. Children are difficult to communicate with: unlike adults who can articulate their predictions, beliefs, and desires, children's cannot. We must therefore use creative measures to investigate what the child knows. Children are inexperienced: the experimenter can take for granted the numerous expectations that adults bring to the testing room, but more detail must be presented to children. This adds additional memory and attentional demands on a task that would otherwise be simple. Children are also shy: while almost any experimenter can sit an adult participant down at a computer and tell them to press the "start experiment" button, children must be approached by a charismatic and articulate actress who can reassure and interest the child throughout the experimental script. So, why bother with children when we can more easily test their older counter-parts?

Despite the various drawbacks, children are an important population to test for numerous reasons. First, children seem to learn about the world at an amazing rate; thus, children may be a good place to look for early developing and sophisticated learning mechanisms. Second, there may be genuine developmental differences and critical periods in learning; those differences can only be explored and explained with comparative populations. Third, unlike adults who often have strong prior theoretical commitments, children have fewer biases. It is easier to control for children's beliefs in experiments. Relatedly, children often have incorrect early beliefs about the world. It is easier to track down the origins of these incorrect beliefs in a population with limited experiences. Finally, children are fun and utterly charming. When adult data goes poorly we find

ourselves frustrated or defeated, but when children provide responses that surprise and confuse us, we laugh and are instead more inspired to pursue science.

Besides drawing inspiration from children, this thesis is influenced by approaches in rational (or sometimes called computational, mathematical, Bayesian[1]) modeling. There are several reasons why a rational modeling approach is appealing. First, rational models have precedent generating interesting predictions of, descriptions of, and explanations for human adults' reasoning. The success of these approaches in describing adult causal induction suggests that we may find similar benefits from models of children's causal inductions. Secondly, formal models force clarity by providing a framework and language with which to consider how reasoning is taking place. In rational frameworks, the modeler must define the problem that the child is solving, identify criteria for solving the problem, evaluate the space of all solutions given the criteria, and choose the best solution. As a result, rational models are easy to falsify because there is only one solution that is 'rational', but many alternative solutions that are not. In a sense, this makes rational models 'simpler' than alternative theories and thus aesthetically and perhaps formally more appealing (e.g. see Bayesian Occam's Razor: Jeffreys & Berger, 1992). Of course, thinking of the learner as rational is also intuitively appealing. How else would learning be simultaneously fast yet accurate and flexible yet conservative? The mind must be approximating some kind of rational process to arrive at a correct representation of the world as often as it does.

In this thesis, I will suggest that the child is both a scientist and a 'rational learner'. If the child is like the scientist, balancing theories and evidence in learning, then a rational model must include a formula for considering how beliefs should be updated as evidence accumulates. A Bayesian framework provides a formula for the interaction of theories and evidence. Here, I use this framework to describe how theories and evidence interact to affect predictions, exploration, and explanations. The framework will make empirical predictions about the kinds of inferences, acts of exploration, and explanations that children should generate, based on their current beliefs and the evidence observed. It will provide a basis to think about the evidence required to change a child's incorrect beliefs, suggesting training that may help children learn more effectively. This integrated approach may even clarify why it's not so surprising to come home and find the fish tank filled with shaving cream.

---

[1] These terms have different meanings and take different names depending on the field of study and the philosophical traditions of a department. Bayesian frameworks, for example, are often considered a subset of other modeling frameworks, sometimes more generally described as rational or computational. I choose to use 'rational' as the descriptor because it implies weighing alternative hypothesis with respect to a goal (which I will argue children are doing in prediction, exploration, and explanation) and also for aesthetic reasons.

Acknowledgments

Before I knew where I would go to graduate school, I knew who I wanted to work with. Laura Schulz took a chance on me, by providing me with a home at MIT before she had the chance to establish one for herself. Despite the challenges that moving and motherhood bring, Laura's contagious enthusiasm and support have always been abundantly felt. She has the rare ability to advise with the joyful enthusiasm of a faculty training her first student, but also advise with the wisdom and patience of one who has seen dozens graduate. She turns our common disagreements (and bets) into learning opportunities when I am wrong, and cedes with grace and thoughtfulness on the (rare) occasions when I am right. Her contributions pervade this thesis; in particular, she collaborated on the design and writing on Chapters Two, Three, Four, and Five. Her contributions have also pervaded my life; in particular, she has helped me to negotiate the challenging and surprisingly delicate social world of science.

Josh Tenenbaum's influence on my thesis and career has also been immense. As an undergraduate, I had the intuition that psychology should be more than surprising results; it should be about models that reverse engineer the mind. But, it wasn't until I met Josh and the clarity of his approach that I began to see how hard questions may begin to have tangible answers. Josh also took a chance on me, by hiring me as a dewy-eyed twenty-one-year-old to help coordinate his lab and run his behavioral research programs. First as his research assistant,

and later as his 'satellite' graduate student, I watched his lab grow from just a handful of us, to dozens and dozens who were similarly inspired by his brilliance and clarity of ideas. Though Josh did not directly collaborate on any of the projects presented here, his influence is in the very foundation of my approach. Besides his theoretical contribution, Josh's sense of humor, lab retreats, music making, and concern for his students have played no small part in establishing a critical base of friends and support here at MIT. Moreover, the high expectations he has of his students have helped me realize strength and drive in myself that I did not know I possessed.

Susan Carey has provided detailed helpful comments and feedback on this thesis and on my work in general; her influence on my thesis is also apparent in every mention of the theory-theory. Moreover, Susan offered me my first position in a development lab while I was working with Josh, which inspired me to think about how cognitive development and computational cognitive science might be bridged. Rebecca Saxe has provided me with copious feedback on the thesis and on my research in general. I have also known Rebecca since my lab coordinator days at MIT and Harvard. She always treated me and my ideas with respect, which gave me the confidence in myself to pursue these questions in graduate school. Tom Griffiths and Tania Lombrozo have also helped shaped the projects presented in this thesis. Tom collaborated on the project discussed in Chapter Two, and Tania was my co-author on the research presented in Chapter Six. Besides their friendship, Tom and Tania's theoretical contributions are great. Tom has the rare gift of taking hard problems, which lack a clear answer, and making the solution seem simple, while Tania taught me that seemingly simple questions are often quite a bit more difficult and interesting that originally perceived.

My first introduction to cognitive psychology came from John Coley, who gave me great scientific freedom when I was an undergraduate in his lab and provided continued support from across the river when I was a graduate student. Fei Xu first introduced me to cognitive development and literally first introduced me to Josh; she also suggested an important control condition reported in Chapter 2. Patrick Winston continues to provide me with the Vision, Steps, News, and Contributions of AI. Liz Spelke's enthusiastic support and feedback has inspired numerous control conditions in various projects. Additionally, conversations with members of the McDonnell collaborative have continued to inspire my work; in particular, thanks to Tamar Kushnir, Susan Gelman, Henry Wellman, Allison Gopnik, David Danks, John Woodward, Clark Glymore, and Christine Legare for providing feedback on various projects over the years.

Since first coming to MIT, I have seen two waves of friends. The first wave was here when I arrived and has since gone on to become influential faculty and researchers at other institutions: Tom Griffiths, Tania Lombrozo, Pat Shafto, Charles Kemp, Lauren Schmidt, Amy

11

was loved and didn't mind too much when I would bring work to our 'vacation time'. Family Meek has provided me with kindness and acceptance and also the all-important outlet for skiing and tennis. Keith and Scott, my Boston family, have been my keystone. Both have inspired me in my work with their rare ability to see clearly in difficult situations, and both have inspired me in my life with their love.

Contents

Chapter 1

Children are powerful causal learners. Before reaching preschool, they have formed predictive theories about biological kinds (such as growth and illness), physical kinds (such as objects and forces), and mental states (such as desires and beliefs). *How* they learn remains largely a mystery. Though computer scientists have be trying to construct human-like intelligence for more than half a century, even the most sophisticated artificial intelligence programs are nowhere near passing the Turing Test. No computer has developed enough 'common sense' causal knowledge about the world to convince a human judge that he is talking with another human rather than a machine.

In defense of computer scientists, even if the causal structure of the world were deterministic, constructing a machine that could learn all about it 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 are not clear-cut. The space of possible models is often vast, and accurate inferences about causal models are plagued by ambiguity in our observations.

Ambiguity arises for a number of reasons. First, 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 accurately perceived. Second,

events frequently occur simultaneously. If the observer doesn't know which variables are a priori the relevant variables, then the observations are confounded and it becomes difficult to replicate the event and control all variables. Additionally, different causal events happen on different time scales; not knowing how much time should transpire between a cause and its outcome greatly increases the confounds. Finally, not all variables are observable. Thus, even given perfect perception without confounds, it may be difficult to be aware of hidden causes or effects.

Some researchers have proposed that despite these numerous ambiguities that learning can precede via bottom up, empirical correlations (e.g. Pearl, 2000; Spirtes et al, 2001). However, a solely-statistical learner would require large sample sizes (very large if we consider that data are often ambiguous); as such, these bottom-up approaches cannot describe how it is that children can learn causal relations from just a few examples so rapidly. In contrast to this covariation-based approach to causal learning, some theorists have argued that at least some causal knowledge is available a priori, (e.g. Spelke et al, 1992). However, without statistical learning, vast amount of innate knowledge would be required to capture children's early proficiencies and could not describe how it is that children flexibly revise beliefs from evidence. Unsurprisingly, n these extremes neither approach succeeds at capturing children's rapid but accurate, flexible but conservative learning. Theory-based approaches tend to focus on the nature of the *structure*, while statistical learning approaches focus on the nature of the *restructuring*.

Indeed, these questions of structure and restructuring are not recent to cognitive development; rather they're arguably responsible for developmental psychology's birth. At its conception, rests Piaget, who formalized the first major theory of cognitive development, and arguably the most influential one for his ideas about knowledge structures and the process of restructuring. Importantly, his theory of learning went beyond traditional learning theories which depended on simple, bottom-up associative statistical processors (e.g. Pavlov, 1927). He stressed a domain general learning process dependent on the current beliefs of the observer (assimilating new evidence to beliefs and accommodating beliefs to new evidence) that were additionally constrained by the child's stage in development. Although the argument for the specific and dramatic changes in developmental stages have been challenged considerably (even by Piaget himself, later in life (e.g. Miller, 2001)), his theories about the process of belief revision (assimilation-accommodation), have remained widely influential.

Since Piaget, there have arisen numerous approaches to provide a more specific description of the nature of knowledge, such as information processing, biological, sociocultural, and theory-theory accounts. These accounts are not necessarily incommensurate with each other;

nevertheless, the theory-theory in particular beckons to our original questions of how knowledge (particularly causal knowledge), arises in different domains, and how beliefs from those domains interact with new evidence to inform learning. The theory theory draws on the idea that knowledge is organized in abstract, often causal, and interrelated concepts. While clearly relevant to scientific theories, this approach suggests that even every day knowledge is organized into intuitive theories (Carey, 1985; Gopnik & Meltzoff, 1997; Murphy & Medin, 1985; Wellman & Gelman, 1992).

Though the theory theory arguably is dramatically shaped by Piaget's theory of development, it does not assume that causal reasoning is independent of the domain. As a result, the theory-theory suffers from a problem of its own—if learning in each domain is different, then cataloging the enormous universe of developmental shifts becomes an unwieldy task. Some psychologists have even suggested that a new theory of learning must be developed for each domain (Gallistel, 2000). One way to maintain a grip on the problem of multiple learning modules is put forth by Carey, who writes that "[an alternative] analysis holds that children represent only a few theory-like cognitive structures, in which their notions of causality are embedded and in terms of which their deep ontological commitments are explicated" (Carey, 1985). That is, the problem is simplified if we consider just a few potential domains that guide learning over all of children's early causal experiences.

Overall, the theory theory provides a compelling account for which to consider how knowledge may be represented; it correctly predicts that the child, like the scientist, engages in on ongoing process of hypothesis testing and revision. However, though this account supports the claim that beliefs are defeasible in light of counter evidence, work remains in characterizing more specifically how and why learning should take place, and it lacks a universal proposal for the interaction of theories and evidence. For example, the theory theory does not specifically speak to whether children make rational interventions during play; whether children learn from the evidence generated during play; how specifically children's explanations connect with their beliefs; when evidence leads to learning, and when it is interpreted in terms of current beliefs. In general, this approach supports the claim that constraints, such as theories and evidence shape inference; but does not provide specific proposals on how evidence and prior beliefs should interact to inform predictions, exploration, and explanations.

Recent advances in artificial intelligence have led to the development of models that bridge the gap between constraint-driven and data-driven approaches, suggesting ways in which naive theories and the ability to learn from evidence might interact. One approach to thinking

about how prior beliefs should interact with statistical evidence involves regarding causal learning as a problem of Bayesian inference. Bayesian statistics describes the same learning problem that children face: it offers a rational account of how the degree of belief in a particular hypothesis should change as evidence accumulates. In Generative 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 – 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. 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).

This Bayesian approach is consistent with the theory theory. On the one hand, it partly solves the problem of 'multiple systems of learning' by offering a universal prescription for how evidence and prior beliefs should interact; on the other hand, theory-based Bayesian approaches also allow for domain-specific contributions from theories, thus offering an account of how learning can differentially proceed across domains. These theory based approaches have met with much success in describing learning in adult populations (e.g. see Tenenbaum, Griffiths, and Kemp, (2006) for a review). And the theory-based approach also simultaneously addresses both the 'nature of the structure' and the 'nature of the restructuring'.

Though the computational perspective formalizes intuitions about prior beliefs and evidence, one should note that these intuitions have long been prevalent to empirically focused research. For example, theory-theorists often describe the learning process as akin to Bayes rule, (without having been exposed the computational framework). For example, take Koslowski on learning:

> "For many events, there is a kind of catalogue of standard causes that, all other things being equal, vary in initial likelihood…However, the eventual likelihood of a cause might reasonably be expected to depend on the interaction between its initial likelihood (its position in the causal catalogue) and the various types of evidence (or other, rational or methodological considerations). …That is…the extent to which additional evidence is taken into account might depend on the

initial plausibility of the explanation to which it is directed." (p. 15, Theories and Evidence, Koslowski, 1996).

Koslowski describes the key components of the Bayesian framework—that hypotheses have different prior probabilities and that the posterior will depend both on those prior probabilities, but also on the evidence (probability of the data given the hypothesis). Providing some computational rigor to these intuitions is appealing, as it helps make more specific predictions of children's behavior.

Formal approaches may describe how the learning process is 'rational'. Part of this rigor requires defining what it means to be rational. As suggested by Chater and Oaksford (1999) this requires specifying the goals of the learner and the nature of the environment. A rational (Bayesian) agent should: act with respect to a goal, evaluate all possibilities given the evidence observed and prior beliefs, and choose the best action. Thus, the rational goal of prediction and explanation is to report the most likely hypothesis given the data observed and the predictor's prior beliefs about the world. The rational goal of exploration is to maximize potential for learning, (e.g. to investigate in order to disambiguate causally ambiguous or uncertain systems. Causal uncertainty arises when the predictive posterior of at least two hypotheses are equal, given the data observed and prior beliefs.)

In this thesis, rational Bayesian models and the theory theory are bridged to explore ways in which children can be described as Bayesian scientists. This approach does not address how theories may be rationally acquired, though theory-based Bayesian approaches have begun to address these questions (see, Tenenbaum, Griffiths, & Kemp, 2006; Kemp, Goodman, & Tenenbaum, 2008; Kemp & Tenenbaum, 2008; Goodman, Ullman, & Tenenbaum, 2009). Instead, I investigate what it means for children to take a rational approach to processes that support learning. In particular, I present empirical studies that show children making rational predictions, exploration, and explanations. I test the claim that differences in prior beliefs or changes in the observed evidence should affect these behaviors. The studies presented in this thesis encompass two manipulations: in some conditions, children's prior beliefs are equal, but the patterns of evidence are varied; in other conditions, children observe identical evidence but children's prior beliefs are varied.

I incorporate an additional approach in this thesis, testing children within a variety of domains, tapping into their intuitive theories of biological kinds, psychosomatic illness, balance, and physical systems. For example, in Chapter Two I look at how evidence and children's strong

beliefs about biological events and psychosomatic illness influence their forced-choice explanations in a story-book task. In Chapter Three I develop a training study to further investigate developmental differences in children's learning from evidence when they have strong prior beliefs. Chapter Four looks at how children's strong differential beliefs of balance interact with evidence to affect their predictions, play, explanations, and learning. Chapter Five looks at children's exploratory play with a jack-in-the-box, (when children don't have strong, differential beliefs), given different patterns of evidence. Chapter Six investigates children's explanations following theory-neutral evidence about a mechanical toy. Chapter Seven concludes the thesis.

By investigating situations where children have strong and weak prior beliefs and situations where children observe strong and weak evidence, we can contrast how differences in theories and evidence affect children's predictions, exploration, and explanations. The difference between 'strong priors' and 'weak priors' is partly of expository convenience. In the studies presented in this thesis children always have theoretical commitments shaping how evidence is interpreted; importantly, cases of 'weak priors' imply that children's beliefs do not differentiate the evidence presented. Testing with a variety of domains provides a more colorful picture of children's early beliefs. The following chapters will suggest that frameworks combining evidence and theories capture children's causal learning about the world. Thus, computational approaches may help explain how children are effective learners.

Chapter 2

As discussed in Chapter 1, the view that children's causal representations resemble scientific theories 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 from domain-specific modules (Leslie, 1994; Scholl & Leslie, 1999) or innate concepts in core domains (Carey & Spelke, 1994; Keil, 1995), while other researchers have focused on children's ability to learn causal relations from statistical evidence (. 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. In one study for instance, Shultz showed children a candle with a screen around it. He turned on a fan and then, five seconds 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 studies, other work suggests that preschoolers can use domain-general information to override domain-specific theories. Research suggests for instance, that four-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, 2006; 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 naïve adults (Chen & Klahr, 1999; Dunbar & Klahr, 1989; 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 relationship consistent with their theories and instances of non-co-occurrence as evidence against the relationship inconsistent 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 mid-air) 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 naïve 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' rule,

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{\sum_{h'} P(D \mid h')P(h')} \qquad (1)$$

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, in press; 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, I look 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 are interested in interactions between naïve theories and evidence, we give children evidence that is formally ambiguous: children observe events in which two candidate causes simultaneously covary with the effect. Children were given a forced choice between the two causes and manipulate the extent to which each cause is 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 two of ten blocks activate the toy (i.e., activating blocks are rare) or they learn that eight of ten blocks activate the toy (activating blocks are common). Children in both conditions then see two novel blocks, red and blue, placed simultaneously 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 three 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 three blocks are causes. Research suggests that children's judgments about activating blocks are sensitive to such base rate information (Tenenbaum, Sobel, Griffiths, & Gopnik, in submission). 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, Tenenbaum, & Gopnik, 2004).

This task is formally similar to the tasks used in these previous studies. Children were read storybooks in which one cause recurs every day and the other cause is always novel (i.e., the evidence is in the form AB→E; CA→E; AD→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 naïve 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 naïve 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 you blush or that feeling frustrated can cause a headache; Notaro, Gelman & Zimmerman, 2001).

Note that this not to suggest that children deny all relations between mental states and bodily events. 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, Hickling, & Schult, 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., tummyaches 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 involuntarily bodily states exclusively to bodily causes (e.g., illness and injury). Thus, following Notaro et al., we investigate children's understanding of only a subset of possible relations between psychological and bodily events: cases where a non-volitional mental state is the cause of an involuntary bodily reaction.[2]

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 Model

In Experiment 1, children are asked "Why does [character] have [symptom]? Is it because of [A] or because of [B]?" The probability that children choose explanation A is modeled as

$$\frac{P(\text{Explanation A} \mid D)}{P(\text{Explanation A} \mid D) + P(\text{Explanation B} \mid D)} \qquad (2)$$

This directly contrasts the two possible explanations given the data observed. 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:

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

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

---

[2]   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 naïve biology (see e.g., Carey, 1985).

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* in 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' 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, Griffiths, and Niyogi (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.

Framework Theory
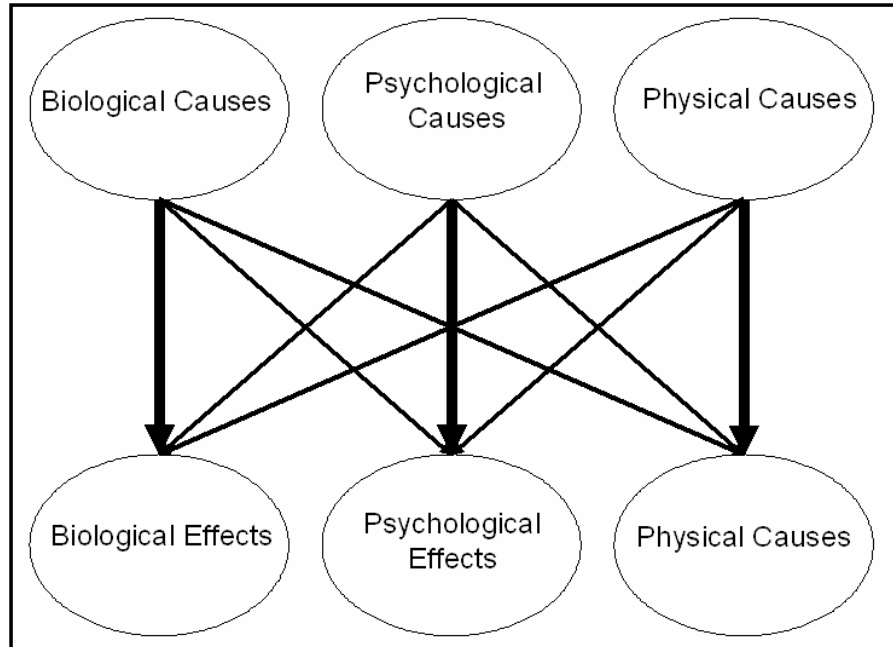
Figure 1: Schematic of framework theory that includes causal connections *within domain* and *cross domains*. Thick lines are modeled with probability *p* (*within domain*) and thinner lines with probability *q* (*cross domains*).

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 $p = .4$[3].

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 mth 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 which influences the effect is present, and ε otherwise, corresponding to a noisy-OR parameterization (Pearl, 1988) where each cause has a strength of 1 and the background has a strength of ε. We assumed that the probability of an effect in the absence of any causes was low, with $\varepsilon = .001$. The probability of the full set of data, D, accumulated over the course of the story is given by

$$P(D\,|\,h) = \prod_m P(d_m\,|\,h)$$

(4)

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

---

[3] The qualitative predictions hold for a wide range of values for p and q as long as $p \gg q$, consistent with our intuitions.

Model Predictions. The predictions of the model given this pattern of evidence are represented in Figure 4. We implemented our intuition of relatively low *cross-domain* probability by setting $q = .1$ and set a higher *within-domain* probability of $p = .4$. As described above, we also assumed a small $\Sigma = .001$. Importantly, the model demonstrates the shift between favoring the within domain candidate cause at baseline to favoring the cross domains candidate cause after evidence. Note that this model makes two clear qualitative predictions. First, children's naïve 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, Tenenbaum, & Gopnik, 2004) suggests that older (four-year-olds) but not younger (three-year-olds) preschoolers are able to integrate knowledge about base rates and patterns of evidence, so we test our predictions about the integration of domain-specific prior knowledge and evidence across a range of ages: three-year-olds, three-and-a-half-year-olds, and four- and five-year-olds.

Experiment 1: *Within* and *Cross Domains* Storybook Task

In Experiment 1, we read children two storybooks: a *Within Domain* story and a *Cross Domains* story. The evidence (presented in the form AB→E, CA→E, AD→E, etc.) is formally identical in the two stories. We predict that children will be more likely to identify A as a cause when A is domain-appropriate than when it is domain-inappropriate. However, we also predict that, for both stories, children will be more likely to think A is a cause after seeing the evidence than at baseline.

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). While 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: three-year-olds (mean age: 39 months; range 36-41 months), three-and-a-half-year-olds (mean age: 45 months; range 42-48 months), and four- to five-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 four- and five-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 four- to five-year-olds and 16 three-year-

29

olds (8 in the older group; 8 in the younger group) were tested before seeing any statistical evidence (the Baseline condition). We tested the three-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."[4] See Figure 2 for details and Appendix A 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 counterbalanced between participants). In the Baseline condition, children were read only the "Sunday" page of

---

[4]   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.

On Monday morning, Bambi runs in the pine grove. Bambi gets excited, Bambi runs in the cat-tails.

Bambi has itchy spots on his legs.

On Monday afternoon, Bambi runs in the cedar trees, and Bambi swings on the rope swing.

Bambi feels great!

Bambi doesn't have any itchy spots.

[Tues...Sat]

On Sunday morning, Bambi runs in the garden. Bambi gets excited, Bambi runs in the cat-tails.

