

**Young children's reasoning about their own and others' cognition**

by

Rachel W. Magid

B.A., Psychology and French  
Wellesley College (2012)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

~~April 2018~~ [June 2018]

© Massachusetts Institute of Technology 2018. All rights reserved.


**Signature redacted**

Signature of Author: .....

Department of Brain and Cognitive Sciences  
April 30, 2018

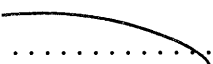
**Signature redacted**

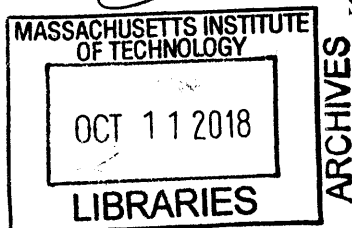
Certified by: .....

  
Laura E. Schulz  
Professor of Cognitive Science  
Thesis Supervisor

**Signature redacted**

Accepted by: .....

  
Matthew A. Wilson  
Sherman Fairchild Professor of Neuroscience and Picower Scholar,  
Director of Graduate Education for Brain and Cognitive Science





# **Young children's reasoning about their own and others' cognition**

by

Rachel W. Magid

Submitted to the Department of Brain and Cognitive Sciences  
on April 30, 2018, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## **Abstract**

This thesis aims to address a central question in cognitive science: how we reason about our own and others' cognition. Representing the self and others as distinct individuals is a fundamental epistemological feature of being human; the richness of these representations underlies our ability to tackle our own objectives and to understand the goals of others. Yet there is much debate about the metacognitive abilities of young children, in particular the extent to which children's estimations of their own and others' knowledge are accurate, whether children's beliefs about their own and others' cognition are influenced by the evidence they observe, and if these beliefs inform effective self-directed learning.

I investigate these questions, examining metacognition and its relationship to learning in 3- to 8-year-olds. Chapter 1 provides an overview of metacognition regarding the self and others. Chapter 2 considers whether young children expect others will learn rationally from evidence. We find that by age 4.5 years, children have a nuanced understanding of how evidence and prior beliefs interact to yield new knowledge. Chapter 3 investigates how children's exploration is influenced by representations of task difficulty, as indexed by the discriminability of alternative hypotheses. We show that there is a precise quantitative relationship between uncertainty and information seeking. Chapter 4 considers how preschoolers use social comparison information to calibrate their self-directed learning, demonstrating that when a task is within children's zone of proximal development, observing evidence that peers perform better increases one's own persistence. Chapter 5 asks how 3- to 5-year-olds integrate representations of their own and others' abilities when allocating roles across contexts. This work demonstrates that children consider who is best suited for a task based on relative ability. Across all four chapters, the results of these studies demonstrate that children have a sophisticated understanding of their own and others' knowledge and skills. In addition, children use information about others to effectively direct their own learning and problem solving. I end by arguing that young children have a theory of individuals' characteristics, of which reasoning about the self is a special case. Taken together, these studies illustrate the importance of considering how reasoning about the self and about others are integrated and are fundamental to our human intelligence.

Thesis Supervisor: Laura E. Schulz  
Title: Professor of Cognitive Science





## Acknowledgements

I owe a large debt of gratitude to the many people who have helped make this thesis possible.

First, I would like to thank my committee members. My advisor, Laura Schulz has profoundly shaped my personal and professional development. She is deeply committed to her students and her passion for our work is evident in our daily interactions— I have treasured my time working with her and will miss being in her lab. Josh Tenenbaum’s humanity and contagious enthusiasm never fails to make doing science fun. Every conversation with Josh improves my work. Rebecca Saxe was the very first faculty member I met in this department, and her curiosity and ability to foster meaningful conversation left quite an impression on me. I continue to admire her rigor and precision in both science and communication. Hyowon Gweon, who I first met when she was a post doc, has been a wonderful role model and source of scientific inspiration. I thank her for her continuing support and sage advice.

Next, I would like to thank my undergraduate mentors at Wellesley College, Tracy Gleason and Jennie Pyers. I am grateful for their support and wisdom, and for shaping me into the researcher I am today.

The studies in this thesis could not have been conducted without the help of our wonderful lab managers, Sammy Floyd and Kary Richardson, my fantastic student collaborators Mary DePascale and Phyllis Yan, and talented UROP research assistants, Andrea Garcia, Priyanka Gaur, Sophie Geoghan, Fatima Gunter-Rahmann, Eleanor Holton, Angela Kim, Rebecca Leibowitz, Sarah Raulston, and Julia Simon. I would like to especially thank the children and parents at the Boston Children’s Museum who generously participated in these studies. Without their interest and curiosity, this work would not be possible.

In graduate school, I have had the opportunity to study, work, and spend time with many fantastic colleagues and friends. I would like to especially thank Julia Leonard, Maddie (and Nellie) Pelz, Max Siegel, Kim Scott, Yang Wu, Pedro Tsividis, Melissa Kline, Julian Jara-Ettinger, Paul Muentener, Kelsey Allen, and Junyi Chu. In addition, I thank fellow cognitive scientists Sophie Bridgers and Jenny Lu, who, while far away, remain close to me.

Over the past two years I have had the opportunity and pleasure of working in the Office of Graduate Education, where I have found wonderful mentors and colleagues on the Diversity Initiatives team. I would particularly like to thank Dean Gloria Anglón, Noelle Wakefield, Sasha Susman, Renée Blackburn, Ashley Trebisacci, Min Hao Wong, Chaney Saephan, and Katherine Mizrahi. They have made being a part of the MIT community a richer and more rewarding experience.

Without my loving family, I would not be where I am today. I would like to thank my parents, Beth and Marty, for their constant support. They are incredible people and parents, who are dedicated to helping their children’s dreams come true. It has been a great blessing to work on my PhD within a stone’s throw of their Cambridge home. I would also like to thank my brother Josh and sister-in-law Jen for their support and cheer in celebrating all the best milestones. While I wish my grandmothers could see the influence they have had on their children and

grandchildren, I am fortunate to have the ongoing support of my grandfathers Bill and Manny. My in-laws, Ruth, Andy, and Liz have provided me with a second home. I am grateful for their love (and for introducing me to Boston Coolers). Finally, I want to thank Annemarie for being my partner in everything.

I would like to thank my sources of funding, the BCS Halis Fellowship, the NSF Graduate Research Fellowship, the Center for Brains, Minds, and Machines (CBMM), funded by NSF STC award CCF-1231216, and NSF Career Award (#0744213) awarded to Laura Schulz.

## Table of Contents

1. Cognition about cognition.....	1
1.1. Understanding of the self and others.....	1
1.2. Reasoning about agency and the mind.....	3
1.3. Reasoning about people’s traits and skills.....	9
1.4. Children’s metacognitive abilities and their relation to self-directed learning.....	14
1.5. Roadmap for the thesis.....	19
2. Changing minds: Children’s beliefs about third party belief revision.....	21
2.1. Introduction.....	22
2.2. Method.....	31
2.3. Results.....	35
2.4 Discussion.....	45
3. Children’s exploratory play quantitatively tracks the discriminability of alternative hypotheses .....	49
3.1. Introduction.....	50
3.2. Experiment 1.....	56
3.3. Experiment 2.....	61
3.4. Experiment 3.....	62
3.5. Experiment 4.....	64
3.6. Joint Analysis.....	65
3.7. General Discussion.....	66
4. When social comparison spurs motivation: Evidence from 3- to 5-year-olds’ persistence.....	68
4.1. Introduction.....	69
4.2. Experiment 1.....	75
4.3. Experiment 2.....	79
4.4. Experiment 3.....	81
4.5. General Discussion.....	85
5. Four and five-year-olds infer differences in relative ability and appropriately allocate roles to achieve cooperative, competitive and prosocial goals.....	89
5.1. Introduction.....	90
5.2. Experiment 1.....	95
5.3. Experiment 2.....	101
5.4. Experiment 3.....	103
5.5. General Discussion.....	105

6. Reasoning about the self and others.....	108
6.1. Summary of Chapters 2-5.....	108
6.2. Limitations and open questions.....	110
6.3. Relationship between reasoning about the self and others.....	112
6.4. Concluding remarks.....	118
References.....	120
Appendix I: Chapter 2 Supporting Information.....	152
Appendix II: Chapter 3 Supporting Information.....	163

# Chapter 1

## 1.1 Understanding of the self and others

Representing the self and others as distinct individuals is a fundamental epistemological feature of being human; the richness of these representations underlies our ability to tackle our own objectives and to understand those of others (Epstein, 1973; Tomasello, 1999). This thesis investigates the role of metacognition—cognition about cognition—in young children’s learning and problem solving. Perhaps the most common perspective on the study of metacognition relates to thinking about our own thought processes. However, metacognition can refer not only to considerations of one’s own cognition, but also our own characteristics as well as others’ thoughts, traits, and learning processes (Flavell, 1979; Wellman, 1985), as illustrated in the following: “You might believe that you (unlike your brother) should use Strategy A (rather than Strategy B) in Task X (as contrasted with Task Y)” (Flavell, 1979, p. 907).

In the study of metacognition, researchers often separately evaluate participants’ understanding of a construct, like uncertainty, in the context of the self and in the context of other people (e.g., Lyons & Ghetti, 2010; Sabbagh & Baldwin, 2001). This dichotomous approach could stem from one or a combination of the following reasons: 1) the methodologies used constrain assessment of reasoning to either the self or others, 2) the primary focus of any given research program is circumscribed (e.g., one researcher is interested in the inferences people make while learning from pedagogy whereas another is interested in self-directed learning) or 3) theoretical positioning regarding metacognitive mechanisms yields separate investigations (e.g., asserting that uncertainty or beliefs is processed differently when considered in relation to the self versus others). While there are a number of studies do that make direct

comparisons between reasoning about the self compared to others (e.g., Bélanger, Atance, Varghese, Nguyen, & Vendetti, 2014; Decety & Sommerville, 2003, Schneider, 1998), studies rarely ask how these two aspects of reasoning might be integrated. The research presented in this thesis explores this distinct, additional facet of metacognition: the relationship between reasoning about others and reasoning about the self. The empirical work described here provides evidence that young learners' metacognition includes the ability to make sophisticated and nuanced inferences about their own and others' cognition and goes on to demonstrate that children also integrate information about others into their understanding of the self to guide their own learning.

In Chapter 1, I review studies of metacognition, broadly defined, as they relate to people's understanding of their own and others' minds, abilities, traits, and learning with the aim of establishing themes and theories relevant to the empirical work reported in Chapters 2-5. In Chapter 6, I discuss the limitations of the work presented in this thesis, future directions, and the question of whether metacognition about the self and others is based on the same underlying system of reasoning. While we do not yet have a definitive answer to this question, the work presented provides some evidence that children engage in metacognitive reasoning about the self and others in tandem. Thus while these processes do not necessarily serve the same functions, there are common terms for thinking about our own as well as others' beliefs, effort, skills, and abilities starting in early childhood.

## **1.2 Reasoning about agency and the mind**

Reasoning about beliefs, desires, and knowledge is critical for social interactions, since understanding the unobservable attitudes of others allows us to interpret and anticipate what they will say and do. In this way, theory of mind and metacognition are highly interrelated. Indeed, some have argued that, “Metacognition consists of a large, multi-faceted theory of mind” (Wellman, 1985, p. 29). Since a substantial literature has been dedicated to the study of this topic (including over 10,000 dissertations), Chapter 1.2 focuses on the understanding of the self and others as agents, the relationship between theory of mind for the self and others, and whether children make nuanced inferences about knowledge in cases that require considering one’s interpretation of evidence.

### **1.2.1 Reasoning about individuals as agents**

From birth, we construe agents as distinct from non-agentive objects and can distinguish one individual from another based on differences in appearance. Studies using looking time measures show that infants expect that agents, but not non-agentive objects, can move of their own accord and interact contingently with their environments (Johnson, Booth, & O’Hearn, 2001; Kinzler & Spelke, 2007; Kulmeier, Bloom, & Wynn, 2004; Watson & Ramey, 1972). In addition, infants understand that certain psychological properties are unique to agents, such as the propensity to act in a goal-directed, efficient manner (e.g., Gergely & Csibra, 2003, Woodward, 1998). Turning to the ability to differentiate one human agent from another, studies have shown that within hours of birth, infants distinguish their mothers’ faces from those of female strangers (Bushnell, 2001; Bushnell, Sai, & Mullin, 1989; Walton, Bower, & Bower, 1992; Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995), and this ability to

differentiate likely broadens as children are exposed to more people. Because metacognitive reasoning is often person-specific, identifying people allows us to reason and make predictions about the actions of a particular individual. Thus, these early-emerging, likely innate, abilities lay the prerequisite groundwork for more complex metacognitive reasoning.

Several lines of research have investigated another core property of our understanding of agency: the recognition of the self. Infants as young as 3 months appreciate their own agency, preferring to interact with a toy that responds contingently to their own actions over one that they do not control (Watson & Ramey, 1972). Moreover, infants as young as 10 months make use of mirrors to adjust their behavior or locate objects near their bodies, suggesting that they represent their physical selves in a broader environmental context (Bertenthal & Fischer, 1978). By 18-20 months, most infants demonstrate explicit awareness of their physical appearance, recognizing themselves in a mirror as well as differentiating their own image from an image of another (see Butterworth, 1992 for review). Following these milestones, children begin to increasingly use self-referential pronouns such as “I” and “me” in their speech, and produce self-evaluative language such as “I can run fast” (Stipek, Gralinski, & Kopp, 1990). Given young children’s understanding of their physical selves and others as agents, we can ask how children reason about others’ minds in richer social contexts.

### **1.2.2 Reasoning about minds**

Theory of mind is called such because many scholars have made the argument that our ability to predict and explain beliefs, desires, and actions is organized as an intuitive theory, with similar structure, function, and dynamics as scientific theories (e.g., Carey, 1985; Gopnik & Wellman, 1992, Wellman & Gelman, 1992, c.f. Harris, 1994). One form of mental state



reasoning that seems to be in place early in life is the understanding that an agent's ability to perceive an event in the world will lead to the agent to have knowledge about that event. Toddlers demonstrate this understanding by the middle of their second year (O'Neill, 1996; Liszkowski, Carpenter, & Tomasello, 2008; Sodian, Thoermer, & Dietrich, 2006). Moreover, infants as young as 15 months recognize and expect that agents who do not have perceptual access to a hiding event will have a false belief about an object's location (Onishi & Baillargeon, 2005; Senju, Southgate, Snape, Leonard, & Csibra, 2011).

Despite this early precocity, there are robust developmental changes in theory of mind reasoning across childhood, as indexed with verbal theory of mind measures. One well-known and reliable change occurs between the ages of 3 years and 4 years, where children succeed on verbal theory of mind tasks around age 4 but not before (see Wellman, Cross, & Watson, 2001 for a review). This discrepancy between the results obtained in implicit, non-verbal tasks with infants and toddlers and explicit, verbal tasks with 3-year-olds have led some to suggest that young infants may have an altogether different theory of mind than older children and adults. Experience and/or maturation might drive developmental changes in the structure, function, and dynamics of children's theory of other people's minds that gradually allow them to explicitly express their expectations about others' desires, knowledge, and beliefs (e.g., Gweon & Saxe, 2013; Gopnik & Wellman, 1992; Richardson, Lisandrelli, Riobueno-Naylor & Saxe, 2018). It is also possible that advances in separable abilities, such as executive functioning, reveal and allow for the construction of a more sophisticated theory (e.g., Moses, Carlson, & Sabbagh, 2005).

The question of whether theory of mind reasoning differs when reasoning about the self and others has been a subject of much interest in developmental psychology and cognitive neuroscience. One debate among scholars concerns the mechanism we use to reason about

mental states— whether we simulate other people’s minds using a model of our own mind, or if we instead use an abstract representation of the human mind, regardless of whether we are reasoning about ourselves or other people (Gopnik, 1993; Perner, 1993; Saxe, 2009). While the mechanism by which we reason about mental states is not the primary focus of the work in this thesis, the results of the empirical studies presented here speak to how theory of mind, which encompasses many dimensions of metacognition, might operate over both the self and others, albeit with different functional applications. I return to this issue in greater detail in Chapter 6.3.

### **1.2.2 Reasoning about how minds interpret ambiguity**

Of course, our reasoning about mental states goes beyond having expectations about whether someone should be knowledgeable or ignorant. When do children develop an understanding of what others can infer based on varying types and amounts of information, rather than only its presence or absence? This question concerns the development of *interpretative theory of mind*: how people think they and others interpret ambiguous or incomplete information (e.g., Astington Pelletier, & Homer, 2002; Carey & Smith, 1993; LaLonde & Chandler, 2002). One of the reasons this aspect of theory of mind has been of great interest is that while children successfully reason about the presence or absence of epistemic access by the second year of life and succeed on tasks probing an understanding of false belief around age 4, they do not pass most interpretative theory of mind tasks until much later. For instance, 8-year-olds, but not 5-year-olds, will go beyond a simple statement like “John’s leg is in plaster because it is broken,” instead producing markers that imply an understanding of *how* they know something to be the case, such as the deductive, “We can tell that John has a broken leg because it is in plaster” (Donaldson, 1986).

Interestingly, on tasks involving deductive inferences, there appears to be some discrepancy between what children report knowing *themselves* and what children think others should know (in contrast to false belief reasoning, where children appear to perform similarly on tasks that probe one's own and others' beliefs; e.g., Gopnik & Astington, 1988). For example, in a task by Sodian and Wimmer (1987), children were asked whether they themselves, as well as another person, would make a deductive inference about the identity of a marble ("Do you know what color marble was in the bag?") drawn from one of three containers, one with all black marble, one with all blue marbles, or one with a mixture of black and blue marbles, when 1) the other person could see the marble that was drawn before it was placed in the bag, and 2) the other person could see which bag the marble came from, but not the marble itself. While both 4- and 6-year-olds tested made the correct inference when asked about their own knowledge, only 6-year-olds correctly said that the other person would have knowledge from perception and deductive logical inference, i.e., that the person would know the color of the marble after seeing the marble itself *and* seeing the bag from which it was drawn (Sodian & Wimmer, 1987).

