Title
Learning by thinking and the development of abstract reasoning

Author
Walker, Caren Michelle

Publication Date
2015

Peer reviewed|Thesis/dissertation

Learning by Thinking and the Development of Abstract Reasoning

by

Caren Michelle Walker

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Psychology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Alison Gopnik, Co-Chair
Tania Lombrozo, Co-Chair
Fei Xu
John J. Campbell

SUMMER 2015

Learning by Thinking and the Development of Abstract Reasoning

Abstract

Learning by Thinking and the Development of Abstract Reasoning

by

Caren Michelle Walker

Doctor of Philosophy in Psychology

University of California, Berkeley

Dr. Alison Gopnik, Co-Chair

Dr. Tania Lombrozo, Co-Chair

As adults, we have coherent, abstract, and highly structured causal representations of the world. We also learn those representations, as children, from the fragmented, concrete and particular evidence of our senses. How do young children learn so much about the world so quickly and accurately? One classic answer points to the similarities between children's learning and scientific learning. In particular, researchers have proposed that children, like scientists, implicitly formulate hypotheses about the world and then use evidence to test and rationally revise those hypotheses. In testing these claims, the vast majority of research in this area has investigated children's developing abilities to draw causal inferences from observed data. However, we know much less about the human ability to build abstract knowledge that extends beyond their observations, simply by thinking. In the current dissertation, I examine a suite of activities that involve learning by thinking in the causal domain, and consider how these activities impose unique, top-down constraints on the processes underlying causal learning and inductive inference. First, in chapter 1, I situate this work within the theoretical context of rational constructivism that has recently emerged in the field of cognitive development. Chapter 2 then presents a series of experiments demonstrating very young children's ability to infer the abstract relations "same" and "different" in a novel causal reasoning task. I conclude this chapter by considering the implications of these findings for our understanding of the nature of relational and causal reasoning, and their evolutionary origins. Chapter 3 extends this paradigm to describe a surprising developmental pattern: younger children outperform older children in inferring these abstract relations. I provide evidence that this failure may be explained by appealing to the role of learned biases in constraining causal judgments. The second part of this chapter explores how prompts to explain during learning facilitate children's ability to override a preference to attend to object properties, and instead reason about abstract relations. Chapter 4 presents empirical findings further examining the particular effects of explanation on the mechanisms underlying causal inference in preschool-aged children. In particular, results demonstrate that explanation prompts children to ignore salient superficial properties and consider inductively rich properties that are likely to generalize to novel cases. Finally, in Chapter 5, I discuss the implications for this body of work as a whole, and suggest a variety of future directions. Taken together, this research contributes to our understanding of the cognitive processes that influence early learning and inference in early childhood.

Acknowledgments

I thank my advisor, Alison Gopnik, who has taught me to be the type of scientist who engages with big ideas – the very ideas that inspired me to pursue research in developmental psychology. Her mentorship has truly been the highlight of my graduate training at Berkeley. In addition to being a leader in the field, she is also a remarkably interesting and genuinely lovable human being. I walk away from every conversation with Alison a little wiser. I am equally grateful to Tania Lombrozo, whose outstanding mentorship is really the product of two characteristics that, in my experience, rarely appear together: she is both a brilliant and productive scholar and an extraordinarily attentive and available advisor. Tania's ability to quickly synthesize abstract ideas and help to translate them into concrete and focused research questions has been invaluable to my progress. I would also like to thank my other committee members – Fei Xu and John Campbell – for their many insightful thoughts and words of guidance throughout the years. I feel extremely grateful to have had such an inspiring collection of minds contribute to this work.

I must also thank my various collaborators and labmates, Sophie Bridgers, Josh Abbott, Joseph Williams, Elizabeth Bonawitz, Stephanie Denison, Jane Hu, Sara Gottlieb, Azzurra Ruggeri, Adrienne Wente, Alex Carstenson, Mike Pacer, Daphna Buchsbaum, and Rosie Aboody (among many others) for all of their ideas, honest criticism, enthusiasm, and social support along the way. Sophie deserves special thanks for spending countless hours writing with me in coffee shops over the years. I have also been extraordinarily lucky to have some of the hardest working, independent, brilliant (and generally adorable) undergraduate research assistants during my five years at Berkeley. Their willingness to learn and tireless enthusiasm have made this time so much easier. It has been my privilege to mentor and learn from each of them.

Nothing about this process would have been possible without the endless love and support from my family. I am incredibly lucky to have two parents who never stopped encouraging me to continue my (seemingly endless) career as a professional student. The path leading to the completion of this dissertation was in no way straightforward – and I would not have made it without their guidance. I thank my Mom for always being available to listen to me – in laughter, tears, and delirium – and for believing every time, without hesitation, that I would be able to accomplish whatever seemingly impossible task I confronted. I thank my Dad for passing along his insatiable curiosity and motivation, for all of his advice along the way, and for every time he spent hours talking through a problem with me. Thanks to my extraordinary siblings, Michael and Lisa, for inspiring me a little bit everyday – each in their own unique way. Finally, I am forever grateful to my husband, Rand, who has been a constant source of love and inspiration to me…I share this Ph.D. with him. He came with me to Berkeley very early in our relationship and has made countless sacrifices to help me along the way, never doubting my ability to succeed and always keeping a smile on my face.

Thanks also to the countless participants and their parents and teachers for volunteering both their time and brainpower to this research. Lastly, this work would not have been possible without the financial support provided by the Lisa M. Capps Graduate award, the McDonnell Foundation, the National Science Foundation, and the Elizabeth Munsterberg Koppitz Fellowship from the American Psychological Foundation.

To the little boy that I have not met, who has kept me company while I write this.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General introduction

The primary project of cognitive development stems from a long-standing question regarding the nature of human knowledge: How it is possible that we ever acquire abstract knowledge about the world, given that the data that we receive from our sensory experience is so unstructured, concrete, and incomplete? Causal learning is a notorious example of this apparent incompatibility. In fact, it was Hume (1748) who originally articulated this difficulty: all we see are contingencies between events – one event follows another. How do we ever know that one event actually caused the other? To make matters more difficult, causal relations are rarely limited to just two events. Instead, dozens of different events are related in complex ways.

Discovering the underlying causal structure in the world is one of the major inductive problems that young learners face during development as they construct and revise early intuitive theories. Despite the apparent complexity of this problem, there has been a great deal of research that suggests that by the age of five, children understand a quite a bit about the causal world, including the principles of everyday physics (e.g., Bullock, Gelman & Baillargeon, 1982; Spelke, Breinlinger, Macomber, & Jacobson, 1992), biology (e.g., Gelman & Wellman, 1991; Inagaki & Hatano, 2006), and psychology (e.g., Gopnik & Wellman, 1994; Perner, 1991). By 2-years of age, children begin to make causal predictions and provide causal explanations for physical phenomena in the world (e.g., Legare, Gelman, & Wellman, 2010; Hickling & Wellman, 2001), for the actions of others (e.g., Wellman & Liu, 2007), and even for imaginary or counterfactual scenarios (e.g., Harris, German, & Mills, 1996; Sobel & Gopnik, 2003).

Beginning in ancient philosophy, and reinvented in the language of psychology, there have been two main solutions to this problem of the origins of human knowledge. In one camp, nativists have proposed that abstract knowledge must exist a priori, and that children build upon innate, domain-specific modules, which serve to impose top-down constraints on incoming information. This solves the problem of knowledge by removing the necessity to build abstract representations from data coming in from the world. In the opposing camp, empiricists have simply denied that truly abstract mental representations exist. Instead, they claim that all knowledge may be understood as a collection of associations that are acquired in a bottom-up manner by domain-general learning mechanisms that track correlations in incoming data. Each of these opposing positions has long suffered from a host of inconsistencies – neither appearing to account for the full complexity of human learning. Indeed, it has long been assumed that the truth lies somewhere between the nativist and empiricist camps. In order to find this middle ground, it is necessary to provide a precise account of learning that provides a means for

combining domain-specific prior knowledge (some of which may stem from innate constraints) with a domain-general process of learning from evidence.

Over the last two decades, developmental psychologists have begun to incorporate formal methods from machine learning to provide a rational framework underlying early learning. This emerging theoretical perspective, often referred to as "rational constructivism," offers a middle ground between nativist and empiricist perspectives (Xu, 2007; Xu, Dewar, & Perfors, 2009; Xu & Griffiths, 2011). The rational constructivist account has grown out of a long tradition in developmental psychology that has proposed that the processes underlying children's knowledge acquisition and development may be analogous to scientific theory and revision. This view, the "theory theory," (e.g., Carey, 1985; Gopnik, 1988; Gopnik & Meltzoff, 1997; Wellman, 1990; Wellman & Gelman, 1998) proposes that knowledge is represented in coherent, hierarchical, causal theories that support prediction, explanation, and control.

1.1.1 Bayesian inference

One feature of the rational constructivist framework is the application of probabilistic models in characterizing the mechanisms underlying learning and inference (Chater, Tenenbaum & Yuille, 2006, Griffiths, et al., 2010; Pearl, 2000; Glymour, 2003). Many of the ideas about probability that underpin these models were first formulated by the philosopher and mathematician, Reverend Thomas Bayes, in the 18th century, and are now being successfully applied to a very broad set of problems in developmental psychology, including the mechanisms underlying early learning (e.g., Glymour, 2003; Gopnik & Schulz, 2007; Tenenbaum, Griffiths, & Kemp, 2006), language acquisition (e.g., Chater & Manning, 2006; Tenenbaum & Xu, 2000; Xu & Tenenbaum, 2007; Niyogi, 2002; Dowman, 2002; Regier & Gahl, 2004), and the development of social cognition (e.g., Goodman, Baker, Bonawitz, Mansinghka, Gopnik, Wellman, Schulz, & Tenenbaum, 2006; Baker, Saxe, & Tenenbaum, 2006), among others. This work has also provided a solution for the problem of causal induction: how we derive rich, abstract representations from the sparse, concrete data that is available in our environment. More specifically, these accounts describe a mechanism that allows theory-like knowledge to be derived from data in our environment while also explaining how prior knowledge constrains the inferences that we make, and the evidence that we choose to attend to.

The idea that serves as the foundation for all of this work is that learning is based upon the *assessed probabilities of possibilities*: we form rational inferences based upon the fact that some possibilities are more likely than others. Bayesian inference provides a formal account of how a learner should update her prior belief in some hypothesis, h, in light of new evidence, d (e.g., Griffiths et al., 2011; Gopnik et al., 2004; Gopnik & Wellman, 2013). Specifically, the learner evaluates the posterior probability of the hypothesis, $p(h|d)$, by applying Bayes' rule: $p(h|d)=p(h)*p(d|h)/p(d)$, where $p(d|h)$ is the "likelihood" of the data given the hypothesis, and $p(d)$ is the probability of the data under all hypotheses in question, h, and alternatives, $\sim h$.

In other words, as we accumulate more evidence about the underlying causal structure of the world, we systematically update the likelihood of various hypotheses. As a result, a very small amount of evidence can effectively support one hypothesis over another. Similarly, if the evidence is strong enough, even the most unlikely possibility can turn out to be true, regardless of our previous experience or currently-held theories. According to this perspective, the process of learning represents a dynamic movement towards more informed inferences that better approximate the truth in a broader range of scenarios. Further, because Bayesian learning uses

structured priors and likelihoods that are drawn both from the learner's background or innate knowledge about causal structure, as well as observed or hypothetical data, variations on this simple algorithm provides a natural framework for combining the strengths from nativist and empiricist accounts of the origins of knowledge.

1.1.2 Learning abstract hypotheses

A related phenomenon that this computational account is able to address is the fact that children use very sparse data to infer abstract causal laws that guide subsequent learning. Previous research has demonstrated that children as young as 15 months of age can learn specific causal relationships from statistical data (Gopnik et al. 2004; Gopnik & Schulz, 2007; Gweon & Schulz, 2011; Gopnik & Wellman, 2012). In the current dissertation, I focus on investigating when and how children are also able to learn more abstract, general, causal principles or "overhypotheses" – that is, hypotheses about which kinds of more specific hypotheses are more likely (Kemp et al. 2007).

Piaget (1930, 1953) originally proposed that children construct abstract laws over extended periods of time, following the acquisition of sufficient evidence. Intuitively, it seems very likely that more abstract hypotheses would be acquired after lower-level, concrete ones that are based on specific features of objects. Indeed, a variety of recent accounts continue to be based on similar claims (e.g., Christie & Gentner, 2014; Gentner, 2010). On the contrary, however, many cases have been documented in which abstract causal laws appear to precede the data. For example, decades of evidence from developmental studies of psychological essentialism (e.g., Gelman, 2005; Keil, 1989) has demonstrated that children assume that animals from similar species are likely to share internal structures. More impressively, they use this assumption in classifying novel cases well before they have any substantial biological knowledge. Children frequently grasp these general principles at the same time, or even *before* they grasp the specific causal relations underlying them (Gelman & Gottfried, 1996; Kemp et al., 2007; Mansinghka et al., 2006; Leher & Schauble, 1998; Rozenblit & Keil, 2002; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Tenenbaum & Niyogi, 2003; Tenenbaum et al., 2006). This type of observation has led many to posit the existence of innate knowledge.

However, theoretical advances drawing on Bayesian accounts of the "blessing of abstraction" (Goodman et al., 2011) combined with empirical research on early learning (Dewar & Xu, 2010; Schulz, Goodman, Tenenbaum & Jenkins, 2008) suggest that children's ability to learn abstract principles does not necessarily depend upon extensive prior experience. In particular, the application of hierarchical Bayesian models has provided a method for learning at multiple levels of abstraction simultaneously (e.g., Tenenbaum, Griffiths, & Kemp, 2006). As a result, abstract learning need not progress in a bottom-up manner. In fact, computational analyses indicate that a learner who is able to simultaneously learn abstract and specific knowledge is nearly as efficient as one who is equipped with an innate theory.

These higher-order generalizations, "framework theories" (Gopnik & Wellman, 1992), or "overhypotheses" (Goodman, 1955), provide the learner with information about the types of specific hypotheses that are likely to be true. Having an overhypothesis, or general principle, leads the learner to assign a higher prior probability to certain types of specific hypotheses, and so constrains children's interpretation of new data (Kemp et al., 2007). This ability to learn abstract and specific relations in tandem helps to explain how it is that children acquire the impressive amount of causal knowledge evident in their early intuitive theories about the world.

1.1.3 Learning by thinking

In testing the claims of this model of learning, the vast majority of research conducted has investigated children's developing abilities to draw causal inferences from their observations, from their physical exploration of the world, and from social information from other people. However, we know much less about how even very young learners are able to acquire abstract representations that extend beyond their experiences, simply by thinking. One of the distinguishing features of human cognition is our ability to generate new ideas by thinking alone. How is learning by thinking possible? What is the role of internal processes in causal learning and reasoning? What does this phenomenon tell us about the nature of mental representations and how they change? Accepting the claim that new learning can occur in the absence of new data acknowledges the incompleteness of current the models, and their inability to account for the full nature of mental representations and how they change.

To begin to answer these questions, it is necessary to first isolate the contributions of our observations from the mechanisms that underlie learning. To this end, the present dissertation focuses on learning contexts that are particularly widespread in childhood, emphasizing learning by analogy, by explanation, and the intersection of the two. The chapters that follow will examine these phenomena as a means to explore the mechanisms underlying children's ability to reason about abstract properties. In particular, I will examine how these activities influence the nature of their developing knowledge representations by imposing unique, top-down constraints on the processes underlying early learning and inference in the causal domain.

1.2 Goals of the present dissertation

The current dissertation investigates the mechanisms underlying children's early capacity to overcome a bias to attend to perceptual features and infer abstract properties. The first line of work (examined in Chapters 2 and 3) focuses on the early development of analogical reasoning, which is characterized by the ability to consider the abstract similarities between objects and events. Analogical reasoning is essential for building abstract knowledge. Children learn to reason about categorical relations (e.g., X is edible), causal relations (e.g. X causes Y), and spatial relations (e.g., X is above Y), to name a few. These relational concepts are critical for inductive inference, and some have speculated that they may be a major contributor to species-specific intelligence (Penn, Holyoke, & Povinelli, 2008). However, relational concepts are not easily accessible among young children, and previous research has shown that analogical reasoning develops gradually (Gentner, 1998). Part of the reason for this difficulty is the fact that learning relations often involves overcoming a bias to attend to objects to consider higher-order properties. For example, thinking about how an atom is like a solar system requires that you set aside the surface properties of the individual objects – like their size – and instead focus on the underlying or abstract structural similarities that exist between the two.

The cognitive ability that forms the foundation of this type of analogical reasoning is the capacity to consider the abstract relations "same" and "different." Chapter 2 will describe three experiments examining 18- to 30-month-olds' relational inferences in a causal version of a relational match-to-sample task that is typically conducted with non-human primates (Premack, 1983). The findings reported in this chapter will suggest that very young children are already able to infer "sameness" and "difference" as relational causal principles from just a few

observations, and use this inference to guide their subsequent actions to bring about a novel causal outcome. The results of this chapter indicate that the seeds of analogical reasoning are in place surprisingly early, emerging spontaneously a few months after the first evidence of the ability to learn about specific causal properties. I will conclude the chapter by considering the implications of these findings for understanding the nature of relational and causal reasoning, and their evolutionary origins.

The results of Chapter 2 appear to contrast with a large literature and long-standing theory that relational reasoning may be late developing, since older children notoriously have difficulty inferring abstract relations (e.g., Gentner, 1998). In Chapter 3, I will therefore explore this apparent incompatibility by examining both the developmental trajectory and underlying mechanisms of children's ability to engage in abstract relational reasoning. In particular, Chapter 3 will describe a surprising developmental pattern: Younger learners are *better* than older ones at inferring abstract causal relations. Across several experiments, I manipulate both the data that children's observe and their search procedure to assess the influence of each of these factors on relational reasoning. To do so, I present the same causal relational task reported in Chapter 2 to both toddlers and older, preschool-aged children. Results of the first experiment reported in Chapter 3 will demonstrate that while younger children (18-30-month-olds) have no difficulty learning these relational concepts, older children (36-48-month-olds) fail to draw this abstract inference. To address this puzzling decline, I will then describe a series of studies assessing the claim that older children have learned the bias to attend to properties of individual objects. In particular, I will test whether the difference in performance found between younger and older children might be the result of an overhypothesis that individual kinds of objects lead to effects. I will discuss these findings in light of recent computational theories of learning, and in particular, that younger children may be more flexible in their commitments about causal systems than older ones.

The final experiment reported in Chapter 3 will explore the extent to which the act of generating explanations may serve as one route to facilitate children's recognition of the relevance of higher-order relations. There is strong theoretical support for a relationship between explanation and abstraction. Within philosophy and psychology, "subsumption" accounts suggest that a good explanation shows how a phenomenon is an instance of a unifying pattern that encompasses a wide range of diverse cases (e.g., Kitcher, 1989; Friedman, 1974; Williams & Lombrozo, 2010). Such patterns will tend to be general, and hence abstract away from the details of individual cases. In previous work, I have found that children tend to discover broad patterns and underlying regularities when prompted to explain in a causal learning task (Walker, Williams, Lombrozo, & Gopnik, 2012; Walker, Lombrozo, Williams, & Gopnik, under review). Explanation may therefore be a particularly valuable tool for guiding children away from appearances to privilege properties that highlight abstract structure. Indeed, the findings reported in the final experiment of Chapter 3 will support the hypothesis that explanation influences how children exercise their representational abilities, potentially scaffolding the transition from a preoccupation with salient surface properties to considering higher-order properties.

Chapter 4 will further assess the claim that the particular constraints imposed by explanation lead children to generate hypotheses that appeal to broad generalizations that highlight abstract structure. In particular, I explore the hypothesis that generating explanations scaffolds the transition from a preoccupation with salient surface properties to considering properties that are more "projectible," i.e., have greater inductive potential to generalize to novel cases. To do so, Chapter 4 reports the results of four experiments with preschool-aged children

testing the hypothesis that engaging in explanation promotes inductive reasoning on the basis of shared causal properties as opposed to salient (but superficial) perceptual properties. First, 3- to 5-year-old children are prompted to explain during a causal learning task to assess whether they are more likely to override a tendency to generalize according to perceptual similarity and instead extend an internal feature to an object that shared a causal property. A second experiment will seek to replicate this effect of explanation in a case of label extension (i.e., categorization). Finally, a third experiment will examine whether explanation improves memory for clusters of causally relevant (non-perceptual) features, but impairs memory for superficial (perceptual) features. In other words, I explore whether the effects of explanation extend to impact lower-level processes, influencing the features children attend to and recall. The data reported in Chapter 4 will support the proposal that engaging in explanation influences children's reasoning by privileging inductively rich, causal properties. This suggests that 3-year-olds already have the conceptual resources to reason on the basis of non-obvious properties, and that explaining facilitates their access to the inductive relevance of those properties.

Collectively, these studies are designed to examine how thinking imposes constraints on learning, impacting the development of abstract knowledge.

Chapter 2

Toddlers infer higher-order relational principles in causal learning

2.1 Introduction

Learning about causal relationships is one of the most important and challenging problems young humans face. Causal knowledge allows you to act on the world – if you know A causes B, you can act on A to bring about B. Studies show that children as young as 16 to 24 months of age can quickly learn causal properties of objects from patterns of statistical contingency and can act on that knowledge to bring about effects (e.g., Sobel & Kirkham, 2006; Meltzoff, Waismeyer & Gopnik, 2012; Gweon & Schulz, 2011; for reviews see Gopnik, 2012; Gopnik & Wellman, 2012).

Much of this research on early causal learning has used a "blicket detector" paradigm (Gopnik & Sobel, 2000), in which children learn which objects activate a novel machine. Children's inferences in these tasks go beyond associative learning, revealing the distinctive profile of causal inference. For example, children will use these inferences to design novel interventions – patterns of action they have never observed – to construct counterfactuals and make explicit causal judgments, including judgments about unobserved features (e.g. Gopnik & Sobel, 2000; Gopnik et al. 2004; Schulz, Gopnik, & Glymour, 2007; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007).

However, we know much less about the development of children's ability to infer higher-order relational causal principles. According to "theory theorists" of cognitive development, children are not only learning particular causal relationships, but also higher-order generalizations *about* causal structure (e.g., Carey, 2010, Wellman & Gelman, 1992, Gopnik & Meltzoff, 1997). Recent computational work also suggests that higher-order generalizations can help children learn new specific relationships from perceptual data more quickly (e.g., Griffiths & Tenenbaum, 2007; Goodman, Ullman, & Tenenbaum, 2011; Kemp, Perfors & Tenenbaum, 2007).

Causal inferences might be more or less abstract, higher-order, or relational in many different ways. Here we focus on just one contrast: between object properties, such as shape or color, and higher-order relations between those properties, such as whether they are the same or different. For example, very young children can learn that red blocks activate a toy. When can they learn that two blocks that are the same (regardless of their color) can do so?

Empirical research using looking-time measures suggests that human infants may be able to recognize patterns of data that involve higher-order relations such as "same" (Dewar & Xu,

2010; Ferry, Hespos & Gentner, 2015; Tyrell, Stauffer & Snowman, 1991). However, there is no evidence to date that infants can use those patterns to make causal inferences or guide subsequent actions.