Bambi has itchy spots on his legs.
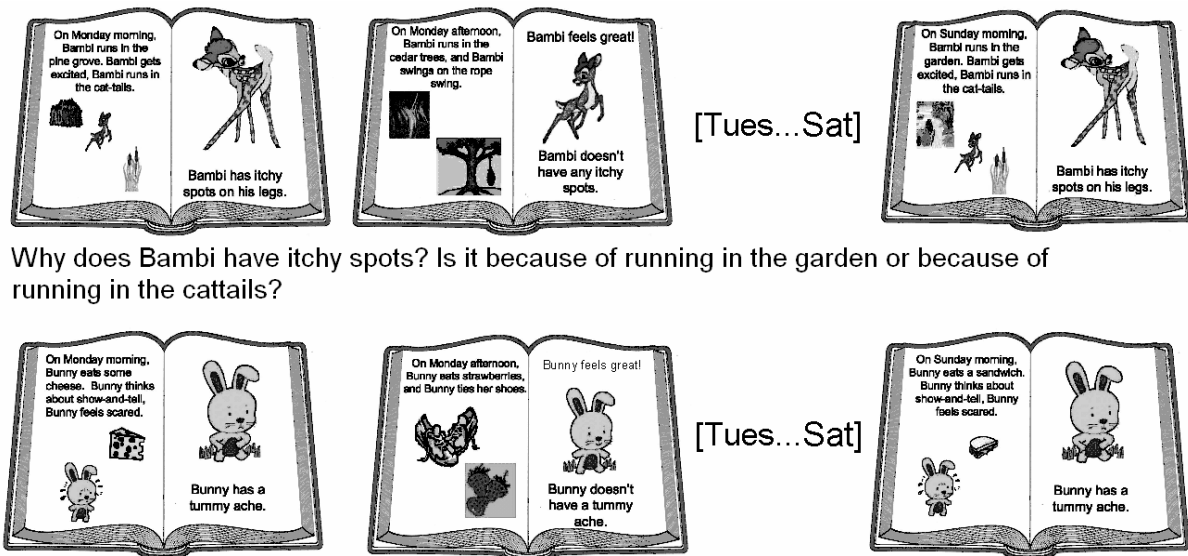
Why does Bambi have itchy spots? Is it because of running in the garden or because of running in the cattails?

On Monday morning, Bunny eats some cheese. Bunny thinks about show-and-tell, Bunny feels scared.

Bunny has a tummy ache.

On Monday afternoon, Bunny eats strawberries, and Bunny ties her shoes.

Bunny feels great!

Bunny doesn't have a tummy ache.

[Tues...Sat]

On Sunday morning, Bunny eats a sandwich. Bunny thinks about show-and-tell, Bunny feels scared.

Bunny has a tummy ache.

Why does Bunny have a tummy ache? Is it because of eating a sandwich or because of feeling scared?

Figure 2. *Sample pages from the storybook used in Experiments 1 and 2*

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 3. An alpha level of .05 was used throughout this chapter, 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 four- and five-year-old age group (4;0–5;6) 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).

31
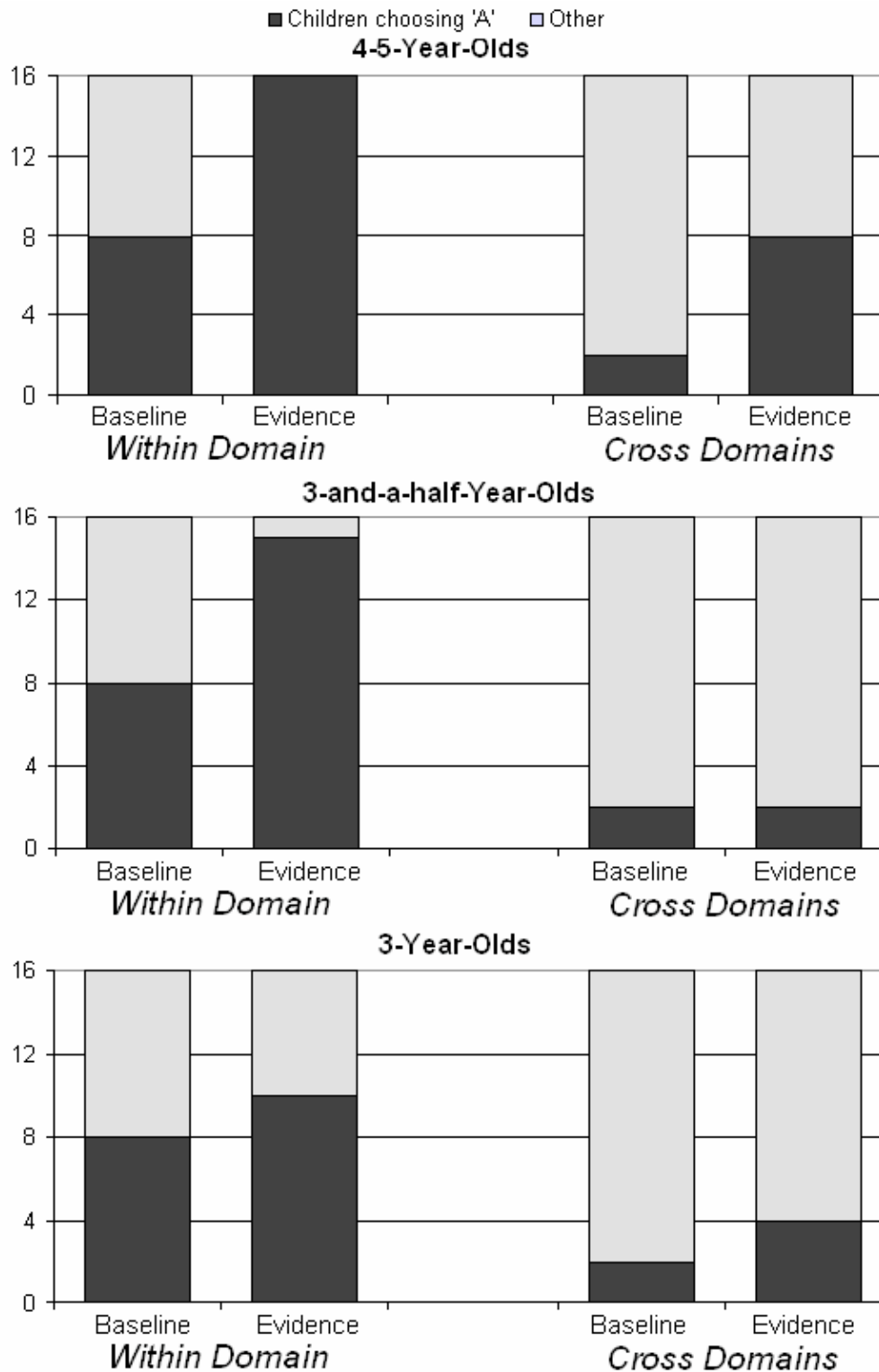
Figure 3. Children's responses to the storybook task in Experiment 1. The vertical axis shows the number of children selecting the different responses.

Children were significantly more likely to choose A after seeing the evidence than at baseline (χ^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 (χ^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: χ^2 (1, N = 32) = 5.24; Evidence: χ^2 (1, N = 32) = 10.67).

In the three-and-a-half-year-old age group (3;6-4;0) 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 (χ^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 (χ^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: χ^2 (1, N = 32) = 5.24; Evidence: χ^2 (1, N = 32) = 21.21).

Seeing the evidence had no effect for the youngest age group (3;0-3;6). 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 (χ^2 (1, N = 32) = .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 (χ^2 (1, N = 32) = .82, p = ns). Children were more likely to choose A in the *Within Domain* than *Cross Domains* task (Baseline: χ^2 (1, N = 32) = 5.24; Evidence: χ^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 by Condition interaction. Removal of the Age Group

factor from the saturated model did not result in a statistically significant increase in lack of fit (χ^2 (2, $N = 80$) = 4.49, p = ns), while removal of Condition or the interaction resulted in a statistically significant increase in lack of fit for both the saturated model (Condition: χ^2 (1, $N = 80$) = 8.93; Age Group x Condition: χ^2 (2, $N = 80$) = 9.48) and the model without the effect of Age Group (Condition: χ^2 (1, $N = 80$) = 8.93; Age Group by Condition: χ^2 (2, $N = 80$) = 7.74). These results indicate that the best model includes Condition and the Age Group by Condition interaction as predictors, supporting the hypothesis that the four- and five-year-olds responded differently to the cross-domain evidence than the three-year-olds.

Discussion of Experiment 1

The results of Experiment 1 suggest both that children's domain-specific beliefs 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 the children were less likely to identify A as a cause when A violated their beliefs. (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, four- and five-year-olds were able to entertain a causal possibility (that being scared might cause tummyaches) that they did not endorse without seeing those data.

By contrast, the younger three-year-olds (3;0 – 3;6) 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 you itch and others might not.) Further research might disambiguate these accounts.

However, for the three-and-a-half-year-olds (3;6-4;0) 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 four- and five-year-olds) while only 12% inferred that A was the cause in the *Cross Domains* task (no different than at baseline).

Why did the three-and-a-half-year-olds respond differently to the *Cross Domains* evidence than the four- and five-year-olds? There are at least three possible explanations. One possibility is that three-year-old children might have difficulty making inferences from ambiguous statistical data. If (as suggested by the failure of the younger three-year-olds to use the evidence at all) 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. Alternatively, the younger children might have a stronger belief in domain boundaries than older children. The data might have been insufficient to overcome three-year-olds' initial inductive bias that psychological causes are unlikely to generate bodily effects. Finally, the older three-year-olds and the four- and five-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. These possibilities are examined in Chapter 3.

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 is 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: Possibility Judgments

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 you could do possible events if you wanted to, but you could not do impossible ones. (Also see Shtulman, in press). In Experiment 2, we use a similar method to look at preschoolers' judgments about the possibility of physical, psychological, and psychosomatic events. We ask 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. While 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 semi-random 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 Judgment condition except that children were first tested on the *Cross Domains* storybook as in the Evidence condition of Experiment 1.

Results and Discussion of Experiment 2

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 (χ^2 (1, $N = 31$) = .125, p = ns) and significantly more than the 12% of children who chose 'being scared' in the Baseline condition of Experiment 1 (χ^2 (1, $N = 31$) = 7.63).

In the possible/impossible picture task, one child in the Baseline Possibility Judgments condition and one child in the Evidence And Possibility Judgments condition answered, "yes" to all six questions and one 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 (χ^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 likely than children who chose 'food' to say that the other psychosomatic events were possible (χ^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

37

Table 1. Children's possible/impossible judgments in Experiment 2

	Number 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 in parentheses.

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 between psychological events and bodily effects (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 tummyaches, headaches, and 'feeling sick' are distinct forms of illness, they might not readily generalize causes of tummyaches 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 tummyaches 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 naïve 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: Free Explanation Task

However, given children's failure to generalize their inferences, one might be sceptical 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 four and five-year-olds and led them to choose at chance.

In Experiment 3, we introduce two measures to distinguish authentic learning from chance performance. First, we modify 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 ask 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 tummyaches), 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 who 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 tummyaches?") might lead the children to believe that we were questioning their original responses. Since 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 are 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 stomachache to worrying in the Evidence And Explanation condition than 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. While 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 of Experiment 3

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', comparable to 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 Judgment 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 (χ^2 (1, $N = 36$) = .09, $p = ns$) and 2 (χ^2 (1, $N = 36$) = .31, $p = ns$). Children were significantly more likely to choose 'being scared' in this experiment than in the Baseline condition of Experiment 1 (χ^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 (χ^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 the Baseline Explanation condition (χ^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,

41

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*). 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 four- and five-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 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 naïve theories, consistent with the predictions of Bayesian inference models. As can be seen comparing the results predicted by the Bayesian model in Figure 4 with the four-year-olds' responses in

Experiments 1 and 2, our model accurately predicted the responses of the oldest children, with a Pearson 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.

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 naïve 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 us in exploring how well children solve this problem. In particular, a Bayesian model can allow us 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 the Appendix provides one such set of predictions, showing that the judgments of the four- and five-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. While 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
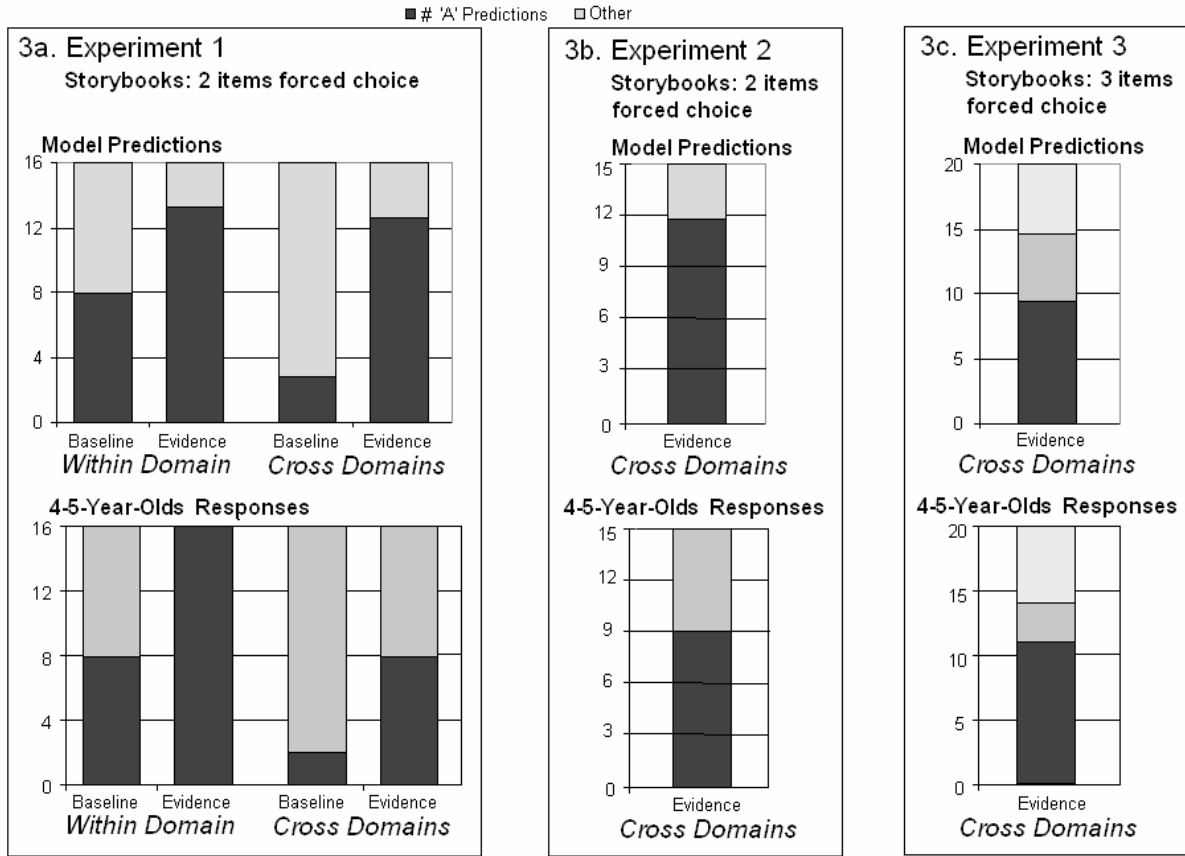
Figure 4. Predictions of the Bayesian inference model compared to four and five-year-olds' responses to the storybook task in Experiments 1, 2, and 3, when q (*cross domains* probability) = .1 and p (*within domain* probability) = .4.

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 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, Tenenbaum, & Gopnik, 2004; Tenenbaum, Sobel, Griffiths, & Gopnik, in submission).

The results of our experiments also raise several other questions. As noted, it is not clear whether the developmental differences between three-year-olds, three-and-a-half-year-olds, and

four- and five-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. Chapter 3 helps disambiguate the role of each of these factors.

Interestingly however, the finding that three-and-half-year-olds did not use the theory-violating evidence to change their judgments is consistent with a wide variety of research suggesting that three-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, three-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 tummyaches. 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 tummyaches 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 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, et al., 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, in press; 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.

Chapter 3

In Chapter 2, I presented a series of studies that tested three groups of children (four- and five-year-olds; mean: 60 months, older three-year-olds; mean: 45 months, and younger three-year-olds; mean: 39 months) on Within and Cross domains storybooks. Consistent with the prescriptions of Bayesian inference, older preschoolers correctly inferred that A was the cause in both cases but were more likely to identify A as the cause in the Within Domain book than the Cross Domains book. However, although the three-and-a-half-year-olds readily identified cause A as the target cause in the Within Domain book (indeed, they were indistinguishable from the older children), they failed to learn at all in the Cross Domains book; that is, they consistently chose the within-domain cause. Finally, the youngest three-year-olds failed to learn from the evidence in either book; they chose at chance in the Within Domain book and chose the within-domain cause in the Cross Domains book.

The performance of the three-and-a-half-year-olds is particularly interesting. It provides a dramatic contrast between children's impressive reasoning (near ceiling) about statistical evidence in neutral contexts and their poor reasoning (near floor) in belief-violating contexts. How might the different accounts explain the discrepancy? Many studies suggest that children have difficulty learning from statistical evidence that conflicts with their prior beliefs (Klahr, Fay,

& Dunbar, 1993; Kuhn, 1989; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Phelps, 1982; Koslowski, 1996; Kushnir & Gopnik, 2006; Masnick & Klahr, 2003; Schauble, 1990; Schulz, Bonawitz, & Griffiths, 2007). Why do children find it so hard to learn from evidence that violates their naïve theories?

One possibility is that children's resistance to anomalous evidence is rational. Arguably, learners *should* require more evidence to learn a priori implausible causal relationships than those that are already well supported. Bayesian analyses suggests that developmental differences in children's prior knowledge might explain differences in children's causal learning (Tenenbaum, Griffiths, & Kemp, 2006; Kemp, Perfors, & Tenenbaum, 2007). Because this view suggests that children's prior beliefs affect their interpretation of the evidence, we will call this the Prior Beliefs account.

An alternative view suggests that children genuinely have difficulty learning from statistical data; that is, children's difficulty learning from theory-violating evidence is a bug not a feature. Consistent with this view, researchers have suggested that strong prior beliefs can lead learners to overlook relevant evidence in favor of irrelevant variables, to fail to understand the relationship between a hypothesis and the evidence that might support or disconfirm it, and to have difficulty generating effective experiments (Klahr, Fay, & Dunbar, 1993; Koslowski, 1996; Kuhn, 1989; Kuhn et al., 1995; Kuhn, Amsel, & O'Loughlin, 1988; Masnick & Klahr, 2003; Schauble, 1990). Because this view emphasizes the vulnerability of children's statistical reasoning, we will call this the Statistical Reasoning account.

In practice, developmental psychologists have tended to emphasize the former perspective and researchers in science education the latter (see Kuhn & Dean, 2004, for review). However, the accounts are not mutually exclusive: children might have strong prior beliefs that rationally affect their interpretation of new data *and* a relatively fragile ability to reason about statistical evidence. One way to look at the relative contributions of prior knowledge and statistical reasoning limitations is to use a paradigm in which children fail to draw accurate inferences from theory-violating statistical data and then intervene selectively on children's prior beliefs or their statistical reasoning abilities. If children's performance improves in one or both training conditions compared to a sham manipulation, we can assess the impact of both rational constraints and process-level limitations (Marr, 1982) on children's reasoning.

Prior Belief Account

The prior belief account suggests that children's inductive biases change over development. In particular, the three-year-olds might have had a stronger belief in some domain

boundaries than the older children.[5] While adults accept that psychosomatic phenomena can cross domain boundaries, preschoolers typically deny the possibility of psychosomatic events (e.g., that feeling embarrassed can make your face to turn red); however, older children are more willing than younger children to accept the possibility of psychosomatic phenomena suggesting that children's beliefs in domain boundaries may change with age and experience (Notaro, Gelman and Zimmerman, 2001).

How might we affect children's prior beliefs about psychosomatic causality? Arguably, young children have relatively limited exposure to psychosomatic events. If children believe the base rate of psychosomatic causality is low, then they will (rationally) resist accepting a psychological cause as the most probable explanation of a bodily effect in our task. Thus we might manipulate children's prior beliefs by affecting their perception of the frequency of psychosomatic events. We will call this the *Prior Beliefs Base Rate* condition.

Alternatively, children might resist psychosomatic causality because they do not understand *how* psychological states affect bodily states. Research suggests that both adults and children are more willing to accept causal relations for which they can imagine plausible causal mechanisms (e.g. Ahn, Kalish, Medin, & Gelman, 1995; Shultz, 1982). Thus we might increase children's acceptance of psychosomatic causality by offering even a relatively shallow explanation (see Keil, 2006; Rosenblit & Keil, 2002) of how emotional states might cause bodily outcomes. We will call this the *Prior Beliefs Mechanism* condition. If either the *Baserate* or the *Mechanism* training is effective, this would suggest that rational inductive biases affect children's ability to reason about a priori unlikely events.

Statistical Reasoning Account

The Statistical Reasoning account emphasizes children's fragile ability to learn from statistical data. The failure of the youngest three-year-olds to learn from the Within Domain book is consistent with this possibility, suggesting that even the neutral task might have been challenging for children. If the older three-year-olds were just beginning to be able to interpret the ambiguous statistical data, they might not have been able to handle the increased task difficulty posed by a conflict with prior beliefs.[6] The Statistical Reasoning account predicts

---

[5] Note that an account appealing to prior inductive biases could also account for the failure of the youngest three-year-olds in the theory-neutral (Within Domain) condition. The youngest children might have believed that nothing distinguished the within-domain variables as causes of the effect; the statistical evidence that one cause was more likely than the others might have been insufficient to overcome this bias.
[6] It is also possible that children's statistical reasoning ability stayed constant but other information processing abilities (e.g., their ability to inhibit the prepotent alternative cause) improved with age. We

that giving three-and-a-half-year-olds additional practice reasoning about ambiguous evidence (i.e., in theory-neutral contexts) should improve the children's ability to reason about a priori unlikely evidence. We will call this the *Statistical Reasoning* condition.

Training Study

In order to investigate these accounts, we designed a two-week training study. Children were included in the study only if they initially endorsed the within-domain cause rather than the statistically likely cause in a Cross Domains pretest book (identical to the book used in Schulz et al. (2007)). The children were assigned to one of four conditions: a *Prior Belief Baserates* training, a *Prior Belief Mechanisms* training, a *Statistical Reasoning* training, and a *Control* condition. At the final test session, children were given a final Cross Domains storybook (formally identical to the initial book but with different specific stimuli).

Children were also given a free explanation task adapted from Schulz et al. (2007), (see Chapter 2). In the free explanation task children were told about a puppy dog who was scared about the first day of school and had a tummy ache; children were asked to explain why the puppy had a tummy ache. Strikingly, we found that at baseline, four-year-olds ignored the only variable mentioned (being scared) and instead spontaneously invented their own domain-appropriate explanations (e.g., "because he fell on his stomach"; "because he ate too much food"). However, four-year-olds who had first been exposed to the evidence in the *Cross Domains* book adopted the psychosomatic explanation. By using the free explanation task as our final dependent measure, we could assess not only whether the training affected the children's responses to the Cross Domains test book itself (responses potentially vulnerable to children's tendency to vary their responses when asked versions of the same question twice, e.g. Poole & White, 1991; Memon, Cronin, Eaves, & Bull, 1993) but also whether the training affected children's willingness to endorse cross-domain causes more generally.