This pattern of interpretive theory of mind performance has been observed across a number of other studies, including the well-known finding that children understand that ambiguous drawings of animals, such as the famous duck-rabbit picture, can be open to different interpretations only at age 6 or 7 (Carpendale & Chandler, 1996; Pillow & Marsh, 1999; Pillow, 1999; 2002). When asked explicitly, young children tend to categorize others' knowledge dichotomously, saying that what can be observed will lead others to have knowledge, and what has to be inferred, thought, or guessed is categorized as a lack of knowledge (Schwanenflugel, Fabricius, & Noyes, 1996). This finding does not necessarily imply children younger than 6 can

never distinguish between inference based on information and random guessing, but rather that they do not show an understanding of this in their verbal explanations.

Indeed, some 4- and 5-year-olds' difficulty with distinguishing inference from guessing in these tasks could stem from problems *articulating* what others know, which may require producing relatively complex linguistic constructions, such as complement clauses (e.g., de Villiers & Pyers, 2002). When asked to provide the relevant response via pointing rather than a verbal explanation as in Sodian and Wimmer (1987), children as young as 5 show improved performance on an inference task (Rai & Mitchell, 2006). Moreover, children are more likely to report that another individual has made an inference when asked what the other person would *think* rather than what they would *know*, implying that, despite earlier failures, children may indeed be sensitive to the degree of surprise and expectation with regard to others' knowledge in a graded way. Chapter 2 takes inspiration from this work, and asks whether children understand how another person's beliefs will change with evidence.

From decades of work in developmental psychology, we know that by the time children start kindergarten, they have a rich understanding of their own and others' mental states. Two key open questions regarding theory of mind germane to this thesis concern 1) the degree to which theory of mind exhibits the same dynamics when reasoning about the self as when reasoning about others and 2) whether young children have a nuanced understanding of knowledge and inference that allows them to make predictions that are sensitive to graded levels of evidence when reasoning about others' beliefs. Chapter 6 discusses the first point in detail, while the empirical work in Chapter 2 aims to address the latter.

### **1.3 Reasoning about people's traits and skills**

#### **1.3.1 Others**

In addition to assessing knowledge and beliefs, our capacity to make inferences about stable characteristics, such as traits, competencies, and skills form the backbone of metacognition. Similar to how beliefs can be used to predict actions, inferences about features like a person's trait or aptitude in a particular domain, can support problem solving and social learning. Adults habitually reason about others' traits, making inferences about people's underlying personality characteristics through observations of their actions. For example, you might infer that an acquaintance who passes you on the street without greeting you is unfriendly (Ross, 1977). Investigations of whether young children make similar behavior-to-trait and trait-to-behavior inferences reveal that children as young as 3 do relate individual actions to broader trait characteristics (Boseovski, Chiu, & Marcovitch, 2013; Boseovski & Lee, 2006; Giles & Heyman, 2003; Hermes, Behne, & Rakoczy, 2015; Heyman & Gelman, 2000; Liu, Gelman, & Wellman, 2007). For example, children expect someone who is labeled as selfish to be unlikely to share goods in the future. However, behavior-to-behavior inferences (extending what you have observed someone do in the past to predict a new behavior in a related domain) are more difficult, and it is not until age 6 do children successfully make these behavior-to-behavior predictions (Liu et al., 2007; Rholes & Ruble, 1984). In sum, young children demonstrate an understanding of psychological traits, in that they can differentiate one kind of trait from another (for example, kindness from intelligence), associate behaviors with related traits, and deploy this understanding to make predictions about and explanations regarding other people.

Knowledge about traits is one source of information that allows children to make predictions about, and guide their interactions with, others. Another source of evidence that

shapes these interactions is the assessment of other people's skills and informativeness. Children are sensitive to whether a particular context is pedagogical and thus offers an opportunity for learning. For instance, children copy novel actions only when they are demonstrated intentionally, and not accidentally (e.g., Bonawitz et al., 2011; Carpenter, Akhtar, & Tomasello, 1998; Csibra & Gergey, 2009; Harris, 2012; Tomasello, 1999). In addition to being sensitive to whether a particular *context* offers an opportunity for learning, children are also sensitive to whether a given *individual* would be optimal to learn from. They selectively choose to associate with and learn from more trustworthy, competent, and knowledgeable people (Koenig & Sabbagh, 2013). Children also use their assessments of individuals' past knowledge to decide if future information they provide is likely to be true and useful. After seeing one person correctly label an object (e.g., a woman calls a shoe "shoe") and another incorrectly labeling the same object (e.g., a man calls a shoe "clock"), children as young as 3 years prefer to use the person who produced the correct label's name for a novel object (Birch, Vauthier, & Bloom, 2008; Corriveau & Harris, 2009; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Sabbagh & Baldwin, 2001).

Moreover, 3- to 5-year-olds report that different positive characteristics may be useful for different tasks, showing an understanding that people may have more expertise in some areas than others (Danovitch & Keil, 2004; Lutz & Keil, 2002). For example, children ask for help from an individual who has the relevant capability (physical strength vs. intelligence) depending on the task they need to solve (Hermes, Behne, & Rakoczy, 2015; Kushnir, Vredenburgh, & Schneider, 2013). Although children (like adults) may be susceptible to halo effects, expecting a positive attribute in one area to predict positive attributes in an unrelated area (Brosseau-Liard & Birch, 2010; Fusaro, Corriveau, & Harris, 2011; Lane, Wellman, & Gelman, 2013), these

findings of sensitivity to expertise show that children can assess which positive traits are relevant for a given situation (e.g., Jara-Ettinger, Tenenbaum & Schulz, 2015). Taken together, these results show that preschool-age children are able to leverage their understanding of traits and skills to identify who would be more or less helpful in the service of their own learning.

### **1.3.2 Self**

If young children apply the same sophisticated understanding of others' traits and skills to themselves, we would expect them to represent themselves as having a variety of psychological traits, and to think about their behavior in relation to higher order traits (e.g., Brim, 1976). However, influential scholars in the area of self-concept development have long advocated the position that children do *not* conceive of themselves as having differentiable traits, and that they do not relate their behavior to enduring traits (Harter, 1982; Harter & Pike, 1984; Nicholls, 1978; 1984). These researchers argue that children instead make simple concrete observations about their current knowledge and behavior, without inferring they have stable characteristics until the age of 7 or 8 years.

This conclusion may have emerged because, generally speaking, young children express relatively positive global self-concepts (Parsons & Ruble, 1977; Stipek & Hoffman, 1980), so studies may not have been able to evoke specific and varying inferences about trait characteristics in a given domain from an action they carried out in that domain. For example, over 85% of 4- and 5-year-olds report a high level of self-evaluation when asked, "Are you good at puzzles?" (Smiley & Dweck, 1994). This notion is supported by the results of studies using verbal self-concept inventories, which show that various dimensions of self-concepts (e.g., mathematical ability, reading ability, athleticism, attractiveness) are more correlated in 5-year-

olds compared to 8-year-olds (Marsh, Craven, & Debus, 1991, 1998), suggesting that self-concepts do undergo development.

However, studies using less verbally demanding inventories have shown that even 4- and 5-year-olds have multidimensional self-concepts, and that their self views in a domain relate to their actual abilities. For example, children's math self-concepts relate more to real-world achievements in math than actual achievement in unrelated domains (Marsh, Ellis & Craven, 2002). That children exhibit these multidimensional self-concepts, representing oneself as skilled in one domain, and less so in another, suggests that even as young as 4, children may be using comparisons with peers or observations of their past success and failures to form multifaceted and potentially veridical views of themselves. Moreover, children incorporate feedback about traits and behavior such as, "You're smart" or "Girls are good at drawing," into their reported self-concepts, indicating that children have concepts of their own intelligence or artistic characteristics that can be influenced by evidence (e.g., Cimpian, Arce, Markman, & Dweck, 2007; Cimpian, Mu, & Erickson, 2012). One important point from these investigations is that lowering task demands as well as finding ways for young children to respond naturally is most likely to reveal their true range of abilities.

In addition to differences in task demands, discrepancies in children's ability to evaluate themselves and others may be due to a difference in the function that they serve. Evaluations of the self may be used, in part, to maintain self-esteem or determine what activities to engage in. Evaluations of others may be useful for comparison to oneself or for decisions about who to select as a partner or teacher. Thus, more accurate self-assessments may emerge in relevant learning contexts, where children have a goal of figuring out how to best learn or allocate their efforts, something I examine in Chapters 4 and 5.



In fact, having these learning goals might be what leads children to be especially overoptimistic about their abilities (Lockhart, et al., 2002; see also Boseovski & Lee, 2006; Droege & Stipek, 1993). They might have a good reason to be overconfident, namely in order to maintain motivation and positive self-esteem (Bjorklund & Green, 1992). Indeed, even adults have a tendency to associate themselves with positive traits and others with negative traits, as well as a tendency to assume positive outcomes are due to features of an individual rather than the situation when the self as opposed to another person is concerned (e.g., see Alicke & Govorun, 2005; Brown, 1986; Ross, 1977). As children develop they are gaining new skills at rapid rate, which can mean that failure on any given task may be temporary. Overoptimism may be an appropriate response to these past experiences when considering increases in their likelihood of success on a given task in the future.

In a similar vein, young children may express overly optimistic predictions about their future performance because they are communicating their wishful thinking, and improvements are also expected as a result of their prior experience with the advancing pace of their own development. In this way, they may be interpreting adults' questions about expectations for the future as questions about desires for the future (Cimpian, 2017). This argument has received some empirical support: children predict that a toy will contain more prizes inside when the child is allowed to keep the prizes than when they are not allowed to keep the prizes (Bernard, Clément, & Mercier, 2016). Because children are less likely to engage in wishful thinking when predicting other people's performance, their predictions for future outcomes are systematically higher for the self than for others (Destan et al., 2017; Schneider 1998; Stipek, 1981; although see Lipko, Dunlosky & Merriman, 2008).

Another, not mutually-exclusive, reason children and adults might over-estimate their own future performance relative to others is that people have access to additional information when making predictions for the self, such as their own level of motivation and anticipated effort. Individuals might assume because they have more perceived control over what they themselves do than what others do, they might also have a higher level of motivation (Heckhausen, 1984). Thus differences in the stakes, as well as sources of information, children consider when making predictions about the self versus others may lead to the aforementioned discrepancies. Chapter 5 investigates children's ability assessments, as they relate to both the self and a partner. The next section concerns how our ability to make person-specific inferences, and inferences about our own thought processes, scale up in the service of effective learning.

#### **1.4 Children's metacognitive abilities and their relation to self-directed learning**

Metacognitive inferences about oneself are crucial for planning and learning (see Bjork, Metcalfe, & Shimamura, 1994), helping us to recognize which aspect of a problem we need to tackle, what steps we should take, and then to evaluate our performance or track errors to improve future problem solving. Therefore, it is critical to know whether young children are able to use introspective metacognitive signals such as uncertainty, as well as evidence they can gather from and about other people, to direct their own learning.

These kind of inferences are particularly relevant in the context of self-directed learning for several reasons, two of which I highlight here. First, while knowledgeable others do transmit information critical to complex activities and cultural practices, most instances of social learning in traditional societies are driven by the learners themselves (Hewlett, Fouts, Boyette, & Hewlett, 2011). Second, self-directed learning often leads to more successful learning outcomes in adults

than when people do not make decisions about what actions to take or what data to observe (Bonwell & Eison, 1991; Freeman et al., 2014; Gureckis & Markant, 2012; Markant, Ruggeri, Gureckis, & Xu, 2016; Ruggeri, Lombrozo, Gureckis, & Xu, 2016). Caregivers, educators, and developmental psychologists have all noted that young children's learning is often self-guided and not formally supervised before schooling begins. Therefore the extent to which young children's learning is effective might be determined by their ability to solve metacognitive challenges, such as making choices about which activities to engage in, who to learn from, and how hard to try.

Despite the ubiquity and value of self-directed learning, it stands that directing one's own actions for learning requires several metacognitive components that are not necessarily invoked in directed learning. To effectively direct one's own actions, an individual must do all of the following: represent what they can do to remedy their knowledge gap, identify where they can seek helpful information, assess the reliability of those sources of information, and connect these representations to those of one's uncertainty (Boekaerts, 1997; Gureckis & Markant, 2012; Zimmerman, 1986). The learner must also consider which of an infinite number of possible actions to take in order for the goal to be achieved, and whether or not they have the capacity to carry out those actions. Representing these components of problem solving in self-directed learning is crucial because trial and error *ad infinitum* is often inefficient or costly in other ways.

Abstracting information from others to influence one's actions can be useful, particularly if you have never encountered a given situation before, and if another person is available to provide guidance. Across a variety of contexts, children use observational learning to direct their own actions (e.g., Buschsbaum, Gopnik, Griffiths & Shafto, 2011; Lyons, Young, & Keil, 2007; Meltzoff, 1995). This pertains not only to learning how to operate a toy, but also how to

“operate” the self. For example, young infants use information from others when deciding how to allocate their efforts: they use the emotional expression of a caregiver to decide whether to attempt to walk down a step slope (Tamis-LeMonda et al., 2008) and integrate an adult’s level of effort to calibrate how hard they should try on a novel task (Leonard, Lee, & Schulz, 2017). In addition, children take into account messages from adults regarding effort and ability when responding to challenges (e.g., Cimpian, 2010; Cimpian et al., 2007; 2012; Gunderson et al., 2015). In this way, children’s self-directed actions and decisions in learning are constrained and guided by evidence from social learning as well as first-person experience of success and failure (e.g., Gweon & Schulz, 2011; Stipek, Recchia, & McClintic, 1992).

As discussed in section 1.3.2, children are sensitive to the extent to which individuals are more or less useful to learn from. However, other people are not always available to serve as guides. Given that children take into account another agent’s confidence when deciding whether to trust the agent (Brosseau-Liard, Cassels, & Birch, 2014), it is reasonable to ask to what degree children gauge their *own* uncertainty for learning. We can begin by considering how previously studied inferences children make about the world have consequences for our understanding of metacognition and self-directed learning. Children’s calibration of their actions in response to uncertainty is demonstrated in contexts where they seek more information. When children do not know which of a number of hypotheses can explain a phenomenon, they will search for additional evidence by exploring, (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Cook, Goodman, & Schulz, 2011; Legare, 2012; Schulz & Bonawitz, 2007; Stahl & Feigenson, 2015) and take actions to specifically address the causal inferences they make from observed patterns of data (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Gweon & Schulz, 2011).

These findings support the notion that children's behavior is guided in a rational way by their uncertainty, and demonstrate that they take actions based on the inferences they can and cannot make from data. This body of work makes a strong case for the role of metacognitive reasoning in learning, since these behaviors require representing one's own current knowledge, including one's level of uncertainty, and the relationship between those gaps in knowledge and possible future actions. What is yet to be investigated is precisely how metacognition and uncertainty monitoring guide exploration in early childhood, and how rational inferential processes support learning not only about the world, but also about the self.

A great deal of the research has measured how children reflect on their own uncertainty, by querying children with questions like "How sure are you that you know this answer?" as well as non-verbal proxies for confidence, such as children's willingness to place a larger or smaller bet on their response being correct. Results from studies across large age ranges suggest that the ability to articulate our thoughts about uncertainty and future learning improves as we develop into young adulthood. Adults' judgments of their own learning or predictions of their future performance on a task are well correlated with actual performance (e.g., Dunlosky & Nelson, 1992, c.f., Sharot, Riccadri, Raio, & Phelps, 2007). Studies with similar methodologies find that children are often much less accurate (Koriat & Shitzer-Reichert, 2002; Lyons & Ghetti, 2010; Sussan & Son, 2007). Because many studies demonstrate poor reflective thinking, poor uncertainty monitoring, and poor performance prediction (Flavell, 1979; Flavell, Green & Eleanor, 2000; Myers & Paris, 1978; Reyna, 1996; Veenman & Spaans, 2005), one prevailing view of young children is that their metacognitive abilities, construed in terms of introspective access to their thinking, are sorely lacking. However, there are at least two reasons to believe these studies have not fairly characterized young children's metacognitive skills in these areas.

One way to account for children's difficulty with these tasks, related to an account for 3-year-olds' poor performance on false belief tasks discussed in Chapter 1.2, concerns to the development of executive functioning (e.g., Bell & Livesey, 1985; Carlson, 2005; Diamond, 2006). Cognitive control has been theorized to play a key role in metacognitive regulation and planning (e.g., Fernandez-Duque, Baird, & Posner, 2000), and thus, young children who do not have extensive cognitive control may fail at metacognitive tasks that are queried through explicit judgments.

Indeed, a host of findings support the argument that difficulties with cognitive monitoring tasks are at least in part due to less well-developed executive control. When queried using tasks that rely less on declarative memory, young children demonstrate discernable metacognitive awareness (Cultice, Somerville, & Wellman, 1983; Hembacher & Ghetti, 2014; Lyons & Ghetti, 2011; Marazita & Merriman, 2004; Winne & Perry, 2000), and before their second birthday, children selectively seek help from an adult, asking for assistance only when they did have epistemic access to a hiding event (Goupil, Romand-Monnier, & Kouider, 2016).