In fact, earlier studies indicate that even preschoolers have difficulty making inferences in higher-order relational reasoning tasks (e.g., Christie & Gentner, 2007, 2010; Gentner, 2010). Children succeeded only when given labels or linguistic scaffolding to point out the pattern of similarity. Indeed, even when explicitly instructed to objects, 3-year-olds' performance was tenuous, dropping significantly below chance when test items were presented sequentially, rather than simultaneously (Christie & Gentner, 2010).

These findings might lead to the conclusion that learning higher-order relations and using them to guide actions depends on direct instruction, language, and cultural input (e.g., Christie & Gentner, 2007; Gentner, 2003, 2010; Gentner, Anggoro, & Klibanoff, 2011). However, these tasks often relied on verbal categorizations of complex, multi-dimensional stimuli (e.g., Christie & Gentner, 2010). One study by Smith (1984) provides a hint that children might do better in a more goal-directed task with simpler materials. In particular, 2½-year-olds showed some understanding of identity matching in a non-verbal game.

Higher-order relational reasoning has also been studied extensively in non-human animals. Chimpanzees, like young infants, are able to spontaneously detect a relational pattern in habituation tasks (Oden, Thompson, & Premack, 1990). However, they have more difficulty with a relational match-to-sample task (Premack, 1983; Oden, Premack, & Thompson, 1988). In these tasks, animals observe a relational pattern – AA', BB', and CC' all lead to a reward. Then they are given a choice between AB (object match) and DD' (relational match). Although A and B have each been associated with the reward, an animal who has inferred the higher-order relational pattern should choose DD'. Premack and colleagues have found that chimpanzees could not solve this relational task without hundreds of trials with feedback (Premack, 1988) or training to use linguistic symbols for "same" (Premack, 1976; 1983; Premack & Premack, 1983; 2002).

Additional comparative studies confirm that this task is especially difficult for non-human primates and other animals (see Penn, Holyoak, & Povinelli, 2008). Moreover, when non-human animals, such as baboons, *do* solve this task, they require extended training and thousands of trials, which may indicate the use of simpler perceptual strategies such as minimizing entropy in a perceptual array (Fagot, Wasserman, & Young, 2001; Wasserman, Fagot, & Young, 2001).

Do human children always require linguistic cues or extensive training to solve relational tasks like the preschoolers and primates in earlier studies? We designed a non-verbal "blicket detector" task to explore when children could use higher-order relations to make causal inferences. In contrast to previous studies of causal inference, the causal effect depended on whether the objects were the same or different, rather than on properties of the objects themselves.

2.2 Experiment 1

In Experiment 1, 21- to 24-month-olds were introduced to a novel toy that played music and 3 unique pairs of identical blocks: AA', BB', and CC'. The experimenter placed blocks on the toy and the toy either activated or did not. Although individual blocks failed to activate the toy alone, pairs of identical blocks produced the effect. Immediately after this brief training, we examined whether children learned the novel relational property (i.e., "same") by asking them to

activate the toy.

2.2.1 Method

2.2.1.1 Participants

A total of 23 21- to 24-month-old toddlers participated in Experiment 1 (*M* = 23.0 months; *SD* = 1.05 months; range = 20.9-25.0 months; 13 girls). Three additional children were tested but excluded for fussiness or for failing to respond. Children were recruited from daycare centers and museums, and a range of ethnicities resembling the diversity of the population was represented**.**

2.2.1.2 Materials

The toy was a 10" x 6" x 4" opaque white cardboard box containing a wireless doorbell. When a block "activated" the toy, the doorbell played a melody. In fact, the toy was surrepticiously activated by a remote control. Six painted wooden blocks in assorted colors and shapes (3 unique pairs of 2 identical blocks) were placed on the toy during the training phase. Six additional blocks were used during the test phase, including 2 novel pairs of identical blocks and 2 novel individual blocks.

2.2.1.3 Procedure

The procedure for Experiment 1 is illustrated in Figure 1. Following a warm-up, the toy was placed on the table. The experimenter said, "This is my toy. Some things make my toy play music and some things do not make my toy play music." Children then observed while the experimenter placed 6 blocks (3 unique pairs of "same" objects: AA', BB', CC') on the table in front of the toy. She said, "Let's try one!", selected a block (A) and placed it on top of the toy. No effect was produced. After a pause, the experimenter again said, "Let's try one!", selected the paired block (A') and placed it next to the first block (A) on top of the toy. This pair of objects (AA') activated the toy. The experimenter smiled and said, "Music!", removed the blocks and returned them to the pile of 6. This procedure was repeated with the two remaining pairs (BB' and CC'). The order of the pairs was randomized. Following all three demonstrations, all blocks were removed.

Next, the experimenter produced 3 test blocks (1 *novel paired block* (D), 1 *familiar block* (A), and 1 *novel distractor block* (E) and placed them in a row on the table. The order of presentation was randomized. She said, "Let's try one!", produced the *target block* (D'), and placed it on top of the toy. No effect was produced. The experimenter then pushed the toy and all 3 test blocks towards the child, and asked, "Can you pick one of these (pointing to the test blocks) to make my toy play music?"

The first test block that the child placed on the toy was recorded. The toy activated if the child correctly selected the *novel paired block* (D). If the child selected the *familiar block* (A) or the *novel distractor block* (E), the toy failed to activate. After this feedback, this procedure was repeated in a second test trial with a new set of test blocks.

If toddlers were acting based on the previous association between the block and effect, they should choose the *familiar block* (A). If they simply preferred to try novel blocks they

should pick the *novel distractor block* (E) as often as the *novel paired block* (D). However, if toddlers were able to learn the higher-order relation, they should select the *novel paired block* (D).

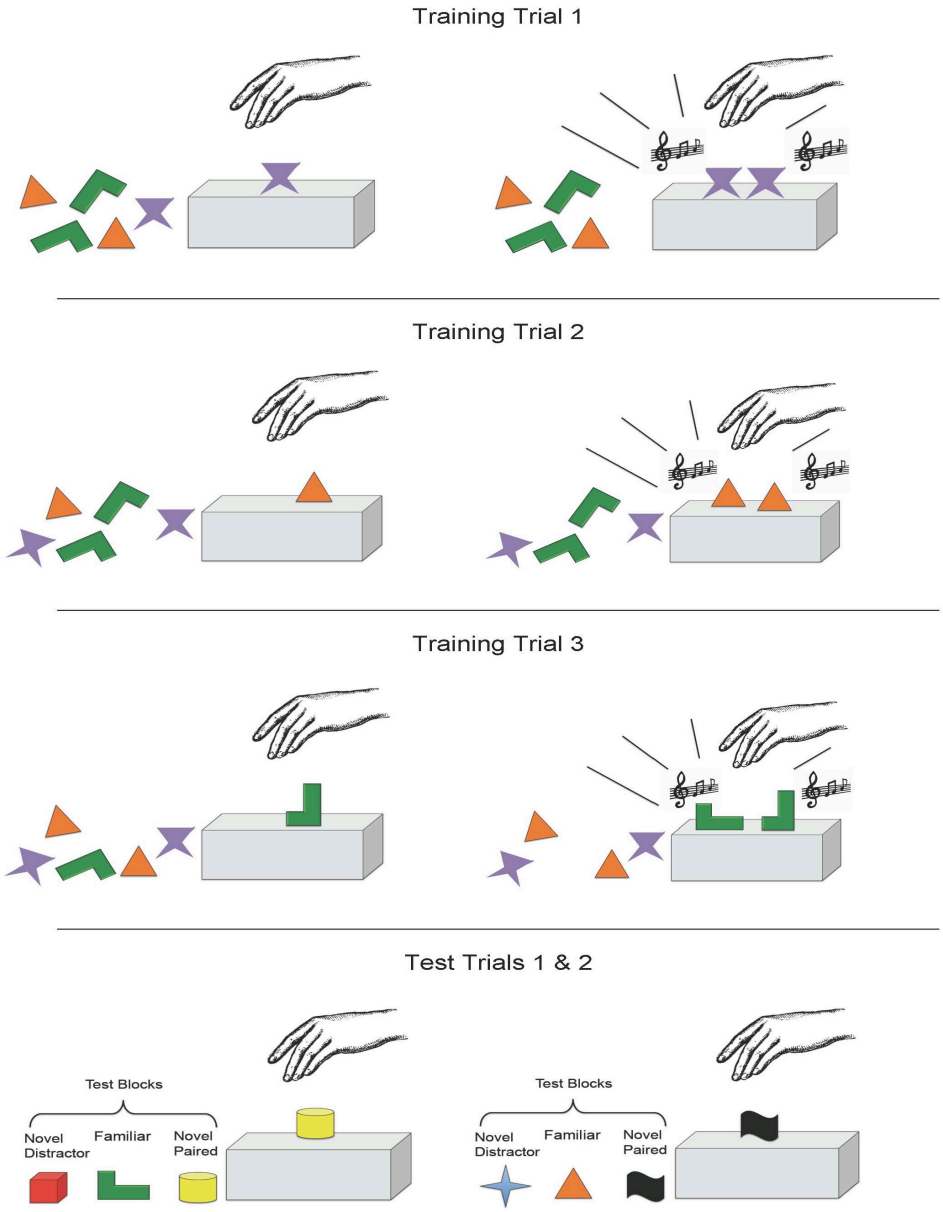

Figure 2.1 Schematic representation of training and test trials in Experiment 1. On each training trial, a single block was placed on the toy (no activation) and then an identical block was added, activating the toy. This was repeated for all 3 training pairs. On each test trial, 3 test blocks (*novel paired block*, *familiar block*, *novel distractor block*) were presented. The experimenter then placed the target block on the toy, yielding no effect. The child was asked to select one test block to activate the toy.

2.2.1.4 Coding and Reliability

Children received 1 point for selecting the *novel paired block* and 0 points for selecting either of the other blocks in each trial. Responses were recorded by a second researcher during the testing session, and all sessions were recorded for independent coding by a third researcher who was naïve to the the hypotheses of the experiment. Interrater reliability was very high; the two coders agreed on 99% of the children's responses. Two minor discrepancies were resolved by a third party.

2.2.2 Results and Discussion

Across the two test trials, children inferred the relational property and selected the *novel paired block* (D) more often than expected by chance ($M = 1.13$, $SD = .82$), $\chi^2(2) = 19.07$, $p < .001$. Fischer exact test revealed no order effects for test trials 1 and 2, $p = .39$. Children chose the *novel paired block* (61%) significantly more often than the *novel distractor block* (20%), $\chi^2(2) = 14.15$, $p < .001$ and significantly more often than the *familiar block* (15%), $\chi^2(2) = 14.09$, $p < .001$. A minority of children (4%) placed more than one block on the toy simultaneously, and were scored as incorrect.

Previous proposals have suggested that children are unable to reason relationally because they tend to focus on the identity of objects that have been previously associated with the outcome (e.g., Gentner, 2010). We show no evidence of this. In fact, only 39% of participants who answered incorrectly on a given trial selected the *familiar block*, with no difference in their selection between the *familiar block* and the *novel distractor*, $\chi^2(2) = 2.43$, $p = .30$. This is particularly surprising, given that this block had been associated with the effect during training.

Results suggest that by 21-24 months of age, toddlers are able to infer the causal principle – "same" – from just a few pieces of evidence and use this inference to bring about a novel causal outcome. However, children might have succeeded on this task by imitating the experimenter's selection or because they preferred to match, regardless of training. Experiment 1a was designed to address these alternatives.

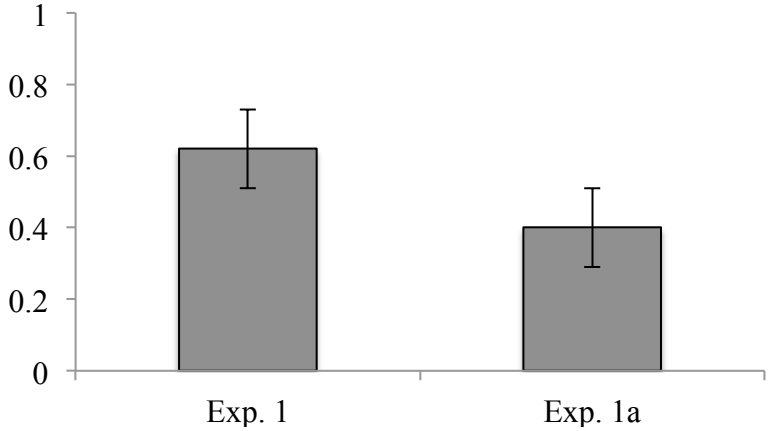

Figure 2.2 Percentage of 21-24 month olds in Experiement 1 and 1a who selected the matched pair (chance = .33).

2.3 Experiment 1a

The procedure for Experiment 1a was identical to Experiment 1, but the second object in the pair was occluded. Because children only observed the first item in each pair, they were given no evidence for the relational property. If children were simply imitating the experimenter or had a preexisting preference for matching, then children's performance should not differ from Experiment 1.

2.3.1 Method

2.3.1.1 Participants

Twenty 21-24-month-olds participated (M = 22.4 months; SD = 1.8 months; range = 20.8-25.6 months; 8 girls). Two additional children were tested but excluded for failing to respond. Recruitment procedures and demographics were identical to Experiment 1.

2.3.1.2 Materials and Procedure

Materials and procedures were identical to Experiment 1. However, children did not observe the second object during the training trials. Instead, the second object was occluded by a 4" x 4" square piece of cardboard. Additionally, only one test trial was administered in order to avoid providing feedback. Therefore, children could receive 0 or 1 points. Interrater reliability for Experiment 1a was 100%.

2.3.2 Results and Discussion

In the absense of evidence for the relational principle, only 40% of participants selected the paired block, [exact binomial test, p = .65, ns], which was significantly different from the percentage of children (61%) of the same age on their first trial in Experiment 1, $p < .05$ by Fischer's exact test (see Figure 2). Childrens' selections were evenly distributed: 40% of children selected the *novel paired block,* 35% of children selected the *familiar block*, and 25% of children selected the *novel distractor*. These results show that the findings from Experiment 1 could not have been the result of imitation or a bias to match.

2.4 Experiment 2

In the earlier primate studies, the canonical *relational* match to sample tasks presented pairs simultaneously during training (e.g., the relation "same" was taught using pairs AA' and BB'), and the animals had to choose between test pairs illustrating "same" (CC') and "different" (DE). Chimpanzees were unable to spontaneously succeed on this task – and had great difficulty even after engaging in trial and error over hundreds of trials. However, chimpanzees *were* able to solve a simpler match-to-sample task. In these tasks, the animals were first taught to match a test object (A) to a target object (A') through multiple positive and negative reinforcement trials over several weeks. They then generalized this pattern to novel objects without additional training (Premack, 1976; Premack & Premack, 1983; 2003; Oden et al, 1988).

Our task in Experiment 1, like the simple primate match to sample task, presented the training objects sequentially, and this may have made the task easier. However, Experiment 1 also differed in several ways from the primate task. Children learned by observation – they did not initially make the responses themselves – and they spontaneously chose the *novel paired block* after observing only three trials. Additionally, they never observed that the mismatching block would "not" produce the effect, so the association between the incorrect *familiar* block and the effect should have continued to be high.

In order to make the comparison to the primate tasks clearer, we designed a causal task that was more directly analogous to the primate relational match to sample tasks, in which both "same" and "different" objects are presented in pairs. This task also allowed us to explore whether children would infer the "different" relation as well as "same." We included toddlers from a broader age range to explore possible developmental differences, recruiting children aged 18 to 30 months.

Participants were randomly assigned to one of two conditions: *same* or *different.* In the *same* condition, children were given two pieces of evidence that pairs of "same" objects (AA', BB') simultaneously placed on the toy produced the effect. We also provided two pieces of evidence that pairs of "different" objects (DE, FG) *failed* to produce the effect. In the *different* condition, children were given the same four pieces of evidence but "different" pairs (DE, FG) produced the effect, while "same" pairs (AA', BB') *failed* to do so.

2.4.1 Method

2.4.1.1 Participants

Thirty-eight 18-30-month-olds participated (M = 25.8 months; SD = 3.8 months; range = 18.0-30.6 months; 21 girls), with 19 children randomly assigned to each condition (*same* and *different*). Seven additional children were tested but excluded: 4 due to failure to complete the study and 3 due to experimenter error. Recruitment procedures and demographics were identical to Experiments 1-1a.

2.4.1.2 Materials

The same toy from Experiments 1 and 1a was used. Eight painted wooden blocks in assorted colors and shapes (2 pairs of "same" blocks and 2 pairs of "different" blocks) were placed on the toy in pairs during training. The "same" blocks were identical in color and shape, and the "different" blocks were distinct in color and shape (see Figure 3). Four additional blocks were used during the test phase, including 1 novel pair of "same" blocks and 1 novel pair of "different" blocks. The pairs of test blocks were placed on 4" x 4" plastic trays.

2.4.1.3 Procedure

The procedure for Experiment 2 is illustrated in Figure 3. Following a warm-up, the toy was placed on the table. The experimenter said, "This is my toy. Some things make my toy play music and some things do not make my toy play music." Children then observed while the experimenter placed all 8 training blocks (A, A', B, B', E, F, G, H) in a random arrangement on

the table in front of the toy. The experimenter said, "Look at these things! We will try them on my toy."

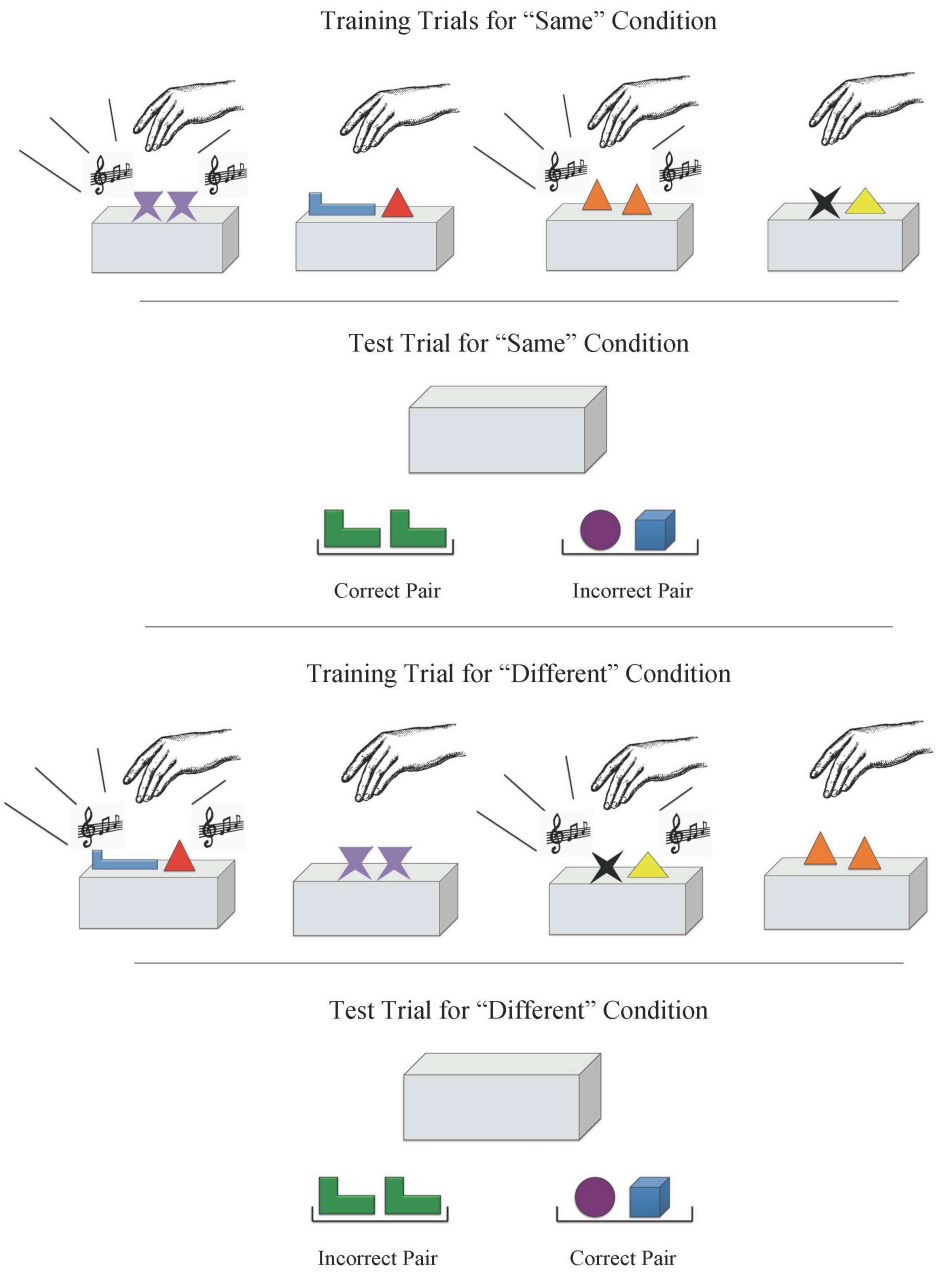

Figure 2.3 Schematic representation of training and test trials in the *same* and *different* conditions of the relational match-to-sample task in Experiment 2. On each training trial, a pair of blocks were placed on the toy. In the *same* condition, the pairs of identical objects activated the machine. In the *different* condition, the pairs of distinct objects activated the machine. Participants observed four pairs (two causal and two inert). On each test trial, 2 pairs of test blocks were presented ("same" and "different"). The child was asked to select the pair that would activate the toy.

Then, the experimenter removed all objects from view, selected the first pair of blocks (e.g., AA'), and placed the blocks simultaneously on the toy. Children in the *same* condition observed the pair of "same" objects activate the toy. The experimenter smiled and said, "Music! Let's try that again!", picked up the pair of blocks, and placed them back on the toy a second time, and children again observed the outcome. After this second demonstration, the experimenter removed the pair, selected another pair – a "different" pair (e.g., EF) – and placed it on the toy. This time, children in the *same* condition observed no effect. As with the first pair, this was demonstrated a second time before moving on to the third pair. This procedure was repeated for all 4 pairs: 2 pairs of "same" objects and 2 pairs of "different" objects. All pairs were placed on the toy twice. Therefore, children observed a total of 8 outcomes (4 positive and 4 negative).

Children in the *different* condition observed the same set of evidence as children in the *same* condition, with one critical change: pairs of "different" objects (e.g., EF) caused the toy to play music, while the pairs of "same" objects (e.g., AA') failed to activate the toy. There were no other differences in procedure. The particular objects included in each pair was randomized, as well as the order that the pairs were presented during training.

Following the training phrase in both conditions, the experimenter said, "Now it is going to be *your* turn. I want you to help me pick the ones that will make my toy play music!" The experimenter produced 2 pairs of test blocks (1 novel "same" pair [JJ'], 1 novel "different" pair [KL]). In order to avoid a novelty preference, *both* test pairs were composed of novel objects. The pairs were presented to the child on plastic trays. The experimenter held up the two trays, shook them to get the child's attention, and asked, "Can *you* pick the ones that will make my toy play music?" She then placed the trays on opposite sides of the table in front of the child. The side on which the correct pair was placed was randomized between subjects. The first tray that the child selected was recorded. Correct selections included pointing to the tray, reaching to the tray, or picking up the objects on the tray.

If they learned the relational property, then children in the *same* condition should correctly select the tray with the novel "same" objects (AA'), while children in the *different* condition should correctly select the tray with the novel "different" objects. Correct selections were given a score of "1" and incorrect selections were given a score of "0". Coding and recording procedures were identical to Experiments 1-1a. Interrater reliability was very high; the two coders agreed on all but one of the children's responses to the test questions.