Methods and Design

Participants. Eighty children (mean: 45mos; range: 39-48mos) were recruited from preschools in a metropolitan area. An experimenter met individually with each child for four twenty-minute

---

emphasize the changes in children's statistical reasoning ability rather than changes for instance, in inhibitory control, because inhibitory control demands do not readily seem to explain the youngest three-year-olds' failure on the theory-neutral task. However, for the purposes of this paper, little changes in our discussion on either interpretation: the Statistical Reasoning account and any more general information processing account are process-level accounts of the developmental differences in children's learning; both contrast with but are not mutually exclusive with the Prior Beliefs account, and both rely on abilities that should be improved by the Statistical Reasoning training (i.e., increasing children's fluency with the basic task should increase the resources available for inhibitory control).
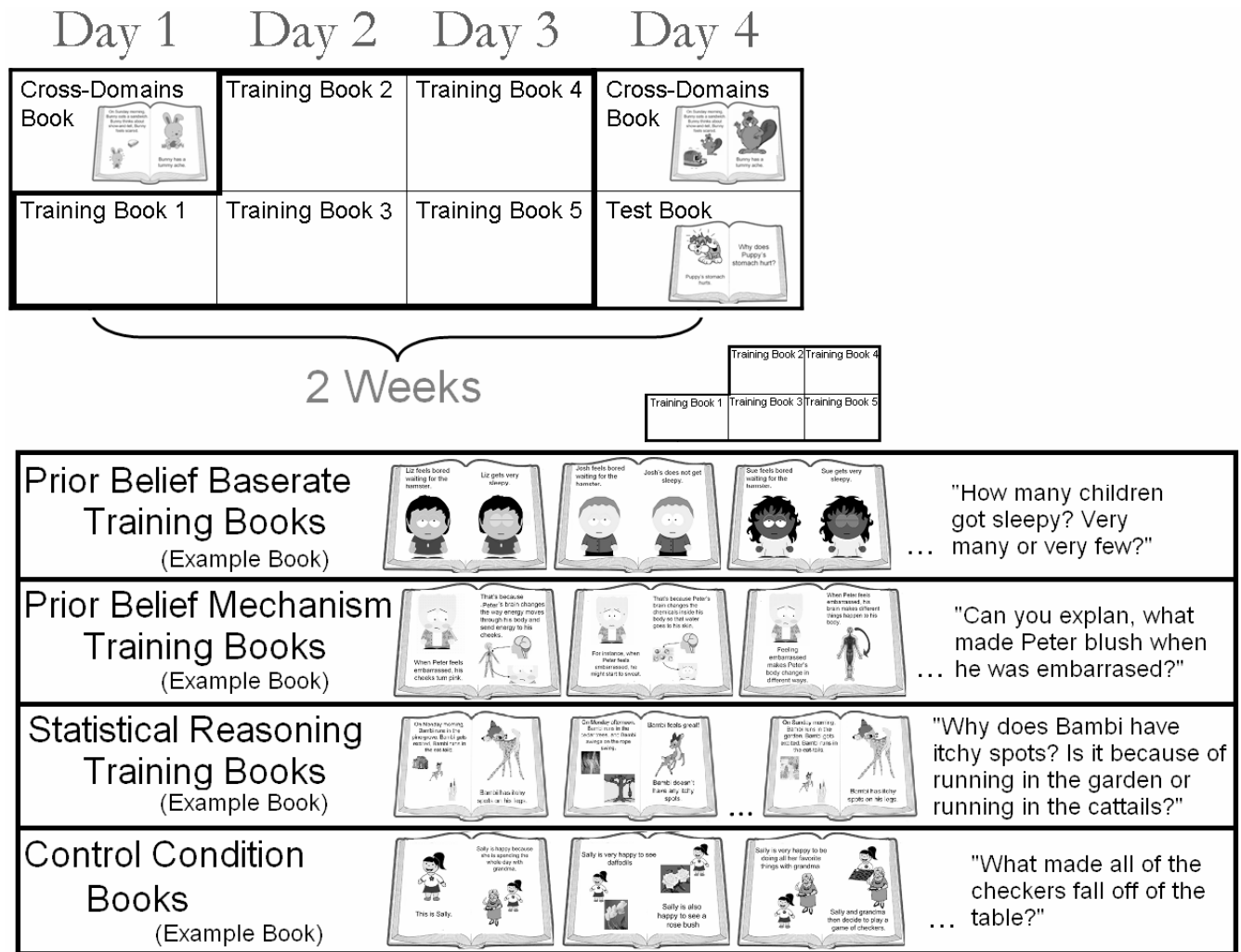
Figure 5: Study design and sample pages from the training books.

sessions over a period of two weeks. No two sessions were on consecutive days. Most of the children were white and middle class but a range of ethnicities resembling the diversity of the population was represented; 54% of the participants were girls. Children who passed the initial Cross Domains test books were dropped from the study and replaced (see below). Children were randomly assigned to a *Prior Belief Baserates Training* condition, a *Prior Belief Mechanisms Training* condition, a *Statistical Reasoning Training* condition, or a *Control* condition.

Materials. Two Cross Domains books and a Free Explanation book were used. Additionally, five different training books were used in each of the four conditions (*Prior Belief Baserate, Prior Belief Mechanisms, Statistical Reasoning,* and *Control),* for a total of 20 training books. The training books were each approximately 20 pages long and had approximately 9 words per page. (See Figure 5.)

53

Cross Domains books: Two books were used, a Bunny book and a Beaver book. The books were identical except for details of the stimuli. In each book, a character (Bunny or Beaver) ate a different food, experienced a recurring psychological cause (feeling worried; feeling scared), and a recurring biological effect (belly ache; tummy hurting) each morning of a seven-day week. Each afternoon, the character ate two different foods and had no ill effect. At the end of the story children were asked a forced choice question about the events of that morning: "Why does (Bunny's, Beaver's) (belly ache? tummy hurt)? Is it because of (feeling worried, feeling scared) or because of eating (the cornbread, the sandwich)?" The order of events (psychological or food) was counter-balanced throughout.

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

Training books: Five books were used in each training condition. Each of the stories involved unique characters and candidate relations.

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

Prior Belief Baserate Training: Each book showed ten characters in a classroom. All ten characters experienced the same emotion (e.g. boredom waiting for a hamster to do a trick). Eight of the ten characters had a bodily reaction (e.g. Sue gets sleepy; Charles gets sleepy; Josh does not get sleepy). At the end of the book children were given a forced choice question asking whether the bodily reaction to the psychological emotion happened to very many or very few characters in the story (e.g., "Can you remind me: did very many students get sleepy or did very few students get sleepy?").

Prior Belief Mechanisms Training: Each book explained that a particular psychological state could generate bodily effects and offered a brief explanation of how this might happen (e.g. "When Peter feels embarrassed, his brain makes different things happen to his body…his cheeks turn pink and he starts to blush. That's because Peter's brain changes the way energy moves through his body and can send energy to his cheeks.") At the end of each book, children were

asked to repeat the explanation for the bodily outcomes in the books (e.g., "Can you explain to me: what made Peter blush?)

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

Procedure. Children were tested individually in a quiet room at their daycare. Participants were first tested on one of the two Cross Domains test books (particular books counterbalanced between children; the other book was then used for the test book at the end of training). Only children who chose the non-psychological (statistically unlikely) cause were included in the training; children who passed were dropped from the study.

Children in the training study were then immediately read the first book from their assigned condition (see Figure 5). The experimenter then met with the child three more times over the course of two weeks. On each of the second and third visits, children were read the two books appropriate to their training condition (Books 2 & 3 on Day 2; and Books 4 & 5 on Day 3). The experimenter gave corrective feedback if the child answered incorrectly during the training sessions (i.e., in the *Statistical Reasoning* training, the experimenter pointed to the recurring variable and showed the child how it occurred each day along with the effect; in the *Base Rates* training, the experimenter pointed to the number of children with the bodily response and observed that it was 'very many' rather than 'very few'; in the *Mechanism* training, the experimenter repeated the explanation for the bodily effect; in the *Control* condition, the children were reminded of the correct information). On the final day (Day 4) the children were first tested on the Cross Domains storybook and then on the Free Explanation test book (the order was fixed so that if children learned from the Cross Domains book, they could use the evidence to answer the Free Explanation question). No corrective feedback was given to the test books.

Results

Replicating Chapter 2 (Schulz et al., 2007), 82% of the three-and-a-half-year-olds tested on the initial Cross Domains storybook failed the task (i.e., chose the theory-consistent rather than the statistically probable cause). These children continued onto the training study; children who passed the initial book were dropped and replaced. There were no age differences among the four conditions ($F(3, 76) = 1.48$, $p=$ NS).
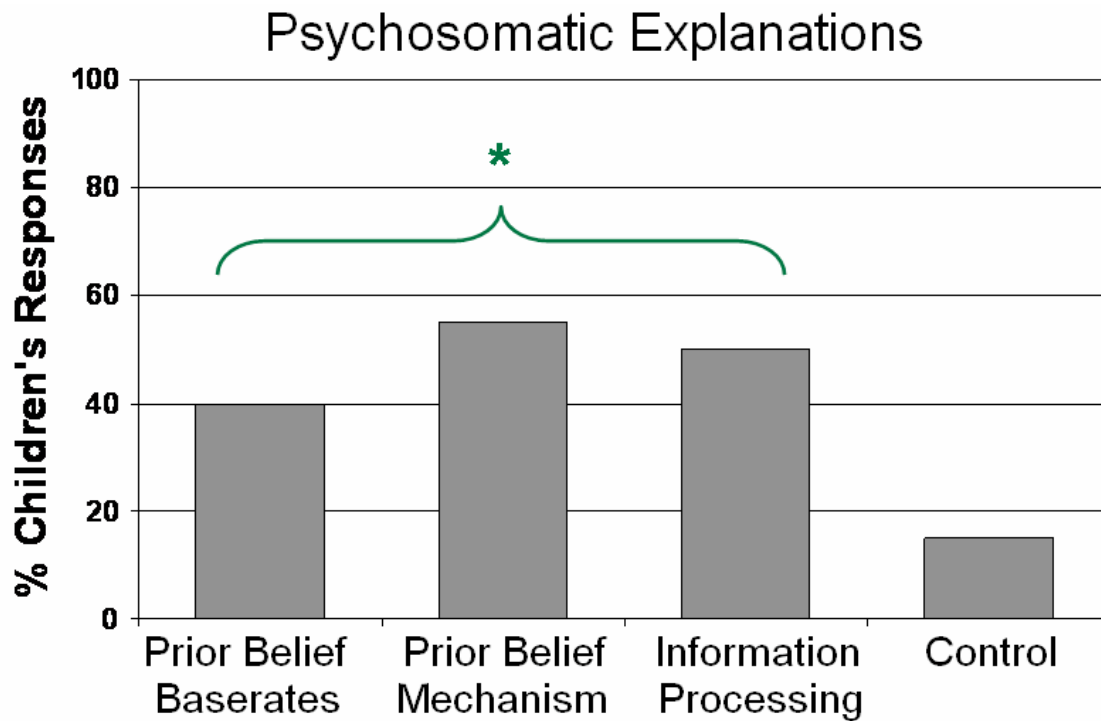
Figure 6: Percentage of psychosomatic explanations on the Explanation Test book.

Across the training period, children's performance on the training books improved. In all three training conditions, children were more likely to answer the prompts at the end of the training books correctly on the last day of training than on the first day (first day 50%, last day 78% in *Prior Belief Baserates*: $\chi2$ (1, N = 60) = 4.66, p < .05); first day 35%, last day 78% in *Prior Belief Mechanism*: $\chi2$ (1, N = 60) = 13.54, p < .01; first day 45%, last day 80% in *Statistical Reasoning*: $\chi2$ (1, N = 60) = 7.54, p < .01). This suggests that the training successfully affected children's performance on the target skill.

Responses of children on the final *Cross Domains* book were coded as either appealing to the recurring psychosomatic cause or to the alternative domain-appropriate cause (i.e., the particular food). Compared to their responses on the original test book (at floor due to the initial inclusion criteria), children were significantly more likely to appeal to psychosomatic causes in all conditions (45% *Prior Belief Baserates*: $\chi2$ (1, N = 32) = 11.61, p < .01; 45% *Prior Belief Mechanism*: $\chi2$ (1, N = 32) = 11.61, p < .01; 35% *Statistical Reasoning*: $\chi2$ (1, N = 38) = 8.49, p < .01; 30% *Control*: $\chi2$ (1, N = 38) = 7.06, p < .01). There were no significant differences between conditions.

What can we make of this across the board improvement? There are several possibilities, including that the course of the study coincided with developmental changes in some of the children's reasoning. However, as noted, research suggests that preschool children tend to vary responses when asked the same question twice, possibly because children interpret the act of asking again as the interviewer's desire for an alternative response (Memon, Cronin, Eaves, and Bull, 1993). Thus children's force-choice responses on the final Cross Domains test book may be a spurious effect of question repetition, rather than a genuine endorsement of psychogenic causality.

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

The pattern of responses on the Free Explanation test book demonstrated a significant effect of the training conditions relative to the control condition: 40% of children appealed to psychological causes in the in the Prior Belief Base Rate condition, 55% in the Prior Belief Mechanism training, 50% in the Statistical Reasoning condition, and 15% Control condition (see Figure 6). Significantly more children appealed to psychological explanations in the *Prior Belief Mechanisms* condition ($\chi2$ (1, N = 38) = 7.03, p < .01) and the *Statistical Reasoning* condition, ($\chi2$ (1, N = 38) = 5.58, p = .02) than the *Control* condition. Marginally more children appealed to psychological explanations in the *Prior Belief Baserates* condition, ($\chi2$ (1, N = 38) = 3.14, p = .08) than in *Control* condition. There were no significant differences between the *Prior Belief Base Rate*, *Prior Belief Mechanisms*, and *Statistical Reasoning* conditions ($\chi2$ (2, N = 57) = .93, p = NS). These results suggest that a brief training, involving only five storybooks, can affect children's willingness to consider unexpected relations as viable explanations for events.

Discussion

What do these findings tell us about the role of prior beliefs and statistical inference in children's reasoning about theory-violating evidence? The success of the *Prior Belief* conditions suggests that either increasing children's perception of the base rate of a target causal relation or increasing their understanding of the target causal mechanism, increases children's willingness to appeal to causal relations that were unlikely with respect to their prior beliefs. Interestingly, the

results in the *Mechanisms* condition were robust while those in the *Base Rates* condition were only a trend, consistent with the claim that children may attend more to mechanism information than correlation information (Ahn et al. 1995; Shultz, 1982). Critically however, both results contrast with the *Control* condition: mere repeated exposure to psychological variables did not increase children's tendency to adopt these variables in psychosomatic explanations.

However, these results fall short of demonstrating that manipulating children's prior beliefs improves their ability to learn from (erstwhile) implausible data. Children in the *Prior Belief* conditions may have been better able to learn from the statistical evidence in the Cross Domains books and bring this evidence to bear on the free explanation task, but it is also possible that the training condition directly increased the children's willingness to appeal to psychosomatic causes. Although several studies have demonstrated the effect of prior knowledge on children's reasoning about evidence (e.g. Bonawitz, Lim, & Schulz, 2007; Legare, Wellman, & Gelman, in press; Schulz & Gopnik, 2004), this work has involved older children (though see Sobel, 2006 for other work suggesting that three-year-olds' ability to draw inferences from statistical evidence may depend on their understanding of the underlying causal mechanisms). Further research however, is needed to isolate the effect of rational constraints from the effects of other task demands and to look at whether changing prior knowledge can change the statistical inferences of even very young children.

By contrast, in the *Statistical Reasoning* condition, the children's only evidence for the psychosomatic causal relations came from the ambiguous evidence in the Cross Domains books. These children were never given any direct information about psychosomatic events. Nonetheless, they were more likely than children in the *Control* condition to adopt psychosomatic explanations.

Not all the children who chose the psychological variable in the ambiguous evidence task endorsed psychosomatic causality in the explanation task (71%) and not all the children who gave psychosomatic explanations chose the psychological variable in the ambiguous evidence task (50%). This is perhaps not surprising, given the different task demands between the two measures. As noted, some of the children's responses in the storybook task might be due to repeated questioning rather than genuine acceptance of psychosomatic causes. On the other hand, some children who chose the within-domain variable in the forced choice task (e.g., when both the psychological and bodily cause were present in the story) might nonetheless have learned enough from the evidence to treat psychological variables as relevant causes for the purpose of free explanation (e.g., when only the psychological variable was given in the story). Overall however, the children taught to reason about statistical data were better able than children in the

Control condition to bring the evidence from the Cross Domains book to bear on an open-ended explanation task.

We do not know however, at what level of abstraction children changed their beliefs. Previous work (Schulz et al., 2007) suggests that children may have changed their inferences only at a specific level (entertaining the possibility only that worrying could cause tummy-aches) rather than at the level of the more abstract theories (revising their beliefs about domain boundaries). It is noteworthy however, that children in both *Prior Beliefs* training conditions were able to generalize from other instances of psychosomatic causation to the test exemplar. This suggests that at least some more abstract inferences were enabled by those training conditions. Further research might establish the degree to which different interventions transform children's ability to reason about theory-violating evidence.

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

Chapter 4

In the previous chapters, we looked at ways in which children's prior beliefs and the evidence they observed affected their choices about candidate causes. However, while these studies showed that children rationally integrate prior beliefs and observed evidence, these studies do not address how children generate evidence in the first place. As Piaget first observed (1930), children begin "experimenting" before they can walk, intervening and testing their beliefs, delighting in their exploration as scientists delight in new data.

However, it is well established that, unlike scientists, children have little metaconceptual awareness of this 'scientific testing' (e.g., Kuhn & Phelps, 1982; Kuhn, Amsel, & O'Loughlin, 1988; Dunbar & Klahr, 1989; Kuhn, 1989). In particular, children (and even adults) often do not explicitly seek disconfirming evidence (e.g. Karmiloff-Smith & Inhelder, 1975; Wason, 1960). The 'child as scientist' account is additionally challenged by the lack of evidence for purposeful play. While much of the literature on children's play has found that children learn from play (e.g., Bruner, Jolly, & Sylva, 1976; Singer, Golinkoff, & Hirsch-Pasek, 2006) these approaches have typically focused on descriptive accounts. In general, the only systematic finding about children's exploratory play is that children (and many other creatures) preferentially explore novel over familiar stimuli (e.g. Berlyne, 1960; Hutt & Bhavnani, 1972; Pavlov, 1927).

Children's remarkable success at learning and their striking failures at designing informative experiments raise questions about the role of theories and evidence in play. If exploratory play is largely unsystematic, how might children generate the type of evidence that could support learning? And when evidence is generated by the child during play, when do beliefs guide how it is interpreted (or discounted), and when does evidence overturn beliefs? In this chapter, I will suggest that prior beliefs and observed evidence rationally drive children's exploration. In particular, we will show that although the specific actions children take in the course of play might not be systematic, children's exploratory play is nonetheless sensitive to whether evidence is surprising with respect to their beliefs. Secondly, I suggest that children can maintain their beliefs in the face of conflicting evidence by appealing to alternative variables to explain away surprising data; children may revise their beliefs only when alternative explanatory variables do not exist.

I test whether children's beliefs and observed evidence interact to affect their exploratory play. In particular, we use the domain of balance, providing children evidence that either conflicts with their beliefs or confirms them; we then test whether children are more likely to explore the familiar block and balance or a novel toy following these types of evidence. After letting the child play freely, we look at whether children have discovered a hidden variable (a magnet) which can be used to explain away surprising evidence and whether children are more likely to appeal to these explanatory variables in conflicting evidence conditions. Finally, we test whether children learn during the course of their own exploratory play, revising their predictions on a final balance attempt. Success on this task would indicate that children are indeed sensitive to the evidence they observe and their prior beliefs; both rationally guiding exploration, which in turn shapes explanation and learning.

Model of Rational Exploration

If children are rational explorers, investigating when there is something likely to be learned, we may expect children to be more likely to explore given causal uncertainty as compared to cases that are relatively more certain. Causal uncertainty can arise for two reasons: because the observer does not have any a priori strong beliefs about an event and evidence fails to disambiguate possible hypotheses; or, because the observer has very strong beliefs about an event, but evidence strongly supports the a priori unlikely belief. Both accounts would require children to be rationally able to integrate prior beliefs with the evidence they observe.

The Bayesian framing of this problem suggests that the posterior probabilities of the candidate causal explanations are equal in uncertain cases, where the posterior is the particular

hypothesis (*h*) about a toy's causal structure, given the observed data (*D*). That is the probability of Hypothesis 1 given the Data over the probability of Hypothesis 2 given the data is approximately equal to x.

$$\frac{P\,(\text{Hypothesis 1} \mid D)}{P\,(\text{Hypothesis 2} \mid D)} = \frac{P\,(D \mid \text{Hypothesis 1})\,P\,(\text{Hypothesis 1})}{P\,(D \mid \text{Hypothesis 2})\,P\,(\text{Hypothesis 2})} \qquad (5)$$

Importantly, we want to understand why children might generate different patterns of play, which depends on their beliefs about the toy's causal structure and the evidence observed. If we express ambiguity in terms of comparing the probability of one causal explanation (Hypothesis 1 in Figure 7 below) with the probability of another causal explanation (Hypothesis 2), we can formally describe this ambiguity.

The first case of causal uncertainty (Figure 7) arises when children have strong prior beliefs but evidence conflicts with those beliefs. In this second case, there is a high prior probability of one explanation (that the child's theory is correct), and a low prior probability of the second (that the theory is incorrect); however, evidence strongly favors the a priori unlikely candidate and disfavors the a priori likely candidate. This interaction leads to equally plausible posteriors. In contrast, if the evidence confirms with beliefs, then the posterior strongly supports a single hypothesis.

The second cases arises when prior beliefs are a priori equally plausible and evidence fails to provide support for one hypothesis over another. In contrast, if evidence was provided that strongly favored one hypothesis (disambiguating potential causal structures), then there would be low causal uncertainty. We come back to the second case of causal uncertainty in Chapter 5.

Case 1 Uncertainty:
...when P (Hypothesis 1) $\gg$ P (Hypothesis 2),
but P (D | Hypothesis 1) $\ll$ P (D | Hypothesis 2)

Case 2 Uncertainty:
...when P (Hypothesis 1) $\approx$ P (Hypothesis 2),
and P (D | Hypothesis 1) $\approx$ P (D | Hypothesis 2)

Figure 7: Two cases of causal uncertainty.

Theories and evidence in play and explanation

Is exploratory play motivated by uncertainty? Identifying and exploring surprising evidence may lead to two benefits: generating evidence that will lead to belief revision, or discovering a hidden variables that can account for the surprising observations. Unobservable causes can support a 'rational discounting' of evidence that conflicts with strongly held beliefs. That is, belief revision is not necessary when alternative variables can be used to 'explain away' data that conflicts with 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, preschool 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) provided preschool children with a false belief task and found that when children where presented with outcomes that were surprising with respect to their theories, they were more likely to appeal to alternative causes then 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.

These studies suggest that children are able to recognize surprising evidence and appeal to hidden explanatory variables. However, none of these previous studies explore how children's sensitivity to uncertainty affects exploratory play. Nor do they connect children's discovery of alternative variables during exploratory play with their ability to appeal to these variables to explain surprising (but not unsurprising) evidence. Finally, they do not examine how both evidence generated during play and the ability to explain away surprising evidence affects learning.

If children are little scientists, integrating causal beliefs and evidence, then the following predictions should hold. First, children should explore a familiar toy (over a novel toy) when evidence conflicts with beliefs, but not when evidence is consistent with beliefs. Second, if children's exploration leads to the discovery of a hidden variable, then children should explain belief-inconsistent, but not belief-consistent, events in terms of that variable. In cases where there are no hidden variables present to explain away the conflicting evidence, then children should revise beliefs.

Getting ahead with a theory of balance

Critically here, we look at cases when children have strong, differential beliefs. If children's play is rationally guided by the desire to resolve causal uncertainty, then the formal models proposed here should predict when and how children's should explore new events. Because the development of children's beliefs have been well established in the domain of balancing blocks, this domain is particularly conducive for investigating the relationship between children's beliefs and their exploratory play. Karmiloff-Smith & Inhelder (1974) demonstrated that younger, "No Theory" children balance blocks by trial and error. Gradually, between six- and eight-years, children 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[7]". Mass 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 such as play that could support such discoveries (though, see Pine and Messer (2000) for the role of explanations in helping children revised beliefs of balance). If prior beliefs and evidence guide children's exploratory play and explanations, then either a change in evidence or a difference in prior beliefs should lead to differential behavior.

Experiment 1

As previously noted, 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). We first presented children with a balance and a set of blocks. Then we showed children that the block balanced either according to their prior beliefs (Confirming Evidence condition), or in conflict

---

[7] We maintain the terms "No Theory", "Center Theory", and "Mass Theory" for consistency; however, we note a) that "No Theory" children do have beliefs about balance; however their predictions are undifferentiated along the bottom of the block. B) The "Mass Theory" children in our study probably do not have an adult concept of mass. Instead, children seem to be able to use the overall size of the object as a cue that weight or size need be distributed evenly at the point of balance. For our purposes here, this difference in Mass Theory is not crucial as our blocks are clearly heavily weighted (and larger) on one side. Thus, predictions from either "Mass Theory" will be to the same location, and "Mass Theory" predictions will be to a different location as compared to Center Theorists. Finally, some have argued that the term 'theory' requires a more complex structured set of representations that are general across a domain, where-as others have suggested that theories are any set of structured, defeasible, (and often causal) beliefs. Both interpretations of 'theory' are consistent with our experimental predictions and discussion, though we remain ambivalent about whether children here have theories of balance or strong beliefs.

with their beliefs (Conflicting 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). To a Center Theorist, a block with a conspicuously heavy side balancing on its geometric center is not in conflict with beliefs; however, this evidence is conflicting to a Mass Theorist. Conversely, to a Center Theorist, a block with one heavy side balancing under its center of mass is conflicting, but that evidence is belief-confirming to a Mass Theorist. Children observing belief-conflicting evidence arguably have more to learn (a conflict to resolve) than children observing belief-confirming evidence (which would not be surprising). After a minute of free play, we returned the block to its initial balancing state and asked the child to explain what made it balance. We predicted that children in the Conflicting Evidence condition (Center Theorists who observe the block balancing at its center of mass and Mass Theorists who observe the block balancing at the geometric center) should be more likely to appeal to an alternative variable to explain away the conflicting data than children in the Confirming conditions. Lastly, we gave children a new block to make a final balance attempt with. We predicted that children who cannot explain conflicting data should be more likely to revise their beliefs.

We also include a younger group of children with 'non-differentiated' beliefs. In the original balancing studies of Karmiloff-Smith and Inhelder (1974), the researchers suggested that children between four- and six-years have not yet developed a theory of balance. If younger children do not have robust theories of balance, neither a conspicuously weighted block balancing at its center of mass nor a block balancing at its geometric center should be particularly surprising; children should show a novelty preference throughout and should be no more likely to appeal to the magnet in one condition over another.