While the literature reviewed in this section provides substantial evidence and reason to argue for metacognitive sophistication in early childhood, the majority of these studies queried children's judgments of what they had learned or could learn without seeing how children direct their own learning based on those judgments. What remains an open question is the extent to which children deploy metacognitive reasoning to *guide their own learning* across a variety of situations, which is particularly relevant given the development of executive control and bias towards optimism that children display when predicting their own performance. It is possible that self-directed learning is informed by metacognition when executive functioning demands are relatively low or when responses are queried using more naturalistic and intuitive measures.

Recent research on children's process of inquiry (see Schulz, 2012 for review), confidence judgments (e.g., Vo et al., 2014), and self-relevant cognition (see Cimpian, 2017 for review), which reduce task demands, all lend support to the argument that metacognition is critical for effective self-directed learning. Chapters 3, 4, and 5 aim to elucidate whether this is the case, taking inspiration from the study of rational self-directed learning and metacognition.

### **1.5 Roadmap to the thesis**

Chapter 1 reviewed the literature on metacognition as it relates to understanding other people and the self, with a particular focus on how cognition about cognition impacts self-directed learning. In Chapter 2, I report an investigation of children's understanding of how people learn from data by eliciting their inferences about third-party belief revision. Children not only integrate data with their prior beliefs, but they expect others to do so as well. Their ability to predict what others will think, even when their prior beliefs differ from their own illustrates that children are able to reflect on the processes that give rise to knowledge. The work in Chapter 3 considered whether children calibrate their exploration based on the informativeness of actions given one's own perceptual limitations and the difficulty of the problem they are trying to solve. The results from this study show that children's exploration is tuned in a fine-grained way to the discriminability of hypotheses they are considering, suggesting they represent their uncertainty in order to direct their learning. Chapter 4 investigates whether children make use of a ubiquitous source of evidence bearing on the self—social comparison – to calibrate their effort. Already by age 3, information about peers' performance influences children's persistence in learning. Finally Chapter 5 reports work investigating whether children invoke assessments of their own and others' abilities to divide labor. This study demonstrates that even young children consider

relative ability when solving problems collaboratively across contexts. Taken together, these papers advance our understanding of the ways in which children consider themselves and others for the purpose of learning. Chapter 6 discusses limitations, future directions, and broader implications we can glean from this work, including whether metacognition for the self and others is based on the same underlying system of reasoning.



## Chapter 2

**The work in this chapter has been published in Magid, Yan, Siegel, Tenenbaum, & Schulz (2017), *Changing minds: Children’s inferences about third party belief revision. Developmental Science, 21*, DOI: 10.1111/desc.12553.**

By the age of five, children explicitly represent that agents can have both true and false beliefs based on epistemic access to information (e.g., Wellman, Cross, & Watson, 2001). Children also begin to understand that agents can view identical evidence and draw different inferences from it (e.g., Carpendale & Chandler, 1996). However, much less is known about when, and under what conditions, children expect other agents to change their minds. Here, inspired by formal ideal observer models of learning, we investigate children’s expectations of the dynamics that underlie third parties’ belief revision. We introduce an agent who has prior beliefs about the location of a population of toys and then observes evidence that, from an ideal observer perspective, either does, or does not justify revising those beliefs. We show that children’s inferences on behalf of third parties are consistent with the ideal observer perspective, but not with a number of alternative possibilities, including that children expect other agents to be influenced only by their prior beliefs, only by the sampling process, or only by the observed data. Rather, children integrate all three factors in determining how and when agents will update their beliefs from evidence.

## 2.1 Introduction

Expectations of rational agency support our ability to predict other people's actions and infer their mental states (Dennett, 1987; Fodor, 1987). Adults assume agents will take efficient routes towards their goals (Heider, 1958; D'Andrade, 1987), and studies with infants suggest that these expectations emerge very early in development (Skerry, Carey, & Spelke, 2013). By the end of the first year, infants can use situational constraints, along with knowledge about an agent's goal, to predict an agent's actions. Similarly, they use knowledge of an agent's actions and situational constraints to infer the agent's goal, as well as knowledge of an agent's actions and goal to infer unobserved situational constraints (Csibra, Bíró, Koos, & Gergely, 2003; Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra, & Bíró, 1995). Such work has inspired computational models of theory of mind that formalize the principle of rational action and successfully predict human judgments (Baker, Saxe, & Tenenbaum, 2009; Baker, Saxe, & Tenenbaum, 2011; Jara-Ettinger, Baker, & Tenenbaum, 2012). Here however, we ask whether learners' expectations extend to the more colloquial meaning of the word "rational": the expectation that other people's judgments and beliefs have a basis in the evidence they observe.

Note that this is distinct from the question of whether children *themselves* draw rational inferences from data. Decades of research suggest that very young children can integrate prior beliefs with small samples of evidence to infer the extensions of word meanings, identify object categories, learn causal relationships, and reason about others' goal-directed actions (see Gopnik & Wellman, 2012; Schulz, 2012; and Tenenbaum, Kemp, Griffiths, & Goodman, 2011 for reviews). However, despite extensive work on children's theory of mind (see Wellman, 2014 for discussion and review), less is known about how children expect others to learn from evidence. Although classic theory of mind tasks look at whether children expect others to update their

beliefs given diverse forms of epistemic access to data – including direct perceptual access (e.g., Wimmer & Perner, 1983), indirect clues (e.g., Sodian, Taylor, Harris & Perner, 1991) and testimony, (e.g., Zaitchik, 1991) – these involve a relatively simple instantiation of the expectation that others will learn based on their observations of the world: children need only understand whether the agent does, or does not have epistemic access to belief-relevant information. Such studies do not ask whether children understand that agents might evaluate evidence differently or draw different inferences from identical evidence.

The studies that do look at children’s understanding of how third parties might evaluate evidence suggest that an “interpretative theory of mind” is a relatively late development (Astington, Pelletier, & Homer, 2002; Carey & Smith, 1993, Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross, Recchia, & Carpendale, 2005; Ruffman, Perner, Olson, & Doherty, 1993). Not until six and seven years do children understand, for example, that an ambiguous line drawing can be viewed as two different kinds of animals (Carpendale & Chandler, 1996) or that iconic symbols are subject to different interpretations (Myers & Liben, 2012). Young children’s failure to understand that agents can reach different conclusions from the same evidence suggests that children might have difficulty understanding how other agents’ prior knowledge affects the interpretation of data.

Arguably however, understanding that evidence is ambiguous and thus open to interpretation may be more challenging than understanding the conditions under which others might be expected to learn from evidence. Relatively little work has looked at what children understand about others inferences from data, and the findings here are mixed. For instance, both four and six-year-olds recognize that an unseen marble must be blue if it is drawn from a bag containing only blue marbles; however, only six-year-olds recognize that a third party (who

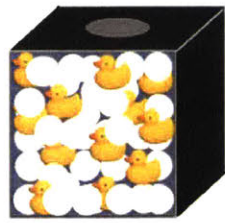
knows the contents of the bag) will make the same inference and thus know the color of the marble (Sodian & Wimmer, 1987). However, four-year-olds do understand that if covariation evidence suggests that one of two causes is correlated with an outcome and the experimenter tricks a puppet by reversing the evidence, the puppet will conclude that the wrong variable is the cause (Ruffman et al., 1993).

Such studies suggest that by four, children are at least beginning to understand that third parties learn from evidence in ways that go beyond mere perceptual access to data. However, they leave open the question of whether children can use patterns of evidence to understand when others will change their minds, and the degree to which children integrate others' prior beliefs in predicting their learning. Do children expect others to update their beliefs from data in cases where learning requires representing not merely an agent's access to evidence but the agent's ability to draw appropriate inferences from the evidence?

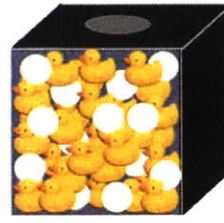
To ask whether children expect others to rationally update their beliefs from data we borrow from two influential tasks in the literature. The first is the classic false belief task (Wimmer & Perner, 1983). The other is derived from work looking at infants and children's understanding of the relationship between samples and populations (e.g., Denison & Xu, 2014; Gweon, Tenenbaum, & Schulz, 2010; Kushnir, Xu, & Wellman, 2010; Xu & Denison, 2009; Xu & Garcia, 2008). Specifically, we show a child and another agent (a Frog puppet) two boxes: one containing more rubber ducks than ping-pong balls (the Duck box) and one containing more balls than ducks (the Ball box). The Frog leaves, and the child watches as the boxes are either moved and returned to the same location (so the Frog has a true belief about the location of each box) or switched (so the Frog has a false belief about the location of each box). At test, the Frog returns, and both the child and the Frog watch as the experimenter reaches into the Duck box and

draws a sample of three or five ducks either apparently at random (without looking into the box) or selectively (looking in and fishing around). After both the child and the Frog see the sample of data, children are asked, “Where does Froggy think the Duck box is now?” See Figure 1.1 for a schematic of the procedure.

Both the ability to reason about others’ false beliefs (see Baillargeon, Scott, & He, 2010) and the ability to recognize when data is sampled randomly or selectively (e.g., Xu & Denison, 2009) emerge relatively early in development. However, children do not reliably provide accurate responses in explicit false belief tasks until later childhood (see Wellman, Cross, & Watson, 2001 for review) and as noted, the ability to understand that identical evidence can be open to different interpretations emerges even later (e.g., Astington et al., 2002; Carey & Smith, 1993, Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross et al, 2005; Ruffman et al., 1993). Because we are interested not in children’s own inferences from the data, but in their inferences on behalf of a third party whose beliefs may differ both from the child’s own and those supported by the observed data, here we focus on 4.5- to 6-year-olds.



Ball box



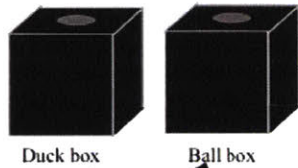
Duck box



I like the Duck box.

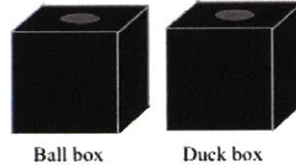
**a) Preference Phase: Frog present**

New Location



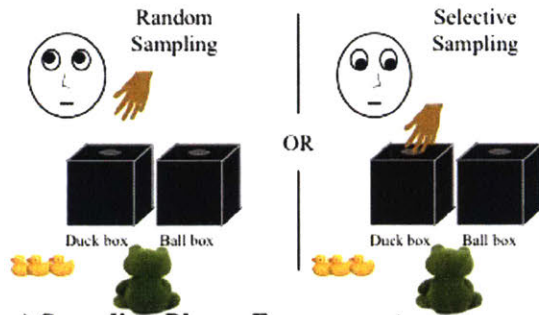
OR

Old Location

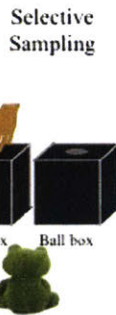


**b) Belief Phase: Frog absent**

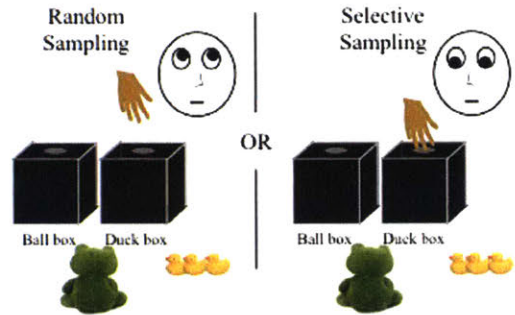
New Location



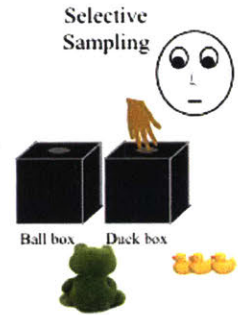
OR



Old Location

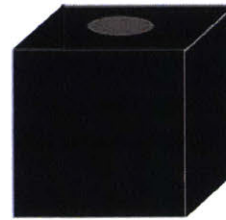
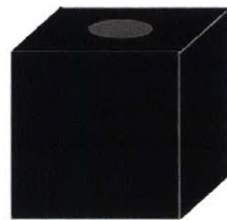


OR



**c) Sampling Phase: Frog present**

Where does Froggy think the Duck box is now?



**d) Test Phase: Frog present**

*Figure 2.1.* Schematic of the procedure. In the Preference phase (a) children are shown the two boxes with different proportions of ducks and balls and ask to identify the Duck box and Ball box based each box's majority object. Then children are introduced to the Frog puppet and his preference for ducks and the Duck box and then learn, along with the Frog, that the boxes can either each move back and forth to stay in the same location or move from one side to the other to switch locations. In the Belief Phase (b) children either see the boxes switch locations (New Location condition) or stay in the same location (Old Location condition) while the Frog is absent. When the Frog returns, he will either have a false belief about the location of the Duck box (New Location condition) or a true belief about the location of the Duck box (Old Location condition). Children are asked two check questions to confirm they have tracked the locations of the boxes and the Frog's belief at the end of the Belief Phase. In the Sampling Phase (c) the Frog returns and the experimenter samples either randomly (Random Sampling condition) or selectively (Selective Sampling condition) from the hidden Duck box. At the Test Phase children are asked where the Frog thinks the Duck box is.

As shown in Table 2.1, if children expect the Frog to update his beliefs from evidence, then the cross between old and new locations and random and selective sampling predicts a pattern of responses distinct from the pattern that would be generated if children adopted many other possible response strategies. We will walk through the predictions of our account intuitively; however, to clarify our proposal, we also include a computational model providing quantitative predictions for both our account and a number of alternatives, in each experimental condition. (See Figure 2.4). The details of the model are not critical to our proposal as our goal here is not to evaluate the Rational Learning model per se. Given that there are only five conditions, and some of them (e.g., both selective sampling conditions) make overlapping predictions, correlations between the model and children's performance may be less convincing than the relative fit of the rational inference model in comparison to the alternative models. That is the analysis we include here. Additionally, it is helpful to consider the qualitative intuitions behind these models insofar as they motivate our predictions and ground our intuitions in a precise statement of what constitutes "rational inference" in this context.

*Table 2.1.* The predictions for the dominant response pattern if children expect other agents to engage in rational learning from data are listed in row a. The \* indicates that the probability that children think the Frog will change his mind should depend on the strength of the evidence the Frog observes. Possible alternative patterns of responses to the test question in each of the four conditions: New Location/Random Sampling (NL/RS); New Location/Selective Sampling (NL/SS); Old Location/Random Sampling (OL/RS); Old Location/Selective Sampling (OL/SS). OLD indicates that the child would point to the original location of the Duck box and NEW that the child would point to the new location.

Response Pattern	New Location Random Sampling NL/RS	New Location Selective Sampling NL/SS	Old Location Random Sampling OL/RS	Old Location Selective Sampling OL/SS
a) Rational Learning	NEW*	OLD	OLD	OLD
b) Actual location (or child's own beliefs)	NEW	NEW	OLD	OLD
c) Frog's beliefs (without updating from data)	OLD	OLD	OLD	OLD
d) Sampled data (without prior beliefs)	NEW	CHANCE	OLD	CHANCE
e) Random-Stay; Selective-Shift	NEW	OLD	OLD	NEW
f) Chance	CHANCE	CHANCE	CHANCE	CHANCE

### 2.1.1 Predictions of the rational inference account

If children expect agents to rationally update their beliefs, they should respond jointly to the type of sampling process and the Frog's prior beliefs about the boxes' locations together with his knowledge that the boxes can move. A sample randomly drawn from a population is likely to be representative of the population. Thus we predict that in the Random Sampling conditions, children should expect the Frog to use the evidence to verify or update his beliefs about the location of the Duck box.



Specifically, randomly sampling three ducks in a row is improbable unless the evidence is sampled from the Duck box. Thus when evidence is randomly sampled from the Old Location (OL/RS), children should infer that the Frog will retain his belief and will continue to think that the Duck box is in the Old location. However, when evidence is randomly sampled from the New Location (NL/RS\_3 ducks and NL/RS\_5 ducks) children should believe that the Frog may now update his former false belief, inferring that the Duck box may have been moved to the New Location. Moreover, the strength of children's inferences should depend, in a graded way, on the strength of evidence they observe: they should be more confident that Frog will change his mind when they see five ducks randomly drawn from the New Location (NL/RS\_5 ducks) than when they see three ducks (NL/RS\_3 ducks) randomly drawn.

By contrast, selectively sampled evidence is uninformative about the population from which it is drawn. The experimenter can selectively draw any sample at all (representative or non-representative) from the population. Indeed, if the experimenter is trying to *guarantee* that she gets three ducks in a row, she should sample selectively regardless of whether she is drawing from the population where ducks are relatively common (the Duck box) or the population where ducks are relatively rare (the Ball box). Since the selectively sampled evidence is consistent with sampling either from box, a rational learner who integrates his prior beliefs with the data should retain his prior beliefs. That is, because both the Old and New Location are consistent with the data and only the Old Location is consistent with the agent's prior beliefs, we predict that children will expect the Frog to say the Duck box is in the Old Location in both the Old (OL/SS) and New Location (NL/SS) Selective Sampling conditions.