2.4.2 Results and Discussion

Results of Experiment 2 appear in Figure 4. Children inferred the relational property and selected the correct pair more often than expected by chance (M = .79, SD = .41; chance = .5), [exact binomial test], p<.02] in both *same* and *different* conditions. In fact, performance was identical in the *same* and *different* conditions, with 15 out of 19 children in each condition selecting the test pair that corresponded with the relation learned during the training trials. Additionally, logistic regression revealed no significant developmental change in performance between 18 and 30 months of age, $\chi^2(1, 38) = .11$, $p = .74$ (ns). The fact that children responded differentially in the otherwise identical *same* and *different* conditions also allowed us to rule out superficial explanations for the results, such as imitation or a preference for same or different pairs – each condition acted as a control for the other condition.

Experiment 2 indicates that toddlers are able to infer the relational causal principles

"same" and "different" from a just few pieces of evidence, and use this inference to intervene to bring about a novel causal outcome.

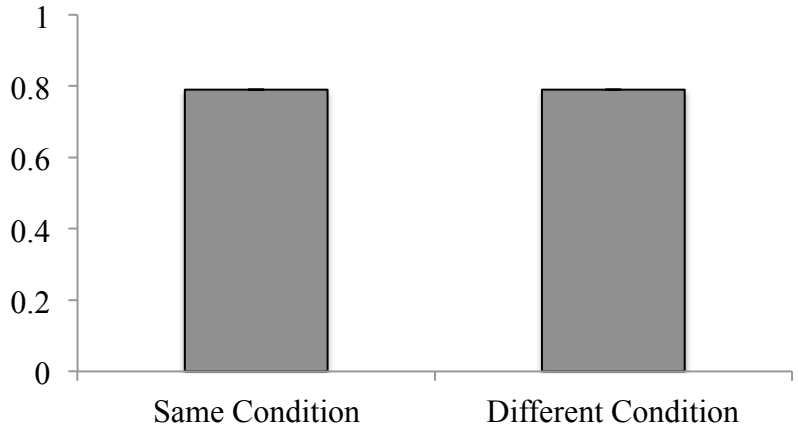

Figure 2.4 Percentage of toddlers in the *same* and *different* conditions in Experiment 2 who selected the correct pair during the test trial (chance = .50).

2.5 General Discussion

These findings show that human toddlers as young as 18 months can succeed on a causal relational match-to-sample task after only a few trials and without explicit linguistic cues, instruction, or reward. This study has implications for our understanding of both causal and relational reasoning. Using this method, toddlers are able to quicky learn higher-order relational causal principles and use them to guide their actions. This ability appears to be in place surprisingly early -- only a few months after the first evidence of the ability to learn about specific causal properties from contingency -- and it may be in place even earlier. This may help explain how children acquire the impressive causal knowledge evident in early "intutive theories" (Gopnik & Wellman, 2012; Carey, 2010).

These findings also contrast with the striking failure of non-human primates to solve similar tasks, even when the relation is associated with a strong pattern of positive and negative reinforcement, and even after hundreds (or thousands) of trials. This finding might support the suggestion that an ability to quickly learn relational causal concepts is a dimension on which humans differ from other primates. This might in turn reflect the broader evolution of higher-order relational cognition (Penn, Holyoak & Povinelli, 2008) or causal cognition in general (Heyes & Frith, 2012; Byrne, 1995; Buchsbaum et al., 2012).

Several questions for further study remain. One is whether the causal nature of this task was critical, or whether other aspects of the task, such as the fact that it involved goal-directed actions, might have made it easier for the children than relational tasks in other studies. It is also possible that children could succeed on this particular task by basing their causal inference on the observed association between the higher-order relational features and the effects. In other

"blicket detector" studies children's inferences go beyond association, but those studies would have to be replicated with the current relational design.

Further, it is possible that the children's success was due to a perceptual heuristic, as has been suggested for non-human primates (Penn et al., 2008; Fagot, et al., 2001; Wasserman, et al., 2001). According to this argument, it is possible to solve relational match-to-sample tasts using the perceptual cue, entropy (i.e., the Shannon entropy of AA' is 0, while that of AB is 1). Several features of the children's behavior weigh against this possibility: children saw pairs of objects (rather than multi-element displays), they observed only two positive and two negative trials, they never acted on the object, and their behavior was never reinforced. Indeed, no other species has come close to demonstrating the first-trial performance of these human children after so few observations (see Penn, et al., 2008). Additionally, although human participants have been shown to be sensitive to entropy, findings suggested that additional processes of categorization likely play a role in the human conceptualization of "same-different" relations (Fagot, et al., 2001). Nevertheless, future research examining this possibility would be informative.

Finally, it will be important to replicate this particular task with non-human primates to determine if, like children, they show greater success, or continue to have difficulty. Our protocol did not require a verbal response, so it may be useful in examining reasoning capacities in both pre-verbal human infants and possibly in non-human animals.

However, the current study does suggest that the ability to infer causal higher-order relations, an ability which could play a crucial role in further learning, is in place in humans from a very early age and does not depend on explicit linguistic cues or cultural scaffolding.

Chapter 3

The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts

3.1 Introduction

A growing literature indicates that children as young as 16 months of age are able to learn specific causal properties from contingency information and can act on that knowledge to bring about novel effects in the world (e.g., Gopnik & Wellman, 2012). But when and how can children learn more abstract principles *about* causal structure? Higher-order generalizations, "framework theories" (Gopnik & Wellman, 1992), or "overhypotheses" (Goodman, 1955), provide the learner with information about the types of specific hypotheses that are likely to be true. Recent computational work suggests that these generalizations about likely causal structures might help children learn new specific causal relationships from perceptual data more quickly and accurately (e.g., Goodman, Ullman, & Tenenbaum, 2011; Kemp, Perfors & Tenenbaum, 2007). The ability to quickly learn abstract and specific relations in tandem might explain how children acquire the impressive amount of causal knowledge evident in their early intuitive theories about the world.

In the current chapter, we examine children's developing ability to infer an abstract causal principle – a relation between objects (i.e., "same" and "different") that causes an effect – from a limited set of observations. Walker and Gopnik (2014) recently demonstrated that toddlers (18-30-month-olds) are surprisingly adept at learning and using the relational concepts "same" and "different" in a causal relational match-to-sample (RMTS) task. In this study, children were assigned to *same* or *different* conditions, and observed as four pairs of objects (two "same" pairs and two "different" pairs) were placed on a toy that played music. In the *same* condition, pairs of identical objects activated the toy while pairs of different objects did not. This pattern of activation was reversed for the *different* condition. During test, children were given a choice between two novel pairs: one pair of "same" and one pair of "different" objects, and asked to select the pair that would activate the toy. Children overwhelmingly selected the pair that was consistent with their training. These results suggest that the ability to reason about abstract relations is in place very early – emerging spontaneously only a few months after the first evidence of children's ability to learn about specific causal properties.

However, Walker and Gopnik's (2014) results with toddlers contrast with a large body of research demonstrating that older, preschool-aged children consistently demonstrate a bias to attend to individual object kinds (e.g., Christie & Gentner, 2007, 2010, 2014; Gentner, 1998

Gentner & Medina, 1998). These robust findings have led some to conclude that active and explicit reasoning about relations tends to develop *after* the establishment of more concrete, object-based representations (e.g., Christie & Gentner, 2010, 2014). This epistemological account indicates that, in general, object-based concepts are likely to be formed before relational concepts when learning in a new domain.

How might we interpret this apparent developmental reversal in which abstract reasoning seems to emerge in the first two years of life, but then disappear in early childhood? First, it is possible that older children failed to engage in relational reasoning in previous studies because of methodological problems – the tasks were simply too difficult. The toddlers in Walker and Gopnik (2014) may have succeeded because the novel causal procedure simply made the task easier. In Experiment 1a, we therefore present participants with exactly the same reasoning task used in Walker and Gopnik (2014). In addition to replicating this previous work with 18-30-month-olds, we also assess performance in older children (ranging from 30-48-month-olds) to detect the presence of a linear developmental trajectory. If the toddlers in Walker and Gopnik (2014) succeeded because of the particular methodological features of the task, then we would expect that older children would succeed as well.

Alternatively, it is possible that younger children succeed because they are relying on some simpler, more perceptual strategy that has been abandoned by older children, rather than making a genuine causal inference (Fagot, Wasserman & Young, 2001; Penn et al., 2008; Wasserman, Fagot & Young, 2001). In Experiment 1b, we therefore assess 18-30-month-olds a second time, using an even more stringent test of toddlers' causal understanding of the relational concepts. In addition to coding which pair of blocks the children selected (by pointing) to activate the machine in the causal RMTS task, we also coded whether the children themselves put the correct novel pair of blocks on top of the toy. This ability to design a new intervention, and to act on a cause in order to produce its effect has been argued to be one benchmark of true causal understanding (Pearl, 2000; Woodward, 2003).

There is at least one reason, however, why younger children might indeed genuinely outperform older children in learning these causal relational concepts. It may be that 3-year-olds have difficulty inferring such relations because they have learned a different "overhypothesis," namely, that individual kinds of objects, rather than relations between them, have causal power. Intuitively, it might seem plausible that more abstract hypotheses would be acquired later than lower-level, concrete ones based on specific features of objects (e.g., Christie & Gentner, 2010, 2014). However, theoretical advances drawing on Bayesian accounts of the "blessing of abstraction" (Goodman et al., 2011) combined with empirical research on early learning (Dewar & Xu, 2010; Schulz, Goodman, Tenenbaum & Jenkins, 2008) suggest that children's ability to learn abstract principles does not necessarily depend on extensive prior experience. In fact, children frequently grasp these general principles at the same time, or even *before* they grasp the specific causal relations underlying them (Kemp et al., 2007; Mansinghka et al., 2006; Schulz et al., 2008; Tenenbaum & Niyogi, 2003; Tenenbaum, Griffiths, & Kemp, 2006).

Hierarchical Bayesian models formalize how it is possible to draw relations among multiple levels of abstraction simultaneously (Tenenbaum et al., 2006). According to these accounts, learning at the most abstract level is surprisingly fast when compared with learning at lower, or more specific levels. As a result, abstract learning need not progress in a bottom-up manner. In fact, computational analyses indicate that a learner who is able to simultaneously learn abstract and specific knowledge is nearly as efficient as one who is equipped with an innate theory (Goodman et al., 2011).

According to Bayesian probabilistic models of cognitive development, learners explain newly observed evidence by searching through a space of potential hypotheses and testing these hypotheses against the data (e.g., Gopnik & Wellman, 2012). To do this, learners combine two probabilities: the "prior" – the probability of a particular hypothesis being true before any data are observed, and the "likelihood" – the probability of the observed data given that a particular hypothesis is true. Combining these two probabilities with Bayes rule produces the "posterior" – the probability of the hypothesis being true given the data. A learner can then compare the posteriors of different hypotheses, settling on the ones with the highest probabilities.

These models predict that if the prior probability of one hypothesis is high, then it will take stronger data to overturn it in favor of another hypothesis. Having an overhypothesis, or general principle, leads the learner to assign a higher prior probability to certain types of specific hypotheses, and so constrains children's interpretation of new data (Kemp et al., 2007). As a result, in order for the learner to consider a specific hypothesis that is inconsistent with the overhypothesis, the learner would need more evidence supporting this competing hypothesis than if she began with no prior expectations and instead assigned all possible hypotheses an equal prior probability (i.e., a "flat" prior).

For example in the case of Walker and Gopnik's (2014) causal reasoning task, an abstract principle of simplicity such as the "Bayesian Occam's razor" (Jefferys & Berger, 1992) might lead toddlers to initially prefer the relational hypothesis, since it proposes fewer causes to account for the data. Indeed, previous work demonstrates that young children show such simplicity preferences (Bonawitz & Lombrozo, 2012). However, if older children have also learned the general principle that individual object kinds are more likely to be causal (which is a robust bias even in adult learners [e.g., Lucas, Bridgers, Griffiths, & Gopnik, 2014]), this may serve to constrain their interpretation of the data, leading them to privilege individual properties over relational ones, in spite of simplicity considerations.

In other words, with increasing knowledge, learners develop expectations that constrain the set of hypotheses they consider. Although this allows learners to more quickly and accurately acquire information consistent with the general principles they have already inferred, it makes learning new information that is *inconsistent* with these general principles more difficult (see Gopnik et al., 2015). In fact, some recent research suggests that in some cases, apparent limitations in children's knowledge and cognitive abilities may lead younger children to be better learners than older children and even adults (Gopnik, Griffiths & Lucas, in press; Lucas et al., 2014; Seiver, Gopnik & Goodman, 2013).

In Experiment 2, we therefore adapt the causal RMTS procedure to test the proposal that older children are able to reason about abstract relations, but have learned the overhypothesis that individual kinds of objects are more likely to be causal. To do so, Experiment 2 provides older children with explicit negative evidence for the causal efficacy of individual objects. Because this evidence is inconsistent with the individual cause hypothesis, it might serve to lower the probability of this alternative. In other words, observing evidence *against* the individual cause hypothesis may lead older children to reject it, even though it is more consistent with their prior knowledge.

Finally, in Experiment 3, we aim to scaffold the relational inference using a different mechanism. Rather than changing the data, we change the way that children search through the hypothesis space. In particular, previous work has demonstrated that asking children to explain patterns of events imposes top-down constraints on their search procedure, leading them to privilege more general and inductively rich hypotheses (e.g., Lombrozo, 2012; Walker,

Lombrozo, Legare & Gopnik, 2014; Walker, Williams, Lombrozo & Gopnik, 2012, under review; Williams & Lombrozo, 2013). If preschool-aged children are already able to reason about relational properties, but assign a higher probability to individual object kind hypotheses, then introducing a prompt to explain may impose the opposite constraint, leading children to privilege abstract properties instead.

Across studies, we test the hypothesis that older children's "failure" on traditional relational reasoning tasks is due to the development of an overhypothesis about the importance of individual object kinds, rather than the inability to represent and reason about relations.

3.2 Experiment 1a

3.2.1 Method

3.2.1.1 Participants

A total of 141 children participated in Experiment 1a, including 56 36-48-month-olds (*M* = 41.6 months; range = 36.0 - 48.2 months), 40 30-36-month-olds (*M* = 33.6 months; range = 30.1 - 35.8 months), and 45 18-30-month-olds (*M* = 25.1 months; range = 18.9 - 29.9 months). Half of the children in each age group were randomly assigned to one of two between subject conditions: *same* and *different*. An additional 10 participants were tested, but excluded. Six children were excluded due to experimenter error or toy failure, and 4 were excluded due to participants' failure to complete the experiment. Children were recruited from local preschools and museums, and a range of ethnicities resembling the diversity of the population was represented.

3.2.1.2 Materials and Procedure

The procedure for Experiment 1a was an exact replication of the procedure used in Experiment 2 of Walker and Gopnik (2014) (see Figure 1).

Children were tested individually in a small testing room, seated at a table across from the experimenter. During the training phase, children saw 4 pairs of painted wooden blocks (2 same and 2 different) placed on top of the toy. The toy was a 10- x 6- x 4-in. opaque white cardboard box that appeared to play music when certain blocks were placed on top. In reality, the box contained a wireless doorbell that the experimenter activated by surreptitiously depressing a button.

In the *same* condition, the pairs that activated the toy consisted of two identical blocks, while in the *different* condition the pairs that activated the toy consisted of two blocks that differed in both shape and color. The experimenter started the training phase by introducing the toy to the child, saying, "This is my toy! Sometimes it plays music when I put blocks on top and other times it does not. Should we try some and see how it works?" The experimenter then took out two blocks, saying, "Let's try these ones!" and placed both blocks simultaneously on the toy, and the toy played music. The experimenter responded to the effect by saying, "Music! My toy played music!" The experimenter then placed the two blocks on the toy a second time and said, "Music! These ones made my toy play music!" Next, the experimenter took out a new pair of blocks in the opposite relation as the first pair. The experimenter placed these two blocks simultaneously on the machine, and the toy did not activate. In response, the experimenter said,

"No music! Do you hear anything? I don't hear anything." The experimenter placed this pair on the machine again and said, "No music. These ones did not make my toy play music." The experimenter then repeated this with two more pairs of blocks, one pair that activated the toy and one pair that did not.

The test phase began after all 4 pairs of blocks had been demonstrated on the toy. In both conditions, the child was given a choice between a novel same pair and a novel different pair to activate the toy herself. The pairs of blocks children observed on the machine and the pairs they were asked to choose between in the test phase were the same across conditions; the only difference between the two conditions was which relation activated the toy. The experimenter said, "Now that you've seen how my toy works, I need your help finding the things that will make it play music. I have two choices for you." The experimenter took out two trays, one supporting a novel same pair and one supporting a novel different pair, saying, "I have these," (holding up one tray) "and I have these" (holding up the other tray). Once the child looked at both trays, the experimenter continued, saying, "Only one of these trays has things that will make my toy play music. Can you point to the tray that has the things that will make it play?" The experimenter then placed both trays on opposite sides of the table just out of reach of the child, and prompted the child to point. The side of the correct pair was counterbalanced between children.

Children's first point or reach was recorded. Children received 1 point for selecting the pair of novel test blocks in the relation that matched their training (same or different) and 0 points for selecting the pair of test blocks in the opposite relation. A second researcher who was naïve to the purpose of the experiment recorded all responses. Inter-rater reliability was very high; the two coders agreed on 94% of the children's responses.

3.2.2 Results

Replicating the results reported by Walker and Gopnik (2014), 18-30-month-olds in Experiment 1a selected the test pair that was consistent with their training, in both *same* (78%), $p = .01$ (two-tailed binomial) and *different* (77%), $p = .02$ (two-tailed binomial) conditions (see Figure 2). By contrast, however, the older children (3-year-olds) failed to select the correct test pair in either *same* (46%), $p = .85$ or *different* (43%), $p = .57$ conditions (see Figure 3), with younger children outperforming older children in both cases (*same*: $\chi^2(1) = 5.37$, $p = .02$; *different*: $\chi^2(1) = 5.99$, $p = .02$). As predicted, the performance of 30-36-month-olds fell between these younger and older groups, selecting the correct test pair marginally above chance (70%) in the *same* condition, $p = .06$ (one-tailed binomial) and at chance (50%) in the *different* condition, $p = 1.0$.

These results demonstrate a surprising decline with age on the causal RMTS task. To provide additional support for this developmental trajectory, we combined children across age groups and conducted a logistic regression, treating age as a continuous factor and correct selection (collapsing across *same* and *different*) as the dependent variable. Results of the logistic regression show a significant decline between 18 and 48 months, $\chi^2(N = 141, df = 1) = 3.88$ (Wald), $p < .05$.

3.3 Experiment 1b

Experiment 1a suggests a surprising decline in older children's ability to learn the

abstract relations "same" and "different." However, one possible explanation for this finding may be that younger children rely upon a simpler strategy, rather than a genuine causal inference, that is later abandoned by older children. In Experiment 1b, we sought to assess 18-30-month-olds a second time, using a more stringent test of causal reasoning: In addition to replicating 18-30-month-olds' selections (by pointing), we also examined the outcome of their own interventions to produce the novel effect. This ability to intervene with the appropriate pair of objects and to act on a cause in order to produce its effect is one benchmark of causal understanding (Pearl, 2000; Woodward, 2003). Would the children who pointed to the correct pair of blocks also actively intervene to activate the toy with those blocks?

3.3.1 Method

3.3.1.1 Participants

Forty 18-30-month-olds (M = 23.6 months; range = 17.9 -31.0 months) were randomly assigned to one of two conditions: *same* (n = 20, M = 24.3 months, range = 17.9 – 30.0 months) and *different* (n = 20, M = 23.1 months, range = 17.9 – 31.0 months). An additional 8 participants were excluded for failing to complete the study. Recruitment methods and participant population was identical to Experiment 1a.

3.3.1.2 Materials and Procedure

The procedure for Experiment 1b was nearly identical to Experiment 1a (refer to Figure 1), except for the following critical change to the test trial. After the child pointed to the selected tray, the experimenter pushed both trays within reach and asked the child to intervene to make the toy play music. When necessary, children were encouraged to use the objects to activate the toy.

As in Experiment 1a, the experimenter recorded children's first point or reach. In addition, the experimenter coded the child's intervention. All children placed a block on the machine at least once. The experimenter coded whether the child initially placed two different blocks or two similar blocks on the machine, or whether they only placed one block on the machine.

3.3.2 Results

In Experiment 1b, 18-30-month-olds again pointed to the test pair that was consistent with their training, in both *same* (80%), p = .02 (two-tailed binomial) and *different* (75%), p = .04 (two-tailed binomial) conditions, replicating the results in Experiment 1a and Walker and Gopnik (2014).

Sixteen children in the *same* condition pointed to the correct tray during their initial selection. Eleven (69%) of these children intervened with a pair of "same" novel blocks (rather than intervening with either the "different" pair or a single block), while only 3 (19%) of the children in the *different* condition did so, with a significant difference between conditions, p = .01 (two-tailed Fisher exact test). Similarly, 15 children pointed to the correct tray in the *different* condition and 10 (67%) of those children intervened with a pair of "different" blocks (rather than intervening with either the "same" pair or a single block), while only 3 (19%) of children in the

same condition did so, with a significant difference between conditions, $p = .01$ (two-tailed Fishers exact test).

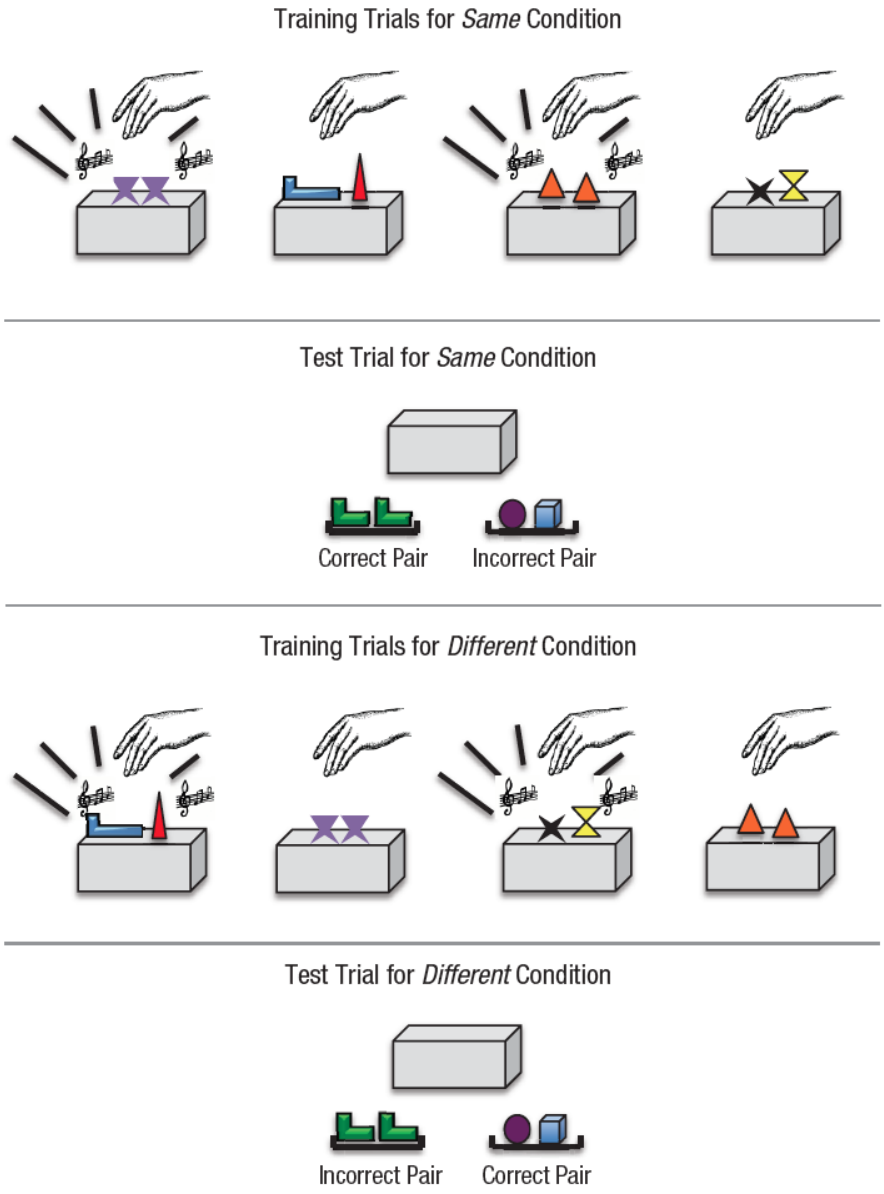

Figure 3.1. Schematic representation of training and test trials in the *same* and *different* conditions in Experiments 1a and 1b. Participants observed four training trials (two causal and two inert). On each test trial, a novel pair of "same" blocks and a novel pair of "different" blocks were presented. The child was asked to select the pair that would activate the toy.