Methods and Design

Participants. Sixty-two six-and-seven-year-olds (M = 7yrs;1mos; range = 72-96mos,) and thirty-two four-and-five-year-olds (M = 5yrs;2mos; range = 51-68mos,) were recruited from a local urban science museum. Approximately equal number of boys and girls participated (46% girls).

Materials. There were four theory-classification blocks, each made of styrofoam and covered with colored tape, (see Figure 8). Additionally there were three familiarization blocks that were identical blue blocks, each with a larger, heavier side. Test blocks balanced by the experimenter were identical to the familiarization blue blocks; however, the two test blocks each contained a magnet in the base located either in the center of the block or off to the side where the block would actually balance. The balancing apparatus consisted of a rod inserted into a rectangular

wooden base. The novel toy was comprised of a metal key ring with several charms; the ring was placed on a pointed rod and base similar to those of the balances. An opaque bag was used to cover the novel toy.

Procedure. See Figure 8.

        Theory Classification. Children were first given a theory-classification task. In this task, children were presented with three of the four classification blocks in random order and were asked to try to balance each block on the post. We coded whether the child attempted to balance the block at its geometric center or towards the center of mass. The experimenter took hold of the block before the child actually set it on the post so children never observed the outcome of their balancing attempts.

        Familiarization. The child was then shown the three familiarization blue blocks, given a chance to explore the blocks for a few seconds, and was then asked to point to the heavier side of each block. Throughout the classification and familiarization trials, the novel toy was on the table, covered so as to be out of the child's view and off to the right or left side (counterbalanced).

        Play. The experimenter said, "I'm going to try to balance my block here very carefully," and 'balanced' the test block either in the geometric center of the block or over the center of mass. Then the experimenter uncovered the novel toy, moving it to a position equidistant with the block to the child, and told the child, "Go ahead and play with which ever toy you want until I come back." Children were given one minute to play.

        Explanation. After 60 seconds of free-play, the experimenter returned to the table and covered up the novel toy. She returned the test block to its original balanced position and asked, "Why is this block staying up? How come it's not falling over?"

        Final Prediction. Following the child's explanation, the experimenter presented the fourth classification block and asked the child: "Can you balance this very carefully for me, so that it does not tip over?"

Design. Six-and-seven-year-olds were classified as Center or Mass Theorists based on where the child attempted to balance the classification block on at least two of the three trials. Center balances included a 10% margin of error around the center of the block (~1 inch radius from center.) All balances towards the heavy side of the block that fell outside of this margin of error were coded as mass balances. The six-and-seven-year-olds were randomly assigned to either a Conflicting or Confirming condition. Note that the block balanced over the geometric center is
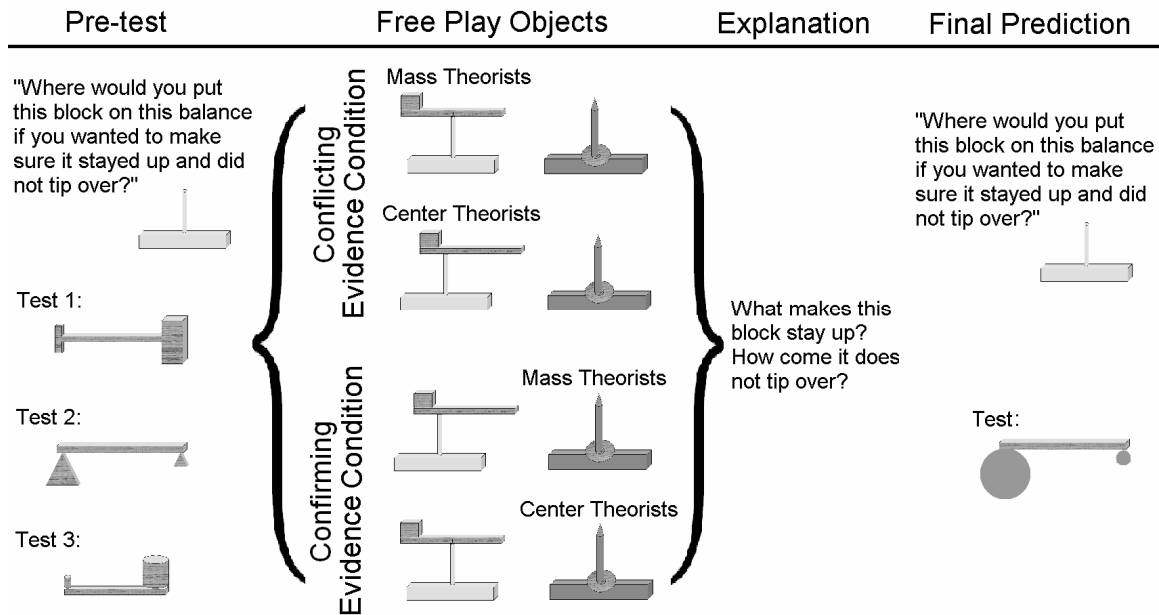
Figure 8: Method and design for Experiments 1 and 3.

the Confirming condition for a center theorist but Conflicting condition for a mass theorist. Conversely, the block balancing over the center of mass is the Conflicting condition for a center theorist and the Confirming condition for a mass theorist. The four-and-five-year-old children were randomly assigned to either the 'Geometric Balance' condition or the 'Center of Mass' condition, (differently named because these children's beliefs did not differentiate between the evidence, and thus neither condition would be considered 'conflicting').

Results of Experiment 1

Four-and-five-year-olds. One child was removed from the study for parental interference. Of the remaining 31 children, 16 were assigned to the Geometric Center condition; 15 were assigned to the Center of Mass condition. There were no age differences between conditions ($t(29) = 0.43$, $p = ns$). The initial predictions of the No Theory children were quite variable, 65% of the children had inconsistent predictions in the classification trials (e.g., picking a different predicted balance location on at least one of three trials) and a third of the children (34%) made at least one prediction that was inconsistent with both Mass or Center Theory (i.e., towards the lighter side of the block.)

68

Play Results. Children were counted as playing with the toys as long as they were touching the toys; we coded the total amount of time each child played with each toy. We analyzed children's play by looking at how long, on average, children played with the balance block.

Overall, there was a significant effect of toy type; children preferred the novel toy in both conditions. We compared how long the children played with each toy in each condition by doing a 2 x 2 mixed ANOVA with play time on each toy as the within subjects variable and condition as the between-subjects variable. Comparisons between the Geometric Center condition and Center of Mass condition revealed a main effect of toy type (averaging across the two conditions, children significantly preferred the novel toy over the balance toy, ($F(1, 29) = 9.43$, $p < .01$)), but no main effect of condition (overall, children played for the same amount of time in each condition ($F(1, 29) = 0$, $p = ns$)), and no interaction ($F(1, 29) = .03$, $p = ns$). This preference for the novel toy held up by condition: children were more likely to play with the novel toy than the balance in the Geometric Center condition ($t(15) = 1.88$, $p < .05$) and marginally more likely in the Center of Mass condition ($t(14) = 1.64$, $p = .06$). In both conditions, a non-significant majority of children played most with the novel toy.

There were no differences across conditions. Children spent the same amount of time playing with the balance toy in the Geometric Center condition as in the Center of Mass condition ($t(29) = .17$, $p = ns$). Additionally, individual children were no more likely to prefer the balance toy in the Geometric Center condition than in the Center of Mass condition ($\chi^2 (1, N = 31) = .02$, $p = ns$).

Explanation Results. Children's explanations uniquely and unambiguously fell into one of four categories: Center Theory consistent explanations (e.g. "It balances because it's in the middle; there's the same length on both sides"); Mass Theory consistent, (e.g.. "There's equal amount of weight on both sides"); appealing to the hidden cause, the magnet, (e.g. "There's something sticky there holding it up, like a magnet"); or Non-Differential (e.g. "It's flat"; "You balanced it slowly and carefully"). A small majority (52%) of No-Theory children gave explanations that were classified as Non-Differential. Of the remaining explanations, appealing to the magnet was the next most likely (29%), with only a minority of children appealing to either the Center Theory consistent (3%) or Mass Theory consistent (16%) explanations. Children were significantly more likely to generate a 'Non-Differential' explanation than chance predicted ((chance = .25), z = 3.21, p(16 or greater) < .01).

There were no differences in explanations comparing across conditions; children were equally likely to appeal to each explanation type (Fisher Exact, $p = ns$). Note this means that children were not more likely to appeal to the magnet in either condition (χ^2 (1, $N = 31$) = .02, $p = ns$).

Final Prediction Data. Children's final balance attempts were also coded as either Mass consistent, Center consistent, or Other. Consistent with Karmiloff-smith and Inhelder's findings that even after a relatively short play period, younger children move from 'No Theory' to 'Center Theory', the majority of final predictive balances were consistent with the Center Theory prediction (83%), though direct comparison of the child's 'change in beliefs' is difficult as children's initial predictions were quite variable. However, the evidence that children observed did seem to play a role in their final predictions. Four of the five 'Mass Balances' that were made as final predictions were generated by children in the Center of Mass condition. And, marginally more children made a Center consistent final balance prediction in the Geometric Center condition (93%) than children in the Mass Consistent condition (67%), (χ^2 (1, N=30)=3.3, $p = .07$).

Six-and-seven-year-olds. Two children were dropped from the study do to parental interference during the experiment. Of the remaining 62 children, 32 were classified as Center Theorists and 30 were classified as Mass Theorists. Consistent with Karmiloff-Smith and Inhelder's (1974) original findings, on average, the Mass theorists (7yrs;5mos) were older than the center theorists (6yrs;10mos), ($t(60) = 4.56$, $p < .0001$). However, there were no differences in age within theory type between conditions, (Center: $t(30) = 1.03$, $p = ns$; Mass: $t(28) = 0.62$, $p = ns$).

Most of the Center Theorists (84%) attempted to balance the block at the geometric center on all three trials, the remaining (16%) children did so on two of the three trials. Also, most of the Mass Theorists (70%) attempted to balance the block closer to the center of mass on all three trials; the remaining children (30%) did so on two of the three trials. Sixteen Center Theorists were randomly assigned to the Confirming condition; 16 to a Conflicting condition; 15 Mass Theorists were assigned to the Confirming condition; 15 were assigned to the Conflicting condition.

Play data. 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 (see Figure 9). To compare the amount of time playing with the blocks, we ran a two-way-between subjects

70
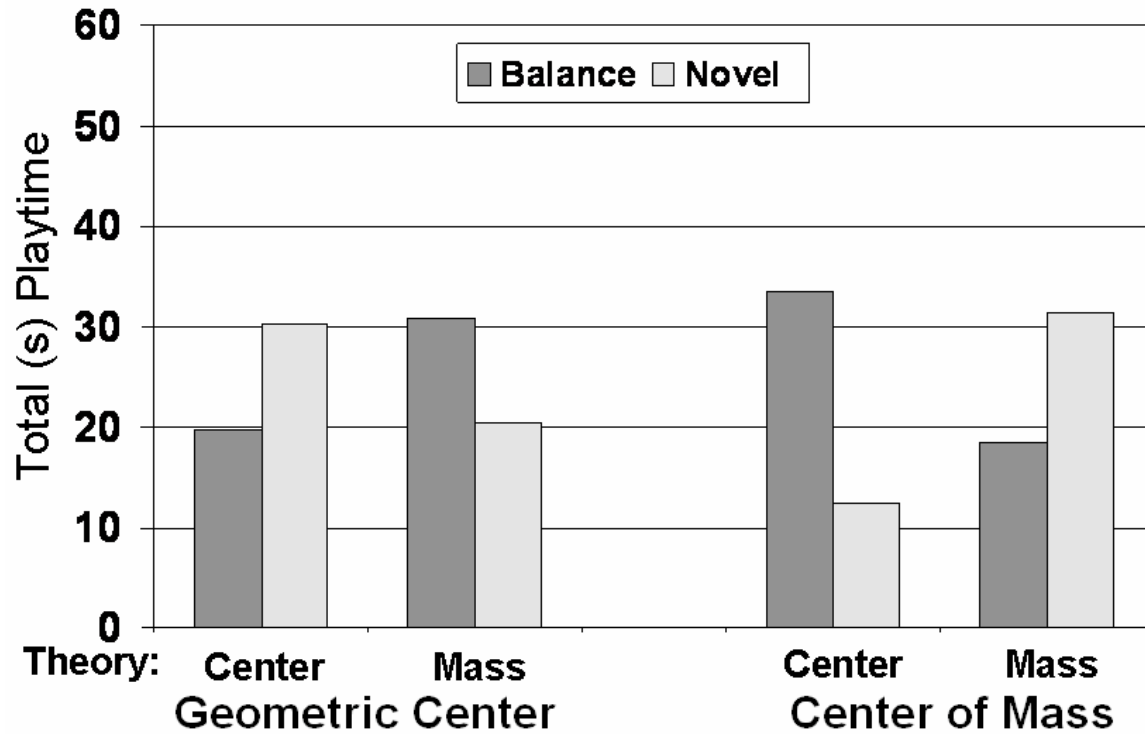
Figure 9: Play results from six-and-seven-year-old children in Experiment 1.

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 ($F(1, 61) = 9.46$, $p < .01$).

We also analyzed whether within condition, children were more likely to play with the novel toy or the familiar balance. Within the Conflicting conditions, Center Theorists played significantly longer with the familiar block than the novel toy ($t(15) = 3.03$, $p < .01$), and Mass Theorists were marginally more likely to play longer with the familiar block than the novel toy ($t(14) = 1.41$, $p = .09$). These results reversed for children Confirming condition, with Mass theorists playing marginally more with the novel toy than the balance block ($t(14) = 1.58$, $p = .07$). Center theorist children in the Confirming condition played equally long with both toys

($t(15) = 1.17$, $p = ns$).   Additionally, more mass theorist children spent the majority of their time playing with the familiar block in the Conflicting condition than in the confirming condition (χ^2 (1, $N = 30$) = 6.7 $p < .01$) and more center theorist children spent the majority of their time playing with the familiar block in the Conflicting condition than in the Confirming condition (χ^2 (1, $N = 32$) = 5.24 $p < .05$).

Explanation Data. Collapsing across conditions, Mass Theorists were marginally more likely to appeal to a Mass Theory consistent explanation than were Center Theorists (χ^2 (1, $N = 62$) = 3.12, $p = .07$) and Center Theorists were more likely to appeal to a Center Theory consistent explanation (χ^2 (1, $N = 62$) = 4.0, $p < .05$). Both Mass and Center Theorists were equally likely to generate Non-differential explanations (χ^2 (1, $N = 62$) = 2.0, $p = ns$) and magnet explanations (χ^2 (1, $N = 62$) = .59, $p = ns$).

Overall, children were significantly more likely to appeal to the magnet as an explanatory variable in the Conflicting condition (61%) then children in the Confirming condition (35%) (χ^2 (1, $N = 62$) = 4.13, $p < .05$).   However, this effect was driven by the Mass Theorists who were significantly more likely to appeal to the magnet in the Conflicting condition (67%) than the Confirming condition (20%), (χ^2 (1, $N = 30$) = 6.62, $p < .01$).   The Center Theorists were equally likely to appeal to the magnet as the explanatory variable in both conditions, Conflicting (56%), Confirming (50%), (χ^2 (1, $N = 32$) = .13, $p = ns$).

Final Prediction Data. Children's final balance attempts were also coded as either Mass consistent, Center consistent, or Other.   One Mass Theorist in the conflicting condition self-terminated the experiment just before the final balance attempt was thus dropped from subsequent analyses. Overall, children were remarkably consistent between initial predictions and their final predictions with the novel block in the Confirming conditions.   Only one child of 31 total changed predictions after observing the confirming evidence.   In contrast, significantly more (12 of the 30 children) changed predictions on the final balance attempt after observing conflicting evidence, (χ^2 (1, $N = 61$) = 12.3, $p < .001$).   These results held up within Theory type: Center Theorists in the Conflicting condition were significantly more likely than Center Theorists in the Confirming condition to learn and make a correct Mass consistent predictions on the final balance attempt, (χ^2 (1, $N = 32$) = 7.4, $p < .01$); and Mass Theorists in the Conflicting condition were significantly more likely than Mass Theorists in the Confirming condition to make Mass inconsistent balances (center or other) on the final balance attempt, (χ^2 (1, $N = 29$) = 5.2, $p < .05$).

Discussion of Experiment 1

Overall, the four-and-five-year-olds' pattern of data support the well established finding that children preferentially explore novel objects over familiar ones. They also support the idea that these younger children do not have a strong (evidence differentiating) belief about balance. Contrasting these results with the six-and-seven-year-old children suggests the influence that children's theories can have in overcoming a preference for stimulus novelty and affecting children's play. Not only do the older children have strong beliefs, but these theories shape their choices in play, suggesting that young children's spontaneous exploratory play is sensitive not just to the perceptual novelty of an object, but also to whether or not observed evidence is consistent with the child's predictions. Six-and-seven-year-olds who observed identical evidence (either the block balancing in the geometric center or the block balancing over the center of mass), showed distinctive patterns of exploratory play dependent on their beliefs. Mass theorists who saw the block balancing in the geometric center found this evidence surprising and thus overcame a preference for novelty and explored the familiar toy more; Center theorists who observed identical evidence did not. Conversely, Center theorists who saw the block balancing at the center of mass found this evidence surprising and explored the familiar toy more than Mass theorists observing identical evidence. Two variables seemed to drive the effect: the child's initial beliefs and the observed evidence.

Additionally, children were able to appeal to the hidden variable (the magnet) discovered during play. Importantly, although the magnet was present in all conditions and all children discovered the magnet during play, Mass theorist children were more likely to appeal to the magnet when the block balanced at a belief conflicting location than when it balanced in a belief consistent location. In fact, of the five mass theorist children who did not directly appeal to the magnet in the Conflicting condition, four were able to explain away the surprising data with other variable explanations consistent with their beliefs about balance such as "Even though this side is smaller, it must weigh the same"; "If this is the middle, it must weight the same on both sides somehow"; and one child notably questioning their observation stating: "Maybe this heavier side is actually closer" (pointing to the center 'balanced' point of the block). The responses from Mass Theorist children support the claim that children are able to spontaneously appeal to alternative variables to explain surprising data—one mechanism that might support conservatism in the face of conflicting evidence.

We believe there are two possible reasons for Center Theorist's failure to differentially appeal to the magnet. First, given that these children were slightly younger than the Mass Theorists and that the magnet is a reasonable explanation for the block's staying up, one

possibility is that the discovery of the magnet during play was so interesting to these younger children that they had more difficulty inhibiting reference to this variable during the explanation phase. Experiment 2 will remove this possible problem by providing Center Theorists with belief conflicting or belief confirming evidence and asking for explanations without the initial free play period that could have lead to discovery and resultant difficulty inhibiting the report of the magnet.

A second possibility, however, is that Center Theorists were learning from their own play, and explanations were thus consistent with a newly acquired belief. That is, explanations were generated after children may have come to different beliefs about balance after play, before explanation. Experiment 2 also controls for this possibility by removing the play period. There are several reasons to consider this proposal that children were learning. First, this explanation is consistent with Karmiloff-Smith and Inhelder's (1974) finding that children rapidly switch beliefs based on the evidence of their own play. Second, if children were learning from play, then the pattern of evidence that children generated during play should predict their explanations. Consider children in the Confirming condition: if these children were learning the correct mass theory during play, then they *should* appeal to the magnet during the explanation phase, when they observe the block balancing at the (now surprising) geometric center. That is, even children in the Confirming condition could generate and learn from conflicting evidence during play *if* they happened to successfully balance the block over the center of mass. Indeed, six children in the Confirming condition successfully balanced the block over the center of mass during play, evidence which could not be accounted for by a magnet (as the magnet was under the geometric center of the block in this condition); all six of these children gave a magnet explanation when the block was rebalanced over the geometric center. In contrast, only 2 of the 10 remaining children (who did not generate surprising evidence, and thus were unlikely to revise their beliefs) gave a magnet explanation when the block was rebalanced over the geometric center. Removing from analyses the six children that generated evidence that they could have learned from, and then comparing only the ten remaining 'evidence consistent' children in the confirming condition to all children in the Conflicting condition, reveals a marginal interaction: more children gave a magnet explanation in the conflicting condition than children who observed only confirming evidence (χ^2 (1, $N = 26$) = 3.3, $p = .07$).

Final predictions also support the claim that children were learning from the evidence, even over the extremely short course of our experiment; six-and-seven-year-old children were significantly more likely to change their predictions on the final balance attempt when the evidence conflicted with their beliefs than when it confirmed their beliefs. One might be

surprised by the changes in final predictions, given that so many children in the Conflicting condition appealed to the magnet as an explanatory variable. However, as previously discussed, note that children also sometimes generated evidence during free play that could not be explained by the magnet. For example, there are two ways in which a Center Theorist in the Conflicting condition could observe surprising data, the block could successfully balance over the center of mass (which could be explained by the presence of the magnet) or the block could fail to balance over the geometric center (which cannot be explained by the magnet). Of the six Center Theorists (all in the Conflicting condition) who changed their final predictions, five attempted to balance the block over the geometric center (and failed), generating evidence that conflicted with their Center Theory that could not be explained away[8]. Of the ten remaining children who did not revise their beliefs, only two children generated theory-inconsistent evidence during play.

In addition to final balances, children's explanations in the Conflicting condition also reflected learning. Half (four of eight) of the Center Theorists in the Conflicting condition, who's initial predictions were Center consistent but who generated Mass consistent evidence during play that could not be explained away, gave Mass consistent explanations. In contrast, none of the eight remaining children, who generated only evidence that was consistent with beliefs or could be explained away, gave Mass consistent explanations. Combined, these results demonstrate that when evidence cannot be explained away, children will genuinely learn to revise their explanations and predictions. In Experiment 3, we more thoroughly test this question by creating Conflicting and Confirming evidence that cannot be explained away with an appeal to magnetism.

Experiment 2: Center Theorist Explanations

The results of Experiment 1 suggest that Mass Theory children are able to explain away surprising evidence by appealing to hidden variables discovered during play. While all children discovered the magnets during play, and magnets were a perfectly good explanation for the blocks staying on the stand, Mass Theorists in the Confirming condition did not appeal to the alternate variable. In contrast, younger, Center Theorist children were equally likely to appeal to magnets in both conditions, (presumably either because they were unable to inhibit reporting their discovery during the explanation phase or because they revised their beliefs from their own play). We hypothesized that if children were made aware of the *possibility* of magnets, but did not

---

[8]   Note that the one remaining child who did not generate a center balance actually indirectly provided conflicting evidence for himself: he successfully balanced the block upside-down over the center of mass, where the magnet was not present and thus could not explain the surprising results.

spontaneously discover the magnets themselves (or generate evidence that could lead to learning), that children would be motivated to appeal to magnetism as an explanation for the block's balancing in Conflicting conditions, but would not appeal to the magnets in Confirming conditions. Thus, children were first given an explanation warm-up task and familiarization with objects that sometimes contained magnets. Then they were shown a block balancing either at its geometric center (confirming) or its center of mass (conflicting) and asked to explain what made the block stay up. If Center Theorist children can appeal to hidden variables to explain away surprising evidence, then they should be more likely to hypothesize that a magnet is present following evidence that conflicts with beliefs than following evidence that confirms beliefs.

Methods and Design

Participants. Forty six-and-seven-year-olds were recruited from a local science museum. Eight children produced Mass Theory consistent initial predictions (see below) and were not included in analyses, resulting in thirty-two 'Center Theorist' children (*M* = 6yrs11mos; range = 72-97mos). Approximately equal number of boys and girls participated (x% girls).

Materials. The balancing apparatus, theory-classification blocks, and blue familiarization and test blocks from Experiment 1 were used; two of the blue blocks had the hidden magnets (one at the geometric center and one at the center of mass) and the other was the inert block. Additionally, three sets of explanation priming toys were used: 2 identical bells (except that one made noise and one did not); 2 small toy cars (one which rolled and one that did not). Additionally 2 identical small cube-blocks (one which was magnetic and one which was not); as well as 3 small mettle clips where used in the magnetic priming.

Procedure. The procedure involved four phases.

    Theory Classification Task. As in Experiment 1, children were first given a theory-classification task and we coded whether the child attempted to balance the blocks at the geometric center or towards the center of mass.

    Explanation Warm-up Task. Children were then given a warm-up task to help them practice generating explanations and to get familiar with the experimenter. During the warm-up, the experimenter first brought out the bells, and showed children that one bell rang but the other did not. The experimenter asked the child "Why do you think this bell rings and this one doesn't?" If the child elicited an explanation then the experimenter moved on to the toy cars. If the child could not explain, the experimenter prompted the child a second time "Can you come up

with any ideas for why this bell works and this one doesn't?" If the child still refused to guess, the experimenter said, "maybe because this one does not have the clapper and this one does, or maybe because it's stuck." The child was then shown that one toy car rolled and one toy car did not roll and was asked "Why do you think this one rolls and this one does not?" Again, if children generated an explanation, the experimenter moved on (to the magnets familiarization) and if not the child was again prompted and finally provided with feedback if no explanation was given following prompting.