### **2.1.2 Alternative accounts**

In contrast to the pattern of responses consistent with third party rational inference (Table 1, row a), there are a number of other ways children might respond to the question “Where does Froggy think the Duck box is now?” Children might respond with the actual location of the Duck box; this is after all, the location from which the Frog sees the ducks sampled, and also of course consistent with what the children themselves believe (row b, Table 1). Alternatively, children might respond with the Frog’s true or false belief about the location of the Duck box, considering whether the Frog saw the boxes moved or not but without expecting Frog to update his beliefs given the sampled evidence (row c, Table 1). Another possibility is that children might expect the Frog to attend to the sampled evidence but not integrate it with his prior beliefs; they may conclude that if the Frog sees randomly sampled evidence he will strongly conclude that the Ducks are in that location, but if he sees selectively sampled evidence, he will recognize that the evidence is uninformative and choose at chance (row d, Table 1). Yet another possibility is that the children think the Frog will attend to the sampled evidence but not as a rational learner would; they might for instance, think that Frog will conclude that random sampling indicates the sample of ducks is pulled from the Duck box and that selective sampling of ducks means the sample is pulled from the Ball box (row e, Table 1). Finally, children might respond at chance, either because they genuinely believe that the Frog will guess or because different children choose different strategies and thus, as a group, generate responses indistinguishable from chance responding (row f, Table 1). Corresponding to the qualitative predictions shown in Table 2.1, 2.4 shows quantitative predictions for the rational inference account and each of the alternative accounts for all of the conditions in our study. The model predictions can be compared with children’s behavioral data (Figure 2.2).

In the experiment to follow, we test these different accounts and predict that children's responses will be best explained as inferring that the Frog will rationally integrate his prior belief about the boxes' locations with the type of sampling process he observes. Note that this specific pattern of responding requires children to track simultaneously the true location of the Duck box, the Frog's belief about the location of the Duck box, and the probability of generating the observed sample from the population. The complexity of the task is necessary to distinguish children's responding to a third party's updating of his beliefs from responses children might make on other grounds. (See Table 1.1) However, given the complexity of the task, we expected a number of children to have difficulty tracking the true location of the Duck box and the Frog's beliefs about the boxes' locations (especially as the boxes were occluded through much of the task and differed only in the relative proportion of their contents). Of course, children can only reason accurately about how the Frog might update his beliefs given the sample if they remember both the true location of the Duck box and the Frog's beliefs about the location. Thus we made an a priori decision to focus our analysis on the responses of the children who successfully answered both check questions.

## **2.2 Method**

### **2.2.1 Participants and Materials**

Two hundred and six children (mean: 66 months; range: 54-83 months) were recruited from an urban children's museum and participated in the study. The testing occurred in three waves in the following order: NL/RS\_3 and NL/SS; OL/RS and OL/SS; NL/RS\_5. Within each wave of testing children were randomly assigned to condition. Testing continued until 30 children passed the check questions in each condition. (See Inclusion Questions to follow.) While most of the children were white and middle class, a range of ethnicities and

socioeconomic backgrounds reflecting the diversity of the Boston metropolitan area (47% European American, 24.4% African American, 8.9% Asian, 17.5% Latino, 3.9% two or more races) and the museum population (29% of museum attendees receive free or discounted admission) were represented.

Two black cardboard boxes (30 cm<sup>3</sup>) were each separated into two sections by a cardboard barrier. The front side of both boxes was a clear plastic panel with a sheet of black felt velcroed over it. Each box had a hand-sized hole in the top. For one box, referred to as the “Duck box,” the front section was filled with 45 rubber ducks and 15 ping-pong balls. For the other box, referred to as the “Ball box,” the front section was filled with 45 ping-pong balls and 15 rubber ducks. (3:1 ratios were chosen because they are easily discriminable by preschoolers and because three consecutive ducks are far more likely to be randomly sampled from the Duck box than the Ball box.) The back sections of both boxes also contained rubber ducks and ping-pong balls, and were hidden from view. Each box was placed on a colored mat. A Frog puppet served as the agent.

### **2.2.2 Design and Procedure**

We crossed the two locations where the Duck box could be at the end of the study (Old and New) and two kinds of sampling processes, sampling three ducks, from the Duck box (Random and Selective), yielding four conditions: the Old Location/Random Sampling (OL/RS) condition, the New Location/Random Sampling (NL/RS\_3 ducks) condition, the Old Location/Selective Sampling condition (OL/SS), and the New Location/Selective Sampling (NL/SS) condition. We also ran a condition in the New Location/Random Sampling case in which children saw a sample of five ducks drawn from the New Location (NL/RS\_5 ducks). We included this condition to ask whether children drew graded inferences that depended on the

amount of randomly generated data the Frog observes (i.e., children should be more likely to think the Frog might change his mind given more randomly sampled data inconsistent with his prior beliefs).

#### **2.2.2.1 Preference Phase.**

In all conditions, the experimenter showed the child the Duck and Ball boxes side-by-side on a table (L/R counterbalanced across participants). Each box was placed on a different colored mat, red or blue, to help children track the identities of the boxes. Initially, the felt hid the boxes' front sections. Children were given a duck and a ball, not drawn from either box to hold briefly. The experimenter then lifted the felt, revealing the front sections of both boxes and said, "One box has mostly ducks, and one box has mostly balls. Which box has mostly ducks? Which box has mostly balls?" If the child answered incorrectly, the experimenter told the child the correct answer and repeated the questions.

Next, the experimenter introduced the agent, "Froggy," saying, "This is my friend Froggy!" The experimenter said, "Froggy likes ducks better than balls." The experimenter then asked the Frog if he wanted to play with the ball. The Frog replied, "No, I only like ducks!" The child was asked to hand the Frog his favorite toy. The Frog's preference for ducks was established to help children track the Frog's goal of locating the Duck box. Next, both the child and the Frog learned that the boxes could move in two ways. The experimenter said, "The boxes can move so that they are in the same place," and "The boxes can move so that they are in different places." (For the former, the experimenter rocked the boxes back and forth three times. For the latter, the experimenter moved the Duck box from the red mat to the blue one and the Ball box from the blue mat to the red one, or vice versa; counterbalanced across participants.) The experimenter then asked the Frog, "Which box do you like best?" The Frog approached the

Duck box and said, “I like this box! I like the Duck box!” The experimenter returned the boxes to their original locations. The experimenter asked the child to point to the box the Frog preferred; all children answered correctly.

#### **2.2.2.2 Belief Phase.**

Next, the experimenter told the child that the Frog was tired and hid him under the table. Children watched as the experimenter re-covered the front of both boxes with the felt. For children in the Old Location conditions, the experimenter rocked the boxes back and forth saying, “I’m going to move the boxes so that they are in the same place.” For children in the New Location conditions, the experimenter switched the locations of the boxes saying, “I’m going to play a trick on Froggy! I’m going to move the boxes so that they are in different places.”

#### **2.2.2.3 Inclusion questions.**

In both conditions, the experimenter then asked children two questions to check that they understood the true locations of the boxes (*location check*) and the Frog’s beliefs about the boxes (*belief check*). The *location check* question was, “Where is the duck box?” The *belief check* question was, “Where does Froggy think the duck box is?”

#### **2.2.2.4 Sampling Phase.**

The experimenter brought the Frog back saying, “Look, Froggy is back!” The experimenter asked the Frog to watch the two boxes and then responded to a pretend phone call saying, “Hello? Oh, you want me to take three (five in the NL/RS\_5 condition) ducks from the box on the red (blue) mat?” (The experimenter always named the actual location of the Duck box.) We included the phone call to dispel any impression that the experimenter was pedagogically sampling from the box in order to teach the Frog (or child) the actual location of

the Duck box. Note that pedagogical sampling is always selective but intentional sampling can be either random or selective: one can intentionally pull objects out at random or intentionally choose particular objects (see e.g., Gweon et al., 2010 for discussion). In the Random Sampling conditions, the experimenter looked over her shoulder (i.e., not into the box) and reached through the hole into the Duck box three times in rapid succession, drawing out a duck each time and counting out “One, two, three (four, five, only in the NL/RS\_5 condition)”. In the Selective Sampling conditions, the experimenter peered through the hole into the Duck box and kept her hand inside the box for approximately two seconds before retrieving a duck. She counted, “One...two...three” after finding each duck. After sampling three ducks from the box and ending the pretend phone call, the experimenter asked, “Froggy, did you see that?” to which the Frog replied, “Yes.”

#### **2.2.2.5 Test Phase.**

In the final phase of the experiment after the sample of three ducks was drawn, children were asked the critical test question: “Where does Froggy think the Duck box is now?”

### **2.3 Results**

#### **2.3.1 Inclusion Questions**

Children’s responses were coded from videotape by the first authors. Forty-seven percent of the data was coded by a second coder, blind to condition and hypotheses.

Inter-coder reliability was high (Kappa = .95, 98% agreement).

We coded children’s responses to the location (“Where is the duck box?”) and belief (“Where does Froggy think the duck box is?”) check questions. Of the 206 children tested, 73% ( $N = 150$ ) answered both questions correctly (“trackers”) and 27% ( $N = 56$ ) answered one or both of the check questions incorrectly (“non-trackers”). The number of

children excluded for failing only the location question, only the belief question or both by condition is as follows: NL/RS\_3: location: 6; belief: 2; both: 4; NL/RS\_5: location: 0; belief: 7; both: 8; NL/SS: location: 1; belief: 12; both: 4; OL/RS: location: 4; belief: 3; both: 0; OL/SS: location: 1; belief: 4; both: 0.

Non-trackers were younger than trackers (non-trackers:  $M=64$  months; trackers:  $M=67$  months;  $t(204) = 2.52, p = .01, d = .39$ ). Non-trackers may have subsequently given responses about the Frog's belief that did not reflect information necessary to make accurate rational inferences on behalf of the Frog so we excluded these children from our primary analysis. This resulted in a final sample of  $N=150$  (53% female<sup>1</sup>) children across the five conditions: NL/RS\_3 ( $n = 30, m_{age} = 65$  mo.; range: 54-78 months), NL/SS ( $n = 30, m_{age} = 66$  mo.; range: 54- 82 months), OL/RS ( $n = 30, m_{age}=66$  mo.; range: 54- 81 months), OL/SS ( $n = 30, m_{age} = 68$  mo.; range: 55- 82 months), and NL/RS\_5 ( $n = 30, m_{age} = 69$  mo.; range: 57- 83 months). Age in months did not differ across conditions ( $F(4, 145) = 1.05, p = .38$ ).

Note that more children failed the inclusion questions in the New Location condition than the Old Location condition (unsurprisingly since the New Location condition involved tracking both a change of location and representing a false belief). On average, 8.67 more children were excluded for failure to track the boxes' locations and/or the Frog's beliefs in the three New Location conditions than the two Old Location conditions (a 33% exclusion rate versus a 17% exclusion rate;  $p = .01$ ). This raises the possibility that the included sample of children in the New Location might differ from those in the Old Location condition in any of a number of ways (e.g., including being

---

<sup>1</sup> Information on the children's gender was available only for 81% of the children; the reported percentage reflects this sub-sample.



more attentive or motivated, having better theory of mind or executive function skills, or differing with respect to other cognitive abilities).

Critically however, the rational learning account does not predict better, or even simply uniformly *different* performance in the New Location conditions than the Old Location conditions (predictions whose investigation could be confounded to the degree that one group of children met more stringent inclusion criteria than the other). Rather, it predicts a precise pattern of responses depending jointly on the Frog's initial beliefs about the box' location, the sampling process, and the amount of evidence observed. That is, this account makes predictions within each condition (where there are no differences in exclusion rates) and also predicts both commonalities and differences across conditions. Neither the prediction that, within each condition, children should be more likely to expect the Frog's beliefs to be informed by randomly than selectively sampled evidence, nor the prediction that children should draw stronger inferences for the Old than the New location condition given randomly (but not selectively) generated evidence, can be accounted for by an overall difference between the two conditions.

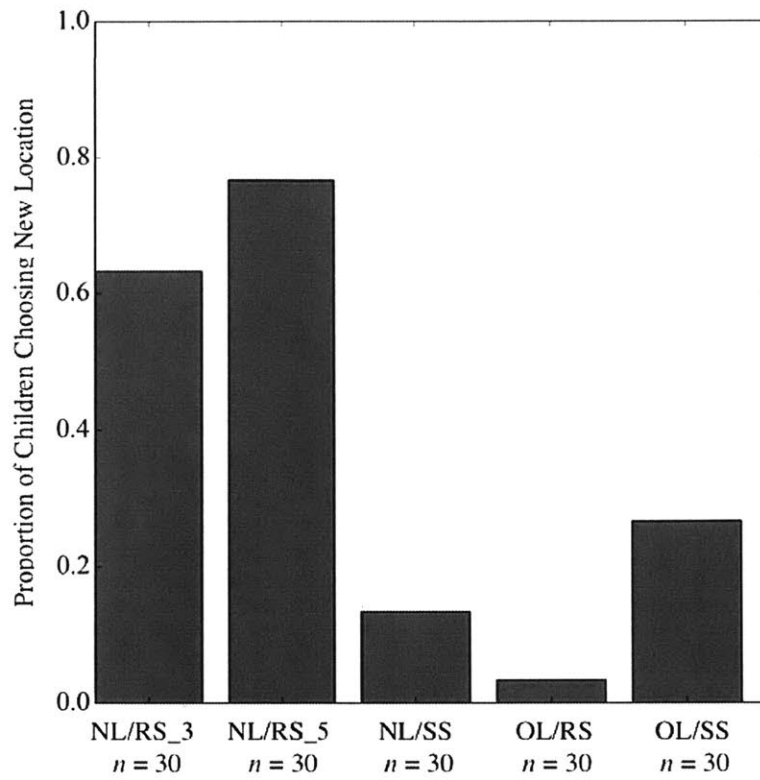
### **2.3.2 Test Question**

Because we had a priori hypotheses about the pattern of results, we performed planned linear contrasts. We formalized the prediction that the responses in the New Location/Random Sampling conditions would differ from the other three conditions, and that the other three conditions would not differ from each other by conducting the analyses with following weights: New Location/Random Sampling with three ducks (3), the New Location/Selective Sampling (-2), the Old Location/ Random Sampling (-2), the Old Location/Selective Sampling (-2), and the New Location/Random sampling with five ducks (3).

For the 150 children who recalled the Frog's belief as well as the boxes' actual locations, the linear contrast was significant ( $F(1, 149) = 19.54, p < .001, \eta^2 = .35$ ). Children were significantly more likely to believe the Frog had updated his belief to the New Location in the New Location/Random Sampling conditions than in the other conditions, (percentage of children choosing New Location by condition: NL/RS\_3: 63%; NL/RS\_5: 77%; NL/SS: 13%; OL/RS: 3%; OL/SS: 27%; see Figure 2.2)

By contrast, for the children who answered at least one of the check questions incorrectly (the non-trackers), the linear contrast was not significant ( $F(1, 55) = 1.78, p = .15, \eta^2 = .12$ ). Instead, children appeared to either respond at chance or respond to the last location where they had seen the ducks. See Figure 2.3. Crucially, these results suggest that the children who met the inclusion criteria were not simply defaulting to some baseline response pattern but were instead responding as predicted: inferring that the Frog would rationally update his beliefs from the data.

We restrict our analyses to children who pass the inclusion criteria because there is no clear way to interpret the responses of children who lost track of the boxes' location or failed to represent the Frog's initial beliefs. However, the linear contrast remains significant if all 206 children are included ( $F(1, 205) = 10.314, p < .001, \eta^2 = .22$ ), suggesting that the results are robust to the exclusion criteria.



*Figure 2.2.* Proportion of children who passed the inclusion criteria (“trackers”) who chose the New location in each condition in response to the test question about the Frog’s belief.

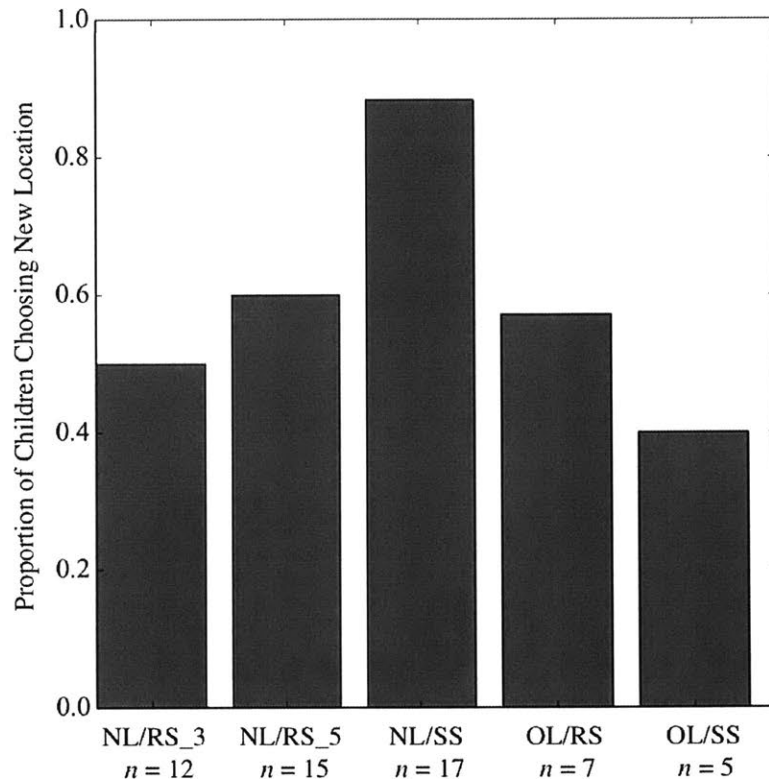


Figure 2.3. Proportion of children who failed the inclusion criteria (“non-trackers”) who chose the New location in each condition in response to the test question about the Frog’s belief.

Looking within each condition, children chose the Old Location significantly more often than chance in all conditions (percentage of children choosing Old Location: NL/SS: 87%,  $p < .001$ ; OL/RS: 97%,  $p < .001$ ; OL/SS: 73%,  $p = .02$ ; by binomial test) except the NL/RS\_5 condition where they chose the New Location above chance (77% of children choosing New;  $p = .005$  by binomial test) and the NL/RS\_3 condition where they chose at chance (63% of children choosing New;  $p = .20$  by binomial test).