These results demonstrate that children are indeed making a causal inference when selecting between the test pairs of blocks – they select the pair they believe will make the toy play music. Children's intervention behavior indicates that they have learned that the relations

between the blocks in our experiment and not the individual blocks themselves carry causal power. These data lend support to the idea that toddlers are using a conceptual strategy, rather than a simplified perceptual one, to solve the causal RMTS task. We discuss this further in the General Discussion.

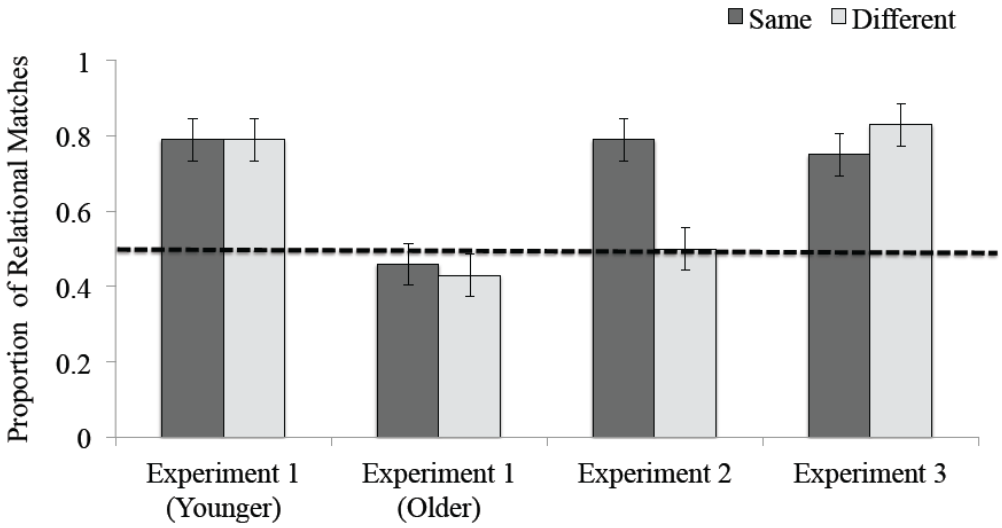

Figure 3.2. Proportion of correct relations selected following the manipulations in Experiments 1-3.

3.4 Experiment 2

Results of Experiments 1a and 1b replicate Walker and Gopnik's (2014) findings that young children are already equipped with the capacity to infer relational properties, though older children fail. We hypothesize that older children may be expressing a learned bias to attend to individual object properties and ignore abstract relations between them. In an effort to assess this claim directly in Experiment 2, we manipulated the data that children observe to provide evidence against the individual object kind hypothesis. In particular, Experiment 2 provided older children with explicit negative evidence that would lower the probability of an individual object kind hypothesis. To do so, 3-year-olds observed the same procedure described in Experiment 1a, with one important change: Before the experimenter placed the pairs of blocks on the toy simultaneously, she first placed each block on the toy one at a time, and children observed that the toy failed to activate (see Figure 3). By providing evidence *against* an individual object cause, these negative observations may prompt older children to override that hypothesis, even though it is more consistent with their prior knowledge, and instead consider the abstract relational principle that is more consistent with the evidence observed.

3.4.1 Method

3.4.1.1 Participants

A total of 56 3-year-olds (*M* = 41.9 months; range = 35.9 - 49.9 months) were randomly assigned to one of two conditions (*same*, *n* = 28, *M* = 41.7, range = 34.9 – 48.9 and *different, n* = 28, *M* = 42.2 months, range = 36.0 – 49.6 months). An additional 4 participants were excluded for failure to complete the study. Recruitment methods and participant population was identical to Experiment 1a and 1b.

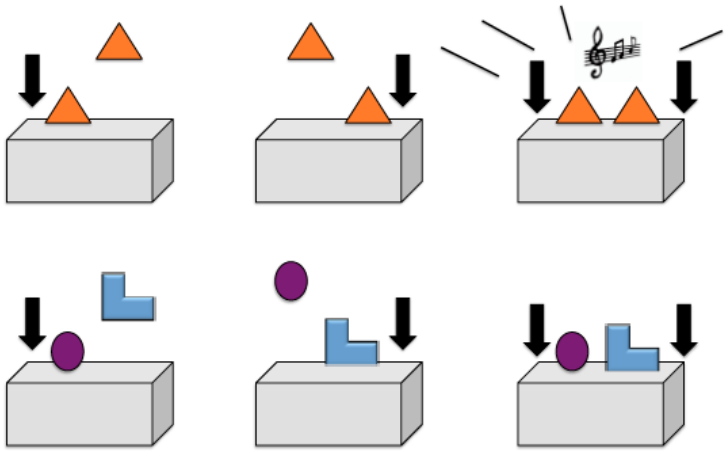

Figure 3.3. Schematic representation of two (of four) training trials in the *same* condition. The pattern of activation was reversed for the *different* condition. All test trials were identical to those used in Experiment 1a.

3.4.1.2 Materials and Procedure

The materials were identical to Experiment 1a and the procedure included the following critical changes. For each pair of blocks, the experimenter first placed each block on the machine *sequentially*, before placing them both on simultaneously (see Figure 3). Therefore, in addition to observing positive evidence that pairs of same or different blocks (depending upon the child's condition) activated the toy together, children also observed negative evidence for the causal efficacy of individual blocks (i.e., each block failed to activate the toy on its own). This training phase was immediately followed by a test phase, which was identical to the test phase in Experiment 1a. Inter-rater reliability was very high; the two coders agreed on 93% of children's responses to the test questions.

3.4.2 Results and Discussion

Results of Experiment 2 are consistent with the proposal that older children have developed a learned bias to attend to individual objects (see Figure 2). Once 3-year-olds were provided with negative evidence for the individual object kind hypothesis, they selected the correct relation significantly more often than chance (64%), *p* = .045 (exact binomial). However,

this overall effect was due to the improved performance of children in the *same* condition, in which 79% of children selected the correct pair, $p = .005$ (exact binomial). This performance was significantly better than children of the same age in the *same* condition in Experiment 1a, $\chi^2(1) = 6.17$, $p = .01$, and no different than the 18-30-month-olds (78%). Children in the *different* condition did not differ from chance performance (50%), $p = 1.0$ (exact binomial), leading to a significant difference between *same* and *different* conditions, $\chi^2(1) = 4.98$, $p = .03$.

How might we explain this emerging asymmetry between the "same" and "different" conditions in older children? It is possible that the data patterns observed in these two conditions interacted differently with the strength of children's beliefs in the "relational" vs. "individual" overhypothesis, leading to differences in how children's beliefs in these hypotheses were updated in light of the evidence. If older children 1) have developed the overhypothesis that individual kinds of objects are causal, 2) assume that the experimenter is randomly sampling blocks, and 3) assume that some fixed proportion of block types activate the toy, then the pattern of data that they observe in the "same" condition has a lower likelihood of occurring than the pattern of data in the "different" condition. Given assumptions 1-3, the probability that the toy will activate on any given trial should be higher when two different kinds of blocks are placed on the toy (i.e., when there are two potential activators), than when two of the same kind of block are placed on the toy (i.e., when there is only one potential activator). In other words, given that there is only one kind of block presented in each positive evidence training trial in the "same" condition, these data offer stronger counterevidence to the individual object kind overhypothesis than the pattern of data in the "different" condition.

However, this asymmetry might not be evident in children who think either that the "relational" or "individual" overhypothesis is much more likely than the alternative. According to Bayes rule, if the prior probability for one overhypothesis is well below the threshold for acceptance and the other is well above it, the difference in likelihoods might have little effect. In an intermediate case, however, where one overhypothesis is slightly more probable than the other, the difference in the likelihood of the two data patterns might lead to a difference in the posterior probabilities for these hypotheses (after observing the data pattern) and thus a difference in performance. In particular, the presentation of negative evidence for individual blocks in Experiment 2 would provide stronger support for the relational inference in the "same" condition than in the "different" condition.

Interestingly, the performance of the 30-36-month-olds in Study 1 also suggests the asymmetry between same and different, although (due to small sample sizes) the difference between the two conditions did not reach significance ($p = .16$). Future work will test this interpretation using a Bayesian model that formalizes the assumptions outlined above.

3.5 Experiment 3

In Experiment 3, we examined whether we could induce relational reasoning another way – not by manipulating the data that children observe, but by introducing a prompt to explain the evidence children observed during the training trials. Experiment 3 contrasted two conditions in which we asked 3- and 4-year-olds to either report *whether* the toy activated in each training trial or to explain *why* the toy did or did not activate in each case. We hypothesized that generating an explanation may motivate a different search procedure (e.g., Lombrozo, 2012; Walker et al., 2014; Walker et al., 2012, under review; Williams & Lombrozo, 2013), increasing the chance that children will accept the relational hypothesis.

3.5.1 Method

3.5.1.1 Participants

Forty-eight 3- and 4-year-olds (*M* = 45.1 months; range = 36.5 -58.9 months) were randomly assigned to one of two conditions (*explain:* n = 24, *M* = 45.9 months, range = 37.0 – 58.9 months; *report:* n = 24, *M* = 44.2 months, range = 37.2 – 58.5 months). Half of the children in each condition (12 per condition) observed evidence that was consistent with the *same* relation and the other half observed evidence that was consistent with the *different* relation. An additional 3 participants were excluded for failing to complete the study. Recruitment methods and participant population was identical to the previous experiments.

3.5.1.2 Materials and Procedure

The procedure for Experiment 3 was nearly identical to Experiment 1a (see Figure 1), except for the following changes. Children in the *explain* condition were prompted for an explanation after the second placement of each training pair on the toy, asking, "Why do you think these ones made/did not make my toy play music?" In the *report* condition, the experimenter asked, "What happened to my toy when I put these ones on it? Did it play music?" (prompting a yes/no response). As in previous work, reporting was selected as a control task because it shares several commonalities with explanation: it draws children's attention to the causal relationship, it requires them to verbalize in a social context, and it roughly matches children's time engaging with each outcome.

In addition to coding children's selections, all explanations were categorized into 3 mutually exclusive types: (1) object-focused (e.g., *"because it's red"*, *"because it has batteries"*), (2) relation-focused *("because they are the same," "because they are not the same"*), and (3) uninformative *("I don't know," "because it played music"*). Inter-rater reliability was again very high; the two coders agreed on 96% of children's responses to the test questions, and 89% of the explanation categories.

3.5.2 Results and Discussion

Three- and 4-year-olds who were prompted to explain during the training trials selected the correct relation significantly more often than chance (79%), $p = .007$ (exact binomial) (see Figure 2). Children in the *report* condition did not differ from chance (42%), $p = .54$, and there was a significant difference between *explain* and *report* conditions, $p = .017$. Unlike in Experiment 2, there was no significant overall difference between *same* (58%) and *different* (63%) relations, $p = .76$. There were also no differences found between *same* and *different* within each condition (*explain*: *same* = 75%, *different* = 83%; *report*: *same* = 42%, *different* = 42%). Comparing the overall pattern of responses of 3- and 4-year-olds who explained to the 18-30-month-olds in Experiment 1a, reveals no significant difference, $\chi^2(1) = 0.02$, $p = .88$, while 3- and 4-year-olds in the *report* condition performed significantly worse than the 18-30-month-olds, $\chi^2(1) = 9.0$, $p = .003$, and no differently from the 3-year-olds in Experiment 1a, $\chi^2(1) = 0.06$, $p = .81$, replicating the developmental pattern in Experiment 1a.

In order to analyze whether the content of children's explanations mattered for this pattern of responses, we classified the type of explanation (i.e., *object-focused*, *relation-focused*, *uninformative*) that each child produced most often, and analyzed their performance on the relational task. Children who provided *relation-focused* explanations as their modal response (N=6) – the most relevant explanation for the task – always selected the correct relational pair (100%). Children who provided *object-focused* explanations (N=9) were also highly likely to select the correct relational pair (89%). However, children who provided *uninformative* explanations or failed to provide an explanation at all (N=9) selected the fewest number of correct relational pairs (56%). The children who provided relevant relational or object-focused explanations were significantly more likely to choose the correct relational pair than children who provided no explanation or uninformative ones (p = .047, 2-tailed Fishers exact test). These data indicate that providing a meaningful explanation (regardless of its content) is sufficient to improve relational reasoning, but that simply being prompted for an explanation may not be.

3.6 General Discussion

Across four experiments, we assessed the influence of both the data that children observed (Experiments 1a, 1b, and 2), as well as their search procedure (Experiment 3) on their abstract reasoning. In Experiment 1a, we replicated Walker and Gopnik's (2014) finding that 18-30-month-olds are able to infer the abstract relations "same" and "different" from very few observations in a causal task. We also included an intervention prompt in Experiment 1b, in which 18-30-month-olds further demonstrated their causal understanding of the relational concept. We also contrasted toddlers' performance with a group of 30-36-month-olds and a group of 3-year-olds. As in previous work, older children failed to learn the relation. In fact, we found evidence for a linear decline in relational reasoning between 18 and 48 months of age.

The findings of Experiment 2 help to further explain this decline. They suggest that children may learn to privilege individual kinds of objects: When provided with evidence against this hypothesis, 3-year-olds were able to infer the relation in the *same* condition. Finally, in Experiment 3, we demonstrated that prompting children to explain during learning leads 3- and 4-year-olds to privilege the abstract relational hypothesis in both the *same* and *different* conditions. These results are consistent with previous work indicating that generating explanations prompts generalization and abstraction in causal reasoning (e.g., Legare & Lombrozo, 2014; Walker et al., 2014).

Discovering when and how children learn relational concepts is important for understanding the processes underlying early causal learning, but it is also important for understanding the development of relational reasoning, both in ontogeny and phylogeny. First, these results indicate that these abilities are in place surprisingly early – emerging spontaneously only a few months after the ability to learn specific causal properties. Although older children often fail to infer the relational hypothesis, this failure can be explained by appealing to the role of prior knowledge in constraining their judgments (see also Gopnik et al., in press).

The earlier literature on the development of relational reasoning invokes a "relational shift" from attending to individual, concrete object features to attending to more abstract, relations *between* objects. This literature attributes the shift to a number of factors, including an increase in relational knowledge (e.g., Gentner & Rattermann, 1991), exposure to relational language (e.g., Christie & Gentner, 2014), and various maturational variables (Halford, 1992; Richland, Morrison & Holyoak, 2006; Thibaut, French & Vezneva, 2010). Our results suggest

that the developmental trajectory of relational reasoning may be better characterized as a "u-shaped curve," in which early reasoning abilities are overshadowed by children's development of conflicting hypotheses (see e.g., Karmiloff-Smith & Inhelder, 1974-1975). In other words, the "relational shift" may not reflect an initial inability or difficulty to formulate or use relational concepts. Instead, it reflects a shift in the probabilities assigned to the individual object kind and relational hypotheses.

This novel proposal also provides an explanation for the well-documented influence of scaffolding on relational abilities. For example, previous research has demonstrated that the use of labels (Christie & Gentner, 2007; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005; Namy & Gentner, 2002; Ratterman & Gentner, 1998; Son, Doumas & Goldstone, 2010; see also Premack, 1983; Thompson & Oden, 2000; Thompson, Oden & Boysen, 1997 for similar findings in chimpanzees) and prompts to compare (e.g., Christie & Gentner, 2014; Gentner et al., 2011; Gick & Holyoak, 1983; Kotovsky & Gentner, 1996) support relational competence. Similarly, we demonstrate (in Experiments 2 and 3) that the individual object kind hypothesis may be overcome in both the *same* and *different* conditions with relatively minimal intervention.

Our findings, however, suggest a different interpretation of these results. In particular, Gentner (2010) has argued that symbolic language abilities support a process of structure mapping, in which close, object-based comparisons potentiate more distant, purely relational ones, through a process of progressive alignment. While we agree that relations are learned through experience, we propose that this learning need not proceed from local properties to more abstract ones. Explicit relational language, comparison, and explanations are not prerequisites for relational reasoning. Instead they serve to make the individual object kind overhypothesis less probable.

These findings are also relevant to the broader evolution of relational reasoning (Penn et al., 2008) and causal cognition in general. There is an ongoing debate in the comparative literature regarding whether differences in relational reasoning indicate a qualitative difference, or merely a quantitative gap between humans and their primate relatives (see Penn et al., 2008). The fact that very young human children already show the relational reasoning advantage, with no explicit prompting or cultural tutelage, may indicate that this is indeed a significant phylogenetic difference. Although it is possible that the younger children's success is due to the use of a perceptual heuristic, as has been suggested for nonhuman primates (e.g., Wasserman et al., 2001), several features of the study design weigh against this possibility: children saw pairs of objects (rather than multi-element displays), they observed only two positive and two negative trials, they never acted on an object, and their behavior was never reinforced. Indeed, no other species has come close to demonstrating the first-trial performance of these human children after so few observations (see Penn et al., 2008). In addition, the inclusion of an intervention task in Experiment 1b provides evidence for a genuine causal understanding of the abstract property. However, additional research is in progress to rule out the possibility that younger children are relying upon a different strategy during learning.

Finally, we propose that these results are consistent with other cases in which younger children are more flexible learners than older ones (Defeyter & German, 2003; Kuhl, 2004; Lucas et al., 2014; Seiver et al., 2013; Werker, et al., 2012). The very fact that children know less to begin with may, paradoxically, make them better (or at least more flexible) learners. In particular, as we acquire abstract knowledge about causal structure, this experience provides a set of inductive biases that are usually quite helpful, allowing the learner to draw quick and accurate conclusions when a new situation is consistent with their past experiences. However,

this experience can also be a double-edged sword – occasionally leading learners away from the correct hypothesis, particularly in cases in which the correct hypothesis is unusual or less consistent with previous observations.

In Bayesian terms, children's flexibility results from a "flatter" initial prior than older children and adults. This flexibility may also reflect different search procedures, as well as different kinds of prior knowledge. For example, in Experiment 3, shifting older children's search procedure by asking them to explain the data led to better performance. There may be a general shift from broader to narrower search procedures as children grow older, independent of their specific knowledge. Developmental differences in both prior knowledge and search procedures may help to explain why very young children are such extraordinarily powerful learners.

Chapter 4

Explaining prompts children to privilege inductively rich properties

4.1 Introduction

The challenge of causal reasoning is to discover the underlying structure of the world to facilitate prediction and action. This is non-trivial task. Despite the often strong correlation between what an object looks like and its causal properties (see Gelman & Medin, 1993), it is not uncommon to observe dissociations. In fact, perceptually similar objects can be endowed with very different causal properties: Poison hemlock may *look* identical to wild carrot, but it is certainly not good to eat. Learning how and when to override perceptual properties as a basis for judgment and action, and to instead favor inductively rich properties (such as causal affordances), is thus an important step in cognitive development.

We propose that the process of seeking, generating, and evaluating explanations plays an important role in encouraging children to recognize and privilege inductively-rich properties as a basis for reasoning, even when those properties are not perceptually salient. In particular, engaging in explanation could help children appreciate causal properties and subtle but reliable cues to causal structure, such as internal parts and category membership. For example, trying to explain why consuming hemlock generates one outcome (namely death) while consuming wild carrots generates another (perhaps pleasure) could help children appreciate that each plant has important internal properties, and that these internal properties are correlated with causal consequences they may wish to prevent (e.g., death) or to predict (e.g., pleasure).

In what follows, we first outline our proposal for the effects of explanation, motivating our hypothesis that explaining leads children to privilege inductively rich properties (i.e., those that facilitate a broad set of useful inferences). We then provide a brief review of prior research on children's inductive generalizations in tasks that require choosing between a salient perceptual property (e.g., an object's color and shape) and a causal property (e.g., activating a machine). This body of research helps lay out the methods and developmental changes that motivate the current experiments.

4.1.1 Explanation and Inference

Accounts of explanation from both philosophy and psychology suggest that explaining past and present observations can foster the acquisition of information that supports future

actions and predictions (e.g., Craik, 1943; Friedman, 1974; Gopnik, 2000; Heider, 1958; Kitcher, 1989; Lombrozo, 2012; Lombrozo & Carey, 2006; Walker, Williams, Lombrozo, & Gopnik, 2012; Walker, Lombrozo, Williams, & Gopnik, under review). These ideas about the *functions or consequences* of explanation are consistent with several accounts of the *form and content* of explanations. In particular, according to subsumption and unification theories, explanations appeal to regularities that subsume what's being explained under some kind of law (e.g., Hempel & Oppenheim, 1948) or explanatory pattern (e.g., Friedman, 1974; Kitcher, 1989). In so doing, they relate the particular fact or observation to a generalization that supports further inferences (Lombrozo, 2006, 2012; Wellman & Liu, 2007). For example, by explaining Socrates' death by appeal to the consumption of a poisonous chemical contained within hemlock (i.e., coniine), one implicitly invokes the generalization that the chemical can cause death in humans. This generalization in turn supports predictions about the consequences of future coniine consumption, provides guidance about how to avoid a particular kind of death (i.e., don't consume hemlock), and even supports counterfactuals about how things could have been otherwise (e.g., if Socrates hadn't consumed hemlock, or if he'd had an antidote to coniine, he would have lived to see another day).

If explanations typically subsume what is being explained under some generalization, then engaging in explanation could influence learning and inference by driving reasoners to form broad generalizations and to consult them as a basis for further reasoning (Lombrozo, 2012). Consistent with this idea, research with adults has shown that prompts to explain can promote the discovery and extension of broad patterns that govern membership in novel categories (e.g., Williams & Lombrozo, 2010, 2013; Williams, Lombrozo, & Rehder, 2013; see also Chi, DeLeeuw, Chiu, LaVancher, 1994). Recent developmental work likewise suggests that when prompted to explain, even young children are more likely to favor broad patterns (Walker et al., 2012, under review) and to develop abstract theories, such as a theory of mind (Amsterlaw & Wellman, 2006), that can accommodate otherwise-puzzling observations (e.g., a character looking for an object in the wrong location). For example, Walker et al. (2012; under review) found that when prompted to explain why particular types of objects activate a machine while others do not, preschool-aged children were more likely to rely on a feature that accounted for all observations (as opposed to a subset) in deciding which new objects were likely to activate the machine.