Magnets familiarization. Children were shown the two blocks and metal clips, one block was held just over the clips and picked them up to show that it was magnetic, when the other block was held over the clips, it did not pick them up. Children were told, "See how this block picks up the clips and this one does not? That because this block has a magnet in it which makes them stick but this one doesn't. I'm going to put the magnet block over here (to the right) and the non-magnet block over here (to the left)" The experimenter then brought out two of the blue blocks (one which had the magnet and one which was inert). With one block the experimenter showed that it picked up the mettle clips (under the position of the magnet) and then showed the other block did not pick up the clips. The experimenter asked, "Can you tell me, which block do you think has a magnet in it and which do you think does not have a magnet in it?" After the child correctly identified the magnetic and non-magnetic blocks, the experimenter asked the child to sort the magnetic block (to the right) and the non-magnetic block (to the left). The magnetic object and non-magnetic object piles remained on the right and left of the table.

Test Phase. After sorting the two blue blocks, the experimenter pulled out the third blue block and showed it to the child, asking "Can you show me, which is the big, heavy side and which is the light side?" Then the experimenter carefully 'balanced' the block on the stand for the child. Half the children saw the block balancing at the center of mass (Conflicting condition) and half the children saw the block balancing at the geometric center (Confirming condition). Children were then prompted with "Can you tell me, why is this block staying up? How come it's not falling over?" Following the child's explanation, the experimenter asked the child "Can do you tell me, which group do you think this belongs in? The group with the magnetic blocks or the group that is not magnetic?"

Results of Experiment 2

Initial prediction results were coded as in Experiment 1. Eight children generated Mass Theory consistent predictions on at least two of the three the initial predictions and were not included in subsequent analyses. Of the remaining thirty-two Center Theorists, sixteen were

randomly assigned to the conflicting condition, and sixteen were assigned to the confirming condition. Of the Center Theorists, 81% of children attempted to balance the block at the geometric center on all three trials, the remaining 19% of children did so on two of the three trials. All children generated responses in the initial explanation phase and were able to successfully sort the initial magnetic and nonmagnetic blocks.

Explanation Data. Children's explanations were coded as in Experiment 1. Children were significantly more likely to appeal to the magnet in the Conflicting condition (63%) than in the Confirming condition (19%), (χ^2 (1, $N = 32$) =6.35, $p = .01$). Children were also significantly more likely to make a Center Theory consistent explanation in the Confirming condition (50%), (χ^2 (1, $N = 32$) =10.67, $p = .001$), but there were no differences between Mass consistent (Confirming: 6%; Conflicting: 25%) or Non-differential (Confirming: 25%; Conflicting: 13%) explanations between conditions (Mass: $\chi^2 =2.13$, $p = ns$; Other: $\chi^2 =.82$, $p = ns$).

Sorting Data. Two children in the Confirming condition did not complete the sorting task and have been dropped from subsequent analyses. Children in the Conflicting condition were more likely than children in the Confirming condition to sort the balanced block with the magnetic objects, (χ^2 (1, $N = 30$) = 4.74, $p < .05$). In the Conflicting condition the majority of children (81%) sorted the block as magnetic, but less than half of the children (42%) sorted the block as magnetic in the Confirming condition.

Discussion of Experiment 2

Overall, the results of Experiment 2 and the results of the Mass Theorists from Experiment 1 support the claim that children's explanations are dependent on their beliefs and the evidence observed. Children can explain away surprising evidence by appealing to hidden variables. Though a block balancing at its center of mass is relevant evidence for a child with the incorrect belief (that the block should balance at the geometric center), children were able to explain away this evidence by appealing to the magnet.

Though Center Theorists in Experiment 1 were as likely to appeal to the magnet in Conflicting and Confirming conditions, we do not know for sure whether it was because their attention was drawn to the magnet due to its surprising discovery or because they had already revised their beliefs by the explanation phase of the experiment. However, when children were unable to spontaneously discover the magnet and unable to generate evidence to revise their beliefs, they rationally appealed to the magnet: explaining away the block's balancing following theory inconsistent evidence but not following theory consistent evidence.

Experiment 3: Center Theorist Learning

The results from Experiment 1 and 2 suggest that children's play and explanations are rationally driven by both their prior beliefs and the evidence they observe. Children's final predictions (and arguably explanations) from Experiment 1 also suggest that children are more likely to revise their beliefs given inconsistent evidence. However, because the magnet was present during exploration, and could be (and was) used to explain away the surprising evidence, children's learning in Experiment 1 may not have been as robust as if they observed only evidence they could not explain away. In Experiment 3, we replicate the procedure of Experiment 1 with Center Theorists[9] who can observe *only* theory consistent or *only* theory inconsistent evidence during play. That is, we use two new L shaped test blocks, one regular block that only balances over it's center of mass and one block which is surreptitiously weighted so that it only balances over its geometric center, (that is, although the block is identical to the other blocks, which clearly have a larger side and a smaller side, this block is subtly weighted so that its center of mass is over the geometric center of the block). Thus, in the confirming condition the block balances at the geometric center and does not balance at the center of mass; in the conflicting condition the block balances at the center of mass and does not balance at the geometric center.

Methods and Design

Participants. Thirty-two six and seven-year-olds (range = 74 to 94mths, *M* = 6yrs;11mths) were recruited from a local urban area science museum. Equal number of boys and girls participated (50% girls).

Materials. The materials were identical to experiment 1a, with the exception of the two test blue blocks. These blocks had a magnet that was placed at the top of the block, so that it would not interfere with balancing or be used to explain away surprising evidence, but so that it would be as equally interesting as the test blocks used in Experiment 1. The blue block for the Confirming condition was additionally surreptitiously weighed so that although it looked heavier on one side, it would actually balance over the geometric center. Additionally, 6 new colored blocks, 3 of which were clearly equally weighed and balanced at the center, 3 of which were clearly unevenly weighed and balanced towards the side were used at the end of the Confirming condition.

---

[9]   Due to the robust learning in Experiment 1, we felt it would be inappropriate to generate a condition that challenged the correctly learned beliefs of Mass Theorist children.

Procedure. The procedure was identical to Experiment 1, with one exception. After the final prediction test block at the end of the experiment, children in the Confirming condition were asked to sort the additional 6 new colored blocks into two piles—blocks that had a big, heavy side and blocks that were not heavier on one side. Children were then asked to sort the test blue block. This allowed us to make sure that children did not discover the surreptitious weighting during play and instead continued to believe the block to be heavier on the larger side (as the block in the Conflicting condition was weighted). All children 'passed' the final sort, sorting the blue block with the other objects that were heavier towards the larger side.

Results of Experiment 3

One child in the Confirming condition was dropped and replaced for failure to play during the play period. There were no age differences between groups ($t(30) = .36$, $p = ns$). Most of the Center Theorists (84%) attempted to balance the block at the geometric center on all three trials, the remaining (16%) children did so on two of the three trials.

Play data. Replicating the results of Experiment 1, children were more likely to explore the familiar toy (the block) when the evidence conflicted with their theories than when it confirmed their theories. We ran a two-way-between subjects ANOVA on play time with type of evidence as the between subjects variables and time spent playing with the blocks as novel toy as the dependent measures. Comparisons between conditions revealed no main effect of condition (children in the Conflicting condition overall played as long as children in the Confirming condition), and no main effect of toy type (averaging across conditions children played as long on average with the blocks as with the novel toy); however, comparisons revealed a significant interaction, ($F(1, 60) = 6.05$, $p < .05$); children spent more time playing with the block over the novel toy when evidence conflicted with beliefs than when evidence was consistent with beliefs. Within the Conflicting condition, Center Theorists played significantly longer with the familiar block than the novel toy ($t(15) = 1.83$, $p < .05$), but Center Theorists in the Confirming condition played equally long with both toys ($t(15) = .83$, $p = ns$).

Explanation Data. Children's explanations were coded as with Experiment 1. Although there was no magnet present in the block at the point of balance, one child in the Conflicting condition did spontaneously explain that the block stayed up because of a magnet. All other responses fell uniquely and unambiguously into the Mass consistent, Center consistent, or Non-differential explanation categories. We conducted a Fisher Exact test on explanation types by condition and found a significant interaction (N = 32; $p < .05$). This was primarily driven by the fact that significantly more children in the Conflicting condition appealed to a Mass consistent

explanation following play (56%) than children in confirming evidence condition (19%), (χ^2 (1, $N = 32$) = 4.8, $p < .05$). In contrast, marginally more children made Center consistent explanations following confirming evidence (31%) than following the conflicting evidence (6%), (χ^2 (1, $N = 32$) = 3.28, $p = .07$). There were no differences between conditions for Non-differential explanations (χ^2 (1, $N = 32$) = 1.2, $p = ns$).

Final Prediction Data. As with Experiment 1, children's final balance attempts were coded as either Mass consistent, Center consistent, or Other. The majority (63%) of Center Theorists in the Conflicting condition changed their final prediction to a Mass consistent prediction. In sharp contrast, no child in the Confirming condition changed predictions on the final balance; all children made a Center Theory consistent prediction. Significantly more children changed predictions on the final balance attempt after observing conflicting evidence than children observing confirming evidence, (χ^2 (1, $N = 32$) = 14.5, $p = .0001$). Interestingly, following the conflicting evidence, the majority of children who made a Mass consistent final prediction also made a Mass consistent explanation (8 of 10 children), while only 1 of the 6 children who made a Center Theory final prediction gave a Mass Theory explanation.

Discussion of Experiment 3

Children's explanations and final predictions support the claim that children were genuinely learning from exploratory play. Following experience with a block which only could balance at its center of mass, and which had no auxiliary variables to explain away the surprising evidence, children were significantly more likely to produce mass consistent explanations and mass consistent final predictions on the blocks than children who only observed evidence that confirmed their beliefs. Importantly, we replicated the finding that both the evidence observed and children's beliefs mediate whether children will continue to explore a familiar toy (the balanced block), or will explore a novel toy.

One might be concerned that final predictions were not indicative of children's learning but rather indicated that children were merely imitating the experimenter's initial balance. That is, did children in the Conflicting condition make a final 'mass consistent' prediction simply because they were emulating what they had just observed? To be sure, witnessing the adult successfully balance the block over the point of mass provides compelling evidence for how the block should balance. However, note that children's explanations were also more likely to be Mass consistent following conflicting evidence; it is difficult to posit a mechanism by which imitation could influence children's explanations.

Additionally, we have strong reason to believe that children were also sensitive to the evidence they generated during play and their ability to explain away the evidence: Center theorists in the Conflicting Condition of Experiment 1 also witnessed the experimenter balance the block over the center of mass. Importantly, seven children in that condition only observed evidence that they could explain away (balanced the block only over the center of mass where there was a magnet, and never over the geometric center, where there wasn't a magnet.) These children should have imitated the experimenter's balance as well; however, none of these seven children made a Mass consistent final prediction. In contrast, the majority of children in the Conflicting condition of Experiment 3 (who could not explain away the evidence) made a mass prediction; this interaction is significant (χ^2 (1, $N = 23$) = 7.7, $p < .01$). This comparison suggests that an explanation appealing only to imitation does not account for children's final mass predictions. Overall, these results do suggest that young children are sensitive to the evidence they observe and generate during play, and can flexibly revise beliefs when conflicting evidence cannot be accounted for.

General Discussion

We began by considering the 'child as a scientist', suggesting that both prior beliefs and evidence should play a critical role not only in supporting causal inferences and explanations, but also in guiding children's exploratory play. Such theory-guided play is arguably a form of optimal exploration: it suggests that children may play more where there is indeed something to be learned: either that there is a hidden variable that might explain the surprising evidence, or that something about the theory is incorrect. Consistent with this claim, we found that children who observed conflicting evidence were more likely to explore a familiar block, (overcoming a novelty bias). This was not the case for children who observed theory-consistent evidence or for children who did not have strong differential beliefs (No theory children in Experiment 1).

If theories support effective exploration, then children may spontaneously discover evidence that can help them revise their causal beliefs. Indeed, children's play led to the discovery of a hidden variable (a magnet) that could explain the block's balance. Importantly, although the magnet was in all cases a reasonable explanation for the block's balancing, children only appealed to the magnet as an explanatory variable when evidence conflicted with their beliefs, but not when it confirmed their beliefs. Additionally, when an alternative variable was not present, (as in Experiment 3), children were more likely to revise their beliefs as compared to children who could explain away the surprising evidence (children in Experiment 1 who generated the same evidence but could account for the balancing by the magnet). Taken together,

these results provide a rational account for the idea that belief revision should be at once flexible (to permit learning) and conservative (to prevent misleading data from overturning strongly held beliefs).

These results may appear to conflict with previous work that argues that children have a relatively impoverished ability to learn from evidence, revise their beliefs, and construct informative interventions (e.g. see Kuhn, 1989). However, the demands of the Kuhn et al. studies required children to be meta-cognitively aware of their theories. While children (and even lay adults) may lack such metacognitive awareness (and thus be unable to design controlled experiments), children may nonetheless, at least implicitly, recognize when evidence conflicts with their prior beliefs. This research suggests that when children do perceive a conflict between their theories and patterns of evidence, they are motivated to explore. Additionally, they can explain away surprising evidence and revise beliefs when no alternative explanations are available.

Looking time paradigms (where infants look longer at novel or surprising events) may also seem somewhat analogous to the work here. For instance, one might be puzzled by the finding that infants as young as 12 months will look longer at an object whose center of mass is not supported (Baillargeon, Needham, DeVox, 1992), yet our subjects, who are more than six years older, do not seem to perceive a violation. In this respect, our study is consistent with many that have found a distinction between children's performance on looking-time and action-oriented tasks (e.g. see Onishi & Baillargeon, 2005). One key difference between the paradigms may be whether evidence is surprising because it is novel or whether evidence is surprising because it violates prior beliefs. An event might be uncommon in the course of everyday experience and lead to longer looking, without requiring the subject to posit any theory of how things *should* be. Recent extensions of this work find that children with differential beliefs of balance will predictively look to different locations of an unevenly weighted block, just before it's balanced (Bonawitz, Brenman, & Schulz, in prep). These results, combined with the myriad of developmental data suggesting that children's apparent understanding of a concept depends on whether the dependent measure involves looking or acting (e.g. Hood, Cole-Davies, & Dias, 2003; Ahmed & Ruffman, 1998) provide a puzzle for future work to explore.

Were younger children really pretheoretical?

Though younger children are able to balance using proprioception, we suggest that there are three reasons to believe the four-and-five-year-old children were genuinely pretheoretical in their explicit understanding of balance (see Karmiloff-Smith, 1992) with respect to the

center/mass distinction. First, the ages of these children align with the ages of children in the original Karmiloff-Smith & Inhelder studies. In these studies, though children gradually learned to make differential balances over the course of the experiment (as did the No Theory children in our study), the initial predictions of children were used to classify their beliefs and the mean ages for these different groups were consistent across experiments. In Karmiloff-Smith and Inhelder, Non-theorists were classified between 4-6yrs, the Center Theorists between 6-7.5yrs, and the Mass Theorists beginning at 7.5yrs; in Experiment 1, our mean ages were similarly 5;2 for the No Theory children, 6;10 for the Center Theorists, and 7;4 for the Mass Theorists.

Secondly, the initial predictions of the No Theory children were significantly more variable than the predictions of the Center and Mass Theorists in Experiment 1, with 66% of the No Theory children generating inconsistent predictions across the three classification trials as compared to only 16% of the Center Theorists and 30% of the Mass Theorists, (No Theory vs. Center: (χ^2 (1, N =63) = 15.7, p < .0001); No Theory vs. Mass: (χ^2 (1, N =61) = 7.28, p < .01)). No Theory children were also significantly more likely to make a prediction that was inconsistent with both Mass and Center theories (e.g. balancing towards the lighter side of the block), whereas only once did this occur in all the balance attempts of all the older children (χ^2 (1, N =93) = 18.6, p < .0001).

Thirdly, while 91% of Center and Mass Theorist explanations appealed to one of the first three classification schemes, the majority of No Theory (52%) children gave explanations that were classified as non-differential. Children were significantly more likely to give Non-differential explanations in the No Theory condition than in the Center Theory condition (χ^2 (1, N =63) = 4.6, p < .05) or the Mass Theory condition (χ^2 (1, N =61) = 7.6, p < .01), and were significantly less likely to give Mass or Center Theory consistent explanations the older children (χ^2 (1, N =93) = 4.7, p < .05) suggesting that the No Theory children genuinely had weaker beliefs about the relevant dimensions of this task than the other groups of children.

Of course, as previously noted, these younger children do have some beliefs about how blocks should balance. No Theory children would certainly be surprised to see a block floating in mid-air, or a block 'balanced' on an extreme edge. In fact, the explanations of many of these children seem to suggest that children might have a theory that contact between flat surfaces is required for balance. Many children gave explanations such as, "It stays up because it's flat", and "You set it on the circle part which is smooth", and "It's even on the bottom." Indeed, a third (31%) of the No Theory children's non-differential explanations involved appealing to flatness. Additionally, work on children's predictions about balance scales (e.g. see Siegler, 1976) suggest that even these younger children may be able to employ rules to help make balancing predictions.

For example, Siegler (1976) found that five-and-six-year-old children had difficulty attending to more than one dimension of the blocks, but could at least make the prediction that as more weight is added to one side of the block, the balance may start to tip. However, this task is importantly different from our task because it draws attention to weight as a relevant variable that is being added to, and therefore changing, the system. In our task, the block is already weighted towards one side, so we do not draw attention to it as a potentially relevant variable. Additionally, rather than making a simple forced choice prediction ("Will the block go up, down, or stay the same?") children must make a prediction about where along the fulcrum the block will balance. Importantly for our study, regardless of the precise nature of these younger children's beliefs, in general both the geometric center and center of mass evidence are equally consistent and therefore equally 'uninteresting' to the youngest children; this leads to the standard novelty bias in exploratory play, which we do not observe for older children in conflicting conditions.

Conclusions

This chapter presented a case of causal uncertainty; evidence strongly favored the a priori unlikely belief and disfavors the a priori likely belief. Because children had strong beliefs about balance, evidence that conflicted with these strong beliefs arguably created uncertainty between candidate beliefs, whereas evidence that confirmed beliefs relieved uncertainty. By contrasting children with initially different prior beliefs, we demonstrated that the *interaction* of observed evidence and specific beliefs led to differential uncertainty and thus differential exploratory play. In the next chapter, we investigate a second case of causal uncertainty, when evidence and prior beliefs fail to disambiguate potential causal structures.

A complete understanding of the processes that support theory development and theory change remains a challenge to the field. However, the hallmarks of good scientific discovery seem to be present even in children's play, explanations, and learning. Theory guided exploration may play an important role in helping children generate relevant evidence and even sometimes leads to the discovery of (and rational appeal to) hidden variables. Although processes as complex and noisy as children's play have rarely seemed amenable to formal principles, the work in this chapter suggests that even in play, children are able to rationally weigh evidence and balance it with prior beliefs.

Chapter 5

As previously noted, although young children do not design controlled experiments, in simple contexts they seem to recognize the difference between informative and uninformative evidence (Masnick & Klahr, 2003; Sodian, Zaitchik, & Carey, 1991) and they can use patterns of evidence to make predictions, interventions, and even counterfactual claims (Gopnik & Schulz, 2004; Gopnik, Sobel, Schulz, & Glymour, 2001; Kushnir & Gopnik, 2006; Schulz & Gopnik, 2004; Shultz & Mendelson, 1975; Siegler & Liebert, 1975; Sobel, 2004; Sobel & Kirkham, in press). In Chapter Four, we demonstrated that children are sensitive, not just to the novelty or perceptual complexity of stimuli, but to evidence that conflicts with strong prior beliefs.

In this chapter, we look at cases when children don't have strong differential beliefs. We hypothesized that children's exploratory play might also be affected by the quality of the evidence they observe. We predicted that preschoolers would distinguish confounded and unconfounded evidence and would engage in more exploratory play when evidence failed to disambiguate the causal structure of events. If children systematically engage in more exploratory play when causal evidence is confounded, then even if even if children do not generate controlled experiments, they might isolate the relevant variables in the course of free play and generate the type of evidence that could support accurate causal learning.
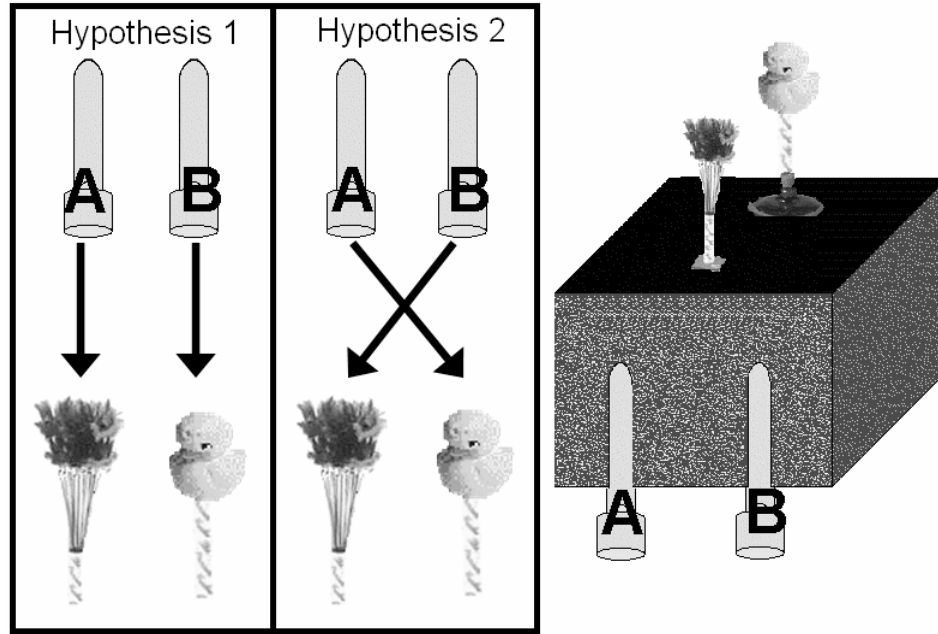
Figure 10: Two a priori equally likely hypotheses to describe the causes of the toy duck and straw puppet. (other hypotheses are also plausible -- for instance, that both levers generate both effects or that the levers interact)

To test this, we created a new toy box with two levers. The two levers can be depressed simultaneously such that both a toy duck and straw puppet pop-up at the same time and the location of the duck and puppet are ambiguous with respect to the levers. A number of possible causal hypotheses may explain the pattern of results, (e.g. either lever might make the duck go, might make the puppet go, one lever might make both toys go, or levers might interact, etc.). Because the position of the toys is ambiguous with respect to the lever, hypothesis one and hypothesis two are arguably a priori equally likely. In contrast, the two levers can also be depressed in alternating sequence, such that the causal structure of the toy (which lever causes which toy) is disambiguated. Children were introduced to this toy and shown either confounded or unconfounded evidence about the causal structure of the toy. We removed the toy and then returned it along with a novel toy. We allowed the children to play freely for sixty seconds.

According the Bayesian rational analysis of this problem, because the two most likely hypotheses (e.g. see Figure 10), are a priori equally plausible and the evidence in the Confounded condition fails to provide support for one hypothesis over the other, there is causal ambiguity. In contrast, the evidence in the Unconfounded condition strongly supports a single hypothesis. This provides a test of the second case of causal ambiguity presented in Chapter Four, when children do not have a prior strong differential beliefs.

Experiment

If children are sensitive to the causal ambiguity, then the kind of evidence they observe should affect their patterns of exploratory play. We predicted that children who observed confounded evidence, (evidence that is equally likely to be observed by either of the two possible causal hypotheses), would preferentially play with the familiar toy. However, children who observed unconfounded evidence (evidence that strongly favored a single causal hypotheses), would show the standard novelty preference and play primarily with the novel toy.

Method

Participants. We recruited 64 preschoolers (mean age: 57 months; range: 48 – 70 months) from the Discovery Center of a metropolitan Science Museum and from urban area preschools. Sixteen children were tested in each of four conditions: a Confounded evidence condition and three Unconfounded conditions, described below. Approximately equal number of boys and girls participated in each condition (45% girls overall). While most of the children were white and middle class, a range of ethnicities and socioeconomic backgrounds reflecting the diversity of the local population were represented.

Materials. Two boxes were constructed from 15 cm x 15 cm balsa boards. One box had a single lever and was covered in yellow felt. The other box had two levers and was covered in red felt. On the yellow box a small (5 cm high) fuzzy, duck toy was attached to a dowel 20 cm in length and 1 cm in diameter that passed through a small hole in the side of the box. The dowel acted as a lever. When the dowel was depressed on the outside of the box, the inside end moved upwards, causing the duck to pop up through a slit in the felt on the top of the box. The construction of the double-lever box was identical, except there was a second lever on the side of the box adjacent to the first lever. On this second lever a small L-shaped bracket was attached to a (7 cm high) puppet made of drinking straws, so that when the second lever was depressed the straw puppet could 'pop-up' without affecting the movement of the first lever. The ends of the two levers were less than 40 cm apart and were easily manipulated, both separately and simultaneously, by preschool children.