Our hypothesis made several key predictions about differences between conditions. First, if children expect the Frog to be sensitive to the distinction between randomly sampled and selectively sampled evidence, then given the same prior beliefs and evidence, they should expect the Frog to draw stronger inferences from randomly sampled evidence than selectively sampled

evidence. Children's inferences did indeed depend on the type of evidence sampled. In the comparison between the New Location/Random Sampling\_3 condition and the New Location/Selective Sampling condition children were more likely to update the Frog's false belief and infer the Duck box was in the New Location in the Random Sampling than the Selective Sampling condition, as warranted, (Fisher's exact,  $p < .001$ ). Similarly, children were more likely to think the Frog would infer that the Duck box was in the Old Location in the Old Location/Random Sampling condition compared to the Old Location/Selective Sampling condition (Fisher's exact,  $p = .03$ ). The fact that children made comparable inferences in both the conditions suggests that the results cannot be explained by differences in children's belief understanding in the two conditions (i.e., as a byproduct of the different inclusion rates in the New and Old Location conditions). Rather, children's tendency to expect the Frog's beliefs to be more influenced by randomly sampled than selectively sampled evidence in both conditions is consistent with the Rational Learning account since indeed, randomly sampled evidence is more informative than selectively sampled evidence about the population from which it is drawn.

Also as predicted, numerically more children said the Frog would update his belief when five ducks were randomly sampled than when three ducks were randomly sampled. The difference between the NL/RS\_3 condition and NL/RS\_5 condition was not significant (Fisher's exact,  $p = .40$ ), however, the graded nature of children's inferences was consistent with the predictions of the rational inference model.<sup>2</sup>

Finally, as predicted, children were sensitive to the Frog's prior beliefs. Given identical evidence and sampling processes, children drew different inferences when the data were sampled

---

<sup>2</sup> Note that although ages did not differ significantly across conditions, the mean age of children in the NL/RS\_5 condition was 69 months, compared to 65 months for children in the NL/RS\_3 condition. We are grateful to an anonymous reviewer for pointing out the possibility that this age difference may have contributed to children's stronger inferences in the NL/RS\_5 condition.

from the Old Location and the New Location. Thus given three ducks randomly sampled from a location, children's inferences about what he would learn from the sample depended on the Frog's prior beliefs about the location of the Duck Box. Children were confident that the Frog would believe the randomly sampled data indicating that the Duck Box was in the old location (97% of children in the OL/RS chose Old) but did not make as strong an inference when the randomly sampled data suggested the Duck Box was in the New Location (63% of children in the NL/RS\_3 chose New; OL/RS vs. NL/RS\_3, Fisher's exact,  $p = .002$ ). The analogous comparison between the selective sampling conditions was also significant. The Rational Learning model predicts that children should choose the Old Location in both selective sampling conditions because selective sampling is uninformative about the population from which it is drawn. As predicted, children interpreted identical evidence differently depending on the Frog's prior beliefs about the location: children were more likely to choose the Old Location in the OL/SS condition (73%) than they were to choose the New Location in the NL/SS condition (13%; Fisher's exact,  $p < .001$ ).

As a further test of the hypothesis that children's judgments on behalf of the Frog reflect an expectation of rational learning, rather than any alternative model (Table 1) we can directly compare the Rational Learning model with alternative models using a Bayes factor analyses (see Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2013). As is clear in Table 2.2 and Figure 2.4, the Rational Learning Model outperforms all of the alternative models in predicting the data.

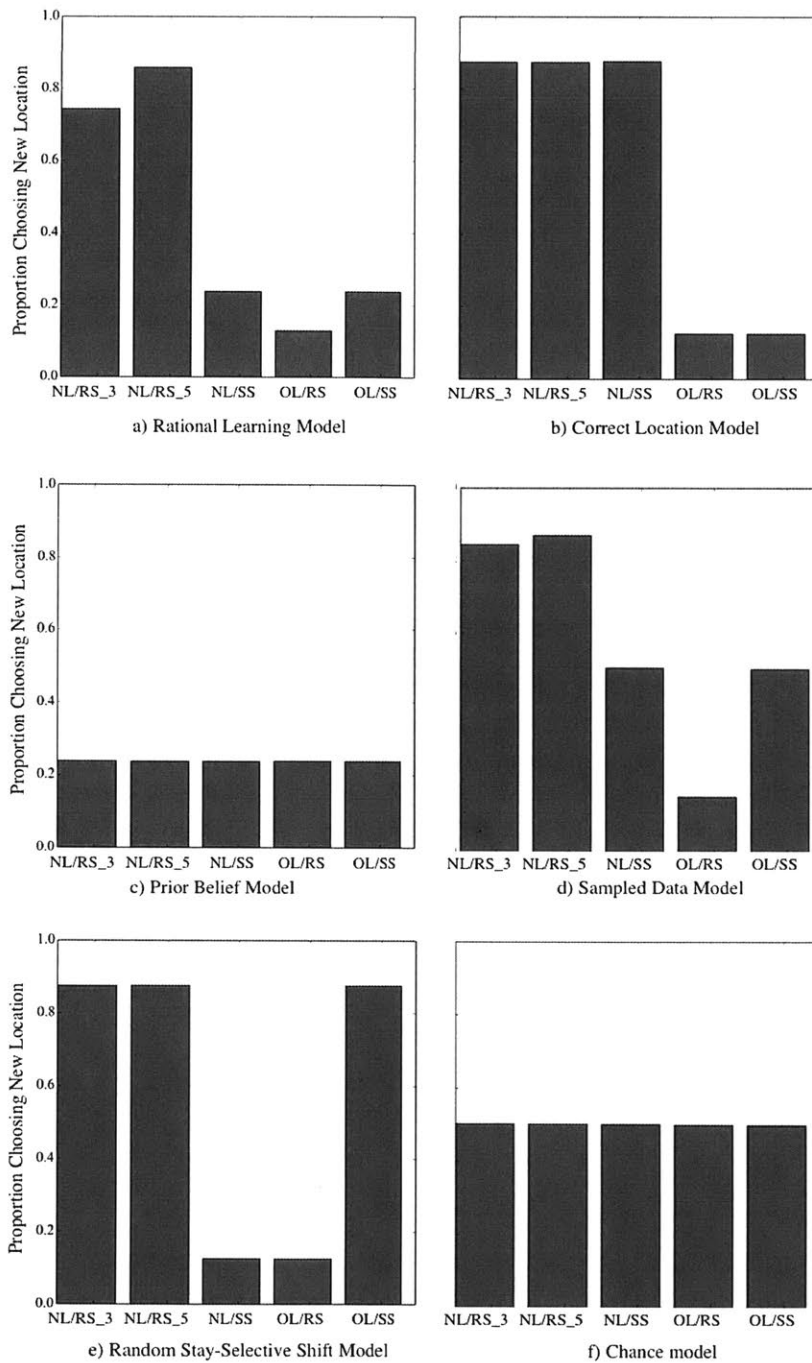


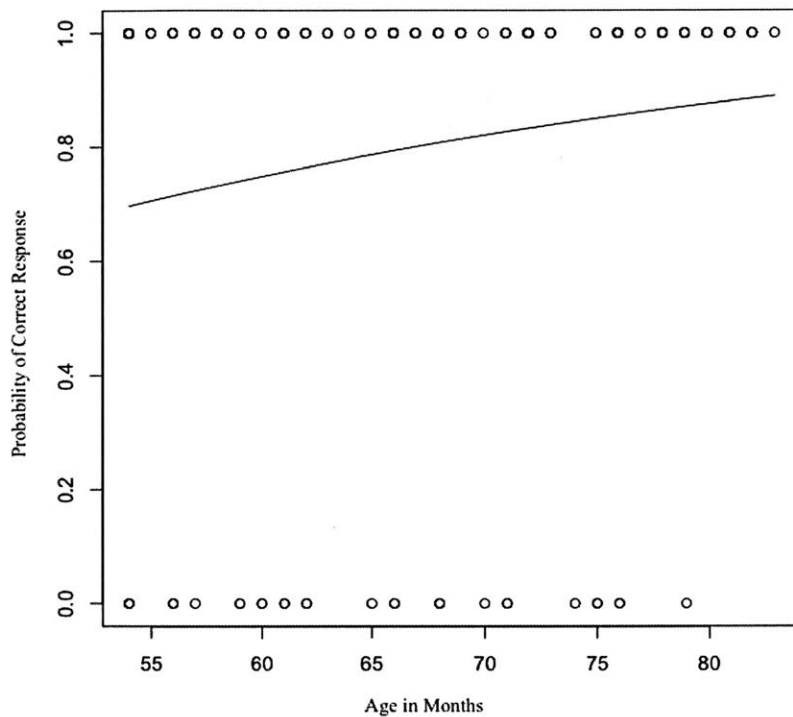
Figure 2.4. Predictions made by the Rational Learning Model for the rational inference model along with the five alternative models (b-f). The Rational Learning Model (a) provides the best fit to the children's responses. (See Figure 2.2 and Table 2.4.)

*Table 1.2* Bayes factor analyses comparing the Rational Learning model with the alternative models.

	Correct Location	Prior Belief	Random-Stay/ Selective-Shift	Sampled Data	Chance
Rational Learning:	33.73 : 1	42.98 : 1	26.32: 1	45.80 : 1	146.20 : 1

Finally, we looked at whether the ability to make the rational inference on behalf of the Frog changed between 4.5 and 6 years. We coded children’s responses as a “1” if they responded with the Old Location in the Old Location/Random Sampling, Old Location/Selective Sampling, and New Location/Selective Sampling conditions and with the New Location in the New Location/Random Sampling conditions and a “0” if they responded otherwise. The logistic regression was marginally significant, suggesting a trend for older children to be more likely to expect others to rationally update their beliefs,  $\beta = 0.043(.024)$ ,  $z = 1.791$ ,  $p = .073$ . See Figure 2.5.





*Figure 2.5.* Children’s responses were coded as a 1 if they were consistent with the expectation of rational learning and a 0 if otherwise. There was a non-significant trend for children’s performance to improve with age.

## 2.4 Discussion

The results of the current study suggest that by four-and-a-half, children not only expect agents to act rationally with respect to their goals (Gergely & Csibra, 2003), they expect other agents to learn rationally from data. To make inferences on behalf of another agent children needed to integrate the agent’s prior beliefs with the evidence the agent observed and the way the evidence was sampled. Children were inclined to believe that the Frog would change his mind only when there was strong evidence against the Frog’s prior belief (i.e., in the New Location/Random Sampling conditions). Children did not expect the Frog to change his mind

when the evidence was consistent with his prior beliefs (Old Location/Random Sampling; New Location/Selective Sampling), or when the evidence may have conflicted with the Frog's prior beliefs but was weak and thus provided little ground for belief revision (Old Location/Selective Sampling).

Although even the youngest children in our sample were able to draw inferences about how a third party would update his beliefs from data, this study provides suggestive evidence that this ability might increase with age. Future research might look both at how children's ability to draw inferences about others' learning changes over development and investigate the origins of this sensitivity earlier in childhood. A basic understanding of how evidence affects others' beliefs (e.g., the understanding that seeing leads to knowing; Onishi & Baillargeon, 2005; Pratt & Bryant 1990; Senju, Southgate, Snape, Leonard, & Csibra, 2011) emerges very early. This knowledge, together with the ability to make predictions about rational action, opens up the possibility that in simpler contexts, even younger children might be able to draw inferences about how third parties might update their beliefs from data. It is also possible that children's representations of the processes that underlie belief revision may support the emergence of broader abilities in interpretive theory of mind (Astington et al., 2002; Carey & Smith, 1993, Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross et al., 2005; Ruffman et al., 1993); future research might investigate the relationship between understanding that evidence conflicts with prior knowledge and understanding that evidence can be ambiguous depending on prior knowledge.

As discussed, children might have made a wide range of other inferences. In particular, they might have assumed the Frog's beliefs would mirror their own; they might have recognized that the Frog's beliefs depended on epistemic access to the location of the box but failed to

recognize that the Frog might update his beliefs based on inferential evidence, or they might have expected the Frog to attend to the sampled evidence but not have expected the Frog to integrate this evidence with his prior beliefs. Yet, children in this study were able to make predictions about what the Frog would think about the location of the Duck box given the evidence, even though they themselves always knew the true location of the Duck box. Moreover, children were able to draw different inferences depending on the ambiguity of the evidence, showing different patterns of responding in the Random and Selective Sampling conditions. This suggests that children can draw inferences that are sensitive both to the distinction between their own and others' prior knowledge, and to the strength of the data that others observe. We believe this finding is broadly consistent with an emerging body of literature suggesting that children make relatively nuanced decisions about when and what to learn from others (Bonawitz, Shafto, Gweon, Goodman, Spelke, & Schulz, 2011; Corriveau & Harris, 2009; Gweon, Pelton, Konopka, & Schulz, 2014; Jaswal, 2010; Jaswal, Croft, Setia, & Cole, 2010; Koenig, Clement, & Harris, 2004; Koenig & Harris, 2005; Stiller, Goodman, & Frank, 2015). Our study extends the literature by suggesting that children also make relatively nuanced decisions about how and when children will expect others to learn.

The current study however, does not indicate how broadly this ability extends, nor does it suggest the conditions under which children might fail to expect others to rationally update their beliefs from data. Here we suggest an account of how children might make normative judgments on behalf of third parties; future research might test the limitations of this account. Also, as discussed, the current study was motivated in part by predictions from an ideal observer model of rational inference. The results are broadly consistent with that account. However, providing a

rigorous test of the quantitative predictions of the rational inference model and alternative accounts remains an important direction for future work.

As adults, we expect other agents to be rational actors not only in terms of the paths they take towards their goals, but also in terms of how they reason about evidence. Here we find that children's developing theory of mind supports the same kinds of inferences. By four and a half, children are able to integrate others' prior knowledge and observed evidence to support predictions about when others will retain their beliefs and when they will change their minds.

## Chapter 3

**This chapter is based on Magid,\* Siegel,\* Tenenbaum, & Schulz, (2017 *CogSci Proceedings*, In prep). Children’s exploratory play quantitatively tracks the discriminability of alternative hypotheses. \*Denotes joint first-authorship.**

Studies suggest that children’s exploratory behavior is sensitive to uncertainty; however, few have approached this phenomenon with sufficient precision to model sensitivity quantitatively. Across four experiments, children ( $N = 96$ , mean age = 5.83 years) were asked to shake a box to identify which of two sets of marbles, differing in numerosity from 1-9, was hidden inside. The contrasts of sets varied in their discriminability indices – the degree to which listeners can distinguish which set is inside the box based on acoustic information generated through exploration. In each experiment, the time children spent exploring varied systematically with the discriminability of the alternative hypotheses they were asked to distinguish, and not the numerosity of the given set they heard. These results show that children represent the uncertainty in their own perceptual discrimination abilities (an ability we refer to as an intuitive psychophysics) and that their exploratory behavior is precisely calibrated to their degree of uncertainty about alternative hypotheses that might explain unobserved causes of perceptual data.

### 3.1 Introduction

Across a variety of domains, children integrate evidence they observe with their prior beliefs to learn (see Gopnik & Wellman, 2012; Schulz, 2012; and Tenenbaum, Kemp, Griffiths, & Goodman, 2011 for reviews). In some cases, the evidence children observe is of their own making: through exploring the physical world, asking questions, and performing interventions on people and things. Studies have shown that children's exploration is guided by rational inferential principles. They selectively explore, for example by testing which blocks will activate a machine, when previously observed evidence violates a strong prior, is confounded, or is ambiguous (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012; Cook, Goodman, & Schulz, 2011; Kidd, Piantadosi, & Aslin, 2012; 2014; Legare, 2012; 2014; Ruggeri & Lombrozo, 2015; Ruggeri, Lombrozo, Griffiths, & Xu, 2016; Schulz & Bonawitz, 2007; Schulz, Standing, & Bonawitz, 2008; Stahl & Feingon, 2015; van Schijndel, Visser, Bers, & Raijmakers, 2015; see Kidd & Hayden, 2015 for review). Thus, children's self-directed information gathering appears to be guided by a sensitivity to their own uncertainty.

To our knowledge, all of the previous studies on young children's tendency to explore more when evidence is confounded, or when exploration is likely to yield new information, have used qualitative measure of uncertainty. Yet untested is the question of whether there are quantitative principles, that are reflective of a more precise process, evident in children's play. The goal of the current work is to understand how children calibrate their exploration in a fine-grained way based on the task difficulty, and therefore their own uncertainty.

In certain conditions, young children seem to use their level of certainty or confidence in planning their behavior, for example in making a decision about how much to wager they will answer a numerical estimation question correctly (e.g., Lyons & Ghetti, 2011; Vo, Li, Kornell,

Pouget, & Cantlon, 2014). While monitoring one's own uncertainty improves with age (Koriat & Shitzer-Reichert, 2002; Lyons & Ghetti, 2010; Sussan & Son, 2007), there is much debate about the metacognitive abilities of young children, in particular the extent to which children's estimations of their own uncertainty are accurate, and whether their behavior reflects these estimations. Here, we investigate children's ability to simulate how difficult a task will be in order to direct their exploration.

In contrast to most contexts that require representing the probability of information gain (e.g., contexts where evidence is more or less surprising or evidence or does not isolate variables) understanding which hypotheses are distinguishable by perceptual data requires that children represent not just something about the world, but also something about themselves: they must know the kinds of discriminations their sensory system can and cannot make. This is the ability we refer to as an "intuitive psychophysics".