Many of the most far-reaching and useful generalizations are those that involve causal relationships, as they support interventions in addition to predictions. Generalizations relating hemlock and death (in the example with Socrates), or beliefs and behaviors (in theory of mind), are cases in point. Some accounts of explanation *require* that explanations be causal (e.g., Strevens, 2008; Woodward, 2005, 2011), but one need not subscribe to a strictly causal theory of explanation to accommodate the observation that explanation and causation are often closely linked: the view that explanations privilege broad and useful generalizations is enough to support the idea that causation will often (if not always) be central to explanations. In line with this idea, previous research with adults has demonstrated that explanations help guide causal inferences (Heit & Rubinstein, 1994; Rehder, 2006; Sloman, 1994). There is also indirect evidence that causation is central to children's explanations (e.g., Hickling & Wellman, 2001). For example, young children's explanations often posit unobserved causes (Buchanan & Sobel, 2011; Legare, 2012; Legare, Wellman, Gelman, 2010; Legare, Wellman, & Gelman, 2009), and Legare and Lombrozo (2014) found that children who explained learned a novel toy's causal (functional) mechanism (i.e., interlocking gears make a fan turn), but not other superficial properties (i.e., the

color of the gears), more readily than children who did not. In the experiments that follow, we focus on causality as a canonical, inductively-rich property that's likely to be privileged in explanation, and we investigate the prediction that prompting young children to explain will help them appreciate and use causal similarities as a basis for learning and inference.

4.1.2 Inductive generalization: a shift from perceptual to conceptual?

A large body of research has examined the role of obvious (perceptual) properties versus non-obvious (hidden or abstract) properties, such as causal affordances, in guiding children's inductive inferences (e.g., Gelman, 2003; Gelman & Markman, 1986, 1987; Gopnik & Sobel, 2000; Keil, 1989; Keil & Batterman, 1984; Nazzi & Gopnik, 2000; Newman, Herrmann, Wynn, & Keil, 2008). This research demonstrates that even young children are able to use both perceptual and non-perceptual properties in categorizing objects (e.g., Gelman & Markman, 1987; Gopnik & Sobel, 2000). Nonetheless, young children tend to spontaneously focus on highly salient surface features. Specifically, while older children and adults often group objects according to complex cues such as common internal properties, labels, and causal affordances, regardless of perceptual similarity (Carey, 1985; Keil, 1989; Medin, 1989; Rips, 1989), young children tend to group objects based on perceptual similarity, and only later shift to favoring other properties (e.g., Gelman & Davidson, 2013; Gentner, 2010; Keil & Batterman, 1984).

To illustrate, consider the findings from Nazzi and Gopnik (2000). In this study, children observed four objects placed on a toy, one at a time. Two of these objects were shown to be causally effective – they made the toy play music – and two were inert. One of the causal objects was then held up and labeled (e.g., "This is a Tib"), and children were asked to give the experimenter the other object with the same label (e.g., the other "Tib"). In conflict trials, the same perceptual properties appeared across causal and inert objects, and performance on such trials revealed a developmental shift: when generalizing the novel label, 3.5-year-olds relied on perceptual cues over causal cues, while 4.5-year-olds relied on causal cues over perceptual cues.

Between the ages of 3 and 5, children also shift how they generalize internal or hidden parts. For example, Sobel, Yoachim, Gopnik, Meltzoff, and Blumenthal (2007) used a procedure similar to that of Nazzi and Gopnik (2000) to demonstrate that older children (4-year-olds), but not younger children (3-year-olds), are more likely to infer that objects have shared internal parts when they share causal properties than when they share external appearance. These examples – and many others (e.g., see evidence from research on psychological essentialism: Keil, 1989; Gelman, 2003) – demonstrate that by 5 years of age, children begin to reliably favor inductively rich properties, such as common causal affordances, over perceptual similarity when generalizing from known to unknown cases.

There have been a variety of proposals for how best to characterize and explain this shift in children's inductive generalizations. For example, one possibility is that children first categorize objects by relying on perceptual or "characteristic" properties, and then shift to a different basis for categorizing objects, one based on more complex or "defining" properties (see Keil & Batterman, 1984). Another possibility is that the basic mechanism underlying children's judgments remains constant, but that the exercise of this mechanism results in different judgments as children gather new evidence. Specifically, properties are often encountered in correlated clusters, with perceptual information serving as a reliable indicator of other properties. As a result, perceptually-based judgments may be quite reasonable until sufficient evidence has been amassed to suggest an alternative (Gopnik & Nazzi, 2000; Gopnik & Sobel, 2000; Keil,

1989; Boyd, 1999; Sobel et al., 2007). From this perspective, even very young children may already be equipped with the conceptual resources to reason on the basis of non-perceptual properties, including causal affordances, even though performance on various tasks can change in the course of development. Consistent with this idea, Gopnik and Sobel (2000) found that when presented with conflicting cues, younger children produced a variety of memory errors that indicated an assumed correlation between different types of properties (i.e., perceptual and causal), even when no such correlation existed in the data. Even looking-time data from infants suggests that by 14- to 18-months, children differentially attend to various perceptual and non-perceptual properties in different tasks (Booth & Waxman, 2002; Mandler & McDonough, 1996; Newman et al., 2005).

In four experiments, we examine the possibility that by 3 years of age, children *already have* the conceptual resources to generalize on the basis of inductively rich properties, and that their failure to do so often results from a failure to access or apply what they know. (For related arguments in other tasks and domains, see, e.g., Hood, Cole-Davies & Dias, 2003; Kirkham, Cruess & Diamond, 2003; Munakata, 2001; Sobel & Kirkham, 2006; Walker & Gopnik, 2013; Zelazo, Frye, & Rapus, 1996.) We investigate whether the process of seeking or generating explanations facilitates access to and application of causal knowledge, supporting children's ability to reason on the basis of non-obvious but inductively rich causal properties as opposed to salient but superficial perceptual properties.

4.1.3 Overview of experiments

In the following experiments, we use a method similar to Nazzi and Gopnik (2000) and Sobel et al. (2007) to examine whether generating explanations makes children more likely to infer that an object's internal parts will be shared by other objects with common causal affordances as opposed to similar appearances (Experiments 1a and 1b), and more likely to believe that objects belong to the same category when they share common causal affordances as opposed to perceptual appearances (Experiment 2). In Experiment 3, we examine whether effects of explanation extend to lower-level cognitive processes, such as attention and memory, and whether they derive from a special relationship between explanation and inductively rich properties or from a global boost in performance. Together, these experiments provide insight into the role of explanation in causal inference in early childhood.

4.2 Experiment 1a

Experiment 1a examines whether explanation influences preschoolers' extension of a hidden, internal property to other objects that share either perceptual or causal properties. Children observed four sets of three objects that were individually placed on a toy that played music when "activated" (see Gopnik & Sobel, 2000). Each set contained three objects: one that activated the toy (*target object*), one that was perceptually identical to the *target object*, but failed to activate the toy (*perceptual match*), and one that was perceptually dissimilar to the *target object*, but successfully activated the toy (*causal match*). After each outcome was observed, children were asked to either explain (*explain* condition) or report (*control* condition) that outcome. Next, children received additional information about the target object: an internal part was revealed. Children were asked which one of the two other objects in the set (i.e., the *perceptual match* or *causal match*) shared the internal property with the *target object*. This

method pit highly salient perceptual similarity against shared causal properties; children could base their generalizations on either one, but not both.

Given the hypothesis that generating explanations encourages learners to favor broad generalizations, and thus to focus on inductively-rich properties such as causal affordances, we predicted that children who were asked to explain each outcome would be more likely than children in the control condition to select the *causal match* over the *perceptual match.*

4.2.1 Method

4.2.1.1 Participants

A total of 108 children were included in Experiment 1a, with 36 3-year-olds (M = 40.9 months; SD = 3.7, range: 35.8 – 47.7), 36 4-year-olds (M = 53.3 months; SD = 3.6, range: 48.5 – 59.8), and 36 5-year-olds (M = 64.4 months; SD = 3.0, range: 60.1 – 70.4). Eighteen children in each age group were randomly assigned to each of the two conditions (*explain* and *control*). There was no significant difference in age between the conditions, and there were approximately equal numbers of males and females assigned to each group. Five additional children were tested, but excluded due to failure to attend to the experimenter or complete the study. Children were recruited from urban preschools and museums, and a range of ethnicities resembling the diversity of the population was represented.

4.2.1.2 Materials

The toy was similar to the "blicket detectors" used in past research on causal reasoning (Sobel & Gopnik, 2000), and consisted of a 10" x 6" x 4" opaque cardboard box containing a wireless doorbell that was not visible to the participant. When an object "activated" the toy, the doorbell played a melody. The toy was in fact surrepticiously activated by a remote control.

Twelve wooden blocks of various shapes and colors were used (see Figure 1). A hole was drilled into the center of each block. Eight blocks contained a large red plastic map pin glued inside the hole; the remaining four blocks were empty. All of the holes were covered with a dowel cap, which covered the opening to conceal what was inside. Each of the four sets of blocks was composed of three individual blocks. Within each set, two blocks were identical in color and shape, and one of these (the *target object*) contained a map pin. The other block (the *perceptual match*) did not. The third block (the *causal match*) was perceptually dissimilar to the other two.

4.2.1.3 Procedure

Children participated in a brief warm-up game with the experimenter. Following this warm-up, the toy was placed on the table. The child was told, "This is my toy. Some things make my toy play music and some things do not make my toy play music." Then the first set of three blocks was brought out and placed in a row on the table. The order of presentation of the three blocks was randomized. One at a time, the experimenter placed a block on the toy. Two of the three blocks in each set (the *target object* and the *causal match*) caused the toy to activate and play music. The *perceptual match* did not. After children observed each outcome, they were asked for a verbal response. In the *explain* condition, children were asked to explain the

outcome: "Why did/didn't this block make my toy play music?" In the *control* condition, children were asked to report the outcome (with a yes/no response): "What happened to my toy when I put this block on it? Did it play music?" After all three responses had been recorded, the experimenter demonstrated each of the three blocks on the toy a second time to facilitate recall.

Next the experimenter pointed to the set of objects and said, "Look! They have little doors. Let's open one up." The experimenter selected the *target object* and removed the cap to reveal the red map pin that had been hidden inside. The experimenter said, "Look! It has a little red thing inside of it. Can you point to the other one that also has something inside?" Children were then encouraged to point to one of the two remaining objects (i.e., the *perceptual match* or the *causal match*) to indicate which contained the same inside part, and this selection was recorded. Children could either select the block that was perceptually identical to the target *or* the object that shared the causal property, but not both.

Following their selection, children were not provided with feedback, nor were they allowed to explore the blocks. Instead, all blocks were removed from view, and the next set was produced. This procedure was repeated for the three remaining sets. Each child participated in a total of four trials, including a total of four unique sets of objects.

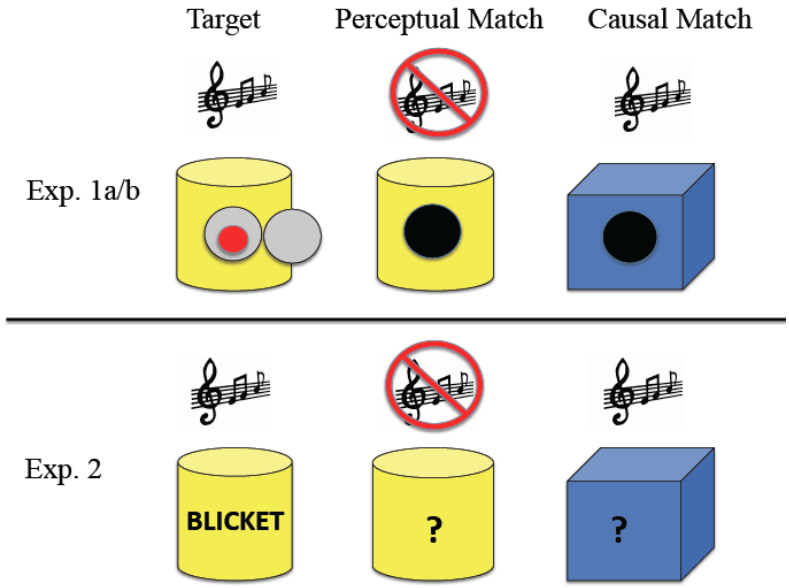

Figure 4.1 Sample set of objects used in Experiments 1a/1b (top) and Experiment 2 (bottom). Each of the objects in Experiment 1a/b included a door (indicated by a black circle), which covered the internal part contained inside. Each row corresponds to a single set of items. There were a total of four sets of stimuli.

4.2.1.4 Coding and Reliability

For each set of objects, children were given a score of "1" for selecting the *causal match* and a "0" for selecting the *perceptual match.* Each child could therefore receive between 0 and 4

points across the 4 trials. The explanations that children provided were also coded into five mutually-exclusive types: 1) appearance (e.g., *"It made the toy play music because it's purple," "...because it's round," "...because it looks like an apple"*), 2) internal parts (e.g., *"...because it has something inside of it," "...because it has a red thing in it," "...because it has batteries," "...because it has a motor"*), 3) kind (e.g., *"...because it's the right kind," "...because it's a music-maker," "...because it's musical"*), 4) other/non-informative (e.g., *"...because it's magic," "...because it wants/likes to," "...because it's special"*), and 5) no guess (e.g., *"I don't know"*). For the few participants who provided explanations that included both perceptual and internal properties, explanations were coded as appealing to internal properties. Because many of the children's explanations were quite minimal (only a couple of words in some cases), we did not examine the quality of children's responses beyond classifying them as belonging to particular explanation type.

Children's responses to the test questions were recorded by a second researcher during the testing session, and all sessions were video recorded for independent coding by a third researcher who was naïve to the the hypotheses of the experiment. Interrater reliability was very high; the two coders agreed on 99% of the children's responses to the test questions and on 91.8% of children's explanations. Disagreements were resolved by a third party.

4.2.2 Results and Discussion

Preliminary analyses revealed no trial-by-trial learning across the four sets of objects; children were no more likely to select the causal match on later trials than on earlier trials, Cochran's $Q(3) = 5.36$, $p = .148$. The data from the four trials were therefore combined to yield a single combined score that ranged from 0 to 4, and the data were analyzed with a 2 (condition) x 3 (age group) ANOVA (see Figure 2). The ANOVA revealed main effects of condition, $F(1, 102) = 50.70$, $p < .001$, and age, $F(2, 102) = 7.34$, $p < .01$, with no significant interaction. Overall, children who were asked to explain ($M = 2.98$, $SD = 1.23$) were more likely than children in the control condition ($M = 1.61$, $SD = 1.58$) to generalize the internal part of the *target object* to the *causal match* as opposed to the *perceptual match*. To better understand the main effect of age, we conducted pairwise comparisons between age groups, which revealed no difference in performance between 3- and 4-year-olds, $p = .86$, but that 3- and 4-year-olds each selected the *causal match* significantly less often than 5-year-olds, $p < .01$.

We also conducted one-sample *t*-tests comparing performance to chance to assess whether explaining prompted children to override a preference to generalize on the basis of perceptual similarity. The 3-year-olds and 4-year-olds in the control condition selected the *perceptual match* significantly more often than chance, $t(17) = -3.69$, $p < .01$, and $t(17) = -2.53$, $p < .05$, respectively, while those in the explain condition selected the *causal match* significantly more often than chance, $t(17) = 3.01$, $p < .01$, and $t(17) = 2.48$, $p < .05$, respectively. Five-year-olds in the control condition performed no differently from chance ($M = 2.61$, $SD = 1.72$), $t(17) = 1.51$, $p = .15$, while 5-year-olds in the explain condition selected the *causal match* significantly more often than expected by chance ($M = 3.39$, $SD = 1.29$), $t(17) = 4.57$, $p < .001$.

These data suggest that in the absence of an explanation prompt, children relied primarily on the target object's salient perceptual features to predict whether a novel object would share an internal property. However, when children of the same age were asked to generate an explanation, they instead privileged the target object's causal efficacy in making inferences about internal properties.

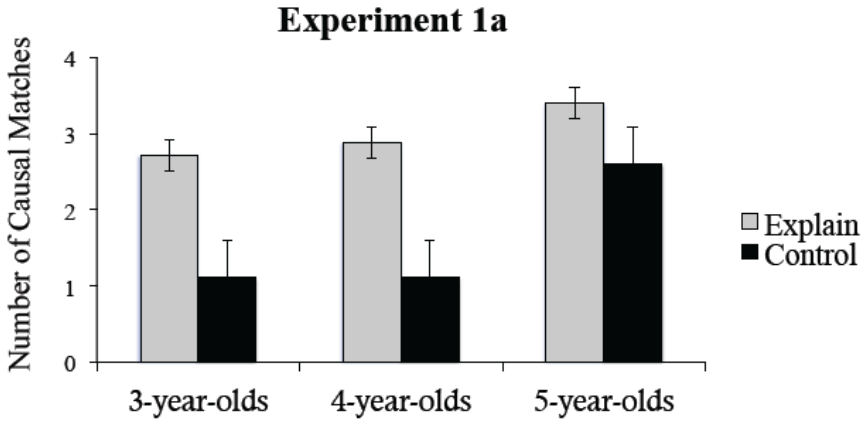

Figure 4.2 Average responses in *explain* and *control* conditions for Experiment 1a. Higher numbers indicate a larger number of trials (of 4) on which an internal part was generalized in line with a shared causal property over perceptual similarity. Error bars correspond to one SEM in each direction.

4.2.2.1 Content of Explanations

The frequencies with which children produced explanations of different types are reported in Table 1.

Baseline explanations for the first set of objects (before receiving any information about the internal properties) most often appealed to appearance (38%), with a minority (5%) appealing to internal properties. After observing the presence of the internal property for the first set of objects, explanations for the second set of objects appealed to appearance (33%) and internal properties (32%) equally often. By the final set, explanations most often appealed to internal parts (38%). An exact McNemar's test comparing the proportion of explanations that appealed to internal parts across the first and last trials revealed a significant difference, $p < .0001$.

Although we did not code the "quality" of children's explanations, we did examine the relationship between explanation type and performance. To do so, we identified the type of explanation that each child produced most often (i.e., the modal explanation for each child; see Table 2) and analyzed generalizations as a function of this designation. Overall, children who provided internal explanations as their modal response – arguably the most relevant explanation in this task – were significantly more likely to select causal matches than the aggregate of other children (80% versus 60%), $t(79) = 1.99$, $p = .05$. Despite the limitations associated with combining all other explanation types in a single group (which was necessary due to the small sample sizes), these results suggest that children who provided the most relevant explanation may have benefited most from the explanation prompt. We also found that 4- and 5-year-olds were each more likely to provide modal explanations that appealed to internal parts (20% and 19% of explanations, respectively) than 3-year-olds (6% of explanations), χ^2 (54, 1) = 5.25, $p < .05$ and χ^2 (54, 1) = 4.29, $p < .05$, respectively.

Table 4.1 *Frequency of Explanation Types for Each Set in Experiments 1a, 1b, and 2.*

	Set 1	Set 2	Set 3	Set 4	Total
Exp. 1a					
Appearance	61	53	38	41	193
Internal	8	51	55	61	175
Kind	8	4	10	11	33
Other	32	23	26	16	97
No Guess	53	31	33	33	150
Exp. 1b					
Appearance	17	14	-	-	31
Internal	2	17	-	-	18
Kind	3	4	-	-	7
Other	16	9	-	-	25
No Guess	16	10	-	-	26
Exp. 2					
Appearance	50	41	45	44	180
Internal	36	15	24	12	87
Kind	13	15	12	11	51
Label	0	9	14	17	40
Other	26	44	40	52	162
No Guess	38	39	28	23	128

We also found evidence that the prompt to explain impacted children's inferences even when the explanations that were generated did not appeal to internal properties. For example, the two children who provided no modal explanation (i.e., children who provided distinct explanation types for each set) and the two children who provided a modal explanation of "no guess" were (numerically) the most likely to select the causal match (88% each). In fact, each category of modal explanation, regardless of type (appearance: 53%, kind: 63%, other: 45%), was associated with a higher proportion of *causal matches* than that observed of children in the control condition (40%). Combining all of the children who provided modal explanations other than insides into a single group and comparing their responses to those of children in the control condition revealed a significant difference, $t(54) = -2.19$, $p = .03$. These data suggest that although children who provided the "correct" (internal) explanation were more likely to generalize according to causal as opposed to perceptual similarity, simply receiving an explanation prompt was enough to impact children's reasoning in this task.

In sum, our data support the proposal that prompts to explain increase children's reliance on inductively rich properties (as opposed to merely perceptual ones) as a basis for inference, and further suggest that effects of explanation are not restricted to children who happen upon the "correct" explanation for the task. There is an alternative explanation for our findings, however, that should be addressed. It is possible that explanation promoted greater projection to the *causal match* because the experimenter revealed the internal property immediately after children were prompted for an explanation, encouraging them to interpret the reveal as the experimenter's

means of providing an answer to the 'why" question the child had attempted to answer. Thus, the design of the task may have signaled to children that the internal part was the reason why the blocks made the toy play music (even if this information was not then reflected in all children's explicit explanations). In Experiment 1b we therefore investigate whether children generalized the internal property to the causal object because the timing and context of the explanation prompt supported a particular pragmatic inference, or because the process of explaining itself directed children to posit or privilege causality as a basis for generalization.

Table 4.2 *Proportion of Causal Matches in Experiments 1a and 2 as a Function of Child's Modal Explanation Type.*

Modal Explanation	Frequency	% Causal Matches
Exp. 1a		
Appearance	13	53%
Internal	24	80%
Kind	9	63%
Other	4	45%
No Guess	2	88%
No Mode	2	88%
Exp. 2		
Appearance	16	33%
Internal	6	33%
Kind	2	88%
Label	4	100%
Other	14	48%
No Guess	11	48%
No Mode	1	75%

Note: The number of children designated in each category is reported under "frequency."

4.3 Experiment 1b

The purpose of Experiment 1b was to rule out a pragmatic account of the findings from Experiment 1a. The procedure in Experiment 1b involved a critical modification from Experiment 1a: the addition of a second experimenter. Rather than having the same experimenter request explanations and reveal the internal properties of the objects, one researcher (R1) demonstrated the causal properties of the objects and provided the explanation prompt, and a second researcher (R2) (who had not observed the previous demonstration or explanation) revealed the internal part and solicited the generalization judgment.

If children in Experiment 1a who were prompted to explain preferentially generalized on the basis of causal properties because they took the researcher's revelation of the internal property as a potential answer to that researcher's why-questions, then changing researchers in this way should block the relevant pragmatic inference, and lead to performance comparable to

the control condition. In contrast, if something about the process of explaining prompts children to privelege causal similarity over perceptual appearance in our task, then this change in task pragmatics should not change children's generalization judgments.

Because we planned to compare children's performance in Experiment 1b to performance in Experiment 1a, and because we found no age differences between 3- and 4-year-old children, we only included one subgroup of children: 4-year-olds in the *explain* condition. By comparing the performance of this new group of children with that of 4-year-olds in the *explain* and *control* conditions from Experiment 1a, we can assess whether the results of Experiment 1a were plausibly an artifact of the pragmatics of the task.