Procedure. Children were tested individually in a quiet corner of their preschools or in the Discovery Center. The experimenter sat next to the child at a table. Both boxes were on a far corner of the table and were covered with a cloth so the child could not see them. The experimenter said, "We're going to play a game today." The experimenter brought the red, two-lever box out from under the cloth and introduced it to the child.

In the Confounded condition, the experimenter said, "You push down your lever and I'll push down my lever at the same time. Ready: one, two, three, down!" When both levers were depressed, a duck and a straw puppet popped out of the middle of the box. The spatial locations of the duck and the puppet were uninformative about their causal relationships with the lever; that is, the objects appeared in the middle of the box so it was not possible, just by looking, to determine which lever controlled which objects. After approximately two seconds, the experimenter said, "One, two, three, up!" The experimenter and the child simultaneously released the levers and the duck and puppet disappeared from view. Counting aloud was an effective means of coordinating the child's actions with the experimenter's so that the onset and offset of events appeared simultaneous; pilot work established that even adult observers failed to perceive temporal cues that would disambiguate the causal structure of the toy. The procedure was then repeated twice more, so that in total, both levers were pushed three times and both effects (the duck and the puppet) occurred three times. Because the two candidate causes were always manipulated simultaneously, the evidence failed to disambiguate the many possible causal structures that might underlie the event (either lever might activate the duck or the puppet, one or both levers might activate both, or the levers might interact).

The Unconfounded/Matched for Effect condition was designed to replicate the effects of the Confounded condition but with an unambiguous causal structure. The Unconfounded/Matched for Effect condition was identical to the Confounded condition except that the child and the experimenter pressed and released their levers simultaneously only twice. On the next trial, the experimenter said, "Let's take turns." (The order of turn-taking and the particular effect was counterbalanced between participants.) "You go ahead, one, two, three, your turn!" The child pushed his lever and just the duck popped up. The experimenter then said, "Now it's my turn." After the child released his lever, the experimenter counted "One, two, three", pushed her lever, and just the puppet popped up. Thus, as in the Confounded condition, each effect (the duck and the puppet) occurred three times, however, this evidence fully disambiguated the causal structure of the toy: the child could see that one lever activated the duck and the other activated the puppet.

Conceivably however, the exposure to the additional trial (as indicated by the "one, two, three" counting ritual) might decrease children's interest in the familiar toy. The Unconfounded/Matched for Trials was designed to control for the possibility that the additional trial bored the children. The condition was identical to the Unconfounded/ Matched for Effect condition except that in this condition, the child and the experimenter pressed and released their levers simultaneously only once. On the second trial, the child pressed his lever by himself; on the third trial, the experimenter pressed her lever by herself. Thus each effect occurred twice, however, as in the Confounded condition, there were three distinct trials. Again, the evidence fully disambiguated the causal structure of the toy.

Alternatively, children might play more with the familiar toy in the Confounded condition than the Unconfounded conditions because they were allowed to play independently with the toy in the Unconfounded conditions but not in the Confounded condition. To control for this possibility, children were tested in an Unconfounded/No Independent Play condition. In this condition, the experimenter and the child pressed and released their levers simultaneously once. Then the experimenter pushed one lever and just the duck popped up. She released that lever and pushed the other lever and just the puppet popped up. The experimenter and the child then pressed and released their levers simultaneously a second time. As in the Confounded condition, the child never had a chance to manipulate the toy without the experimenter, however, this evidence fully disambiguated the causal structure of the toy. There was no significant difference in the length of time children were exposed to the effects of the familiar toy in the Confounded condition and any of the Unconfounded conditions (Confounded: mean = 12.1 seconds; Unconfounded/Matched for Effects: mean = 13.5 seconds; Unconfounded/Matched for Trials: mean = 11.6 seconds; Unconfounded/No Independent Play: mean = 13 seconds; for each comparison $t(30)$, $p = ns$).

After the child observed the evidence, the experimenter returned the red box to the far end of the table and uncovered the novel yellow box. The experimenter then rotated the table so that the boxes were just out of arms' reach of the child (so the child had to stretch to reach either box). The boxes were located approximately two feet apart from each other (left/right position of the boxes counterbalanced between children). The experimenter said, "I'll be back in just a minute. Go ahead and play" and walked out of the child's line of sight. After 60 seconds, the experimenter returned, thanked the child for participating and ended the experiment.
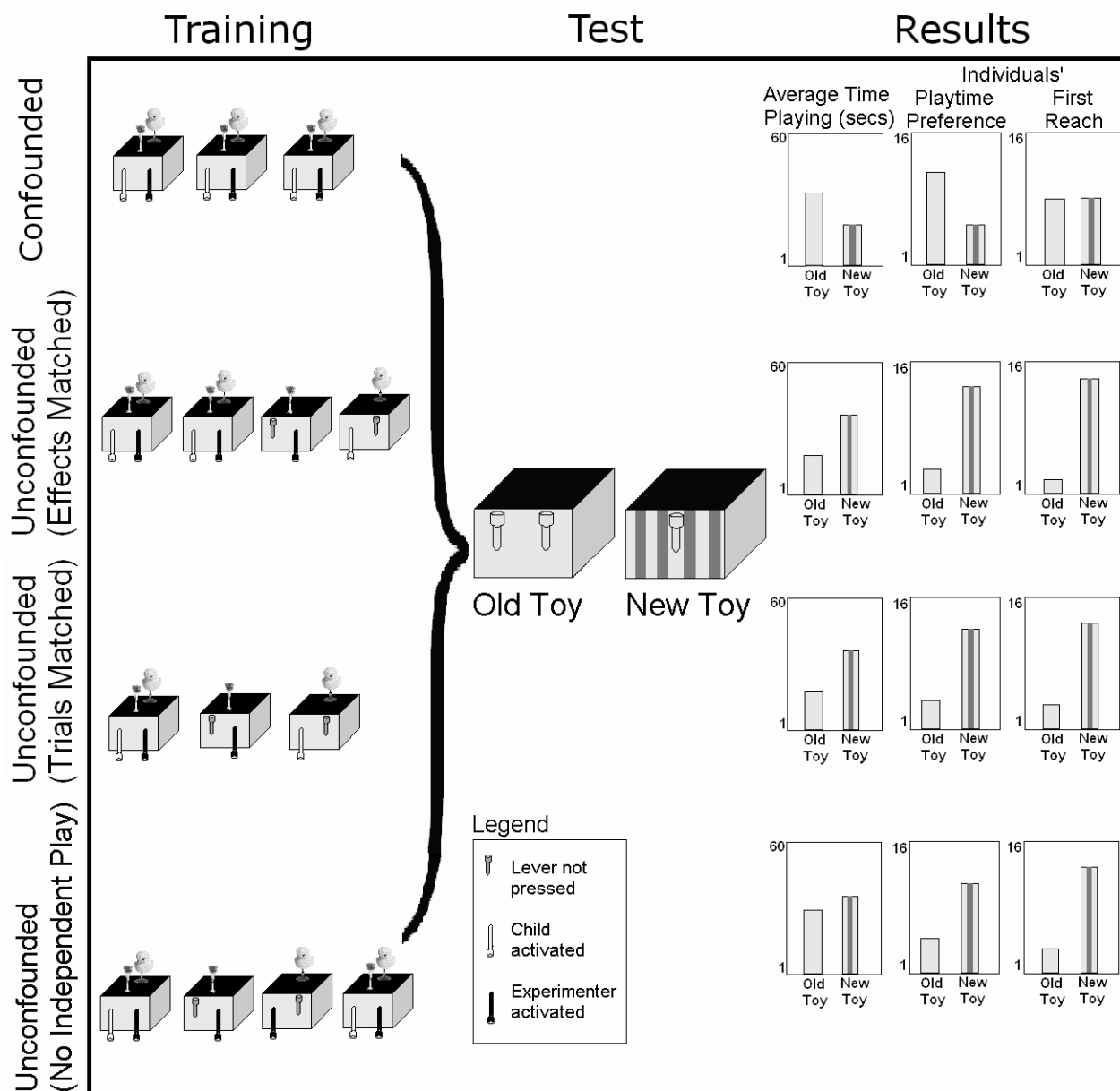
Figure 11: Method, design, and results of experiment.

Results and Discussion

Children were counted as playing with a box as long as they were touching the box and we coded the total amount of time that each child played with each box. We analyzed children's exploratory play in three ways: we looked at whether, on average, children played longer with the familiar box or the novel box; we looked at how many individual children preferentially played with each box, and we looked at whether children's first reach was to the familiar box or the novel box. Additionally, we coded children's actions in the Confounded condition to see whether children who played with the familiar toy spontaneously disambiguated the evidence. Children were counted as fully disambiguating the evidence if, in the course of their free exploratory play, they depressed and released each lever separately at least once. If children only isolated one of the two levers, if they only ever moved both levers together, or if they engaged only in unrelated exploratory play (e.g., reaching in the box; shaking the box), they were counted as failing to fully disambiguate the evidence.

All data were coded by the second author and recoded by a blind coder[10]. Inter-coder agreement on children's play time was high across all conditions (r = .949); coders agreed perfectly on children's first reach and whether or not children fully disambiguated the evidence (% agreement = 100). If children played for less than 15 seconds overall, they were dropped from the study and replaced. One child was replaced in the Confounding condition; no children were replaced in any other condition.

By all three measures, children were more likely to explore the familiar toy in the Confounded condition than in the Unconfounded conditions; there were no differences between the three Unconfounded conditions on any analysis (see Figure 11). We compared how long children played with each toy in each condition by doing a 2 x 4 mixed ANOVA with play time on each toy as the within-subjects variable and condition as the between-subjects variable. There was an interaction between condition and toy preference, $F(3, 60) = 11.64$, $p < .0001$. There was also a main effect of toy type, $F(3, 60) = 6.53$, $p < .05$ suggesting that, collapsing across the four conditions, children preferred the novel toy. There was no main effect of condition, suggesting that children in each group played for the same amount of time overall (mean playtime across conditions was 27.6 seconds per toy).

To follow-up on the omnibus ANOVA, we did pairwise analyses of the four conditions. Each analysis was a 2 x 2 mixed ANOVA with play time on each toy as the within-subjects

---

[10]   The clips for two children in the Unconfounded/No Independent Play condition were lost due to technical error after the original coding. These two children were not recoded.

variable and condition as the between-subjects variable. Comparisons between the Confounded condition and each Unconfounded condition revealed no main effect of play time (averaging across the two conditions, children did not prefer one toy to the other) and no main effect of condition (overall, children played for the same amount of time in each condition), but did reveal a significant interaction: children spent more time playing with the familiar toy in the Confounded condition than in the Matched for Effects condition ($F(1, 32) = 17.32, p < .001$), the Matched for Trials condition ($F(1, 32) = 13.86, p < .001$), and the No Independent Play condition ($F(1, 32) = 10.83, p < .01$).

Additionally, more children spent the majority of their time playing with the familiar toy in the Confounded condition than in the Matched for Effects condition ($\chi^2 (1, N = 32) = 8.13, p < .01$), the Matched for Trials condition ($\chi^2 (1, N = 32) = 6.15, p < .025$), and the No Independent Play condition ($\chi^2 (1, N = 32) = 4.5, p < .05$). Finally, children were more likely to reach first for the familiar toy in the Confounded condition than in the Matched for Effects condition ($\chi^2 (1, N = 32) = 5.24, p < .025$) and there was a similar trend for children in the Matched for Trials condition ($\chi^2 (1, N = 32) = 3.46, p = .06$) and the No Independent Play condition ($\chi^2 (1, N = 32) = 3.46, p = .06$).

Within the Confounded condition, children played significantly longer with the familiar toy than the novel toy ($t(15) = 2.79, p < .01$). These results reversed for children in the Matched for Effects ($t(15) = 3.1, p < .01$) and Matched for Trials conditions ($t(15) = 2.48, p = .01$). Children played equally long with both toys in the No Independent Play condition ($t(15) = 1.06, p = ns$). In the Confounded condition, a non-significant majority of children played most with the familiar toy ($p = ns$ by binomial test; one-tailed throughout) but more children played most with the novel toy in the Matched for Effects ($p = .01$ by binomial test) and Matched for Trials conditions ($p < .05$ by binomial test) and were marginally more likely to play with the novel toy in the No Independent Play condition ($p = .07$ by binomial test). Finally, in the Confounded condition, children's first reach was just as likely to be for the familiar toy as the novel toy ($p = ns$ by binomial test), whereas children were significantly more likely to reach first for the novel toy than the familiar toy in all Unconfounded conditions (Matched for Effects: $p < .01$ by binomial test; Matched for Trials: $p = .01$ by binomial test; No Independent Play: $p < .01$ by binomial test).

It is possible that the children were simply more interested in the simultaneous effects than the separate effects. However, if children were more interested in the three simultaneous effects of the Confounded condition than, for instance, the two simultaneous effects of the Unconfounded/Matched for Effects and No Independent Play conditions, one might also expect children to play more with the familiar toy in those conditions than in the Unconfounded/Matched

for Trials condition where the effects occurred simultaneously only once. In fact, there were no significant differences among the Unconfounded conditions. This suggests that it is the absence of disambiguating evidence rather than the presence of simultaneous effects that encourages children's exploration.

We also looked at the actions children performed on the familiar box in the Confounded condition. In the course of their free play with the familiar box, children often manipulated the levers simultaneously. Critically however, 12 of the 16 children (75%) also manipulated each lever separately, fully disambiguating the evidence[11]. This suggests that children's free exploratory play could, in principle, generate the type of evidence that would support accurate causal learning.

General Discussion

Our findings suggest that preschoolers' spontaneous exploratory play is sensitive, not just to stimulus features such as novelty and perceptual salience, or to cases when evidence strongly conflicts with prior beliefs, but also to formal properties of evidence, like confounding. Note that in all four conditions, children were familiarized with the same toy and children's exposure to the toy's effects and affordances was closely matched across conditions. A single manipulation seemed to drive the effect: if the two levers were moved separately, on just a single trial, the children spent most of their free time playing with the novel toy; if the two levers were always moved simultaneously, children spent most of their free time playing with the familiar toy. Children appear to recognize confounded evidence and are motivated to explore stimuli whose causal structure is ambiguous.

In this study we relied on an implicit measure of children's understanding of confounding – spontaneous exploratory play. Because previous research suggests that children have a poor metacognitive understanding of confounding and experimental design (Chen & Klahr, 1999; Inhelder & Piaget, 1958; Kuhn, 1989; Kuhn, et al., 1988; Koslowski, 1996; Masnick & Klahr, 2003), we expected that children might not be aware of their own motivation for exploration.

---

[11]   We subsequently coded children's actions on the familiar box in the Unconfounded conditions. In the Matched for Effects condition, 31% of the children manipulated each lever separately; in both the Matched for Trials and No Independent Play conditions, 50% of the children manipulated each lever separately (this percentage does not include the two children in the No Independent Play condition whose clips were lost, see footnote 2). Of course, the children played longer with the familiar box in the Confounded condition. However, it is interesting that the children were, if anything, more likely to manipulate the levers separately in the Confounded condition than in the Unconfounded conditions, where they had actually observed the separate manipulations.

However, it is possible that at least some of the children might have been able to say that they were more curious about one toy than another or articulate why they wanted to play with one toy more than another. Further research might investigate the extent to which preschoolers' explicitly recognize confounded evidence.

Do children actually learn causal relationships from the evidence of their own interventions? Though results from Chapter Four suggests that this is the case, the results presented in this chapter do not address this directly, although in simple cases (i.e., when each lever either does or does not cause an effect) it seems probable that children would. However, we do not want to suggest that in all cases children's free exploratory play reliably leads to accurate causal learning. There is every reason to believe that in many contexts children's spontaneous exploratory play might not suffice for correct causal inferences. Children might be inaccurate for many reasons: because they are unable to disambiguate the relevant variables, because they fail to disambiguate the relevant variables, or because they fail to attend sufficiently to the evidence they generate in exploratory play. Nonetheless, children's tendency to selectively explore confounded events could be advantageous for causal learning; whether or not children learn from their explorations in any particular instance, overall, they would be more likely to explore where there is something to be learned.

Importantly however, research presented in Chapter four, and other studies suggests that children do indeed learn from exploratory play in some contexts. Preschoolers for instance, were able to use disambiguating evidence generated by their spontaneous play with a gear toy to distinguish causal chain and common cause structures (Schulz, Gopnik, & Glymore, 2007). Such findings are consistent with the possibility that children's selective exploration of confounded evidence might support causal learning. Future research might investigate the generality of this hypothesis by looking at the range of contexts and ages in which children are sensitive to and selectively explore confounded evidence (e.g., whether such findings hold for toddlers and infants) and by looking at the extent to which children's free exploratory play generates informative evidence.

Children's exploratory play is a complex, dynamic phenomenon, indubitably affected by many factors (e.g., the child's temperament, the child's comfort and energy level, and the perceived cost or benefit of various actions in terms of effort expended, knowledge gained, and external reinforcement). However, these results suggest that children's normative understanding of evidence and their curiosity about the causal structure underlying observed evidence play a significant role in their decision to explore. At least in simple cases, preschool children distinguish confounded and unconfounded evidence and rationally choose to selectively engage

in more exploration when the causal structure of events is ambiguous. The exploratory play of even very young children appears to reflect some of the logic of scientific inquiry.

Chapter 6

In Chapter Four, we looked at whether children appeal to an unobservable variable to 'explain away' evidence that conflicts with their prior beliefs. However, this involved children's explanations following two extremes, when evidence strongly conflicts with beliefs and when evidence is in support of those beliefs. One additional thing to investigate is whether children demonstrate a graded sensitivity to ambiguous evidence as probability information becomes stronger or weaker, and importantly, whether this graded sensitivity interacts with children's prior beliefs. Because it's difficult to generate 'ambiguous' evidence in a balance paradigm, this task explores a new paradigm where children have strong prior beliefs and where probability information can be easily manipulated.

The process of seeking, generating, and evaluating explanations plays a crucial role in learning and development (Lombrozo, 2006; Keil, 2006). Research in education, for example, has found that explaining – even to oneself – can facilitate learning and generalization (e.g. Chi et al., 1994; Siegler, 2002). Recent work in cognitive development further demonstrates that explaining can prompt discovery and scaffold causal learning (Legare, Gelman, & Wellman, in review; Legare, Wellman, & Gelman, in press; Wellman & Liu, 2007; Bonawitz & Schulz, in prep.), even in complex domains (e.g. Wellman & Lagattuta, 2004). Explanations and the understanding they

foster give us power over the world: to the extent we can accurately explain, we can better predict and control (Lombrozo & Carey, 2006; Heider, 1958).

The ability to evaluate competing explanations plays a special role when the evidence we seek to understand is ambiguous. Is the day/night cycle best explained by an orbiting sun or an orbiting earth? Did Sally look for her marble in the basket by accident or because she had a false belief about its location? Inferring the truth is challenging in such cases because it is underdetermined: multiple explanations are possible, requiring that an inference be constrained by more than consistency with data. A potential solution is to use an explanation's "loveliness" as a guide to its "likelihood" (Lipton, 2002), a strategy known as *inference to the best explanation* (Pierce, 1998; Harman, 1965; Lipton, 2002). More precisely, one chooses among candidate hypotheses by considering which hypothesis, if true, would best explain the evidence in question. Occam's razor, the well-known stricture not to multiply entities beyond necessity, offers a compelling criterion for evaluating which explanation is best: simplicity (Baker, 2004).

This chapter explores the hypothesis that children as young as four years old engage in a process of inference to the best explanation, and that simplicity is a factor in establishing which explanation is best. Even more than adults, children require strategies for coping with the problems of underdetermination. Children must not only draw inferences from ambiguous evidence, but must do so on the basis of skeletal domain knowledge at best and no domain knowledge at worst. A domain-general principle for choosing among competing explanations could thus play a crucial role for naïve learners, and simplicity provides a viable candidate for such a principle.

Previous research on simplicity in explanation evaluation has focused on adults. This work has quantified simplicity in terms of the number of propositions (Thagard, 1989; Read & Marcus-Newhall, 1993; Lagnado, 1994) or causes (Lombrozo, 2007) invoked in an explanation. For example, Lombrozo (2007) examined whether adults prefer explanations involving fewer causes, and whether this preference informs assessments of probability (see also Read & Marcus-Newhall, 1993; Lagnado, 1994). Participants learned about an alien with two symptoms that could be explained by appeal to one disease or two, and were asked to identify the most satisfying explanation. Importantly, participants also received information about the baserates of these diseases, and in some conditions the complex explanation (two diseases) was more likely than the simpler alternative (one disease).

Lombrozo (2007) found that adults were sensitive to probability information, but only preferred the complex explanation when it was much more probable than the simpler alternative. Moreover, participants who committed to a simple but unlikely explanation overestimated the

baserate of the disease invoked in that explanation, suggesting that simplicity was used as a guide to probability. However, both Lagnado (1994) and Lombrozo (2007) found that when participants were explicitly told that the complex explanation was most likely rather than having to infer this on the basis of baserate information, simplicity did not influence judgments. These findings suggest that in the face of probabilistic uncertainty, adults employ simplicity as a basis for gauging probability, as might be expected from a process of inference to the best explanation.

If adults appeal to simplicity to constrain inferences when faced with uncertainty, children have all the more reason to do so. However, little is known about how children evaluate competing explanations, and whether they engage in a process like inference to the best explanation. Research on children's explanations has generally focused on explanations' content (e.g. Wellman, Hickling, & Schult, 1997), not on the roles of probabilistic evidence or inference. Research on causal inference, in contrast, has demonstrated that even young children are sensitive to probability (Gopnik et al, 2004; Schulz & Gopnik, 2004; Kushnir & Gopnik, 2007) and sampling information (Xu & Garcia, 2008; Kushnir, Xu, & Wellman, 2008), and use prior beliefs to inform judgments (e.g. Koslowski, 1996; Schulz, Bonawitz, & Griffiths, 2007). However, this research has not examined explanatory virtues like simplicity.

To illustrate, consider a study in which children were trained that blocks that activate a machine, called 'blickets,' are either rare or common (Sobel, Tenenbaum, & Gopnik, 2004). Children were then presented with a backwards blocking paradigm: children were shown that two objects (A & B) activated the machine together, and that one object (A) activated the machine by itself. Is the other object (B) a blicket? When children were taught that blickets were rare, they made the correct backwards blocking prediction and only extended the blicket label to object A, even though B was associated with the machine's activation. However, when children were taught that blickets were common, they categorized the uncertain object, B, as a blicket as well. Backwards blocking may reflect a principle like Ockham's razor: why assume both blocks are potential causes of the machine's activation if A alone is a sufficient explanation? However, this study does not address whether and how simplicity trades-off with probabilistic evidence, as children's judgments in this task need only rely on probability information to generate the observed pattern of responses.

To our knowledge, we present the first study to examine whether children engage in a process of inference to the best explanation, and in particular whether simplicity – quantified as the number of causes invoked in an explanation – plays a role in constraining children's inferences. If simplicity is a useful constraint on learning, children should mirror adults in integrating a preference for simplicity with probabilistic evidence. But doing so requires that

children effectively evaluate simplicity, track probabilistic information, and integrate these two sources of constraint.

Simplicity Experiment

In a task adapted from Lombrozo (2007), children and adults chose between a simple explanation (involving one cause) and a complex alternative (involving two causes), where both explanations accounted for the data being explained. To make the task engaging and appropriate for 4-year-olds, we modified Lombrozo (2007) to involve a live event, and solicited spontaneous explanations. Specifically, we designed a toy for which the simplicity and probability of candidate explanations could be independently varied. Rather than diseases and symptoms, participants learned about colored chips that generated one or two effects when placed in the toy's activator bin: blue chips activated the toy's light and fan, red chips activated the light, and green chips activated the fan. Participants were asked to provide an explanation for an event in which a bag of chips accidentally tipped into the toy's activator bin, and the fan and light both activated. Citing a single blue chip was the simplest possible explanation, but citing a red chip *and* a green chip was a viable complex alternative. Varying the numbers of different chips in the tipped bag was equivalent to varying the baserates of diseases in Lombrozo (2007). This task thus allowed us to examine whether and how children and adults integrate information about simplicity and probability in making an inference to the best explanation.

Methods

Participants & Design. Eighty-five children and sixty-four adults were randomly assigned to one of four conditions: a 1:1 probability condition (in which the probability of the complex explanation was equal to that of the simple explanation), a 1:2 condition (in which the probability of the complex explanation was twice that of the simple explanation), a 1:4 condition, and a 1:6 condition. Eighteen children (R=48m-70m; M=58.7m) participated in the 1:1 condition; eighteen (R=47m-69m; M=57.4m) in the 1:2 condition; twenty-one (R=49m-70m; M=58.0m) in the 1:4 condition; and twenty-five (R=49m-72m; M=59.9) in the 1:6 condition. Sixteen adults, ranging in age from 18-23 years, were tested in each of the four conditions.