In a series of previous experiments, we found that children aged 3-5 years simulate their own perceptual experience to select evidence that would be more informative given a particular problem (Siegel, Magid, Tenenbaum, & Schulz, 2014). When given the option to discriminate more perceptually similar stimuli (the sound of 6 versus 8 marbles being shaken in a box) or more perceptually dissimilar stimuli (the sound of 2 versus 8 marbles being shaken in a box), children preferred the more easily discriminable pair. This ability to accurately determine which of two tasks would be easier is presumably only possible if children have an intuitive theory of how physical objects interact to generate sound and can make choices in accordance with that intuitive theory. These findings demonstrate that children have a sense of what they are more and less likely to be able to do, and they strategize about what interventions on the world would be the most informative given their assessment.

In the current study, we use a novel task, which allows us to perform a quantitative analysis of uncertainty or problem difficulty, which we call the “box shaking” task (Siegel, Magid, Tenenbaum, & Schulz, 2014). Children are shown two sets of marbles varying in numerosity (between 1 and 9 marbles), and then, out of their sight, one of the two marble sets is placed into a box. Children are allowed to shake and manipulate the box to see if they can use the sounds of the marbles to decide which set of marbles is inside. Critically, children might hear exactly the same set of marbles in a box (e.g., 8 marbles) in the context of either an easy discrimination problem (when they know that either 8 or 2 marbles were placed in the box) or a difficult discrimination problem (e.g., when they know that either 8 or 6 marbles were placed in the box). In the objective language of psychophysics, the degree of uncertainty in this task can be represented by a discriminability index for the two distributions of sounds generated by shaking boxes with the two different sets of marbles. Analogous to the representation of visual number the discriminability of two different sets of marbles depends both on their statistical separation and the spread of the distributions (Halberda, Mazocco, & Feigenson, 2008).

Here, however, children hear only a single set of marbles in the box. They have no way of judging the objective discriminability of two set sizes. Rather, the degree of discriminability depends on children’s ability to simulate the contrast between the sounds they hear and the sounds they would have heard had the alternative set of marbles been in the box. The current studies look at whether children’s self-directed exploration is guided in a precise way by this intuitive sense of the perceptual discriminability for a given pair of marble set sizes. If children hear the sound of 8 marbles and think that this would be relatively easy to distinguish from the alternative possibility (e.g., 2 marbles), they might shake and listen only briefly. However, if children hear the sound of 8 marbles and think it would be hard to distinguish this from the



alternative (e.g., 6), they might spend more time shaking the box. Thus we can investigate the precise relationship between the discriminability of sets and exploration time, in experiments that are designed to investigate whether children's exploration of the box depends on what children actually hear, or on the discriminability of what children imagine they could hear under the two alternatives.

Unlike other ways one could modulate task difficulty, this task provides consistent indexing of a level of "problem" difficulty. The difficulty of the marbles task is clearly greater for large number differences (e.g. 9 or 1) compared with small differences (e.g. 4 or 3). To quantify objective task difficulty, we use psychophysical data obtained from a different task with adult participants (Siegel, Tenenbaum, & McDermott, *submitted*). This experiment measured participants' ability to estimate the number of marbles in a box, from the sound made while the box was shaken; the number of marbles ranged from 1 to 9. From the distribution of participant responses, we computed the discriminability index ( $d'$ ) for each contrast presented in this work. We modeled children's intuitions about task difficulty as proportional to this objective measure derived from adult task performance. This assumes that children's percepts are similar to adults, and that children can mentally simulate the sounds of shaking different numbers of marbles, not necessarily with high fidelity but just well enough to make relative judgments of how discriminable two numerosities might be. Across these experiments we present children with two sets of marbles that range in their  $d'$  from high to low discriminability.

Pilot data suggested that testing children on more than four contrasts in a single session was impractical. Because we wanted to test children on a range of discriminability contrasts, we ran three separate experiments consisting of four contrasts each. The experiments differed only in the contrasts presented, except for Experiment 3, where the actual contents of the boxes were

fixed such that participants heard the same number of marbles inside the box (8 and 3) during the two easy discrimination trial (e.g., 8 or 2 and 3 or 9) as the hard discrimination trials (8 or 6 and 3 or 4). We predicted that despite clear differences in the sensorimotor experience of shaking more (8) or fewer (3) marbles in a box, children's exploration would depend not on the actual contents of the box but on the difficulty of discriminating the actual contents from the alternative hypotheses under consideration. Experiment 4 was a replication, and was pre-registered on the Open Science Framework.

We analyze children's exploration within the context of the four trials within each experiment. However, across the four experiments, there are a total of 16 contrasts. Thus we also pre-registered a set of analyses to look at all 16 contrasts and ask in a fine-grained way whether graded differences in children's exploration are systematically related to the contrasts' discriminability. We predicted a negative relationship between discriminability and exploration time (i.e., children should explore more when the contrasts are relatively less discriminable). We were primarily interested in children's exploration, rather than the accuracy of children's guesses about the contents of the box. Nonetheless, and even given compensatory exploration, one would expect children to be more accurate on easier discrimination problems than on difficult ones. For brevity, we report accuracy results in the Joint Analysis at the end of all four experiments and in Fig. 2b.

To our knowledge, no previous work has looked at whether children represent their own ability to discriminate perceptual stimuli, let alone whether this kind of representation informs children's exploration. In the absence of any clear developmental predictions, we sampled a relatively broad age range (from age 4 to 8) centered on 5- and 6-year-olds consistent with previous work on children's exploratory play in response to uncertainty (Bonawitz, van

Schjndel, Friel, & Schulz, 2012; Cook, Goodman, & Schulz, 2011; Legare, 2012; Schulz, Standing & Bonawitz, 2008).

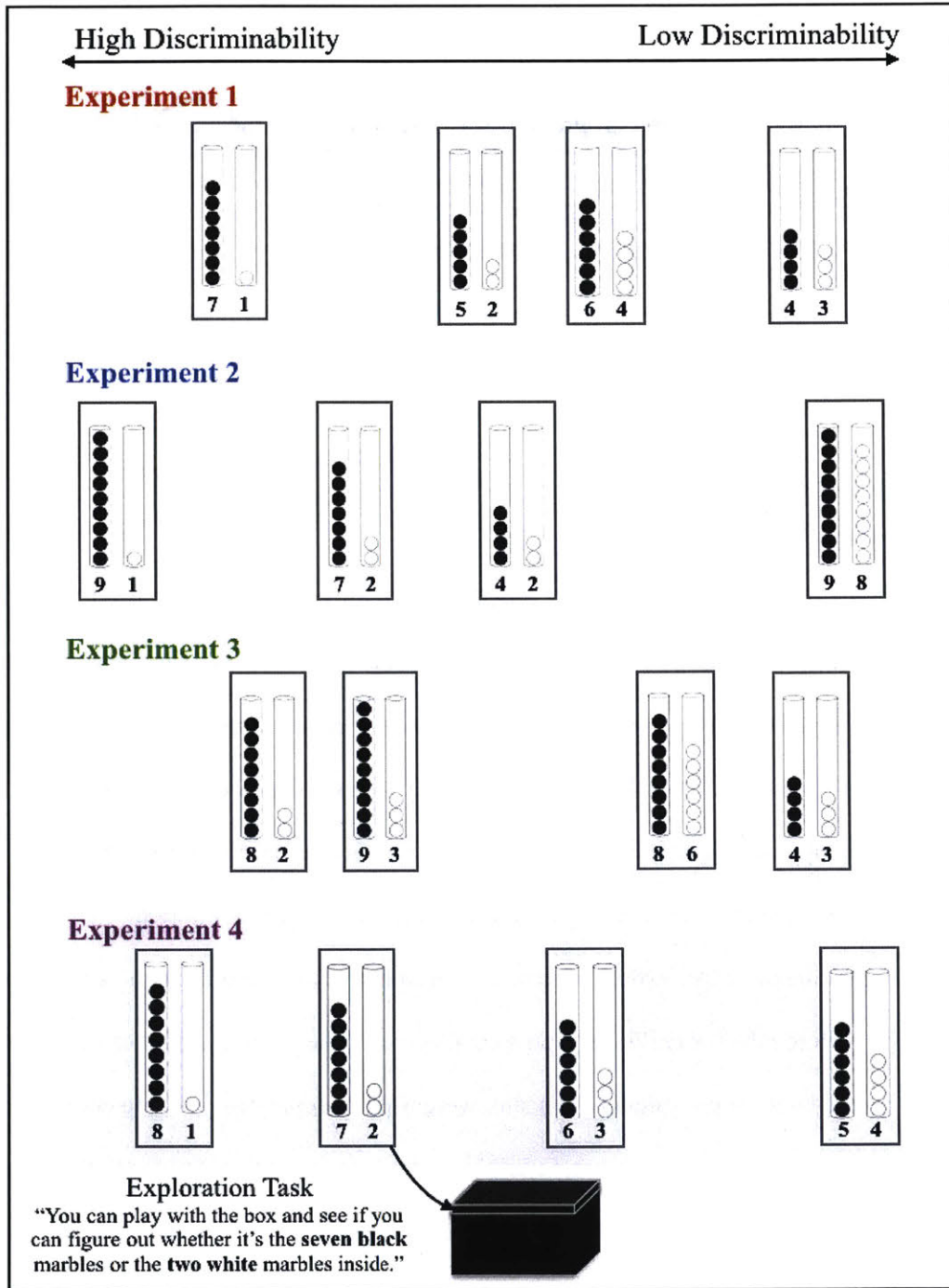


Figure 3.1. Schematic of task in Experiments 1-4. Placement of contrasts represents relative discriminability. Contrasts, as well as sets within contrasts, were presented in a randomized order within each experiment.

## **3.2 Experiment 1**

### **3.2.1. Method**

Participants were recruited from an urban children's museum. Based on previous experiments using the same method, we estimated the effect size ( $f$ ) for a single experiment was .29. We used the power calculation program, G\*Power, to calculate the planned sample size of for this experiment using an effect size = .29, alpha = .05, and power = .80. The projected sample size using these values is 24 participants. Twenty-four children (mean age = 5.72 years; range 4.10-8.18 years) were included in the final sample. One additional child was excluded because they did not explore before providing a response on one or more trials (see Procedure for details).

#### **3.2.1.1 Materials**

A box covered with black electrical tape (18 cm x 16 cm x 12 cm) was used. Four objects were used in the practice trials: a plastic duck, a star-shaped pillow, a flat glass bead, and a cotton ball. For the test trials, standard-size glass marbles in eight colors and eight translucent cylindrical tubes were used. The tubes were pre-loaded with the appropriate number of marbles and sealed at the top; although children believed the tubes of marbles were poured into the box, marbles were in fact added quietly by hand to ensure that no sound was emitted.

A large cardboard screen (80 x 60 cm) was used both as an occluder and as an answer board with six Velcro tabs for children to provide their responses. Laminated pictures with Velcro tabs on the back, approximately to scale, were used to depict the possible contents of the box for both the practice trials and the test trials. A button was used to activate "hiding music" (the Jeopardy theme song) from a portable speaker, to mask the sound of marbles being placed into the hiding box.

### 3.2.1.1 Procedure

Children were introduced to the task as a guessing game in which their goal was to figure out what was hidden in the box. Two practice trials were used to teach children: 1) there were two possibilities for what could be hidden inside the box, 2) that these would be represented by the laminated pictures; 3) that children could not open the box but could shake the box or explore it in any other way they liked; 4) that they could make a guess by affixing one of the two pictures to the answer board and 5) that they would not get feedback on every trial but would get feedback at the end of a set of trials (i.e., on the second of the two practice trials and on the last experimental trial).

The experimenter explained the practice task by introducing one set of practice objects (order counterbalanced). She said, “We’re going to play a guessing game. See these two toys? Do you want to feel them? I’m going to hide one of these toys inside the hiding box. Then you’re going to shake it and listen and see if you can figure out what’s inside. Remember, I’m going to hide either the (pillow or duck; bead or cottonball) and you’re going to figure out what’s inside without opening the box!” Then the experimenter set up the answer board/occluding screen and placed the pictures of the two possible contents of the box on two Velcro tabs on the bottom of the screen facing the child. She pointed to each of the pictures in turn while reminding the child “I’m going to hide either the (pillow or duck; bead or cottonball) inside the box.” The experimenter then moved behind the occluding screen and placed one of the two objects into the box out of the child’s line of sight. To mask any acoustic cues made by the experimenter, the “hiding music” was played while the experimenter loaded the box with one set of marbles (counterbalanced). The experimenter reminded the child of what could be inside of the box and indicated the location on the screen where the child could point the picture corresponding to his/her guess, and

then handed the child the box. Children were allowed to shake or explore the box in any way they liked for as long as they liked until they made a verbal guess or touched a picture on the board.

Table 3.1. Contrasts used across Experiments 1-4, ordered from most discriminable to least discriminable. Trial order was counterbalanced, as was the order of introduction of the tubes of marbles, and the actual hidden contents of the box (e.g., whether 1 or 7 marbles were hidden inside on the 7 vs. 1 trial) except in Experiment 3, per features of the design.

Experiment	Contrast 1		Contrast 2		Contrast 3		Contrast 4	
	Sets	$d'$	Sets	$d'$	Sets	$d'$	Sets	$d'$
Exp. 1	7 v 1	3.01	5 v 2	1.90	6 v 4	0.67	4 v 3	0.58
Exp. 2	9 v 1	4.71	7 v 2	2.40	4 v 2	1.69	9 v 8	0.45
Exp. 3	8 v 2	2.85	9 v 3	2.54	8 v 6	0.84	4 v 3	0.58
Exp. 4	8 v 1	3.49	7 v 2	2.40	6 v 3	1.25	5 v 4	0.24

Children did not receive any feedback on their guesses on the first practice trial. After the second practice trial, children were told that they were done with the first part of the game. The experimenter revealed the contents of the second box, and the children received a sticker for guessing correctly. (A minority of children guessed incorrectly on the second practice trial but were told they received the sticker for guessing correctly on the first box.)

Test trials were administered in the same manner as the practice trials, except that test trials consisted of contrasts of sets of marbles. The experimenter began each test trial by introducing two tubes of marbles. The contents of each tube differed from each other in color and each tube had a different number of marbles inside. (See Figure 3.1.) The experimenter asked the child to count the number of marbles in each tube. The contrasts used for each

Experiment are displayed in Table 3.1. Trial order was counterbalanced, as was the order of introduction of the tubes of marbles, and the actual hidden contents of the box (e.g., whether 1 or 7 marbles were hidden inside on the 7 vs. 1 trial). As in the practice trials, children were allowed to shake or manipulate the box in any way they liked for as long as they liked until they made a guess about the contents of the box.

### 3.2.2 Results

All study sessions were filmed for offline coding from videotape using the software program VCode. Although children occasionally paused or tapped the box with their fingers, children spent virtually all their time shaking the box. Thus a coder blind to contrasts coded cumulative exploration simply as the time starting from the time when children first made contact with the box until they provided a verbal response by saying either the number or color of one of the sets of marbles, or touched one of the pictures, whichever came first.

To normalize for individual differences in children's exploratory behavior, we computed the time each child spent exploring on each trial as a proportion of the child's total playtime across all four trials, and multiplied this proportion the number of trials in the experiment:  $\text{Trial } i / \sum \text{trials } i - k * k$ . In the current study,  $k=4$ . Thus, a proportion less than 1 represents less playtime than would be expected if length of exploration were deployed at chance, and a proportion greater than 1 represents more playtime that would be expected at chance. Although we transformed playtime on each trial into a proportion of the child's total playtime to control for individual differences, the results of all the model comparisons hold when using untransformed playtime reported in seconds.

We estimated the difficulty of each contrast from data collected from adult participants in a related task (Siegel et al., *in prep*). Adults were asked to estimate, from sound recordings, the

number of marbles that were shaken in a box. We calculated the mean and standard deviation of participant responses for each of 1-9 marbles, and calculated the discriminability index

$$d'(i, j) = \frac{\mu_i^2 - \mu_j^2}{\sqrt{\frac{1}{2}(\sigma_i^2 + \sigma_j^2)}}$$

for each  $i, j$  numerosity contrast. Using the R programming language, the data were submitted to linear mixed-effects regression models, with subject as a random effect. An example of our model specification (with discriminability as a predictor variable) in the common lme4 syntax is as follows: `Playtime ~ Discriminability + (1 | subject)`. We ran four models with the following predictors: 1) Model 1: Discriminability; 2) Trial order; 3) Discriminability + Trial order; 4) Discriminability + Trial order + Number of marbles inside the box), see Table S1. To assess which of these variables predicted significant variance, we ran three model comparisons using the ANOVA function. This allowed us to obtain p-values from likelihood ratio tests of the full model with the effect in question against the model without the effect in question.<sup>3</sup> Comparing Models 1 and 3, we found that trial order also had a significant effect on exploration time, where children on average explored for less time as the task progressed,  $\chi^2(1) = 5.944, p < .015$ . Comparing Models 2 and 3, we found that discriminability affected children's exploration time, where the less discriminable the contrast, the more children explored,  $\chi^2(1) = 17.076, p < .0001$ . This model comparison shows that discriminability explains variance above and beyond the effect of trial order. Comparing Models 3 and 4, we found no effect of the number of marbles inside the box, suggesting children's exploration time was not affected by what they actually heard, but rather by the discriminability of the two sets,  $\chi^2(1)=0.2635, p=.608$ . In addition, we bootstrapped 95% confidence intervals of mean exploration time to assess overlap across the

---

<sup>3</sup> A detailed description of the analyses are available on the Open Science Framework.



four contrasts. We found that the most discriminable contrast's confidence interval did not overlap with the intervals of the two least discriminable contrasts. The second most discriminable contrast overlapped with the other three contrasts (See Figure 3.2a). These results provide preliminary evidence that children's exploration is well-calibrated to the discriminability of the hypotheses under consideration.

### **3.3 Experiment 2**

Experiment 2 was identical to Experiment 1 except for the set of contrasts used, see Table 3.1. Twenty-four children ( $m = 5.94$  years; range 4.06-7.96 years) participated.