4.3.1 Method

4.3.1.1 Participants

Eighteen 4-year-olds were included in Experiment 1b (M = 53.14 months; SD = 3.1, range: 48.8 – 59.4). All children were assigned to the *explain* condition. There was no significant difference in age between the 4-year-old children included in Experiments 1a and 1b, p = .84, and there were approximately equal numbers of males and females. Two additional children were tested, but excluded due to experimenter error. Recruitment procedures and demographics were identical to Experiment 1a.

4.3.1.2 Materials

Matericals were identical to those used in Experiment 1a.

4.3.1.3 Procedure

The procedure was similar to the one used in the *explain* condition in Experiment 1a, with two exceptions. First, one researcher (R1) provided explanation prompts, while a different researcher (R2) revealed the hidden properties and solicited the generalization judgments. Second, there were only two trials (rather than four) to avoid the concern that repeatedly switching experimenters could make the experimental situation too implausible or complex.

After children observed the first set of three objects placed on the toy and provided explanations for each one to R1, R2 entered the testing room. R2 said, "Hey, cool! Can I look at those?" R1 consented and walked away from the table. R2 examined the blocks on the table, saying, "Look! They have little doors. Let's open one up." R2 then selected the *target object* and removed the cap to reveal the red map pin that had been hidden inside, saying, "Look! It has a little red thing inside of it. Can you point to the other one that you think also has something inside?" As in Experiment 1a, children were encouraged to point to one of the two remaining objects (i.e., the *perceptual match* or the *causal match*) to indicate which contained the same inside part, and this selection was recorded. Following their selection, children were not provided with feedback, nor were they allowed to explore the blocks. Instead, R1 returned to the table, R2 departed from the testing room, and all blocks were removed from view. This two-experimenter procedure was repeated for one additional set of blocks.

4.3.1.4 Coding and Reliability

For each set of objects, children were given a score of "1" for selecting the *causal match* and a "0" for selecting the *perceptual match.* Each child could therefore receive between 0 and 2 points across the two trials. Explanation coding procedures were identical to Experiment 1a. Two coders agreed on all of the children's responses to the test questions and on 94.4% of children's explanations; disagreements were resolved by a third party.

4.3.2 Results and Discussion

As in Experiment 1a, children in Experiment 1b did not perform significantly differently across trials, Cochran's Q (1) = .143, p = .705. Data from both trials were therefore combined into a single score from 0 to 2, and the scores from this group were compared with the combined score from the first two trials of the 4-year-old participants in the *explain* and *control* conditions from Experiment 1a.

A univariate analysis of variance (ANOVA) with combined score as the dependent variable and condition (3: Exp. 1a *control*, Exp. 1a *explain*, Exp. 1b *explain*) as the independent variable revealed a main effect of condition, $F(2, 54) = 7.79$, $p = .001$. Children who were asked to explain in both Experiments 1a ($M = 1.3$, $SD = .77$) and 1b ($M = 1.56$, $SD = .62$) were each more likely than controls ($M = .61$, $SD = .85$) to generalize the internal part of the *target object* to the *causal match* as opposed to the *perceptual match*, $p < .01$ and $p < .001$, respectively. Pairwise comparisons revealed no difference in performance between 4-year-olds in the *explain* conditions of Experiments 1a and 1b, $p = .379$. We also conducted a one-sample t-test comparing children's performance to chance. Children in Experiment 1b selected the *causal match* significantly more often than chance, $t(17) = 3.83$, $p < .01$.

These data suggest that children in Experiment 1a were *not* simply interpreting the experimenter's revelation of the internal property as an answer to the "why?" question that the experimenter had previously posed. In Experiment 1b, the experimenter who provided the explanation prompt was different from the experimenter who revealed the hidden property, so the relevant pragmatic inference was disrupted. Instead, it appears that children in the *explain* condition privileged the target object's causal efficacy in making inferences about internal properties as a consequence of something about the very process of explaining.

4.3.2.1 Content of Explanations

Frequency data for each explanation type are reported in Table 1. Explanations were divided into the same five categories as in Experiment 1a. Baseline explanations for the first set of objects (before receiving any information about the internal properties) most often appealed to appearance (32%), with a minority (4%) appealing to internal properties. After observing the presence of the internal property, explanations for the second set of objects most often appealed to internal properties (32%), with explanations appealing to appearance dropping slightly (30%). An exact McNemar's test comparing the proportion of explanations that appealed to internal parts across the first and second (last) trials revealed a significant effect, $p < .0001$. Because there were only two trials, an analysis of children's modal explanation was not conducted.

4.4 Experiment 2

The purpose of Experiment 2 was two-fold. First, we were interested in whether the effect of explanation on children's inferences is restricted to generalizations concerning the relationship between causal properties and internal (or hidden) parts, or whether it extends to other properties as well. Second, we were specifically interested in whether explanation would affect how children extend novel labels. An effect of explanation on label extension would suggest that the process of explaining changes how children form categories, potentially shifting them from categories formed on the basis of perceptual properties to those tracking non-obvious, inductively rich causal properties. The ability to override perceptual similarity is an important hallmark of both scientific and everyday categories, as highly salient perceptual properties can be good predictors of category membership, but they can also be deceptive. For example, a dolphin may resemble a large fish, but dolphins are actually warm-blooded mammals. When such properties appear in conflict with one another, category membership is often based on non-obvious cues (e.g., internal biological properties) rather than surface appearance (e.g., having a tail).

Previous research demonstrates the importance of labels as indicators of category membership and guides to inference (e.g., Carey, 1985; Diesendruck, Markson, & Bloom, 2003; Gelman, 2003; Gelman & Markman, 1987; Gelman & Medin, 1993; Keil, 1989; Legare, Gelman, & Wellman, 2010; Medin, 1989; Nazzi & Gopnik, 2000; Rips, 1989). For example, Gelman and Coley (1990) found that in some cases, even 2-year-old children answered questions in line with category membership over appearances when labels were provided. But in the absence of labels, judgments are typically dominated by perceptual similarity. In fact, some have argued that children's categories are driven by low-level perceptual mechanisms that lead them to focus on object shape and other surface features (e.g., Landau, Smith, & Jones, 1988; Smith, 1999). However, other findings suggest that children extend labels differently depending on their intuitions about the kinds of object being classified, or on the nature of the task, and that classification is not always perceptually driven (e.g., Carey, 1985; Diesendruck, Markson, & Bloom, 2003; Keil, 1989). Finding that a prompt to explain leads children to extend labels on the basis of common causal properties would further suggest that even young children are able to form categories that disregard appearances, and that explaining helps them do so.

In sum, Experiment 2 used a method similar to Experiment 1a to examine whether the effects of explanation would extend to children's generalization of a novel *label* from a target object to an object that was either perceptually similar or causally similar. We predicted that explaining would make children more likely to attend to the causal powers of objects, which in turn would make it more likely for children to use causal properties as a basis for extending category labels to novel objects.

4.4.1 Method

4.4.1.1 Participants

A total of 108 children were included in Experiment 2, with 36 3-year-olds (M = 42.1 months; SD = 3.8, range: 35.9 – 48.0), 36 4-year-olds (M = 54.0 months; SD = 3.0, range: 48.4 – 59.9), and 36 5-year-olds (M = 65.0 months; SD = 3.8, range: 60.6 – 70.9). Eighteen children in each age group were randomly assigned to each of the two conditions (*explain* and *control*). There were no significant differences in age between the conditions, and there were

approximately equal numbers of males and females in each. Eight additional children were tested, but excluded due to failure to complete the study or failure to respond to the experimenter. Two more children were excluded due to experimenter error. Children were recruited from urban preschools and museums, and a range of ethnicities resembling the diversity of the population was represented.

4.4.1.2 Materials

The toy from Experiment 1 was again used in Experiment 2. Twelve wooden blocks of various shapes and colors were also used. There were a total of four sets of objects, each containing three blocks. As in Experiment 1, two of these blocks (the *target object* and the *perceptual match*) were perceptually identical (same color and shape) and one of these blocks (the *causal match*) was distinct (see Fig. 1).

4.4.1.3 Procedure

The procedure for Experiment 2 was identical to Experiment 1, with one exception: Instead of revealing a hidden internal property, the experimenter held up the *target object* and labeled it, saying, "See this one? This one is a blicket! Can you point to the other one that is also a blicket?"

4.4.1.4 Coding and Reliability

Coding for Study 2 was identical to Study 1, with children receiving a "0" for generalizations to the perceptual match and a "1" for generalizations to the causal match, resulting in a score of 0-4 points across the four sets. Interrater reliability was very high; the two coders agreed on > 99% of the children's responses to the test questions and 96.8% of children's explanations The few minor discrepancies were resolved by a third party.

4.4.2 Results and Discussion

Preliminary analysis revealed no significant differences across trials, Cochran's Q (3) = .60, p = .896. Data from all four trials were therefore combined into a single score from 0 to 4 and analyzed with a 2 (condition) x 3 (age group) ANOVA (see Figure 3). This analysis revealed a main effect of condition, $F(1, 102) = 13.51$, $p < .001$, and no additional significant effects. Overall, children who were asked to explain (M = 1.91, SD = 1.83) were more likely than children in the control condition (M = .72, SD = 1.47) to generalize the label to the *causal match* as opposed to the *perceptual match,* regardless of age.

We next considered the data against chance responding. One-sample t-tests revealed that 3-, 4-, and 5-year-olds in the control condition selected the *perceptual match* significantly more often than chance, $t(17) = -2.93$, $p < .01$, $t(17) = -3.69$, $p < .01$, and $t(17) = -3.10$, $p < .01$, respectively. In the explanation condition, the average of children's selections did not differ significantly from chance, $t(17) = .12$, $p = .90$, $t(17) = -1.26$, $p = .23$, and $t(17) = .375$, $p = .712$, respectively. However, examining the distribution of selections across the four trials revealed that approximately half of the children in the explanation condition selected the causal match on three or more trials (50% for 3-year-olds, 44% for 4-year-olds, and 56% for 5-year-olds). This

distribution differed significantly from that expected by chance in all age groups, $\chi^2(4) = 84.26$, $p < .001$, $\chi^2(4) = 66.49$, $p < .001$, and $\chi^2(4) = 83.97$, $p < .001$, respectively.

Because responses in Experiment 2 were not normally distributed, we conducted a non-parametric test comparing the performance of children across conditions. Children who selected the causal match on three or four trials were designated as "causal reasoners," and all others as "perceptual reasoners" (see Sobel et al., 2007). Results of a chi-square test replicate the findings reported in the parametric tests above, revealing a significant effect of condition, $\chi^2(1) = 8.28$, $p < .01$, with children in the explanation condition more likely to be designated "causal reasoners" (50%) than children in the control condition (19%).

In sum, like the younger children in Experiment 1a, children in the control condition in Experiment 2 relied primarily on a target object's salient perceptual features to predict whether a novel object would share a category label. This is particularly surprising given that the same label was provided across all four trials, during which the perceptual features of the target object varied from trial to trial. However, when children of all ages were asked to generate an explanation for the evidence that they observed, they considered the target object's causal efficacy significantly more often in making inferences about shared labels.

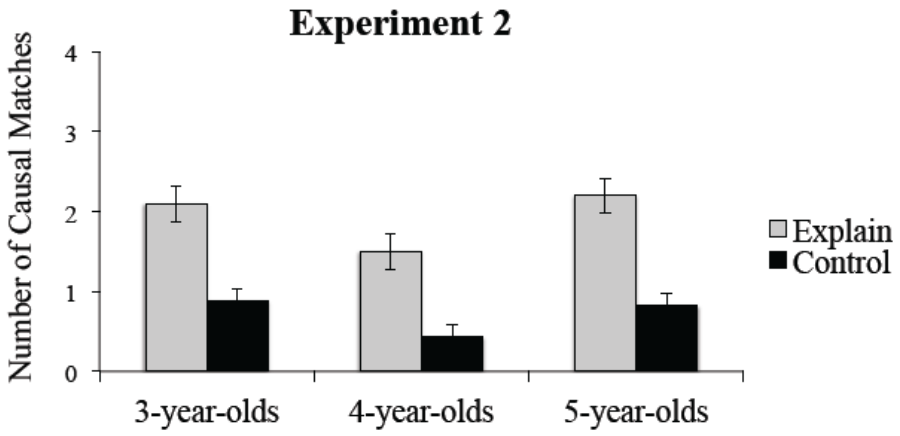

Figure 4.3 Average responses in *explain* and *control* conditions for Experiment 2. Higher numbers indicate a larger number of trials (of 4) on which a label was generalized in line with a shared causal property over perceptual similarity. Error bars correspond to one SEM in each direction.

4.4.2.1 Content of Explanations

Explanations were coded as in Experiments 1a and 1b, with one additional explanation type for those children who appealed to the label (e.g., *"It's a blicket"*) (see Table 1). Appearance explanations were most common overall (28% of all explanations); however, there was an increase in explanations that explicitly mentioned the label across trials, with 0% in the first set and 11% in the final set. An exact McNemar's test comparing the proportion of label explanations across the first and last sets revealed a significant difference, $p<.0001$.

To analyze the relationship between explanation type and performance in Experiment 2, we again calculated a modal explanation for each child, reflecting the most common explanation

type that the child provided (see Table 2). Children who most often provided an explanation that referred to the label also privileged causality in generalizing the label more often (100% versus 39%). However, so few children appealed to labels as their modal explanation (N=4) that there were no significant differences in performance as a function of modal explanation type.

Also as found in Experiment 1a, simply being prompted for an explanation was enough to affect children's inferences. Each modal explanation, regardless of type (appearance: 33%; internal: 33%; kind: 88%; other: 48%; no guess: 48%), was associated with a greater probability of selecting the *causal match* than in the control condition (18%). Combining all of the children who provided modal explanations other than labels into a single group and comparing their responses to those of children in the control condition revealed a significant difference, $t(90) = 2.39$, $p = .02$. As in Experiment 1a, these data suggest that although providing the most relevant explanation type (in this case, an appeal to the category label) leads to a special boost in performance, simply receiving an explanation prompt is enough to influence reasoning.

4.4.2.2 Comparing Experiments 1 and 2

To examine differences across our two experiments, we analyzed the data from Experiments 1a and 2 in an ANOVA with experiment as a between-subjects factor. This analysis revealed a significant difference in children's performance in Experiments 1a and 2, with a greater number of causal responses in Experiment 1 ($M = 2.3$; $SD = 1.6$) than Experiment 2 ($M = 1.31$; $SD = 1.8$), $F(1) = 22.41$, $p < .001$. There were also significant effects of age, $F(2) = 4.74$, $=p<.02$, and condition, $F(1) = 38.0$, $p<.001$, but no significant interactions. In other words, despite a greater baseline tendency to privilege perceptual features when reasoning about labels than about insides, the effect of explanation – increasing causal responding – did not differ across our two experiments, nor across age groups.

The observed difference in children's baseline responding across our two experiments is in line with previous research (Gopnik & Sobel, 2000; Sobel et al., 2007), which has found that children are more willing to privilege causality over appearances when extending internal parts than when extending labels. This pattern could also reflect a tension between more conceptual uses of labels, such as reference to essences or causes, and the more perceptually-based "shape-bias" found in noun labeling (e.g., Gelman & Markman, 1986; Gelman, 2003; Landau, Smith, & Jones, 1988; Jones & Smith, 1993; Imai, Gentner, & Uchida, 1994; Smith, Jones, & Landau, 1996). Nevertheless, explanation has a similar effect in promoting more conceptual (as oppose to perceptual) generalizations for both insides and labels. In effect, children in Experiment 2 were categorizing differently, depending on whether they explained or not. These results show that explanation guides children to attend to causal properties as an important but non-obvious basis for category membership.

4.5 Experiment 3

The findings from Experiments 1a, 1b, and 2 confirm our prediction that explanation encourages children to favor inductively rich properties (i.e., causality) as a basis for generalization. In Experiment 3 we hoped to bolster and further develop our interpretation of these novel findings by investigating three specific questions. First, the previous experiments demonstrate that explanation encourages children to privilege causal properties over perceptual properties when it comes to generalizing insides or labels. We propose that this is because the

process of generating explanations prompts learners to seek broad, generalizable patterns, and that this in turn should privilege properties that feature in such generalizations – by definition, those that are inductively rich. In Experiment 3, we investigate whether effects of explanation are restricted to inductive generalizations, or additionally manifest in lower-level processes that might be prerequisites to inductive inference, such as memory for object properties. In particular, might prompting children to explain make them more likely to attend to, and therefore effectively *remember,* an object's causal properties? And will benefits for memory be restricted to causal properties, which is directly related to what's being explained (i.e., an effect or its absence), or will they extend to other inductively rich properties that might figure in the explanations themselves, such as insides and category label?

Second, if we do find that explanation improves memory for properties such as insides and labels, it raises a question about the selectivity of explanation's effects (see also Legare & Lombrozo, 2014). In particular, the findings from the preceding experiments are consistent with the idea that prompts to explain result in an indiscriminate increase in children's overall attention or engagement, which could potentially account for more adult-like performance without needing to posit a special relationship between explanation and inductively rich properties. This account, like ours, would predict that children who are prompted to explain would have better memory for object insides and labels than those in a control condition, but would additionally predict that children who explain should have better memory for a property that is *not* inductively rich. In Experiment 3, we introduce such a property in the form of a sticker that is not correlated with any other object properties. Our hypothesis suggests that effects of explanation are selective – as opposed to indiscriminate – and predicts improved memory for object insides and labels (which are correlated with causal properties in both the task and in the world), but not for an uncorrelated perceptual property like the sticker.

A final question addressed by Experiment 3 is whether explanation-induced advantages for inductively rich properties come at the *expense* of memory for other kinds of properties. In particular, it could be that explainers simply fail to remember an uncorrelated sticker any better than controls, or that they actually show *impairment* in memory for this feature relative to controls. The latter possibility is consistent with previous research involving both children (e.g., Legare & Lombrozo, 2014) and adults (e.g., Hegarty, Mayer, & Monk, 1995; Needham & Begg, 1991) in which increased focus on an important abstract principle decreases memory for surface features.

To test these ideas, children in Experiment 3 were asked to explain or report causal outcomes after observing four unique objects, two of which activated the toy. After each object was placed on the toy, three properties were revealed: an internal part, a label, and a sticker (added to the object). The internal parts and the labels correlated with the toy's activation (i.e., all and only objects that activated the toy had a particular inside part and label) while the sticker did not. Children then completed a memory task in which they were asked to report the properties of each object. Because we did not observe age differences in the effects of explanation in Experiments 1-2, Experiment 3 was restricted to 4-year olds.

4.5.1 Method

4.5.1.1 Participants

A total of 36 4-year-olds were included in Study 3 (*M* = 53.8 months; *SD* = 3.7 months; range = 47.9 – 59.7). Eighteen children were randomly assigned to each of two conditions (*explain* and *control*). There were no significant differences in age between the conditions, and there were approximately equal numbers of males and females in each. Three additional children were tested but excluded due to experimenter error. Children were recruited from urban preschools and museums, and a range of ethnicities representative of the diversity of the population participated.

4.5.1.2 Materials

Experiment 3 used the same toy as in the previous experiments. A different set of test blocks was used, however, which consisted of 4 *unique* blocks – i.e., each block was distinct in color and in shape (see Table 4). As in Experiments 1a and 1b, all blocks had a hole drilled into the center. Two of the blocks had a red, round plastic map pin glued inside and two of the blocks had a white, square eraser glued inside the hole. Four stickers were used during the experiment (two heart stickers and two star stickers). Several small cards were constructed as memory aids for use during the test phase of the experiment. Half of the cards had an image of a black music note (placed in front of the objects that children believed activated the toy), and half of the cards had an image of a black music note crossed out with a red "X" (placed in front of the objects that children believed did not activate the toy). Four additional cards were constructed: one with a red circle, one with a white square, one with a heart sticker, and one with a star sticker. These cards were used to facilitate the forced-choice test.

4.5.1.3 Procedure

As in the previous experiments, the experimenter introduced the toy. The experimenter then produced a single block and placed it on the toy. The child observed as the block did or did not cause the toy to play music. As before, children in the *explain* condition were asked to explain the outcome for each of the blocks and children in the *control* condition were asked to report the outcome with a "yes/no" response. After the response was recorded, the experimenter repeated the demonstration a second time.

The experimenter then provided three additional pieces of information about the object: the type of internal part was revealed ("Look! It has a little door on it! Let's open it up. Look, there is a [red]/[white] thing inside."), a label was provided ("See this one? This one here? This one is a [Fep]/[Toma]!"), and a sticker was placed on the bottom ("Now I am going to put a sticker on it! I am going to put a [heart]/[star] sticker on the bottom, see?"). The experimenter repeated each property twice, and then the block was removed from view. The entire procedure was repeated for the three remaining blocks, one at a time. All children observed the causal property first. The order of the remaining three properties was counterbalanced.

Next, the experimenter placed all four objects on the table in front of the child in random order, and told the child that they would now play a "memory game." Children were asked a baseline causal memory question first, and then three additional property memory questions in randomized order. To assess baseline recall for the causal property of each object, the experimenter produced two cards – one with a music note, and one with a crossed out music note. The experimenter asked the child to point to the card that indicated whether the block did or did not play music. The child responded once for each of the four objects. Depending upon

the child's response, the experimenter would then place an additional card (with a music note or a crossed-out music note) in front of the object, which would remain throughout.

To assess recall for the internal part, the experimenter produced two cards – one with a red circle and one with a white square. The experimenter asked the child to point to the card that indicated the type of thing inside the block. The child responded once for each of the four objects. To assess recall for the label, the experiment said, "Some of these blocks were called 'Tomas' and some of these blocks were called 'Feps'. What was this one called, a 'Toma' or a 'Fep'?" The child responded once for each object. The order of presentation was counterbalanced across trials.

Finally, to assess recall for the type of sticker added to the block, the experimenter produced two cards – one with a heart sticker and one with a star sticker. The experimenter asked the child to point to the card that indicated the type of sticker added to the bottom of the block. The child responded once for each of the four objects.

Table 4.3 *List of properties for objects used in Experiment 3.*

	Object 1	**Object 2**	**Object 3**	**Object 4**
Causal	Yes	No	Yes	No
Internal	Red	White	Red	White
Label	"Toma"	"Fep"	"Toma"	"Fep"
Sticker	Heart	Heart	Star	Star

4.5.1.4 Coding and Reliability

Memory for internal parts, labels, and stickers was solicited in the same order as the corresponding properties were presented to that child in the demonstration phase of the experiment. For each property, children were given a score of "1" for accurate recall and a "0" for inaccurate recall. Because there were a total of four objects, each child could receive between 0 and 4 points for each property.

4.5.2 Results and Discussion

Because the causal property was always presented first during the observation phase, and always assessed first during the testing phase, memory for the objects' causal properties was analyzed separately with a one-way ANOVA. Results of this ANOVA revealed that while the majority of children in both conditions were able to recall the causal property of each object, children in the explain condition were significantly more accurate ($M = 3.93$, $SD = .24$) than controls ($M = 3.39$, $SD = .78$), $F(1, 34) = 8.42$, $p < .01$.

A repeated measures ANOVA with the other object properties (internal part, label, sticker) as the repeated measure and condition (explain, control) and order of presentation (label-sticker-insides, insides-label-sticker, sticker-insides-label) as the between subjects variables revealed a main effect of object property, $F(2, 60) = 7.05$, $p < .01$, as well as the predicted interaction between object property and condition, $F(2, 60) = 8.23$, $p<.002$ (see Figure 4). Children who were prompted to explain were significantly more accurate than controls in

reporting the labels, $F(1, 34) = 9.34$, $p<.01$, but *less* accurate than controls in recalling the sticker type, $F(1, 34) = 5.16$, $p<.05$. Although children who explained were numerically more accurate in recalling the internal part than controls ($M = 3.06$, $SD = 1.3$ and $M = 2.78$, $SD = 1.0$, respectively), this difference was not significant, $F(1) = .536$, $p = .47$.