Materials. The task involved a toy with a red bulb and a green fan, both of which spun and lit up when activated. These effects were apparently generated by placing colored chips in an activator bin, but were in fact controlled with a hidden switch.

Procedure. There were three phases: a demonstration, a memory check, and an explanation event, (see Figure 12).

        Demonstration. Participants were shown that placing a red chip in the activator activated the globe, placing a green chip in the activator activated the fan, and placing a blue chip in the activator generated both effects. The experimenter then asked the participant to help count chips into a clear container. In all conditions, only one blue chip was added, while the numbers of red and green chips varied[12] (see Table 2).

        Memory check. Participants were asked what happened when each chip went in the activator and to predict what would happen if both a red and green chip went in simultaneously. This served as a basis for eliminating individuals unable to attend to the task, and also ensured that participants realized that the blue chip or a red chip *and* green chip could activate both parts of the toy. The experimenter also asked the participant how many chips of each color were placed in the container.

Explanation Event. After mixing the chips in the container, the experimenter poured them into an opaque, rigid bag that was placed next to the toy's activator. The experimenter then 'accidentally' knocked the bag towards the activator and away from the participant, and the globe and fan immediately activated. 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?" Explanations were recorded.

The number of chips required to achieve each condition's probability ratio was computed by assuming that accidentally tipping the bag was equally likely to result in one chip or two chips falling. This assumption was reinforced by having the experimenter explicitly note that "one or two chips" fell in the activator bin. To calculate the number of chips of each color required to generate each condition's probability ratio, we assumed that in the tipping event, the chips were sampled uniformly without replacement. We also counted outcomes *containing* the blue chip as simple, whether or not there was a second chip. That is, chip combinations consisting of blue and red, blue and green, and blue alone were all counted as "simple," as this results in a more conservative estimate for the role of simplicity. Note, however, in coding participants' explanations we used a more stringent criterion (detailed below) to be consistent in our definition of simplicity as appealing to a *single cause*.

        Adults were informed that the procedure was designed for young children; the adult and child procedures were otherwise equivalent.

---

[12] Pilot work suggested that increasing the total number of chips beyond 40 (as would be required if we increased the number of blue chips to 2 while maintaining the probability ratios) was too cumbersome.
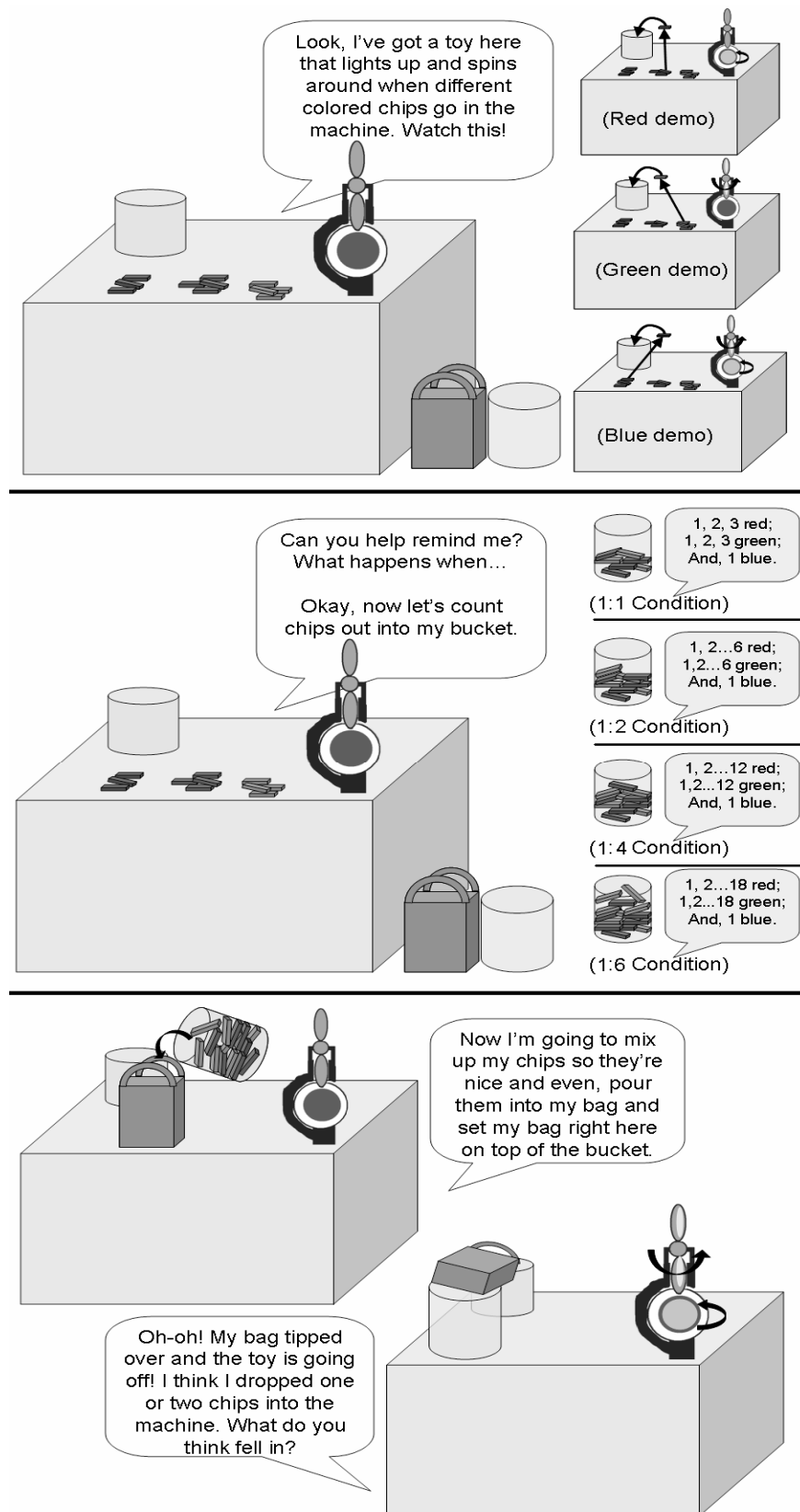
Figure 12: Schematic illustration of the experimental procedure.

Results

　　Data from Children. Children's responses on the memory check were coded. Children failing any portion of the check were excluded from analyses, resulting in sixteen children per condition. There were no significant age differences across conditions (Kruskal-Wallis, k=4; h(3)=1.33, *n.s.*). Before considering children's relative preference for simple and complex explanations, we note that children overwhelmingly provided explanations that were adequate in the sense that they accounted for both observed effects. Of the 64 explanations generated, only 4 failed to account for both effects (e.g. just a red chip).

　　We next examined the roles of simplicity and probability in explanation evaluation. All children generated a single explanation, and responses fell unambiguously into one of three categories: simple (blue chip only), complex (red & green chips only), or other[13] (see Table 2). Collapsing across all four conditions, children were no more likely to choose the simple explanation or the complex explanation ($\chi^2(1)$=1.17, *n.s.*). However, 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 (Figure 13). Yet when these explanations were equally likely (1:1 condition), significantly more children selected the simple explanation ($\chi^2(1)$=7.58, p<.01). Even in the 1:6 condition, over 30% of children chose the simple explanation.

　　These data suggest that children were sensitive to the probability information conveyed by the relative numbers of chips, but had a baseline preference for the simpler explanation. That simplicity and probability both influenced explanation choices suggests children effectively integrated these competing explanatory demands, and treated simplicity as commensurate with frequency information.

　Could these data be explained without appeal to simplicity? We reject an alternative explanation, 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, which was not found ($\chi^2(3)$=5.69, *n.s*).

---

[13]　　Note that only 1 child (in the 2:1) condition generated an explanation with all three chips (red, green, and blue), which we coded as 'other,' though coding it as 'complex' does not affect the results.

Table 2: Types of explanations generated as a function of condition. The condition labels indicate the ratio of the probability of the simple explanation (blue) to that of the complex explanation (red & green).

Condition:	1:1	1:2	1:4	1:6
# Chips used:	1 Blue 3 Red 3 Green	1 Blue 6 Red 6 Green	1 Blue 12 Red 12 Green	1 Blue 18 Red 18 Green
Children's Responses				
# Simple	8	10	6	5
# Complex	1	3	9	10
# Other	7	3	1	1
# Failed Memory	2	2	5	9
Adults' Responses				
# Simple	6	2	3	1
# Complex	10	14	13	15

Another possibility is that children simply believed one chip was more likely to fall out of the bag than two chips, and that this alone governed responses. The most compelling reason to reject this alternative is that the ratio of simple responses monotonically decreased as a function of the probability ratio, a trend that could only be accounted for on this alternative explanation if assumptions about the probability of one chip falling likewise changed, a hypothesis we rejected above. As additional evidence that children judged a single chip about as likely to fall as two chips, note that across all four conditions, the frequency of one-chip and two-chip explanations, 33 and 30 respectively, did not differ from each other ($\chi^2(1) = .28$, n.s.) or from the presumed value of 50% (binomial tests, n.s.).
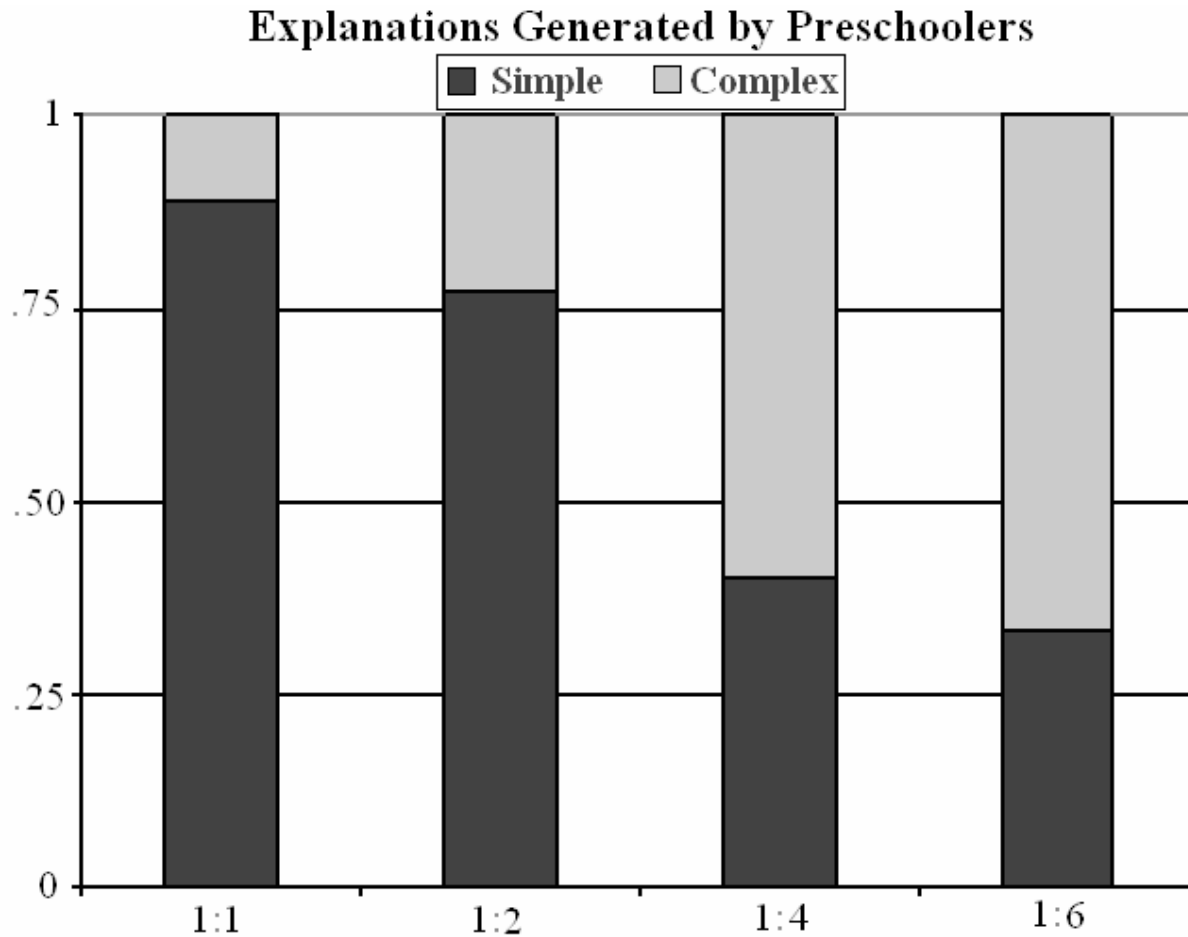
Explanations Generated by Preschoolers

Figure 13: Proportion of children generating simple and complex explanations as a function of condition, excluding the small number of children who generated "other" explanations. The condition labels indicate the ratio of the probability of the simple explanation (blue chip) to that of the complex explanation (red & green chips).

Data from adults. All adults passed the memory check and generated only simple or complex explanations (blue chip only; red & green chips only), (see Table 2). Collapsing across all four conditions, adults were significantly more likely to choose the complex explanation over the simple one ($\chi^2(1)=50.0$, p<.001). Mirroring the trends with children, adults were more likely to favor a complex explanation as it became more probable: they were significantly more likely to provide the simple explanation in the 1:1 condition than in the 1:6 condition ($\chi^2(2)=4.57$, p < .05). However, there were no other significant differences between conditions.

Overall, adult responses differed significantly from children's responses. In all conditions adults were more likely than children to generate the complex explanation (1:1, $\chi^2(2)=4.57$,

p<.05; 1:2, $\chi^2(2)=15.18$, p<.01; 1:4, $\chi^2(2)=4.8$, p<.05; 1:6, $\chi^2(2)=4.57$, p<.05). And while children were significantly more likely to choose the simple explanation over the complex alternative in the 1:1 condition, adults were not (binomial test, n.s.). The adult response pattern can thus be accounted for by probability alone.

General Discussion

Using a novel method, we find that children's early explanations are sensitive both to frequency information and to simplicity. Three points about these findings are worth emphasizing. First, the fact that children's spontaneous explanations vary as a function of frequency information contributes to the emerging literature suggesting that children are savvy probabilistic reasoners, but does so using a quite different task from those previously employed. Second, the fact that children's explanations vary in response to simplicity – even when frequency is held constant – suggests that children prefer simpler explanations by virtue of their simplicity, and not (as in previous experiments) because they are also more likely on the basis of frequency information. Finally, the fact that children are able to integrate both probability and simplicity as sources of inferential constraint suggests that children engage in a process of inference to the best explanation, using simplicity as a basis for assessing the probability of competing explanations.

Unlike Lombrozo (2007), which found that adults used simplicity to inform explanation choices when faced with probabilistic uncertainty, the current experiment found no evidence of a preference for simpler explanations among adults. However, Lombrozo (2007) and Lagnado (1994) both found that adults typically ignored simplicity when the most probable explanation was clearly identified. In the current task, our adult population may have been able to compute relative probabilities in the course of the task, reducing uncertainty concerning which explanation was most likely. It's thus possible that the observed differences between children and adults reflect expertise rather than a developmental change. Specifically, simplicity may be used as a guide to probability only when more reliable bases for assessing probability are unavailable. This could arise when physical systems are not well understood (as could be the case with the children in this experiment) or when there is uncertainty about how to assess probability (as with adults in some Lombrozo (2007) conditions). Future work could examine how simplicity and probability interact given more complex or natural conditions, as well as conditions in which mechanism information is well understood.

Why might children and adults rely on simplicity in the face of probabilistic uncertainty? In philosophy, statistics, and computer science, a role for simplicity in inference has been defended

on normative grounds (Akaike, 1974; Rissanen, 1978; Li & Vitanyi, 1997). A preference for simpler explanations reduces the risk of "over-fitting" data, and has also been advocated as a rational consequence of Bayesian inference (Jeffreys & Berger, 1992). Most normative justifications for simplicity do not apply to the task presented here, but if a role for simplicity in inference is sometimes warranted, our findings could reflect an over-extension of that role. Future work could evaluate how simplicity influences children and adult's reasoning when simplicity is quantified in ways more consistent with these normative approaches.

In the course of learning and development, children are constantly faced with situations for which more than one explanation is possible. This occurs not only in explaining isolated events or properties, but also in constructing explanatory frameworks like a theory of mind (e.g., Wellman & Lagattuta, 2004) or a mental model of the earth (e.g. Vosniadou & Brewer, 1992). Prior beliefs provide one way to leverage limited experience in the service of inference, but our findings suggest an additional resource available to children and adults under uncertainty: domain-general constraints that inform judgment by playing a role in the evaluation of explanations. Specifically, we've provided evidence for a principle of parsimony like Occam's razor, and for the claim that children, like adults, engage in a process of inference to the best explanation.

Chapter 7

We began with the proposal that children may be rationally guided in their predictions, exploration, and explanation by the interaction of prior beliefs and observed evidence. Rationality was defined using the prescription of the Bayesian framework, which specifies how prior beliefs and new evidence should interact to select the most likely hypothesis given the data. To test this model, we explored cases where children had strong prior beliefs about the observed evidence, guided by intuitive theories in several domains, and cases where they had no differential beliefs. We also looked at children's responses with a variety of methods, force-choice tasks, exploratory play, spontaneous exploration, and predictions.

Note that if children are like scientists, their decisions about the correct hypothesis must be influenced both by their prior theories of the world and by the observed evidence. That is, when comparing two likely candidate causal hypotheses, choices should reflect a graded interaction between the prior probability of the hypotheses (before observing the data) and the likelihood of observing the data given the hypotheses. Similarly, spontaneous explanations of observed events should also result in choosing the best hypothesis from all possible candidate hypotheses. Additionally, if children are rational explorers, investigating and intervening when there is indeed something to be learned, then exploration should follow when there is uncertainty

Table 3. Examples of children's sensitivity to both prior beliefs and observed evidence in preceding chapters.

	Equal Prior Beliefs Differential Evidence	Differential Prior Beliefs Equal Evidence
Forced-choice Variables	**Chapter 2:** Different predictions across the Baseline and Evidence conditions	**Chapter 2:** Different predictions across the within domain and cross domains conditions **Chapter 3:** Training to alter prior beliefs altered endorsement of a priori unlikely cause
Exploration	**Chapter 4:** Different exploration of familiar and novel toy across the confounded and unconfounded evidence conditions.	**Chapter 5:** Different exploration of familiar block and novel toy across the type of theory (Mass & Center). No differential exploration in younger, No-differential theory children.
Explanation	**Chapter 6:** Different appeal to simple and complex explanations as evidence varied.	**Chapter 5:** Different appeal to magnet as explanatory variable, dependent on prior beliefs and evidence observed.

about the correct hypothesis. To test these claims, the studies presented in this thesis encompass two manipulations: in some conditions, children's beliefs were a priori equal, but children received varied patterns of evidence; in other conditions, children observed identical evidence but they had varied prior beliefs (see Table 3). We now consider these examples in more detail.

Summary

Forced-choice variables

In Chapters 2 and 3, I looked at the ways in which theories and evidence interacted to affect children's choice about the most likely causal variable. In order to investigate the interaction between these factors, preschoolers were presented with stories pitting their existing theories against statistical evidence (Chapter 2) in cases when they had strong prior beliefs (cross domains) about the likely causal variable, and in cases when they did not (within domain). Children were randomly assigned to either a *Baseline* condition or an *Evidence condition* and were read two stories in which two candidate causes co-occurred with an effect. In one story, all variables came from the same domain; in the other, the recurring candidate cause, A, came from a different domain (i.e. A was a psychological cause of a biological effect). Children in the *Baseline* condition read stories where the event and two candidate causes occurred only once (AB→E). In the *Evidence* condition, evidence was presented in the form: AB →E, AC → E, AD → E, etc. After reading the book, children were asked to identify the cause of the effect. Consistent with the predictions of a Bayesian model, both prior beliefs and evidence played a role

for children's causal predictions. Four-year-olds were more likely to identify 'A' as the cause in the *Evidence* conditions then in the *Baseline* condition. Results also showed a role of theories in guiding children's predictions. All children were more likely to identify A as the cause within domains than across domains. However, while the four- and five-year-olds learned from both the within- and cross-domains evidence, three-and-half-year-olds learned only from the within-domain evidence and three-year-olds failed altogether.

I presented three possible explanations for younger children's failure to update their beliefs: one (Prior Knowledge account) suggests that in some domains, younger children have stronger prior beliefs and thus require more evidence to overturn them; the second (Statistical Reasoning account) suggests that three-year-olds have a fragile ability to reason about statistical evidence; the third (Joint Factors account) suggests that both of these limitations play a role. To distinguish these accounts, we conducted a two-week training with three-and-a-half-year-olds (Chapter 3). Children participated in either one of two *Prior Belief Trainings* (*Baserates* or *Mechanisms*), a *Statistical Reasoning Training*, or a *Control* condition. Relative to the *Control* condition, children in all three training conditions showed an improvement in their ability to reason about theory-violating evidence. These results suggest both that statistical reasoning limitations need to be addressed to improve young children's ability to use statistical data for belief revision, but also that children's prior beliefs and the evidence they observe interact in a rational way, as prescribed by Bayesian frameworks, to guide children's choices of the most likely candidate causal variables.

Exploration

In Chapters 4 and 5, I presented evidence that suggests that theories and evidence interact to affect children's choices in exploration in cases when children have strong prior beliefs and in cases when they do not. First, I argued for the important contribution children's beliefs about balance make in their exploratory play (Chapter 4). We provided both Mass Theorist and Center Theorist children with either conflicting or consistent evidence with respect to their beliefs. Importantly, conflicting evidence for Mass Theorists (a block balancing in the geometric center, despite uneven weighting) is identical to confirming evidence for the Center Theorists. And the reverse is also true: conflicting evidence for Center Theorists (a block balancing off towards the side of greater weight) is identical to confirming evidence for Mass Theorists. Play time comparisons between conditions revealed no main effect of theory and no main effect of evidence type. 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. That is, children overrode a novelty preference to continue exploring the balance toy when the evidence conflicted with their theories, but not when it confirmed their theories. The results suggested that children's exploratory play is guided by both prior beliefs and novel evidence.

Secondly, we looked as how varying the evidence changes children's choices in play when they do not have strong prior beliefs distinguishing the likely causal hypothesis (Chapter 5). We introduced children to a toy and showed them either confounded or unconfounded evidence about the causal structure of the toy. We removed the toy and then returned it along with a novel toy allowing the children to play freely for sixty seconds. We found that children who observed confounded evidence preferentially played with the familiar toy but children who observed unconfounded evidence showed the standard novelty preference and played primarily with the novel toy, suggesting that children's normative understanding of evidence and their curiosity about the causal structure underlying observed evidence played a significant role in their decision to explore. The results of this work demonstrated that the play of even very young children appears to reflect some of the logic of scientific inquiry.

Explanation

In Chapters 2 and 3, I demonstrated than when provided with a forced-choice alternative, children are able to choose the mostly likely explanation. However, Chapters 4 and 6 also provided evidence that children's spontaneous explanations rationally reflect a sensitivity to evidence and prior beliefs, in cases when children have strong prior beliefs and in cases when they do not. First, I presented evidence in the balancing block studies of Chapter 4 that children's strong prior beliefs guide explanation. During the course of free play with the balance blocks, both Mass and Center Theorist children were able to discover a magnet which held the block in place in both *Conflicting* and *Confirming* conditions. However, while the magnet is always a *sufficient* explanation for the block sticking to the platform, children in the conflicting condition were significantly more likely than children in the confirming condition to appeal to the magnet as an explanatory variable. This signifies that children who observed evidence that went against their initial beliefs responded by trying to 'explain away' the evidence. Note that evidence alone (i.e. the block balancing in the geometric center as compared to the block balancing in the center of mass) is not sufficient to describe this pattern of results, nor is the theory alone (i.e. is the child a Mass theorist or a Center theorist.) Rather, the interaction of the child's theory with the evidence observed determined whether or not children appealed to the magnet as an explanatory variable.

Chapter 6 investigated whether young children prefer explanations that are simple, where simplicity is quantified as the number of causes invoked in an explanation, and how this preference is reconciled with probability information. Preschool-aged children were asked to explain an event that could be generated by one or two causes, where the probabilities of the causes varied across conditions. Children preferred explanations involving one cause over two, but were also sensitive to the probability of competing explanations. That is, as evidence increased favoring the complex explanation, children were more likely to also favor the complex explanation. These data suggest that children are sensitive to evidence when evaluating competing causal explanations but also employ a principle of parsimony like Occam's razor as an inductive constraint. This constraint may be employed when more reliable bases for inference are unavailable.

Remaining Questions

Uncertainty and Exploration

In this thesis, I have suggested that causal uncertainty will lead to greater exploration. Causal uncertainty arises because the posterior odds of the two hypotheses are approximately equal. This can occur in two cases: Case 1 (Chapter 4) is when one hypothesis has higher prior probability, but the other hypothesis has a higher likelihood, so the two terms cancel each other out; Case 2 (Chapter 5) is when the prior probability and likelihood are both approximately equal. However, one could imagine creating a scenario where children are faced with competing cases of causal uncertainty (e.g. on one side of the child, a block surprisingly balanced, and on the other side, a confounded jack-in-the-box). How the children choose between exploring these two cases of uncertainty remains both an empirical question and an open area for computational accounts to capture.