#### **3.3.1 Results**

Data were coded as in Experiment 1. Again, to normalize for individual differences in children's exploratory behavior, we computed the time each child spent exploring on each trial as a proportion of the child's total playtime across all four trials. The same models were used as in Experiment 1, see Table 3.S2. Unlike in Experiment 1, did not find that trial order had a significant effect on exploration time,  $\chi^2(1) = 0.106, p = .7449$ . Our key prediction, that discriminability predicts children's exploration time replicated in Experiment 2  $\chi^2(1) = 19.114, p < .0001$ . Once again, we found no effect of the number of marbles inside the box,  $\chi^2(1) = 0.286, p = .5926$ . Comparing bootstrapped 95% confidence intervals of mean playtime, we found that the most discriminable contrast's confidence interval did not overlap with the intervals of the two least discriminable contrasts. The second most discriminable contrast overlapped with the other three contrasts (See Figure 3.2a). These results again suggest that children's exploration is closely matched to the difficulty of the discrimination problem.

### **3.4 Experiment 3**

The same procedure as in the preceding experiments was used except for the contrasts (from most to least discriminable 8 vs. 2; 3 vs. 9; 8 vs. 6; and 3 vs. 4), and rather than counterbalancing the number of marbles in the box, there were always either 8 or 3 marbles hidden in the box. This provides a strong test of whether children's exploration is driven primarily by the salience or ancillary sensory properties of the stimuli, then children might spend more time exploring the box when it contained more versus fewer marbles. If instead, children's exploration tracks not the actual contents of the box but the discriminability of the actual contents from the alternatives, then children should spend proportionately less time exploring on the two easy contrasts (8 vs. 2 and 3 vs. 9) than the two hard ones (8 vs. 6 and 3 vs. 4). Participants were recruited from an urban children's museum. Twenty-four children ( $m = 5.74$  years; range 4.07-7.68 years) were included in the final sample. Three additional children were excluded because of family interference ( $n = 1$ ) and issues with video recordings ( $n = 2$ ).

### **3.4.1 Results**

Data were coded as in previous experiments. Again, to normalize for individual differences in children's exploratory behavior, we computed the time each child spent exploring on each trial as a proportion of the child's total playtime across all four trials. The same models were used, see Table 3.S3. As in Experiment 1, we found that trial order also had a significant effect on exploration time,  $\chi^2(1) = 13.800, p = .0002$ . As in Experiments 1 and 2, we found that discriminability was a significant predictor of children's exploration time,  $\chi^2(1) = 12.788, p = .0003$ . Experiment 3 provided a strong test of whether the number of marbles heard inside the

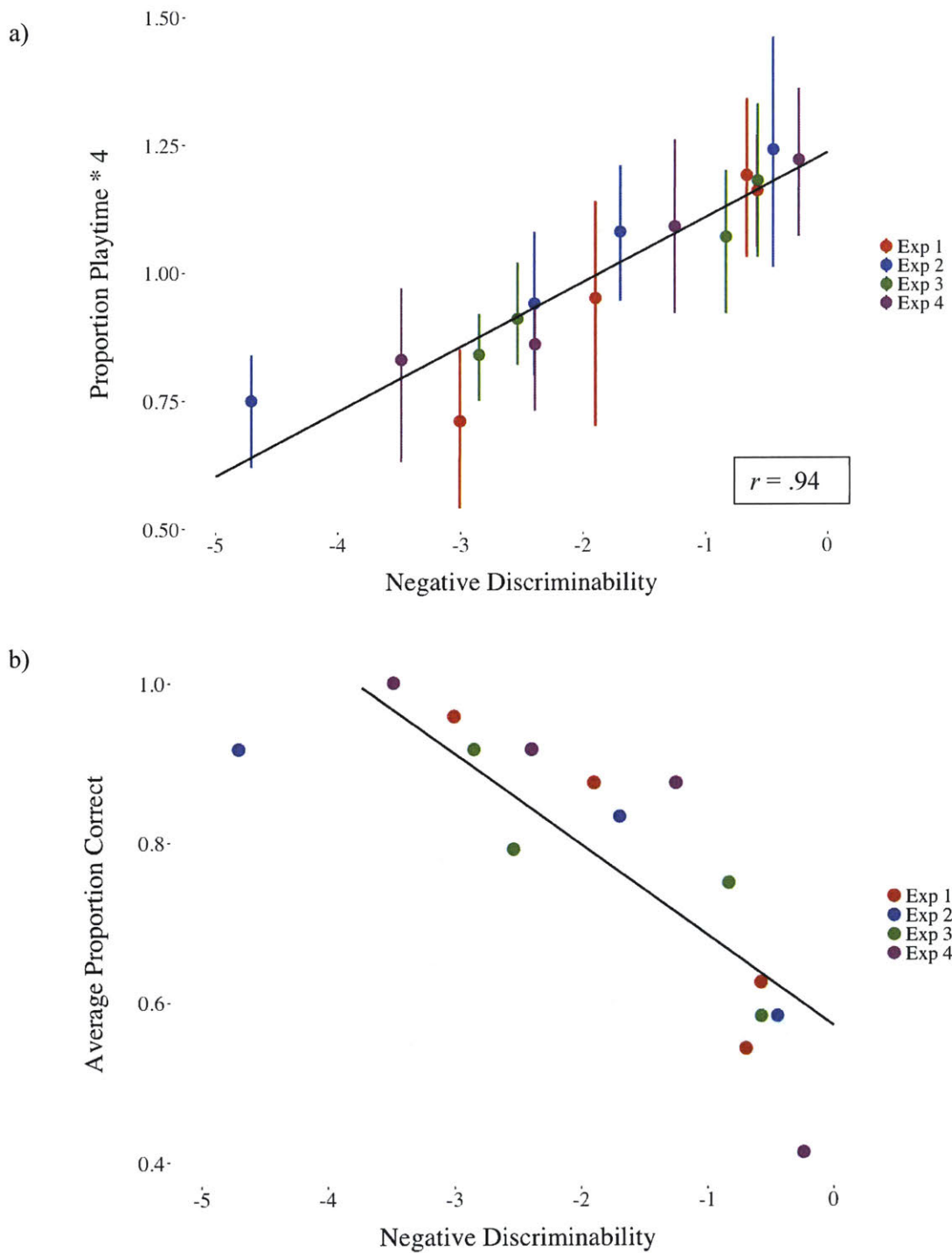


Figure 2. (A, top) Children’s proportional exploration time as a function of the negative discriminability of each contrast across Experiments 1-4, showing bootstrapped 95% confidence intervals. (B, bottom) Children’s accuracy on the same trials, as a function of their (negative) discriminability.

box affects exploration time since two hard discrimination trials (8 vs. 6 and 3 vs. 4) and two easy discrimination contrasts (8 vs. 2 and 3 vs. 9), were matched for the number of marbles inside the box. We found no effect of the number of marbles inside the box,  $\chi^2(1)=1.213$ ,  $p=.271$ . In addition, we bootstrapped 95% confidence intervals of mean exploration time to assess overlap across the four contrasts. We found that the most discriminable contrast's confidence interval did not overlap with the intervals of the two least discriminable contrasts. The second most discriminable contrast overlapped with the other three contrasts (See Figure 3.2a).

### **3.5 Experiment 4**

To establish the robustness of the pattern of results in Experiments 1-3, we pre-registered all methods and analyses on the Open Science Framework for Experiment 4 and the Joint Analysis. The same procedure as in the preceding experiments was used, see Table 1 for contrasts. Participants were recruited from an urban children's museum. Twenty-four children ( $m= 5.91$  years; range 4.21-7.68 years) were included in the final sample. One additional child was excluded due to attention issues.

#### **3.5.1 Results**

Data were coded and normalized as in previous experiments, and the same models were used, see Table S4. Unlike in Experiments 1 and 3, but similar to Experiment 2, trial order had no effect on exploration time,  $\chi^2(1) = 0.0172$ ,  $p = .8956$ . As in Experiments 1 and 2, we found that discriminability was a significant predictor of children's exploration time,  $\chi^2(1) = 15.999$ ,  $p < .0001$ . We found no effect of the number of marbles inside the box,  $\chi^2(1)=0.208$ ,  $p=.6477$ . In addition, we bootstrapped 95% confidence intervals of mean exploration time to assess overlap

across the four contrasts. We found that the most discriminable contrast's confidence interval did not overlap with the intervals of the two least discriminable contrasts. The second most discriminable contrast overlapped with the other three contrasts (See Figure 3.2a).

### 3.6 Joint Analysis

Now, we turn to the analysis of the 16 contrasts across all four experiments, all of which were pre-registered on the Open Science Framework. First, given the broad age range tested, we looked to see if there was a correlation between children's age and how rationally they explored or how accurately they performed. As a simple measure of rational exploration, we looked at children's tendency to explore more on the hardest contrast than on the easiest one; there was no relationship with age,  $\beta=-0.041(.21)$ ,  $z = -0.197$ ,  $p = .822$ . Accuracy was weakly predicted by children's chronological age,  $\beta=0.031(.02)$ ,  $t = 1.862$ ,  $p = .068$ .

However, the critical question is whether there are graded differences in the extent of children's exploration across the 16 contrasts and the contrast's discriminability. To assess this, we constructed the same models used to analyze exploration time in Experiments 1-4, but added an additional random effect for Experiment. The discriminability of the contrast, derived from adults' psychophysics judgments (Siegel et al., *submitted*) predicted children's exploration time across the full range of contrasts,  $\chi^2(1) = 62.555$ ,  $p < .0001$  (Fig. 2a); we found a strong correlation,  $r > .939$ . In addition, we found that trial order had a significant effect on exploration time,  $\chi^2(1) = 8.750$ ,  $p = .003$ , suggesting that as children became more familiar with the task, they explored for less time. Importantly, there was no overall effect of the number of marbles that children heard,  $\chi^2(1)=0.366$ ,  $p=.545$ . In a remarkably fine-grained way, children's exploration of the box tracked the difficulty of the contrast they were trying to discriminate. See Table S5 for linear mixed effects models. As expected, there was also a correlation (Fig. 2b)

between the (negative) discriminability of a given contrast and children's accuracy in identifying the correct number of marbles in the box,  $\beta = -0.897(.14)$ ,  $z = -6.221$ ,  $p < .0001$ .

### 3.7 Discussion

Across four experiments encompassing 16 numerosity contrasts, we evaluated the hypothesis that young children's exploratory play is modulated by the difficulty of the task. We found a striking correspondence between the amount of time that children play and an independent measure of contrast discriminability, even though we did not present children with data from both alternatives to be discriminated. This suggests that children's sense of task difficulty must have come from an internal computation, a sense of the uncertainty in their perceptual data (an "intuitive psychophysics") evaluated relative to the alternative hypotheses they were asked to evaluate. Our findings suggest that some forms of metacognition related to uncertainty monitoring, often considered to be late-developing (Lipowski, Merriman, & Dunlosky, 2013), may be available to young children after all – at least in familiar task domains and perceptual contexts.

The current study does not ask whether children know, a priori, that they will need more information to answer less discriminable problems. Some evidence suggest that they can represent the dimension of problem difficulty before exploring at all (Siegel et al., 2014), but the future work should explore whether the precise relationship between discriminability and the degree of planned information seeking could be observed prior to exploration.

What is most informative, or how much information is sufficient to understand a problem, is a function of both the problem to be solved *and* the learner's capabilities. To be effective, one must flexibly reason jointly about the self in the current moment and the level of difficulty of the task one is trying to solve. Future work might investigate the relationship

between an individual's abilities (in perception or reasoning) and the degree of information that individual needs to solve a problem. For example, one could establish whether individual children's accuracy on a number identification task like the one used in Siegel & McDermott (in prep) predicts the amount of information that particular child requires to discriminate between alternative hypotheses in the box shaking task used in the current study.

Just as infants demonstrate an understanding of their physical capabilities when deciding whether they can cross a bridge (Berger & Adolph, 2003; Berger, Adolph & Kavookjian, 2010), preschoolers demonstrate sensitivity to their perceptual capabilities by gathering more information when freely exploring to solve problems of varying difficulty levels.

## Chapter 4

**This chapter is based on Magid & Schulz (2015 *CogSci Proceedings*, In prep). When social comparison spurs motivation: Evidence from 3- to 5-year-olds' persistence.**

Comparing oneself to peers affects adults' self-evaluations, performance, and motivation across a variety of circumstances. While most previous research finds that children younger than 7 are immune to social comparison information, the extent to which young children's self-directed learning is influenced by observing the performance of peers is currently unknown. The current study asks whether children's ( $N = 220$ , mean = 4.46 years, range: 3.00-5.97 years) persistence is influenced by social comparison information. We investigated this question in two domains relevant to early education: counting and naming shapes. The results from the current study suggest that children as young as 3 capitalize on social comparison to calibrate effort, persisting more when they believe peers have performed better, and less when they believe the opposite is true. Of note, social comparison evidence only matters when the task is in the child's zone of proximal development. The inferences children make from others affect their behavior in a way that could have cascading effects on opportunities for learning in early childhood.



## **4.1 Introduction**

Many readers are well acquainted with the proverb by educator Thomas H. Palmer: “If at first you don’t succeed, try, try, try again.” Rather than trying hard on everything, though, we are selective about how we spend our time. Armed with rich representations of our current knowledge, abilities, and traits, we selectively allocate time and effort across different activities, spending more time learning things we have not yet mastered (Metcalf & Kornell, 2005; Nelson, Dunlosky, Graf, & Narens, 2004). How do we come to have these representations? One particularly salient and ubiquitous source of information that influences adults’ self-evaluations and effort allocation comes from social comparison: interpreting one’s own performance relative to others (Festinger, 1954; Morse & Gergen, 1970; see also Avila, Chiviawowsky, Wulf, & Lewthwaite, 2012; Lewthwaite & Wulf, 2010; Swallow & Kuiper, 1992; Wheeler & Miyake, 1992).

### **4.1.1 The effects of social comparison**

Social comparison information can have a marked influence on people’s self-assessments and behavior. Any given response to social comparison is shaped by a number of factors, including: the direction of the comparison, the learner’s own goals or mood, and the domain of comparison. When adults observe upwards social comparison (observing better performing individuals), they often respond by expressing more negative self-evaluations, and when they receive downwards social comparison (observing poorer performing individuals), they express more positive self-evaluations (Morse & Gergen, 1970). However, when an adult’s goal is to reach a certain threshold of performance rather than do as well as possible, they are less affected by social comparison information (Schwartz et al., 2002). In addition, the domain of the task

influences whether the comparison bolsters or chips away at adults' self-evaluations. If adults believe a task requires a great deal of innate talent, such as producing art or solving mathematical equations (e.g., Patterson, Kravchenko, Chen-Bouck, & Kelley, 2105), seeing other people outperform them is likely to have a demotivating effect and reduce the effort they are willing to expend on the task. However, in domains where improvement is achievable through effort, observing others perform well leads people to try harder in order to meet a higher standard (e.g., Alcott, 2011; Buunk & Brenninkmeijer, 2001; Buunk, Ybema, Gibbons, & Ipenburg, 2001). For example, learning that others in one's neighborhood have conserved more energy in the past month, something that doesn't require inherent talent, leads individuals to conserve 5-20% more energy themselves (Allcott, 2011). In this way, social comparison causes people to adjust not only how they feel about themselves in a particular domain, but also affects how they make decisions and allocate resources. Thus, while adults are influenced by observing peers' performance and the plausibility of improvement and relevance to one's current goals play a key role in determining how someone will respond (Cohn, 2004).

Given that social comparison information can be motivating or discouraging, in the current study, we ask whether even very young children use information about how their peers perform to calibrate their effort. We investigate the effects of social comparison on persistence behavior in children who have yet to enter traditional schooling environments, since explicit comments regarding children and their peers are frequently highlighted by adults in academic settings (Eccles, Midgley, & Addler, 1984). By investigating responses to social comparison in this age group, we can address questions regarding how fundamental this type of peer effect is to learning.

#### **4.1.2 Accounts for children's sensitivity to social comparisons**

Most previous studies suggest that children are insensitive to social comparison information, whereas a few capture sensitivity under certain circumstances. By and large, children younger than age 7 appear to be unaffected by social comparison information when queried about how they feel about themselves in a particular domain, reporting equivalently high self-evaluations both after relative failure and after relative success (Boggiano, Feldman, & Loebel, 1980; Boggiano & Ruble, 1979; Lapan & Boseovski, 2017; Pomerantz, Ruble, Frey, & Greulich, 1995; Ruble, Boggiano, Feldman, & Loebel, 1980; Ruble, Eisenberg & Higgins, 1994). Yet a handful of studies have documented children's use of social comparisons. By age 3, children remark on their own performance relative to others saying things like, "I can run faster than you" (Frey & Ruble, 1985; Mostache & Bragonier, 1981). Children of this age also express distress after doing worse than another (Stipek, Recchia, & McClintic, 1992) and children ages 5-6 years been shown to incorporate social comparison performance feedback into their self-representations in certain circumstances (Butler 1998; Morris & Nemcek, 1982),

There are two broad possible interpretations that can account for these data, and thus how preschoolers might respond to social comparison information. One interpretation of the above findings is that young children do not represent social comparison information as evidence that can inform self-evaluations, perhaps because they do not interpret others' relative performance as relevant to their own. However, in light of the fact that children mention relative performance information, a second possibility may be more explanatory: children may understand upwards and downwards social comparison evidence, but not make use of that information when evaluating themselves (Gentile, Twenge, & Campbell, 2010; Ruble et al., 1994). In this way,

children might be processing social comparison evidence in some forms, without updating their beliefs about the self.