These data address all three of the questions raised in Experiment 3. First, explanation does have an influence on memory for different object properties, and is not limited to inductive generalizations. Second, the findings challenge the idea that engaging in explanation simply improves overall engagement or attention in an indiscriminate manner. Instead, these data support the proposal that children who explain are more likely to *selectively* recall inductively rich, correlated cluster of properties (causality, internal part, label). Finally, we also found that children who were prompted to explain were significantly *less* likely to recall a superficial, perceptual property that did not correlate with other features, suggesting that benefits of explanation can come at a cost.

It is noteworthy that children who explained could not have been simply *ignoring* superficial perceptual features altogether, since the only way to track which properties corresponded to which objects in our task was to recall the unique color and shape of each of the blocks. Instead, explanation appears to impair memory for uncorrelated properties – those that are unrelated to other properties and therefore unlikely to support generalizations.

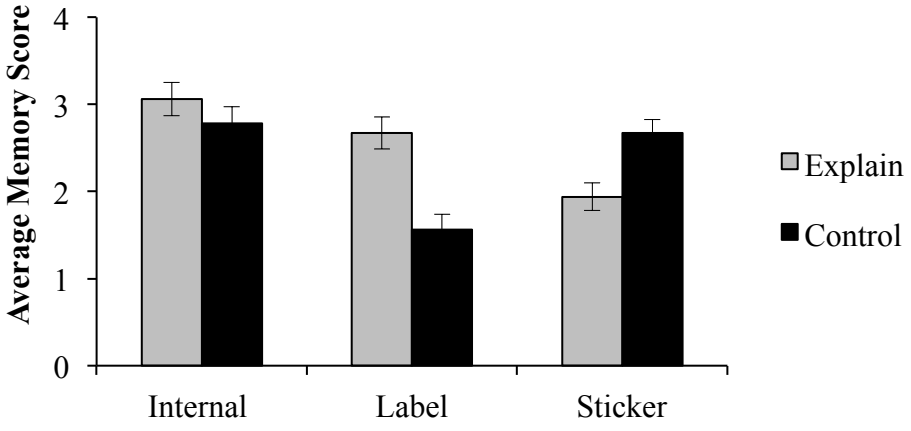

Figure 4.4 Average memory score (out of 4 trials) for each property assessed in Experiment 3. Error bars correspond to one SEM in each direction.

4.5.2.1 Content of Explanations.

Children's explanations were coded according to the categories generated in Experiment 2, with the addition of a new possible category: appeal to the sticker. Combining explanation data from all four objects (a total of 76 individual explanations), there were a total of 24 explanations that appealed to appearance, 27 explanations that appealed to the internal part, 4 explanations that appealed to the kind of object, and 2 explanations that appealed to the label. Notably, however, *none* of the children's explanations appealed to the presence of the sticker. This provides additional support for the claim that explanation selectively increased attention to

those properties that were inductively rich. An exact McNemar's test comparing the proportion of internal property explanations across the first (0%) and last sets (53%) revealed a marginally significant difference, p=.06 (one-tailed).

4.6 General Discussion

Our data demonstrate that prompting young children to explain makes them more likely to privilege inductively rich, non-obvious causal properties over salient surface similarity in making novel inferences. Children in the *control* conditions, who were not prompted to explain, instead based their judgments on perceptual similarity. These effects of explanation cannot be explained by the pragmatics of the task, as explanation produced the same effect when a two-experimenter design was employed (Experiment 1b). Moreover, these effects of explanation were not restricted to a particular kind of inference, as comparable effects were observed across two quite different judgments: the generalization of hidden, internal parts (Experiment 1a) and inferences about category membership (Experiment 2). Finally, these effects were not restricted to a particular age group: we found comparable effects of explanation across our 3-, 4-, and 5-year-old participants.

The results of Experiment 3 provide additional support for the idea that explanation privileges inductively rich properties, demonstrating improved memory for a correlated cluster of such properties (not just for causal affordances) in children prompted to explain. Importantly, Experiment 3 also provides evidence that effects of explanation are selective: Children who explained had *impaired* memory for an uncorrelated superficial property (the sticker). This challenges one possible alternative interpretation of the results: that explanation produces a general benefit for learning by globally and indiscriminately increasing engagement or motivation (see also, Legare & Lombrozo, 2014; Legare, 2012), and additionally suggests that the benefits of explanation are not without costs (see also Williams, Lombrozo, & Rehder, 2013).

The present findings suggest that children as young as 3 years of age have the conceptual resources to reason on the basis of non-obvious properties, such as causal affordances. These findings are therefore consistent with others suggesting children's early competence (e.g., Booth & Waxman, 2002; Gopnik & Sobel, 2000; Mandler & McDonough, 1996; Newman et al., 2005). Nonetheless, young children tend to privilege perceptual features over these less obvious alternatives under most conditions (e.g., Gelman, 2003; Keil, 1989; Wellman & Gelman, 1992; Sobel et al., 2007), and our findings go beyond prior work to identify a novel process that helps children overcome this tendency: namely engaging in explanation. In other words, engaging in explanation appears to facilitate children's access to (or ability to use) knowledge concerning the inductive relevance of causal properties.

Our findings have additional potential implications for our understanding of conceptual development. Experiments 1 and 2 deliberately spanned the age range (3 to 5 years) over which prior studies – which involved no explanation prompts – found developmental changes in children's tendency to generalize on the basis of perceptual versus causal properties (Nazzi & Gopnik, 2000; Sobel et al., 2007). While we did find age-related changes in children's baseline tendency to generalize one way or the other, the effects of explanation were uniform across ages. That is, we did not find interactions between the explanation manipulation and age group. One possibility is that the differences *within* age-groups observed across our experimental groups were driven by distinct mechanisms from those governing the changes observed *across* age-groups in our study and in others. For example, while the experimental effects were driven by

explanation, the developmental effects could have been driven by general improvements in executive function or inhibitory control, or by different intuitive theories at different points in development. Another possibility, however, is that older children were more likely than younger children to engage in explanation *spontaneously* (i.e., in the absence of a prompt), shifting performance towards causal inferences in the control condition, and to generate more effective explanations when prompted to explain, leading to a comparable shift in the explanation condition. Consistent with these ideas, Legare and Lombrozo (2014) found that older children were more likely than younger children to generate explanations in response to an ambiguous verbal prompt, suggesting that self-initiated explanation increases over this age range. And in Experiment 1, we found that older children were more likely than younger children to provide explanations that appealed to internal parts, suggesting an age-related boost in explanation quality. Age-related changes in explanation frequency and quality could be driving part of the developmental shift in children's baseline tendency to generalize according to perceptual versus non-perceptual properties. The current data cannot adjudicate between these possibilities, but do raise them as promising hypotheses for future research.

We have discussed effects of explanation in Experiments 1 and 2 as favoring causal similarity over perceptual similarity, however, it is worth returning to the ideas about explanation that motivated our initial predictions, as they suggest a more nuanced view. We propose that explanations tend to subsume what is being explained under a pattern or regularity, and that in so doing, the act of explaining could lead children to recognize or formulate broad generalizations that in turn support inference to new cases (Legare, 2014; Lombrozo, 2012; see also Walker et al., 2012, under review; Wellman & Liu, 2007; Williams & Lombrozo, 2010, 2013; Williams, Lombrozo, & Rehder, 2013). On this view, explanation drives learners towards broad generalizations, not towards causal properties (or away from perceptual properties), per se. However, children may already have formed higher-level generalizations (Dewar & Xu, 2010; Kemp, Perfors, & Tenenbaum, 2007) suggesting that certain types of properties, such as insides and category labels, are more likely to track common causal properties than superficial perceptual ones.

Consistent with this idea, some existing findings support the proposal that internal properties have a special status relative to a superficial perceptual property, such as a sticker, even when their correlational structure is matched within the context of a specific task. Beyond our own findings from Experiments 1a and 2, Sobel et al. (2007, Experiment 3) report an experiment in which the researcher presented a target object that produced an effect and revealed *two* properties of the object: an internal part and a sticker affixed to its back. Four-year-olds, but not 3-year-olds, inferred that another object with the same internal part was more likely to produce the effect than an object that shared the same sticker. In other words, older children spontaneously favored an internal property over a temporary perceptual one as a basis for generalizing a causal property, even in the absence of explicit evidence that the internal property was more likely to be correlated with causality in the context of the experimental task. This suggests that children form and apply higher-order generalizations about the *kinds* of properties that are likely to be inductively rich. In fact, recent computational formulations of the "theory theory" of cognitive development have proposed that learners represent generalizations at multiple levels of abstraction, creating "overhypotheses" (Goodman, 1983/1955) that enable learners to learn quickly and generalize effectively to novel cases. Building on these ideas, the act of explaining could encourage children not only to favor properties that support broad generalizations in a given task, but also the *kinds* of properties that are typically reliable guides

to particular inferences.

One open question – both in experimental and real-world contexts – relates to the role of pedagogical cues in fostering the benefits of explanation on inductive inference. Although the two-experimenter paradigm used in Experiment 1b ruled out certain pragmatic inferences that might have occurred as a direct result of the experimental procedure, children may have still interpreted the interactions pedagogically. Pedagogical learning does not necessarily require formal teaching, but rather a teacher's intent to communicate information to a learner in a context in which there exists some epistemic distance between those individuals (Shafto, Goodman, & Frank, 2012). Recent research suggests that children's interpretation of evidence may vary depending on whether learning occurs in pedagogical or non-pedagogical contexts (Bonawitz et al., 2011; Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Rhodes, Gelman, & Brickman, 2010; Shafto, et al., 2012). In particular, previous research has shown that, like explanation, pedagogical cues can promote attention to inductively rich features (Csibra & Gergeley, 2006; 2009). While the pedagogical cues in the current studies were well matched across *explain* and *control* conditions, it is certainly possible that *both* explanation *and* pedagogical cues may play a role in the effects reported here. The role of natural pedagogy in mediating or moderating effects of explanation on learning represents an important and novel avenue for future research.

4.6.1 Conclusions

Our data demonstrate that children as young as 3 years of age have the conceptual resources to reason on the basis of inductively rich properties, and that explanation facilitates their ability to avoid perceptually-bound judgments. In the current experiments, children had to decide whether to favor causal similarity or perceptual similarity in generalizing a hidden property or category membership from one object to another. Perceptual properties are often a reasonable basis for generalization, however, "insides" and category membership (labels) are more reliably associated with causal properties than with superficial, perceptual ones across many real-world cases. We propose that the process of explaining supports the construction and consultation of higher-order generalizations concerning such clusters of associated properties, in turn supporting inferences to new cases. By prompting children to favor inductively rich regularities, explanation encourages children to look beyond immediate observations to consider higher-order generalizations that support abstract knowledge.

Chapter 5

Conclusions

5.1 Conclusions and implications of the empirical work

Together, these empirical studies illustrate children's ability to build abstract representations that extend beyond their direct experience. By isolating the contributions of various learning mechanisms from the data that children observe, this work provides a novel perspective on the early development of causal reasoning and inference.

First, Chapter 2 demonstrates that children as young as 18 months are able to successfully learn the abstract relations "same" and "different" in a causal learning task from very few trials and without explicit instruction. This set of studies carries important implications for our understanding of the development of both causal and relational reasoning. First, the ability to learn abstract relations appears to be in place very early. Combined with recent computational work on the "blessing of abstraction" (Goodman, et al., 2011), this early competence may help to explain how children acquire complex causal representations that are early "intutive theories" (Gopnik & Wellman, 2012; Carey, 2010). These findings also contrast with previous research demonstrating the failure of non-human primates to solve these type of tasks, suggesting that relational reasoning may be a dimension along which humans differ from other primates (Penn, Holyoak & Povinelli, 2008).

In Chapter 3, I first replicated the success of toddlers in the relational reasoning task, and then contrasted this performance with an older group of preschool-aged children, who failed to infer the abstract relations they observed. The remaining experiments then assessed the influence of older children's observations, as well as their search procedure, on their abstract reasoning. When these older children were provided with evidence against the hypothesis that individual object properties are causal, they were able to infer the easier of the two relations: "same". In addition, simply providing older children with a prompt to explain their observations during training trials led them to override their bias to privilege object properties and to infer both of the abstract relations. Therefore, although older children often initially fail to infer relational hypotheses (in these experiments and others), this failure may be explained by appealing to the role of prior knowledge and learned overhypotheses in constraining their subsequent causal judgments (see also Gopnik et al., in press). This suggests that young children are equally able to reason about higher-order relations and object properties—they begin with a flat prior. However, as children learn more about the causal powers of objects, they form a general principle, or overhypothesis, such that hypotheses concerning objects take priority over those concerning relations. Later, over the course of development, this priority changes again, and children come to realize that relations also have predictive powers.

The experiments outlined in Chapters 2 and 3 carry important implications for our interpretation of the previous literature on relational reasoning. For instance, the vast majority of the earlier literature on the development of relational reasoning proposes that children shift from an initial focus on object properties to more abstract concepts, following processes of knowledge acquisition (e.g., Genter & Rattermann, 1991), language learning (e.g., Christie & Gentner, 2014), and other maturational advances (Halford, 1992; Richland, Morrison & Holyoak, 2006; Thibaut, French & Vezneva, 2010). Instead, the results outlined here suggest that the development of relational reasoning need not proceed in this bottom-up manner, from local properties to more abstract ones. The reason that relational language (Christie & Gentner, 2007, 2010) prompts to compare (Christie & Gentner, 2014), and other types of scaffolding have proven effective may be reinterpreted as a means for overriding a learned overhypothesis to attend to individual object properties. In fact, taken together, the results of Chapters 2 and 3 are consistent with research in other domains in which the relative lack of prior knowledge and flexibility sometimes results in infants and younger children being better learners than older children and adults (Defeyter & German, 2003; Kuhl, 2004; Lucas et al., 2014; Seiver et al., 2013; Werker, et al., 2012).

Finally, in Chapter 4, findings further demonstrate that simply prompting children to explain leads them to override attention to highly salient perceptual cues in favor of causal properties that are more inductively rich. In these studies, 3-year-olds who explained were more likely to infer that two objects with a common causal function, as opposed to common appearance, would share internal parts and category membership. These effects also extend to impact lower-level processes, influencing the object features that children attend to and subsequently recall. This suggests that 3-year-olds already have the conceptual resources to reason on the basis of non-obvious properties, and that explaining facilitates their access to the inductive relevance of those properties. These findings not only support the general proposal that prompts to explain can systematically change learning and inference, but also shed light on the underlying mechanisms by which explanation may produce these effects. In particular, these results suggest that explaining serves to direct learners to those hypotheses and aspects of their environment that support good or satisfying explanations. The act of explaining encourages children to not only favor those properties that support broad generalizations in a given task, but also the *kinds* of properties that tend to be reliable guides to particular inferences, and as a result, allows them to discover a novel inference.

In sum, Chapter 4 sheds light on the mechanisms by which explanation informs and constrains causal learning in early childhood. First, our findings help us to understand prior work demonstrating that generating explanations can influence belief revision. In addition, these data have important implications for our understanding of the nature of conceptual development during the preschool years. Like the data presented in Chapters 2 and 3, we demonstrate in Chapter 4 that even very young children already have the conceptual resources to reason on the basis of more inductively rich properties: they are not perceptually-bound. In this case, the cognitive process that is prompted by explaining supports the construction and consultation of higher-order generalizations, supporting inductive inference.

Across all chapters, the empirical findings demonstrate that applying a particular framework to the process of knowledge construction and causal inference results in changing the nature of the representation. Given this theoretical picture, knowledge is constructed due in large part to the constraints that are imposed on the learning problem, and therefore often carries unique implications for the learning outcome.

5.2 Remaining questions and future directions from work on analogical reasoning

Together, the results of Chapters 2 and 3 suggest a surprising decline in children's ability to learn the abstract relations "same" and "different" over the course of early development. These data provide evidence that this apparent decline is likely the result of a learned bias to attend to object properties in causal reasoning tasks. However, there are a variety of open questions that remain, prompting different avenues for future work.

For example, one possible alternative explanation for this decline may be that younger children (unlike older children) are relying upon a simpler or more implicit strategy, rather than forming a genuine causal inference or performing operations over abstract relational representations. Indeed, infant research using looking time methods (Dewar & Xu, 2010; Ferry, Hespos, & Gentner, 2015) provides some evidence for the existence of implicit mechanisms supporting the early development of relational reasoning. It is possible, therefore, that a perceptual strategy relying upon simple featural proxies for "same" and "different" that allows toddlers to succeed on this task is later abandoned.

While the data presented in the current dissertation are unable to completely rule out this alternative, several features of the study design serve to lower the possibility of the use of a perceptual strategy. For example, much of the previous research on the use of perceptual strategies has been conducted in non-human primates (Penn et al., 2008; Fagot, et al., 2001; Wasserman, et al., 2001), and relies upon the perceptual cue of entropy in the presented stimuli (i.e., "same" stimuli are lower in entropy than "different" stimuli). However, in the current studies, children observed pairs of objects across very few trials, rather than the type of large, multi-element arrays typically used in perceptual entropy research. In addition, children's pattern of intervention behavior in Chapter 3 indicates that they have learned that the relations between the blocks in our experiment and not the individual blocks themselves carry causal power. These data lend support to the idea that toddlers are likely using a conceptual strategy, rather than a simplified perceptual one, to solve the causal RMTS task. Nevertheless, it remains an intriguing possibility the toddler's success might be due to the use of a perceptual heuristic (e.g., Wasserman et al., 2001). In future research, it will be important to test this possibility directly. As a result, in ongoing work, I am currently exploring this alternative in a causal relational match-to-sample paradigm that controls for the amount of perceptual entropy that appears in the sets of stimuli that younger children observe.

In future work, it will also be important to explore the specific role played by the simplicity bias in children's ability to infer an abstract, relational hypothesis. The developmental decline that is described in Chapter 3 relies upon the idea that younger children lack a bias to privilege object properties, and are therefore left with a "flat prior." However, if this is the case, then toddlers should weigh both the relational and object hypotheses with equal probability. How might we explain younger children's tendency to privilege the relational hypothesis?

It is possible that an abstract principle of simplicity such as the "Bayesian Occam's razor" (Jefferys & Berger, 1992) might lead toddlers to initially prefer the more abstract hypothesis, since it proposes fewer causes to account for the data. Indeed, previous work demonstrates that young children (as well as adults [Pacer & Lombrozo, under review]) express this type of preference for fewer causes (Bonawitz & Lombrozo, 2012). I speculate in Chapter 3 that this initial bias is likely overridden in older children who have also learned the general principle that individual object kinds are more likely to be causal. This would lead them to

privilege individual properties over relational ones, in spite of simplicity considerations. However, future work should test this claim directly, and consider how the overhypothesis for simplicity may interact with the development of other early biases.

A third area of future research should further explore the role of language in the development of abstract relational reasoning – both the role of linguistic abilities in general, as well as relational language in particular. For example, some researchers have proposed that language is essential for the ability to engage in relational reasoning (e.g., Gentner, 2010). In line with this proposal, it has been demonstrated that language-trained chimpanzees that are taught to use the linguistic symbols for "same" and "different" are better able to succeed at relational match to sample tasks (Premack & Premack, 2003). Similarly, previous work with human children has shown that the use of labels scaffolds the ability to achieve relational insight (Christie & Gentner, 2007, 2010).

In ongoing work, I am currently addressing whether infants' developing linguistic representations are linked to the early acquisition of these relational concepts. This exploration is based upon a proposal regarding the close relationship between semantic and cognitive development: a variety of connections have been found between specific linguistic and conceptual achievements (e.g., Gopnik 1981, 1982, 1984; Gopnik & Meltzoff, 1987). For example, previous research suggests that children often use the word "more" to indicate similarity between objects and events (Gopnik, 1981). It is therefore possible that infant production of the word "more" may correlate with the appearance of early relational reasoning abilities. Preliminary work indicates the presence of a relationship between language production and relational reasoning in 14-15-month-olds (Walker, Hubacheck, & Gopnik, in prep). Future work will aim to strengthen the claim that abstract concepts may be reflected in, and linked to specific linguistic representations.

5.3 Remaining questions and future directions from work on explanation and learning

The work presented in the Chapters 3 and 4 provide evidence that the act of explaining prompts children to consider hypotheses at higher levels of abstraction. However, one interesting open question is whether this is the result of explanation influencing whether children behave in a manner that is more or less optimally Bayesian. For example, in related work, I have found that explanation leads children to consider hypotheses with broad scope, which results in (at least) two distinct effects on learning: explanation can make learners more sensitive to evidence, or more likely to rely on prior beliefs (Walker, et al., 2012, under review). Depending on which effect dominates, explanation can lead to either an increase (e.g., Brown & Kane, 1988; Rittle-Johnson, 2006; Siegler, 1995; Wellman, 2011) or a decrease (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Chi et al., 1994; Chinn and Brewer, 1998; Lombrozo, 2006) in belief revision, relative to children who don't explain.

Whether explaining indeed serves to increase fidelity to Bayesian conditionalization or not, it may nonetheless be possible to provide a formal account of explanation's effects in Bayesian terms. For example, as noted above, the process of explaining may recruit a set of evaluative criteria for what constitutes a *good* explanation (Lombrozo, 2012; Walker, et al., under review; Williams & Lombrozo, 2010). As a result, explaining could encourage learners to formulate and privilege hypotheses that exhibit certain features, or "explanatory virtues" (Lipton, 2001, 2004), that they may not have otherwise considered. These explanatory considerations

may then influence how Bayesian inference is approximated, even if it does not always lead to greater accuracy.

In particular, recent work has explored the idea that, at an algorithmic level, both children and adults approximate ideal Bayesian inference by using various "sampling" procedures. In these procedures, learners generate a few hypotheses to test at a time, adjusting the probabilities of those hypotheses as they acquire more data (Bonawitz et al 2014a, b). Explaining could potentially influence how this sampling process occurs, especially at the stage of hypothesis *generation*, which can lead to systematic effects on Bayesian inference (e.g., Bonawitz & Griffiths, 2010; Gopnik & Wellman 2012; Schulz, 2012; Ullman, Goodman & Tenenbaum, 2012). In future work, I hope to develop more formal approaches to the effects documented in our studies (see also, e.g., Pacer et al., 2013; Schupbach, 2011; Schupbach & Sprenger, 2011).

Relatedly, future research should explicitly consider how these findings relate to previous proposals regarding the role of explanation for learning. Much of the evidence for the benefits of explanation comes from research on the "self-explanation effect," the finding from educational psychology that prompting students to explain can improve learning (e.g., Fonseca & Chi, 2010) and foster transfer to novel problems (e.g., Nokes, Hausmann, VanLehn & Gershman, 2011; Renkl, Stark, Gruber & Mandl, 1998; Rittle-Johnson, 2006; Roy & Chi, 2005). Researchers have proposed a variety of plausible mechanisms that could underlie the effect. For example, Siegler (2002) suggests (among other things) that one consequence of explaining is a general increase in attention and engagement, and several researchers have suggested that explanations invoke prior beliefs (e.g., Ahn, Brewer & Mooney, 1992; Chi, 2000; Chi et al., 1989, 1994; Lombrozo, 2006; Williams & Lombrozo, 2013). Additional proposals include the ideas that explaining improves metacognitive monitoring, encourages learners to draw novel inferences, and helps learners form effective procedures (e.g., Chi, 2000; Chi et al., 1989, 1994; Fonseca & Chi, 2010; Johnson-Laird, Girotto & Legrenzi, 2004; Siegler, 2002).