Library of Alexandria. I presented cases of uncertainty when there were just a few likely candidate hypotheses. That is, the children had to have enough knowledge to recognize that there was uncertainty in the likely causal models of the toy. But what happens as the number of potentially plausible explanations increases, such as when the constraints on plausible hypotheses provided by prior belies become weaker? On the one hand, this would provide a lot of uncertainty about the system and one might expect children to explore more. On the other hand, empirically and intuitively this does not seem to be the case. The novel toy used in both play studies arguably offers significantly more uncertainty than the familiar toys; however, our design shows that, though children will show a standard novelty preference when the novel toy is pitted

against a familiar toy with unambiguous causal structure, children will not explore the novel toy when it is pitted against a familiar toy with ambiguous causal structure. Intuitively, we can also see that we are not driven to explore the things we have most uncertainty about. This is perhaps best captured by an analogy provided by a friend and colleague, Noah Goodman[14]:

> God, your adviser, and a dog are standing outside the gates of the Library of Alexandria. Who goes into the library to read the books? God does not go into the library, because God already knows everything there is to know and thus does not need to read the books. The dog does not go in, because the dog does not know enough to know that it could learn from the books. The only person to go into the library is your adviser. She knows enough to know that she still has much to learn[15].

This example suggests that children must have some knowledge to get them started; if we all start like Noah's dog outside the library gates, then how might we ever know to go inside to explore and learn? The choice not to explore (even when aware of uncertainty), may still be captured by rational models. To be sure, a goal in exploring uncertain situations is to gain knowledge and thus also gain predictive and explanatory control over the world. If we are content with the amount of control we have over a situation, even if we know we have a veil of explanatory depth pulled over our eyes (Rozenblit & Keil, 2002), then perhaps we have a rational means for ignoring uncertainty, so that we can instead explore the things that do require our understanding.

Pedagogy and play. Additional modeling work can also explain how the *source* of observed evidence should be interpreted and consequently explored. For example, how do children interpret evidence generated by a knowledgeable teacher as compared to evidence that is unintentionally demonstrated? Here I have suggested that, in the Piagetian tradition (1929), self-guided play serves as an important vehicle for learning both inside and outside the classroom. However, research in the Vygotskyean tradition (1978) has placed relatively less emphasis on children's self-directed exploration and more emphasis on how children learn from social interactions and cultural transmission. Research investigating learning through social interaction suggests that even young children are sensitive to whether information was generated intentionally or accidentally (Xu & Tenenbaum, 2007), by a reliable or unreliable teacher

---

[14]   Via personal communication.
[15]   Arguably my joke version is a better telling of the exploration problem "…Who goes into the library to read the books? No one. God does not go into the library, because God already knows everything there is to know and does not need to read the books. The dog does not go in, because the dog does not know enough to know that it could learn from the books. Your advisor does not go in because, though she would like to go in, she has once again lost her keys and so cannot unlock the gates."

(Koenig & Harris, 2005; Kushnir, Wellman, & Gelman, 2008), or in a neutral vs. a pedagogical setting (Gergely, Kiraly, & Egyed, 2007; Tomasello & Barton, 1994). However, these projects have not looked at how these different contexts might affect children's exploratory behavior, and in particular, how observations of pedagogical evidence and play interact to affect children's learning.

Recently, some scientists have developed models of Bayesian inference operating over complex knowledge structures or theories (e.g. Tenenbaum et al., 2007). One such model contrasts the effects of pedagogical and non-pedagogical settings on learners' inferences. Shafto and Goodman (2008) formalize pedagogical learning as Bayesian inference based on the assumption by the learner that the teacher is being helpful; this expectation may facilitate learning in novel situations. For example, learning the causal relationships of a novel artifact is challenging because for any object, there are an unknown, and potentially large, number of causal properties. If a knowledgeable teacher explicitly demonstrates one action and a novel effect results, the model predicts both that the learner can make a strong inference that that there is a causal relationship between the action and the effect, and if the teacher demonstrates only the single action/outcome relation, the model predicts that the learner infers that other potential actions afforded by the object are less likely to generate novel or interesting effects. Intuitively, it captures the learner's inference about the teacher: "Teacher is helpful and knowledgeable, so why show me just that action, if there were other things to learn?"

While previous work provides first steps towards modeling adult causal learning, children face special challenges when it comes to learning from others. To successfully make inferences from evidence generated by a teacher, children must infer: a) that the teacher has a different state of beliefs about the world than them b) whether or not the teacher is knowledgeable (and thus whether demonstrated actions are representative and informative); c) whether or not they are in a pedagogical context. Finally, children must be able to integrate these sources of information to infer likely causal structures and to make decisions that guide their actions. Though some research has looked at children's imitation following accidental and intentional actions (Carpenter, Akhtar, & Tomasello, 1998), no research has investigated whether pedagogical information affects children's exploratory play and learning. These empirical questions offer an important connection to the exploration work presented here.

Levels of Explanation

The study of explanation has a rich history in philosophy (e.g. see Salmon, 1989) and social psychology (e.g. Malle, 2004). However, it's only been relatively recently that explanation

has been a focus of interest in cognitive development. Part of the reason for this has been the development of and increasing recognition for the Theory-Theory, which has placed explanations as central to its claims. That is, theories explain phenomena; to the extent that children's knowledge representations are theory like, so should be their reliance on explanations. Explanation also has a special relationship with causal reasoning. Indeed, many contemporary accounts define explanations as statements that identify at least some of the causes of the explanandum (Salmon, 1984; Woodward, 2003). However, there are other numerous accounts of what constitutes an explanation, dating back to Aristotle's 'modes of explanation'.

In this thesis, explanation is used as a means to evaluate children's inductive inferences about the correct hypothesis (cause) given the data. In each case presented here, this takes the form, "What caused E, A or B (or C)?" Sometimes the variables are explicitly presented, (as in Chapter 2 & 3, "What caused bunny's tummy ache, sandwich or worrying?"), and sometimes the variables are implied, and inference is left to the child, (as in Chapter 4, "What caused the block to stay up, [the magnet or the block balancing at that point]", and in Chapter 6, "What caused the toy to go off, [the blue chip or the red and green chip]?") This construction was deliberate because it provided a means by which to test predictions of the Bayesian framework: given the data and some prior beliefs about likely causes, do children rationally integrate this information to infer the most likely explanation?

In many cases, however, explanation appears to be at a different level of causal analyses. That is, the explanation often follows observation of the causal variable (rather than potential causal variables being obscured or ambiguous). We might have instead asked the child, "Why did worrying make bunny's tummy ache?" or, "Why did the blue chip make the toy go?" In these cases, explanations may naturally appeal to mechanisms (e.g. because the brain sends signals to the stomach that make it clench up and cramp) or conceptual generalization (e.g. because the blue chip is a blicket, and blickets make machines go). The study of what serves as a sufficient or satisfying explanation, and the relationship between this phenomenal criteria and Bayesian models, remains surprisingly understudied and thus poorly understood.

Additionally, the connection between children's explanations and these two literatures raises additional questions. How are explanations generated by children, and how does this ability develop? Hierarchical Bayesian models and specifying children's intuitive theories seem like a promising starting point, but then lead to questions like: under what conditions do children spontaneously seek or generate explanations? Recent research (Legare, Gelman, & Wellman, in review) suggests that children will generate explanations for surprising events over unsurprising events. However, note that explanations are also pedagogical and psychosocial in nature. To

what extent are children sensitive to other's knowledge states when generating explanations? (e.g. Do children provide reasonable 'explanatory depth' for explanation seekers? And, are children aware of other's mental states when choosing relevant explanandums? How do these abilities develop and on what types of knowledge do they depend?) Finally, explanation seems fundamental to children's understanding and belief revision (e.g. see Wellman and Liu (2007) for a review). What role, specifically, does explanation play in this process of learning and development? By rigorously describing the problem at the computational level and by simultaneously characterizing children's developing knowledge can we can address these important questions.

Developmental Change

What changes in development? Clearly, one thing that changes in development is children's working understanding of the world. In this thesis, I proposed that children's rational explanation and exploration is a vehicle for belief revision. That is, just as science depends on generating and modifying working hypotheses, taking into account the strength of a theory and the evidence for or against it, so does children's learning depend on theory and evidence. However, much work has also focused on possible information processing limitations that children face, such as limited working memory (e.g. Cowan, 1997). We came across one possible example of this developmental change in Chapters 2 and 3.

Modeling Developmental Change. New approaches in computer science and machine learning can also begin to capture why children's ability to reason from theory-violating evidence changes in development, as described in Chapters 2 and 3. One way to consider information processing limitations is to begin with a computational level theory and then consider how the learner is approximating those optimal computations. The formal models presented in this thesis employ Bayesian principles which provide normative prescriptions dictating how the learner should update their beliefs given a predetermined hypothesis space of possible causal models. However, more traditional cognitive modeling paradigms focus on the algorithmic level (Marr, 1982), presenting an account of how the mind may be approximating these inferences. That is, arguably in practice, a learner does not represent an infinite space of possible causal relationships and then rationally evaluate each one; instead the learner must approximate the rational solution using whatever limited information processing resources they have available. Machine learning has developed Monte Carlo methods that can approximate computationally large problems like this, using techniques such as Markov chain Monte Carlo, particle filters, rejection sampling,

likelihood weighting, and parallel tempering (see Robert and Casella, 2004, for a review).

Griffiths and colleagues have explored how these ideas from computer science and statistics can be used to develop psychological models that incorporate limitations on information processing. For example, Sanborn, Griffiths, and Navaro (2006) use particle filters to approximate rational statistical inferences for categorization. Shi, Feldman, and Griffiths (2008) show that importance sampling corresponds to exemplar models, a traditional process-level model that has been applied in a variety of domains. These examples illustrate how Bayesian inference can be approximated using a small number of samples, with the number of samples being allowed to vary to reflect available memory. This approach is consistent with the host of developmental evidence suggesting that memory capacity and strategies for remembering improve with age and experience (e.g. see Cowan, 1997; Schneider & Bjorklund, 1998; and Schneider & Pressley, 1997 for a review). By adjusting the available number of samples in the process models, we can explore how processing limitations in computation align with possible processing limitations that the child faces, resulting in age-related differences in the ability to reason about statistical evidence. Only by simultaneously working at both computational and algorithmic levels of analyses can we come to fully understand the developmental progressions that we see in children's learning from evidence.

Conclusions

We started with the problem of learning: without some constraints to shape the interpretation of evidence, the space of potential solutions to any particular causal problem remains large. Inspired by both the Theory-theory and Bayesian models, I have suggested that children may begin to constrain the space by being rational agents, integrating theory and evidence to evaluate beliefs. The results presented in this thesis support this claim, suggesting that children rationally integrate observed evidence with their prior beliefs to come to decisions about likely candidate explanations and to stimulate their exploratory play. Though they do not design carefully controlled experiments, children are more likely to explore when there is something to be learned. They generate evidence and revise beliefs during exploratory play. However, children also show conservatism in learning, maintaining strongly held beliefs if surprising evidence is weak or can be explained away, but rationally revising those beliefs when evidence strongly supports an alternative explanation. While there is still much work to be done, this work demonstrates the explanatory leverage gained by combining behavioral developmental studies with the computational description of how prior beliefs and evidence interact.

References

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a non-search task? *Developmental Psychology, 34*, 441–445.

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(3),* 299-352.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Anderson, J. R. (1990). *The Adaptive Character of Thought.* Hillsdale, NJ: Erlbaum.

Baillargeon, R., Kotovsky, L., & Needham, A. (1995). The acquisition of physical knowledge in infancy. In D. P. D. Sperber (Ed.), *Causal cognition: A multidisciplinary debate.* Symposia of the Fyssen Foundation; Fyssen Symposium, 6th Jan 1993, Pavillon Henri IV, St-Germain-en-Laye, France. New York, NY: Clarendon Press/Oxford University Press.

Baillargeon, R., Needham, A., & DeVos, J. (1992). The development of young infants' intuitions about support. *Early Development and Parenting, 1,* 69-78.

Baker, A. (2004). Simplicity. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* URL=<http://plato.stanford.edu/archives/win2004/entries/simplicity/>.

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, (4),* 946-964.

Berlyne, D. E. (1954). An experimental study of human curiosity. *British Journal of Psychology, 45(4),* 256-265.

Berlyne, D. E. (1960). *Conflict, Arousal and Curiosity.* New York: McGraw Hill.

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

Bonawitz, E.B. & Schulz, L. (in preparation). Balancing theories and evidence in children's exploration, explanations, and learning.

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, Tennessee.

Bruner, J., Jolly, A., & Sylva, K. (1976). *Play-Its Role in Development and Evolution.* New York: Basic Books, Inc.

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, NY: 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*; Based on a conference entitled "Cultural Knowledge and Domain Specificity," held in Ann Arbor, MI, Oct 13-16, 1990 (pp. 169-200). New York, NY: Cambridge University Press.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). 14-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21,* 315-330.

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.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104(2),* 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, US: The MIT Press.

Chi, M.T.H., de Leeuw, N., Chiu, M.H., LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Cowan, N. (1997). The development of working memory. In N. Cowan (Ed.), *The development of memory in childhood.* Hove, UK: Psychology Press.

Dember, W. N. & Earl, R.W. (1957). Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review, 64(2),* 91-96.

Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think. *Psychological Science, 16(4),* 291-297.

Dunbar, K. & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr and K. Kotovsky (Eds.), *The 21st Carnegie-Mellon symposium on cognition: Complex information processing: The Impact of Herbert A. Simon.* Hillsdale, NJ: Lawrence Erlbaum.

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, Serial 24, 60(1).*

Gelman, S. A. & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition 38(3),* 213-244.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition 38(3),* 213-244.

Gergely, G., Király, I., & Egyed, K. (2007). On pedagogy. *Developmental Science, 10:1,* 139-146.

Goodman, N., Ullman, T., & Tenenbaum, J. (in review). Learning a Theory of Causality.

Gopnik, A. & Schulz, L. E. (2004). Mechanisms of theory-formation in young children. *Trends in Cognitive Science, 8(8),* 371-377.

Gopnik, A. & Schulz, L. E. (2007). *Causal learning: Psychology, Philosophy, and Computation.* New York: Oxford University Press.

Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind and Language, 3(3),* 197-217.

Gopnik, A., & Meltzoff, A. (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., 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., 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 patters of variation and covariation. *Developmental Psychology, 37(5),* 620-629.

Griffiths, T. L. (2005). *Causes, coincidences, and theories.* Unpublished doctoral dissertation, Stanford University, Stanford CA.

Halford, G.S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002) Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology, 81,* 417-445.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74,* 88-95.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61(3),* 233-259.

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

Heider, F. (1958). *The psychology of interpersonal relations.* New York, NY: John Wiley & Sons, Inc.

Henderson, B. & Moore, S.G. (1980). *Children's Responses to Objects Differing in Novelty in Relation to Level of Curiosity and Adult Behavior.* New York: Dover.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology, 37(5),* 668-683.

Hood, B., Cole-Davies, V., & Dias, M. (2003) Looking and Search Measures of Object Knowledge in Preschool Children. *Developmental Psychology, 39(1)*, p. 61-70.

Hume, D. (2000). *A Treatise on Human Nature.* New York: Oxford University Press.

Hutt, C. & Bhavnani, R. (1972). Predictions from play. *Nature, 237(5351),* 171-172.

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind body distinction. *Child Development 64(5),* 1534-1549.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

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.

Jeffreys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist, 80,* 64-72 (Erratum, p. 116).

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

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

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

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

Keil, F.C. (2006). Explanation and Understanding. *Annual Review of Psychology. 57*, 227-254.

Kemp, C., & Tenenbaum, J.B. (2008) The Discovery of Structural Form. *Proceedings of the National Academy of Sciences. 105(31),* 10687-10692.

Kemp, C., Goodman, N. D. & Tenenbaum, J. B. (2008). Theory acquisition and the language of thought. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Kemp, C., Perfors, A. & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10(3),* 307-321.

Kirkham, N., Cruess, L., & Diamond, A. (2003) Helping children apply their knowledge to their behavior on a dimension switching task. *Developmental Science, 6(5),* 449-476.

Klahr, D. & Dunbar, K. (1988). Dual Space search during Scientific Reasoning. *Cognitive Science, 12,* 1-48.

Klahr, D., Fay, A.L., & Dunbar, K. (1993). Developmental differences in experimental heuristics. *Cognitive Psychology, 25,* 111-146

Koenig, M., & Harris, P.L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261–1277.

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. & Dean, D. (2004) Connecting Scientific Reasoning and Causal Inference. *Journal of Cognition and Development, 5(2)* 261-288.

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

Kuhn, D., & Phelps, E. (1982) The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior (Vol. 17*, pp. 1-44). New York: Academic Press.

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

Kuhn, D., Garcia, M., Zohar, A., & Andersen, C. (!995) Strategies of knowledge acquisition. *Society for Research in Child Development Monographs, 60(4,* Serial No. 245).

Kushnir & Gopnik (2005). Children infer causal strength from probabilities and interventions. *Psychological Science, 16(9),* 678-683.

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, 44,* 186-196.

Kushnir, T., Wellman, H. M. & Gelman, S. A.(2008). The role of preschoolers social understanding in evaluating the informativeness of causal interventions. *Cognition. 107 (3),* p. 1084-1092

Kushnir, T., Xu, F. & Wellman, H. (2008) Preschoolers use statistical sampling information to infer the preferences of others. In V. Sloutsky, B. Love, & K. MacRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Lagnado, D. (1994). *The psychology of explanation: A Bayesian approach.* Masters Thesis. Schools of Psychology and Computer Science, University of Birmingham.

Legare, C.H., Gelman, S.A., & Wellman, H.M. (in review). Inconsistency with prior knowledge triggers children's causal explanatory reasoning.

Legare, C.H., Wellman, H.M., & Gelman, S.A. (in press). Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology.*

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;* Based on a conference entitled "Cultural Knowledge and Domain Specificity," held in Ann Arbor, MI, Oct 13-16, 1990 (pp. 119-148). New York, NY: Cambridge University Press.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25(3),* 265-288.

Li, M. & Vitanyi, P.M.B. (1997). *An Introduction to Kolmogorov Complexity and its Applications.* New York, NY: Springer-Verlag.

Lipton, P. (2002). *Inference to the best explanation.* New York, NY: Routledge.

Lombrozo, T. & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99,* 167-204.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10,* 464-470.

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

Malle, B.F. (2004) *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction.* MIT Press.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. 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(1),* 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.

Memon, A., Cronin, O., Eaves, R., & Bull, R. (1993) The cognitive interview and child witnesses. In G.M. Stephenson & N.K. Clark (Eds*), Children, Evidence, and Procedure. Issues in Criminological and Legal Psychology.* No. 20. Leicester, UK: British Psychological Society.

Munakata, Y., & Yerys, B.E. (2001) All together now: when dissociations between knowledge and action disappear. *Psychological Science, 12,* 335-337.

Normandeau, S., Larivee, S., Roulin, J., & Longeot, F. (1989). The balance scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology, 150,* 237–250.

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.

Onishi KH, Baillargeon R. (2005) Do 15-month-old infants understand false beliefs? *Science. 308(5719):*255-8.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* New York: Dover.

Pazzani, M.J. (1991) A Computational Theory of Learning Causal Relationships. *Cognitive Science, 15(3):* 401-424

Peirce, C.S. (1998). *The Essential Peirce: Selected philosophical writings,* 1893-1913. Bloomington, IN: Indiana 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.

Piaget, J. (1929) *The Child's Conception of the World.* New York: Harcort, Brace.

Piaget, J. (1930). *The child's conception of physical causality.* New York: Harcourt, Brace & Company.

Piaget, J. (1951). *Plays, Dreams, and Imitation in Childhood.* London: Routledge and Kegan Paul Ltd.

Poole, D.A. & White, L.T. (1991) Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27,* 975-986.

Read, S.J. & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65,* 429-447.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14,* 465-471.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods, 2nd ed.* Springer, New York.

Rozenblit, L.R. and Keil, F.C. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, 26,* 521-562.

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.

Salmon, W. (1984) Scientific *Explanation and the Causal Structure of the World.* Princeton University Press.

Salmon, W. (1989) *Four Decades of Scientific Explanation.* University of Minnesota Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*

Saxe, R., Tenenbaum, J.B., & Carey, S. (2005). 10 and 12-month-old infants' capacity for causal attribution. *Psychological Science, 16,* 995-1001.

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.

Schneider, W., & Bjorklund, D. (1998). Memory. In W. Damon (Series Ed.) & D. Kuhn & R.S. Siegler (Vol. Eds.) *Handbook of child psychology: Vol. 2. Cognition, perception, and language (5th ed.).* New York: Wiley.

Schneider, W., & Pressley, M. (1997). *Memory development: Between two and twenty (2nd ed.)* Mahwah, NJ: Erlbaum.

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 & Sommerville, (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development, 77(2),* 427-442.

Schulz, L. E. & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40(2),* 162-176.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science 10*(3), 322-332.

Schulz, L. E., Kushnir, T., & Gopnik, A. (in press). Learning from doing: Interventions and causal inference. In A. Gopnik & L. E. Schulz (Eds.), *Causal Learning; Psychology, Philosophy and Computation.* New York: Oxford University Press.

Schulz, L., Bonawitz, E.B., & Griffiths, T.L. (2007). Can being scared give you a tummy ache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology, 43,* 1124-1139.

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

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

Shafto, P. & Goodman, N. (2008) Teaching games: Statistical sampling assumptions for pedagogical situations. *Proceedings of the 30th annual meeting of the Cognitive Science Society.*

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology, 37(1),* 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, US: Academic Press, Inc.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shi, L., Feldman, N. H., & Griffiths, T. L. (2008). Performing Bayesian inference with exemplar models. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Shtulman, A. (in press). The development of possibility judgment within and across domains. *Cognitive Development.*

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

Shultz, T. R. & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development, 46,* 394-399.

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

Siegler, R. S. & Liebert, R.M. (1975). Acquisition of formal scientific reasoning by 10 and 13-year-olds: Designing a factorial experiment. *Developmental Psychology, 11,* 401-412.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). New York: Cambridge University.

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

Singer, D. G. , Golinkoff, M. R., & Hirsh-Pasek, K. (2006). *Play = Learning: How play motivates and enhances children's cognitive and social-emotional growth.* New York: Oxford University Press.

Sobel, D. (2006). When Mr. Blicket Wants It, Children Are Bayesian. *Proceedings of the Cognitive Science Society.* 810-816.

Sobel, D. M. & Kirkham, N. Z. (in press). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology.*

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(3),* 303-333.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62,* 753-766.

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

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7(6),* 337-342.

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

Tenenbaum, J. B., Sobel, D. M., Griffiths, T. L., & Gopnik, A. (in submission). Bayesian inference in causal learning from ambiguous data: Evidence from adults and children.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science, 10,* 309-318.

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

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435-467.

Tomasello, M. and Barton, M. (1994). Learning words in nonostensive contexts. *Journal of Developmental Psychology, 30(5)*:639--650.

Vosniadou, S. & Brewer, W.F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 535-585.

Vygotsky, L. (1978) *Mind in Society.* Cambridge, MA. Harvard University Press.

Watson, J. S., & Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly, 18(3),* 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. New directions for child development No. 75* (pp. 7-25). San Francisco, CA: Jossey-Bass/Pfeiffer.

Wellman, H.M. & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L.E. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 261-279). Oxford University Press.

Wellman, H.M., & Lagattuta, K.H. (2004). Theory of mind for learning and teaching: the nature and role of explanation. *Cognitive Development, 19,* 479–497.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.

Xu, F. & Garcia, V. (2008) Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 5012-5015.

Xu, F. and Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review 114(2).*

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.

Appendix A

Text of *Within Domain* storybook used in Chapter 2, Experiment 1

Title: <u>Bambi's Adventures</u>

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.

Text of Cross Domains storybook used in Chapter 2 Experiment 1

Title: <u>Bunny's Big Week</u>

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

Appendix B

Stimuli for *Far Transfer* study of Chapter 2, Experiment 2.
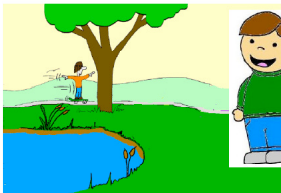
Physically possible: This is Alex. Alex is at the playground. Alex is throwing a ball. Alex is throwing the ball near a lake. Alex' friend tells Alex that if he throws the ball into the water it will make a big splash. Can that happen? Can Alex make a splash by throwing the ball into the water?

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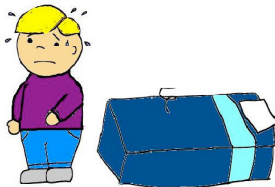
Psychogenic sickness: This is Jordan. Jordan is in his bedroom. Jordan is upset and nervous because he has to stay with a babysitter. Jordan feels very upset about the babysitter. Jordan's friend tells Jordan that if he keeps being upset and nervous he will start to feel sick. Can that happen? Can Jordan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