The work that is outlined in the current dissertation builds upon these accounts by demonstrating how explaining can moderate type of hypotheses that are considered, changing the nature of children's representations. In particular, I propose that by highlighting generalizable patterns, explanation also serves to abstract away individual details. Explanation therefore appears to be a particularly valuable tool for guiding children away from appearances to consider properties with greater inductive potential. In ongoing and future work, I plan to extend this examination to other areas of learning. For example, although children's literature is often used as a pedagogical tool in early childhood, the ability to spontaneously extract underlying themes from narrative develops late. Research in education has proposed that children fail to represent the problem at the optimal level of representation – one that prioritizes abstract generalizations and understates surface features (van den Broek, et al., 2003). Given the current work demonstrating that explaining leads children to privilege hypotheses that highlight abstract structure, explanation may facilitate children's ability to extract the underlying moral of a story.

A related question relates to the role of pedagogical cues in fostering the benefits of explanation on inductive inference. Research suggests that children's interpretation of evidence may vary substantially depending on whether learning occurs in pedagogical or non-pedagogical contexts (Bonawitz et al., 2011; Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Rhodes, Gelman, & Brickman, 2010; Shafto, et al., 2012), and, like explanation, pedagogical cues can promote attention to inductively rich features (Csibra & Gergeley, 2006; 2009). Future work should therefore examine the role of pedagogical context in mediating and moderating the influence of explanation.

Finally, the interpretation outlined thus far has focused primarily on the impact of prompts to explain on the formation of children's causal inferences. However, if it is the case that children already have the conceptual resources necessary to reason on the basis of abstract and inductively rich properties, it is also worthwhile to consider the performance of children who were *not* prompted to explain. Williams and Lombrozo (2013) suggest that explanation can guide learners to effectively consult prior knowledge that would otherwise remain inert or under-utilized. These results of Chapters 3 and 4 can also be interpreted in line with Rozenblit and Keil's (2002) "illusion of explanatory depth," or the bias to overestimate one's own explanatory understanding of causal mechanisms (e.g., how a bicycle works), which has also been found in young children (Mills & Keil, 2004). If children erroneously believe that they already possess an adequate explanation, they may not feel it necessary to explain presented observations, and therefore fail to capitalize on the resources that are recruited by explaining.

Another possibility, of course, is that explaining may not always be beneficial. In fact, the work described in Chapter 4 has suggested that explanation can have associated costs: children prompted to explain why blocks activate a machine are less likely to remember superficial properties of each block than are those in a control condition. Related work suggests that, under some conditions, explaining may actually result in causal inferences that are less appropriate than those that children make in the absence of explanation (Walker et al., under review), and children prompted to explain how a gear toy works are less likely than controls to remember the colors of particular gears (Legare & Lombrozo, 2014). With adults, Williams, Lombrozo, and Rehder (2013) report cases in which prompting adults to explain can impair learning by leading to errors of overgeneralization. Future work should consider whether children are indeed selective when choosing to which phenomena to explain.

5.4 Concluding remarks

The results of the research presented above provide a more complete picture of how children learn and form abstract representations, with theoretical implications for developmental and cognitive sciences and practical implications for education and artificial intelligence. These findings may therefore be useful in informing educational practices and policies, which are increasingly moving towards a model of early childhood education that incorporates child-directed and inquiry-based learning.

While the current findings contribute to our understanding of the role of abstract reasoning and explanation for learning in particular, they also shed light on the nature of learning in general. When "learning by thinking," the learner gains new knowledge by engaging with information that they already have. This phenomenon therefore challenges a simple data-driven view of knowledge acquisition, one in which children's learning is simply a function of observations, exploration, and social information. Instead, the current findings provide evidence for a more complex picture of learning, one in which processes such as explaining to oneself – which does not involve new data or testimony from others – influence the way in which data and currently-held theories inform judgments. Understanding how these processes influence early learning therefore contributes to a more complete understanding of how knowledge is acquired and revised.

References

Ahn, W.K., Brewer, W.F., & Mooney, R.J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 391-412.

Baker, C., Saxe, R., & Tenenbaum, J.B. (2006). Bayesian models of human action understanding. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems,* 18, 99-106.

Bonawitz, E., Denison, S., Griffiths, T.L., & Gopnik, A. (2014a). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in Cognitive Sciences, 18,* 497-500.

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T.L. (2014b). Win-stay, lose-sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology, 74*, 35-65.

Bonawitz, E.B., & Griffiths, T. (2010). Deconfounding Hypothesis Generation and Evaluation in Bayesian Models. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2260-2265). Austin, TX: Cognitive Science Society.

Bonawitz, E.B. & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, 48(4), 1156-1164.

Bonawitz, E.B.*, Shafto, P.*, Gweon, H., Goodman, N.D., Spelke, E. & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction affects spontaneous exploration and discovery. *Cognition, 120*, 322-330.

Bonawitz, E.B., van Schijndel, T.J.P., Friel, D., Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*, 215-234.

Booth, A. E., & Waxman, S. R. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology, 38*, 948–957.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523.

Buchsbaum, D., Gopnik, A., Griffiths, T.L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, 120, 3, 331-340.

Buchsbaum, D., Bridgers, S., Weisberg, D.S., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367,* 2202-2212.

Bullock, M., Gelman, R. & Baillargeon, R. (1982). The development of causal reasoning. In W.J. Freidman (Ed.), *The developmental psychology of time* (pp. 209-254). New York: Academic Press.

Byrne, R.W. (1995). *The thinking ape: Evolutionary origins of intelligence.* New York: Oxford University Press.

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

Carey, S. (2010). *The origin of concepts*. USA: Oxford University Press.

Chater, N., Tenenbaum, J. & Yuille, A.L. (2006). Probabilistic models of cognition: Where next? In *Trends in Cognitive Neuroscience*, 10(7), 292-293.

Chater, N. & Manning, C.D. (2006). Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences,* 10, 287-291.

Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating

inferences and repairing mental models. In R. Glaser (Ed.), Advances in Instructional Psychology, Hillsdale, NJ: Lawrence Erlbaum Associates. 161-238.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145–18

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

Chinn, C.A. & Brewer, W.F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1): 1-49.

Christie, S. & Gentner, D. (2007). Relational similarity in identity relation: The role of language. In: Vosniadou, S. & Kayser, D. (Eds.). *Proceedings of the Second European Cognitive Science Conference.* London, UK: Taylor & Francis.

Christie, S. & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition & Development, 11*, 356-373.

Christie, S. & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, 38(2), 383-397.

Craik, K.J.W. (1943). *The nature of explanation.* Cambridge University Press (pp. viii-123).

Csibra, G., & Gergely G. (2006). Social learning and social cognition: The case for pedagogy. In: Munakata Y, Johnson M.H. (Eds.) *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI* (pp. 249-274). Oxford: Oxford University Press.

Csibra, G. & Gergeley, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153.

Defeyter, M.A. & German, T. (2003). Acquiring an understanding of design: Evidence from children's insight problem solving. *Cognition*, 89(2), 133-155.

Dewar, K.M. & Xu, F. (2010). Induction, overhypotheses, and the origins of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, 21(12), 1871-1877.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*(2): 164-168.

Downman, M. (2002). Modeling the acquisition of colour words. In: *Proceedings of the 15th Australian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence,* 259-271. Springer-Verlag.

Fagot, J., Wasserman, E.A., & Young, M.E. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *Journal of Experimental Psychology, 27*(4), 316-328.

Ferry, S. Hespos, S. & Gentner, D. (2015). Prelinguistic relational concepts: Investigating analogical processes in infants. Manuscript in press at *Child Development.*

Fonseca, B. & Chi, M.T.H. (2010). The self-explanation effect: A constructive learning activity. In Mayer, R. & Alexander, P. (Eds.), The Handbook of Research on Learning and Instruction. Routledge Press.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy, 71,* 5-19.

Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. *Cognition, 23,* 183-209.

Gelman, S. & Markman, E. (1987). Young children's induction from natural kinds: The role of

categories and appearances. *Child Development, 58*, 1532-1541.
Gelman, S.A. & Coley, J.D. (1990). The importance of knowing a Dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*(5), 796-804.
Gelman, S. & Wellman, H. (1991). Inside and essence: Early understandings of the non-obvious. *Cognition, 38*, 213-244.
Gelman, S. & Medin, D.M. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8*, 157-167.
Gelman, S.A., & Gottfried, G.M. (1996). Children's causal explanations of animate and inanimate motion. *Child Development, 67* (5), 1970–1987.
Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford Press.
Gelman, S. (2005). *The essential child: Origins of essentialism in everyday thought.* (Oxford Series in Cognitive Development). Oxford: Oxford University Press.
Gelman, S.A. & Davidson, N.S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology, 66,* 327-353.
Gentner, D. & Rattermann, M.J. (1991). Language and the career of similarity. In: Gelman S.A., Byrnes J.P. (Eds.) *Perspectives on thought and language: Interrelations in development*. London: Cambridge University Press.
Gentner, D. (1998). Analogy. In: Bechtel W, Graham G.A. (Eds.) *Companion to Cognitive Science* (pp. 107-113). Oxford: Blackwell.
Gentner, D. (2003). Why we're so smart. *Language in mind: Advances in the study of language and thought,* 195-235.
Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*, 752-775.
Gentner, D. Anggoro, F.K., Klibanoff, R.S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development, 82*, 1173-1188.
Gick, M.L. & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1-38.
Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Science, 7*, 43-48.
Goodman, N. D. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.
Goodman, N. (1955). *Fact, Fiction, and Forecast.* Indiana: Hackett Publishing Co.
Goodman, N.D., Baker, C.L., Bonawitz, E.B., Mansinghka, V.K., Gopnik, A., Wellman, H., Schulz, L., & Tenenbaum, J.B. (2006). Intuitive theories of mind: a rational approach to false belief. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* Mahway, NJ: Erlbaum.
Goodman, N., Ullman, T. & Tenenbaum, J.B. (2011). Learning a theory of causality. *Psychological Review*, *118*(1), 110-119.
Gopnik, A. (1981). The development of non-nominal expressions in 15-21 month old children. In P. Dale and D. Ingram (Eds.) *Child language: An international perspective.* Baltimore: University Park Press.
Gopnik, A. (1982). Words and plans: Early language the development of intelligent action. *Journal of Child Language, 9,* 617-733.
Gopnik, A. (1984). The acquisition of gone and the development of the object concept. *Journal of Child Language, 11,* 273-292.

Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind and Language* 3(Autumn), 197-216.
Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory-formation system. In: Keil F. & Wilson R.A. (Eds.), *Explanation and cognition* (pp. 299-324). Cambridge, MA: MIT Press.
Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*, 1623-1627.
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 3–32.
Gopnik, A., Griffiths, T.L. & Lucas, C.G. (in press). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*.
Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development, 58*(6), 1523-1531.
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
Gopnik, A. & Schulz, L. (2007). *Causal Learning: Psychology, Philosophy, and Computation.* Oxford: Oxford University Press.
Gopnik, A. & Sobel, D. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. Child Development, *71*(5), 1205-1222.
Gopnik, A., Sobel, D.M., Schulz, L.E. & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620-629.
Gopnik, A. & Wellman, H.M. (1992). Why the child's Theory of Mind really is a theory. *Mind and Language*, *7*(1-2), 145-171.
Gopnik, A., & Wellman, H. M. (1994). The theory theory. In: Gelman, S.A. & Hirschfeld, L.A. (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. 257-293). New York, NY, US: Cambridge University Press.
Gopnik, A. & Wellman, H. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*, 1085-1108.
Griffiths, T. & Tenenbaum, J. B. (2007). Two proposals for causal grammars. In: A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 323–345). New York, NY: Oxford University Press.
Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14(8), 357-364.
Gweon, H., Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332,* 1524.
Halford, G.S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, *35*(4), 193-217.
Harris, P.L., German, T. & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition* 61, (3) (Dec.): 233-59.
Hegarty, M., Mayer, R.E., and Monk, C.A. (1995). Comprehension of arithmetic word problems:

A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87* (1), 18-32.

Heider, F. (1958). *The psychology of interpersonal relations.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Heit, E. & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 411-422.

Hempel, C.G. & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135-175.

Heyes, C. & Frith, U. (Eds.) (2012). New thinking: The evolution of human cognition (Theme Issue). *Philosophical Transactions of the Royal Society, B,* 367.

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

Hochmann, J.R., Mody, S. & Carey, S. (under review). Infants' representations of same and different in match- and mismatch-to sample.

Hood, B., Cole-Davies, V. & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Dev Psychol*, 39(1), 61.

Hume, D. *An enquiry concerning human understanding*. Oxford world's classics. Oxford, England; New York: Oxford University Press.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45-75.

Inagaki, K., and Hatano, G. (2006). Young children's conception of the biological world. *Current Directions in Psychological Science* 15, (4) (Aug.): 177-81.

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

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

Johnson-Laird, P.N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3): 640-661.

Jones, S.S. & Smith, L.B. (1993). The place of perception in children's concepts. *Cognitive Development, 8*, 113-139.

Karmiloff-Smith, A. & Inhelder, B. (1974-1975). If you want to get ahead, get a theory. *Cognition*, 3(3), 195-212.

Keil, F.C. & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*, 221-236.

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

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

Kirkham, N.Z., Cruess, L. & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449-467.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In: P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410-505). Minneapolis: University of Minnesota Press.

Kotovsky, L. & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Dev*elopment, *67*(6), 2797-2822.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews*

Neuroscience, 5(11), 831-843
Landau, B., Smith, L.B., & Jones, S.S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321.
Legare, C.H., Wellman, H.M., & Gelman, S.A. (2009). Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology, 58*, 177-194.
Legare, C.H., Gelman, S.A., & Wellman, H.M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929-944.
Legare, C.H. (2012). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development, 83*, 173-185.
Legare, C.H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives, 8*, 101-106.
Legare, C.H. & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Experimental Child Psychology*, 126, 198-212.
Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: Children's conception of gears. *Journal of Research in Science Teaching*, *35* (1), 3–25.
Lipton, P. (2001). Is Explanation a Guide to Inference? In G. Hon and S. Rackover (Eds.), *Explanation: Theoretical Approaches* (pp. 93-120). Kluwer.
Lipton, P. (2004). *Inference to the best explanation*. Psychology Press.
Loewenstein, J. & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50(4), 315-353.
Lombrozo, T.L. (2006). The structure and function of explanations. *Trends in Cognitive Science*, 10: 464-470.
Lombrozo, T. (2012). Explanation and abductive inference. In: Holyoak K.J., Morrison R.G. (Eds.) *Oxford Handbook of Thinking and Reasoning* (pp. 260-276). Oxford: Oxford University Press.
Lombrozo, T. & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167-204.
Lucas, C.G., Bridgers, S., Griffiths, T.L. & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, 131(2), 284-299.
Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition, 59,* 307–335.
Medin, D.M. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.
Meltzoff, A., Waismeyer, A. & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-olds infants. *Developmental Psychology,* Online First Publication.
Mills C, & Keil FC. (2004). Knowing the limits of one's understanding: the development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology, 87,* 1-32.
Namy, L.L. & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Experimental Psychology General*, 131(1), 5-15.
Nazzi, T. & Gopnik, A. (2000). A shift in children's use of percpetual and causal cues to categorization. *Developmental Science, 3*(4), 389-396.
Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition,* 19, 543–557.

Newman, G.E., Herrmann, P., Wynn, K., & Keil, F.C. (2008). Biases towards internal features in infants' reasoning about objects. *Cognition, 107*, 420-432.

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In: *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (W. Gray & C. Schunn, Eds., 697-702). Mahwah, NJ: Erlbaum.

Nokes, T. J., Hausmann, R. G. M., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: The case of self-explanation prompts. *Instructional Science, 39*(5): 645-666.

Oden, D. L., Premack, D., & Thompson, R. K. (1988). Spontaneous Transfer of Matching by Infant Chimpanzees (Pan-Troglodytes)**.** *Journal of Experimental Psychology-Animal Behavior Processes, 14(2)*, 140-145

Oden, D.L., Thompson, R.K., Premack, D. (1990). Infant chimpanzees spontaneously perceive both concrete and abstract same/different relations. *Child Development, 61*(3), 621-31.

Pacer, M. & Lombrozo, T. (under review). Ockham's Razor cuts to the root: Simplicity in causal explanation.

Pacer, M., Williams, J., Xi, C., Lombrozo, T., & Griffiths, T. L. (2013). Evaluating computational models of explanation using human judgments. *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence.*

Pearl, J. (2000). *Causality.* London: Cambridge University Press.

Penn, D.C., Holyoak, K.J., & Povinelli, D.J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109-178.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Piaget, J. (1930). *The child's conception of physical causality*. London: Kegan Paul.

Premack, D. (1976). *Intelligence in ape and man***.** Hillsdale, N.J.: L. Erlbaum Associates

Premack, D. (1983). The codes of man and beasts. *Behavioral & Brain Sciences, 6,* 125-167.

Premack, D. & Premack, A.J. (1983). *The mind of an ape*. New York: W.W. Norton.

Premack, D. (1988). Minds without language. In: L. Weiskrantz, (Ed.), *Thought without language* (pp. 46-65). Oxford: Clarendon Press.

Premack, D. & Premack, A. J. (2003). *Original intelligence: unlocking the mystery of who we are***.** New York: McGraw-Hill.

Rattermann, M.J. & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In: Holyoak K, Gentner D, Kokinov B. (Eds.) *Advances in Analogy Research: Integration of Theory & Data from the Cognitive, Computational, and Neural Sciences*. Sophia: New Bulgarian University.

Regier, T. & Gahl, S. (2004). Learning the unlearnable: the role of missing evidence. *Cognition*, 93: 147-155.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition, 34*(1), 3-16.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*(1), 90-108.

Rhodes, M., Gelman, S.A., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science, 13*(3), 421-429.

Richland, L.E., Morrison, R.G. & Holyoak, K.J. (2006). Children's development of analogical

reasoning: Insights from scene analogy problems. *Experimental Child Psychology*, 94(3), 249-273.

Rips, L.J. (1989). Similarity, typicality, and categorization. In: S. Vosniadou & A. Orton (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rittle-Johnson, B. (2006) Promoting transfer: the effects of direct instruction and self-explanation. *Child Development, 77*, 1–15.

Roy, M. & Chi, M.T.H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), Cambridge Handbook of Multimedia Learning (pp. 271-286). Cambridge Press.

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

Schulz, L.E. (2012). Finding new facts; thinking new thoughts. In F. Xu & T. Kushnir (Eds.) *Rational Constructivism in Cognitive Development.* Advances in Child Development and Behavior, 43. Waltham, MA: Academic Press.

Schulz, L., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science (special section on Bayesian and Bayes-Net approaches to development)* 10(3), 322-332.

Schulz, L.E. , Goodman, N.D., Tenenbaum, J.B. & Jenkins, A.C. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, 109(2), 211-223.

Schupbach, J.N. (2011). Comparing probabilistic measures of explanatory power. *Philosophy of Science, 78*, 813-829.

Schupbach, J.N., & Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science, 78* (1), 105–27.

Seiver, E., Gopnik, A. & Goodman, N.D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development*, 84(2), 443-454.

Shafto, P., Goodman, N.D., & Frank, M.C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, 7, 341-351.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225-273.

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.

Sloman, S.A. (1994). When explanations compete: the role of explanatory coherence on judgments of likelihood. *Cognition*, 52, 1-21.

Smith, L.B. (1984). Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development, 55*(2), 363-380.

Smith, L.B., Jones, S.S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*, 143-171.

Sobel, D.M. & Gopnik, A. (2003). Causal prediction and counterfactual reasoning in young children: Separate or similar processes? Unpublished Manuscript.

Sobel, D. Kirkham, N.Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*, 1103-1115.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development, 8,* 159-182.

Son, J.Y., Doumas, L.A.A. & Goldstone, R.L. (2010). Relational words as handles: They bring along baggage. *Problem Solving*, 3(1), 52-92.

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

Stevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Tenenbaum, J.B. & Xu, F. (2000). Word learning as Bayesian inference. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*: 517-522.

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

Thibaut, J.P., French, R. & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Experimental Child Psychology*, 106(1), 1-19.

Thompson, R.K., Oden, D.L. & Boysen, S.T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. *Experimental Psychology of Animal Behavior Processes*, 23(1), 31-43.

Thompson, R.K.R. & Oden, D.L. (2000). Categorical perception and conceptual judgments by Nonhuman primates: The paleological monkey and the analogical ape. *Cogn Sci*, 24(3), 363–396.

Tyrrell, D.J., Stauffer, L.B., & Snowman, L.G (1991). Perception of abstract identity/difference relationships by infants. *Infant Behavior & Development, 14*, 125-129.

Ullman, T.D., Goodman, N.D., & Tenenbaum, J.B. (2012). Theory acquisition as stochastic search. *Cognitive Development.* http://dx.doi.org/10.1016/j.bbr.2011.

Van den Broek, P., Lynch, J.S., Naslund, J., Ievers-Landis, C.E., & Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology, 95*, 707-718.

Walker, C.M., Hubachek, S., & Gopnik, A. (in prep). Language acquisition and the onset of relational reasoning in infants.

Walker, C.M., Williams, J.J., Lombrozo, T. & Gopnik, A. (under review). The role of explanation in children's causal learning.

Walker, C.M., Williams, J.J., Lombrozo, T. & Gopnik, A. (2012). Explaining influences children's reliance on evidence and prior knowledge in causal induction. In: Miyake N, Peebles D, Cooper R.P. (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1114-1119). Austin, TX: Cognitive Science Society.

Walker, C.M. & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, 25(1), 161-169.

Walker, C.M., Lombrozo, T., Legare, C.H. & Gopnik, A. (2014). Explanation prompts children to privilege inductively rich properties. *Cognition*, 133(2), 343-357.

Wasserman, E.A., Fagot, F., & Young, M.E. (2001). Same-different conceptualizations by baboons (*Papio papio*): The role of entropy. *Journal of Comparative Psychology, 115*(1), 42-52.

Wellman, H.M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Wellman, H.M. (2011). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives*, 5(1): 33-38.

Wellman, H.M. & Gelman, S.A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337–375.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In: W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 523-573). New York: Wiley.

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

Werker, J.F., Yeung, H. H., & Yoshida, K. (2012). How do infants become experts at native speech perception? Current Directions in Psychological Science. 21(4), 221-226.

Williams, J.J. & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34*, 776-806.

Williams, J.J. & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, 66(1), 55-84.

Williams, J.J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General.* Advanced online publication doi:10.1037/a0030996

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

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2011). "Scientific Explanation", in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), URL = <http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/>.

Xu, F. (2007). Rational statistical inference and cognitive development. In: P. Carruthers, S. Laurence, and S. Stich (eds.), The innate mind: foundations and the future, Vol. 3 (pp.199-215). Oxford University Press.

Xu, F. & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114, 245-292.

Xu, F., Dewar, K. & Perfors, A. (2009). Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. In: B. M. Hood & L. Santos (Eds.), The origins of object knowledge (pp. 263-284). Oxford University Press.

Xu, F. & Griffiths, T.L. (2011). Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. Cognition. (Introduction to the special issue "Probabilistic models of cognitive development").