Why might children *not* use social comparison evidence when making explicit self-evaluations? One hypothesis is that because young children hold mostly positive self-concepts from an early age, and self-congratulatory behavior is quite common in kindergartners (Parsons & Ruble, 1977; Stipek & Hoffman, 1980), the beliefs that one is capable of things beyond one's current abilities may serve as a buffer against negative relative comparisons. Indeed, this kind of over-optimism may be an appropriate strategy given children's rapidly developing and changing abilities (Lockhart, Chang & Story, 2002; see also Boseovski & Lee, 2006; Droege & Stipek, 1993), serving as a protective mechanism that allows children to maintain their motivation in the face of challenges when many tasks attempted for the first time will end in failure but can be achieved eventually (Bjorklund & Green, 1992).

When a peer of an opposite gender is shown to perform better than the child at a speeded drawing task, children as young as 4 report feeling less competent at a task (Rhodes & Brickman, 2008). While this finding is, at first glance, seemingly at odds with children's bias to predict positive outcomes and report positive evaluations social comparison evidence is sufficient to override children's over-optimism in cases when highly salient and invariant traits, such as social category membership, are invoked. Therefore, the promise of future improvement may be important for young children, as it is for adults (e.g., Buunk & Brenninkmeijer, 2001).

Given that children notice differences between themselves and others (Mostache & Bragonier, 1981) and adjust their self-evaluations when these differences are less likely to change (Rhodes & Brickman, 2008), they may very well use social comparison information to guide their self-directed learning. Below we discuss a proposal that is sufficiently nuanced to

account for children's use and apparent lack of use of social comparison information documented in the previous literature, and test some of the proposal's implications in the current study.

Given the right opportunity, we predict that preschoolers *will* use social comparison evidence to calibrate their level of effort when they believe they can improve on the task for which they have evidence about their peers, namely the task is in their zone of proximal development (Vygotsky, 1978), and when the domain is relevant to their learning goals (Bandura 1981; 1986; Pepitone, 1980; Pomerantz, Ruble et al., 1995). Specifically, if an individual is interested in improving in a particular domain, for example, learning how to climb a tall structure on the playground or improve one's accent in a foreign language, they might engage in social comparison for the purpose of figuring out how those who are better have achieved what they have achieved (Lockwood, Jordan, & Kunda, 2002; Taylor, Neter, & Wayment, 1995). Thus, when people engage in upward comparisons for tasks in their zone of proximal development, watching better-performing peers in areas where they have an attainable learning goal will increase motivation, as children are generally interested in gaining new skills. However, when individuals are not interested in improving in an area or have mastered the task they are observing their peers perform at, social comparison information is likely to be much irrelevant (Cohn, 2004). This account allows for children's explicit self-evaluations to remain constant (due to their optimism) even if they receive evidence that others performed better.

The current study tasks a novel approach to ask whether young children's effort allocation is affected by social comparison (in cases that do *not* invoke category membership) by measuring persistence on a task rather than eliciting self-evaluations after observing downwards and upwards social comparison, as well as at baseline. For children under the age of 6, observing

better-performing peers might provide evidence about what is possible given enough effort whereas for older children, observing better-performing peers might impact self-evaluations of dimensions like aptitude or ability.

### 4.1.3 Current study

For the experiments in this study, we used the following tasks: counting and comparing apparently endless pairs of sets (which is difficult but possible for 4- and 5-year-olds), and identifying and sorting endless pairs of shapes (which is difficult but possible for 3-year-olds and relatively trivial for 4- and 5-year-olds). In this way, we are equipped to measure children's responses to social comparison information for domains that are both within their zone of proximal development, as well as in a domain outside this zone.

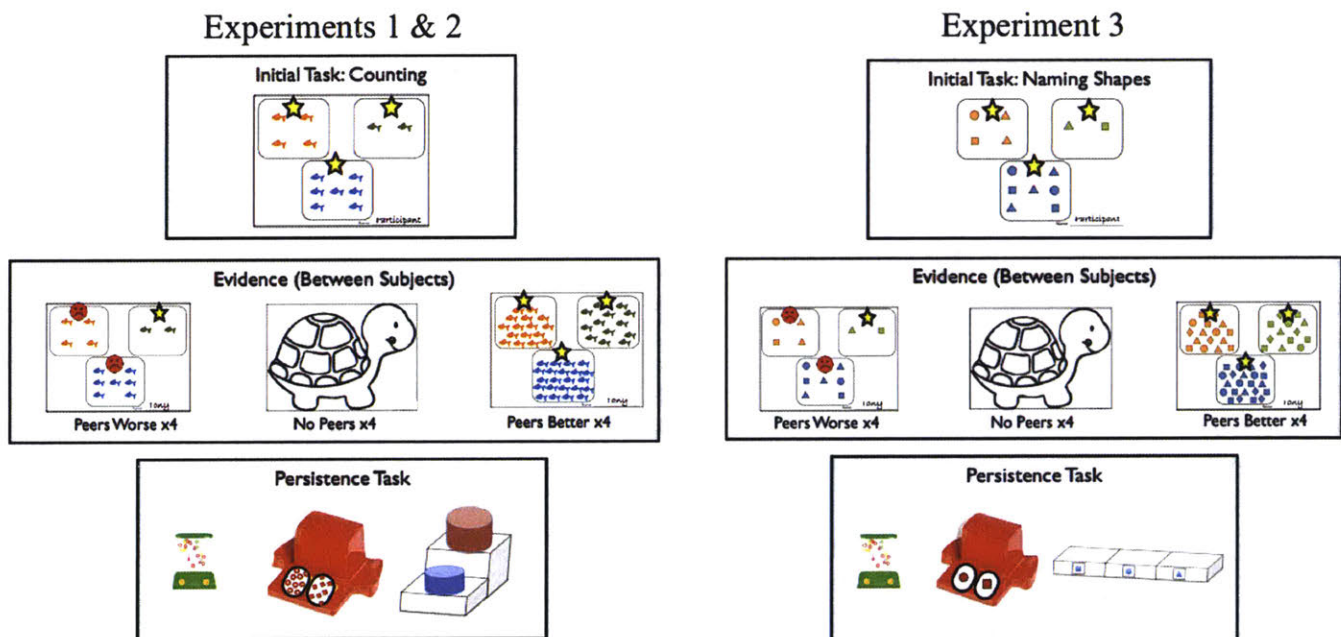


Figure 4.1. Design of experiments showing initial tasks (writing letters, counting, and naming shapes), an example of evidence shown in each condition (Peers Worse, Peers Irrelevant/No Peers, and Peers Better), and test tasks measuring children's persistence.

## 4.2 Experiment 1

### 4.2.1 Method

#### 4.2.1.1 Participants

Children were recruited from an urban children's museum. Sixty-six children were included in the final sample ( $m_{\text{age}} = 5.00$  years, range: 4.00-5.95 years, 53% girls) and randomly assigned ( $n = 22$  per condition; the necessary sample size was calculated based on the preliminary experiment with an estimated effect size of  $f = .75$ ,  $\alpha = .05$ ,  $\beta = .80$ ) to a *Peers Better*, *Peers Worse*, or a *No Peers* condition. Age was equivalent across the three conditions ( $p = .36$ ). Eleven additional children participated but were excluded from further analysis due to experimenter error ( $n = 1$ ), technical difficulties with stimuli ( $n = 2$ ), family interference ( $n = 5$ ), voluntarily withdrawing from the study ( $n = 1$ ), and inability to count the initial sets, which had a maximum of 7 objects ( $n = 2$ ).

#### 4.2.1.2 Materials<sup>4</sup>

In the Initial Task and Evidence Phase, children saw pieces of paper with sets of fish outlined in a rectangle. In the Evidence Phase of the *No Peers* condition, line drawings of four animals were shown to children. For the Persistence Task, we modified a commercially available game for children, *Zingo*, that dispenses two plastic tiles (4 x 3 cm) at a time. Each tile was covered with a sheet of paper depicting a set of 6-11 shapes printed in a single color. The 56 tiles in the set were sorted into 28 pairs, including one pair for practice, so that the set sizes on each pair of tiles differed by only 1 or 2 shapes, ensuring that careful counting was required; tile pairs were loaded into the dispenser in a fixed randomly generated order. Two cups (5 cm in height) for sorting the tiles were placed on small cardboard staircase to place one was higher than the

---

<sup>4</sup> All stimuli and data will be available upon publication on the Open Science Framework.

other. The smaller (diameter 6.5 cm) was placed on the lower stair, raised 2 cm from the table. The larger cup (diameter 8 cm) was fixed to the top stair, raised 6 cm from the table. A small handheld water toy (16.5 x 10.2 x 3.8 cm) was used as an alternative choice. See Figure 4.1 for a schematic of the stimuli.

#### 4.2.1.3 Design and Procedure

**Initial Task:** children were given a sheet with pictures of three sets of fish and asked to count the fish in each set (set sizes were 4, 2, and 7). The experimenter put a star next to each set to indicate that the child had counted that set correctly and said, “You counted the orange fish right. I’m going to give you a star for that.”

**Evidence Phase:** Four sheets identical to the Initial task (purportedly completed by four peers) were used in the *Peers Worse* condition; two sheets had only two stars and two had a single star indicating that only one or two sets were counted correctly. The experimenter said, “This is Calvin. He counted the orange fish right and the blue fish right, but not the green fish.” The remaining three pieces of evidence were similarly introduced. Stickers with red frowning faces indicated which sets had been counted incorrectly. In the *Peers Better* condition, all four sheets had three stars and the set sizes were larger (16, 13, and 22 fish) and the Experimenter said, “This is Calvin. He counted the orange fish right, the blue fish right, and the green fish right.” In the *No Peers* condition, children were shown line drawings of animals as a filler. See Figure 4.1.

**Persistence Task:** We modified a commercially available game for children, *Zingo*, which dispenses two plastic tiles at a time. Each tile was covered with a slip of paper printed with between 6 and 11 shapes. Tile pairs were loaded into the dispenser in a fixed random order. The



experimenter said, “Here I have a machine with some cards inside. Let me show you how it works.” The experimenter dispensed two cards, counted the number of shapes on each, and together with the child she placed the card with more shapes in a large cup on a high shelf and the card with fewer shapes in a small cup on a low shelf. Children were told they could count and sort as many cards as they wanted and that they could play with a water toy placed next to the machine whenever they finished. If the child stopped counting for 20 second but did not play with the water toy, the experimenter asked the child if they were done counting in order to establish whether the Persistence Phase has been completed.

#### 4.2.2 Results

Each pair of tiles sorted counted as a trial. We counted the number of pairs of tiles children counted before stopping. A naïve coder blind to condition and hypotheses recoded a randomly selected sample of 25% of the data, and the agreement between coders was high ( $r=.99, p < .001$ ). Across conditions, the number of trials completed by participants differed significantly, Kruskal-Wallis test:  $\chi^2(2) = 13.09, p = .001$ , (mean *Peers Worse*: 5.55 trials, 95% CI [2.50, 7.96]; Mean *No Peers*: 8.73 trials, 95% CI [5.00, 12.05]; Mean *Peers Better*: 13.86 trials, 95% CI [10.55, 17.27]).<sup>5</sup> In follow-up pairwise comparisons, we found, as we predicted, that children in the *Peers Worse* condition persisted for fewer trials than children in the *Peers Better* condition ( $p < .001$ ), and children in the *Peers Better* condition persisted for more than

---

<sup>5</sup> All confidence intervals were bootstrapped with 10,000 samples. We used non-parametric statistics because our outcome variables did not adhere to a normal distribution (see ElBasslouny, 2013). Non-parametric tests present a more conservative approach to analyzing ordinal data than the equivalent parametric tests, although parametric tests yield the same pattern of results across all analyses.

children in the *No Peers* condition ( $p = .025$ ). There was no significant difference between the *Peers Worse* and *No Peers* conditions ( $p = .143$ ). See Figure 4.2.

### 4.2.3 Discussion

In Experiment 1, children showed an effect of social comparison in the *Peers Better* condition (persisting more than in the control condition). The difference between persistence in the control condition and the *Peers Worse* condition did not reach significance. These findings lend initial support for the account that children use social comparison information to calibrate their persistence. In a domain that is relevant for their learning, namely counting, 4- and 5-year-olds responded by finding out that their peers completed a harder version of a task by ramping up their own efforts, as compared to those children who saw no social comparison evidence, or who saw that they in fact performed better than their peers.

However, two different (not mutually exclusive) accounts could explain children's increased persistence relative to baseline in the *Peers Better* condition of Experiment 1. Children might have persisted because they were worried the experimenter thought badly of them compared to their (fictitious) peers and they wanted to impress the experimenter. Alternatively, children might have persisted because social comparison evidence may have motivated children to allocate more effort to the task at hand, independent of a concern for their reputation.

To explore these possibilities, we conducted another experiment that was similar to the *Peers Better* condition of Experiment 1 in Experiment 2, with two changes. First, to avoid any implication that the experimenter had deliberately chosen an easy task because she doubted the child's abilities, we ensured that the Initial Task was (apparently) chosen by the child herself. Second, we switched experimenters halfway through so that the experimenter present during the

Persistence Task did not know that other children had (apparently) done better at the task. If children are motivated primarily by reputation maintenance, they should be less likely to persist in the *Peers Better* condition of Experiment 2 than Experiment 1. However, if the social comparison evidence leads to a recalibration of effort for the purposes of allowing more opportunities for learning when merited, the results of the *Peers Better* condition of Experiment 1 should replicate in Experiment 2.

### **4.3 Experiment 2**

#### **4.3.1 Method**

##### **4.3.1.1 Participants**

Children were recruited as in Experiment 1. A total of 22 children were included in the final sample ( $m_{\text{age}} = 4.98$  years, range: 3.98-5.89 years, 59% girls). An additional four children were excluded from further analysis due to experimenter error ( $n = 1$ ) and voluntarily withdrawing from the study ( $n = 3$ ).

##### **4.3.1.2 Materials, Procedure and Design**

Experiment 2 had a single condition (*Peers Better*) and was identical to Experiment 2, except as follows. First, children were given a choice of a red and a green folder (both actually containing the same Initial Task) so they would not believe that the experimenter had deliberately selected an easy task. Second, after explaining how the Persistence Task worked, the first experimenter said, “Oh, I just remembered I have to go run an errand. I’m going to have my friend come sit with you.” A second experimenter entered the room and administered the remainder of the study as in Experiment 1.

### 4.3.2 Results and Discussion

The data were coded as in Experiment 1. Reliability between coders was high ( $r=.99, p < .001$ ). Children in Experiment 2 completed a mean of 13.45 trials, 95% CI [8.82, 17.91]). This average was equivalent to their persistence in the Experiment 2 *Peers Better*, condition (13.86 trials, 95% CI [10.55, 17.27]), Mann Whitney independent samples test:  $W = 257, p = .73$ , Bayes Factor in favor of the null = 3.34. This suggests that children's increased persistence was not primarily driven by reputation maintenance; rather the evidence of better-performing peers appeared to affect children's motivation in the task independent of concerns for their reputation. See Fig. 2.

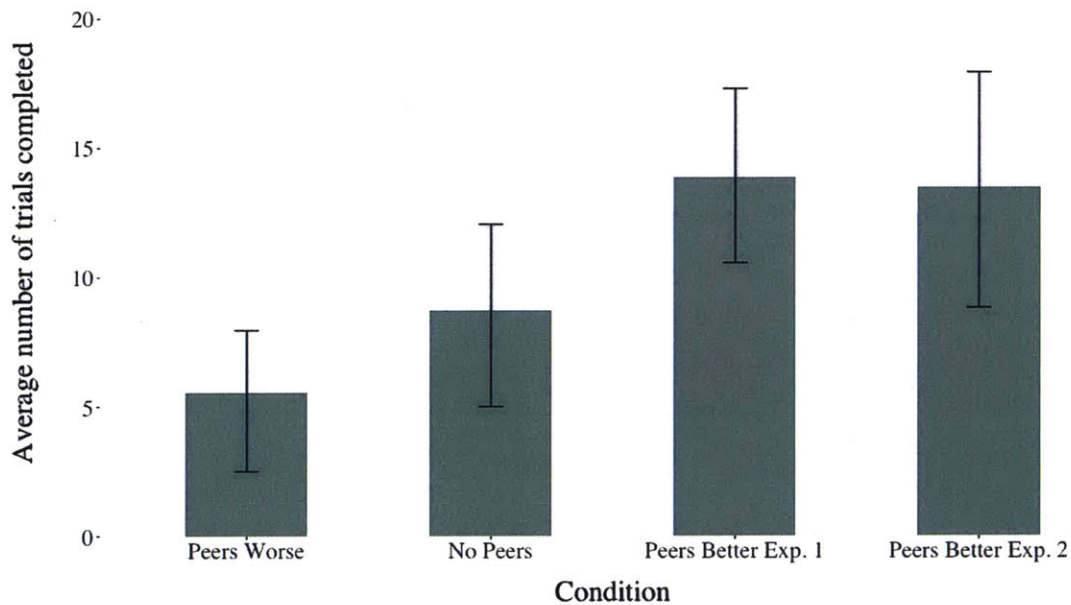


Figure 4.2. Results of Experiments 1 and 2. Error Bars represent bootstrapped 95% Confidence Intervals.

## 4.4 Experiment 3

In Experiments 1 and 2, we found that children's persistence was affected by social comparison evidence. In a task that 4- and 5-year-olds were still mastering, counting and comparing sets, they used information about better-performing peers to increase the effort they allocated to a task in the same domain. In general, a small amount of new evidence should affect one's beliefs, and therefore potentially one's behavior, when one's prior beliefs are relatively uncertain. Here we suggest that when children's prior beliefs about their abilities and ability to perform a task are very strong, a small amount of evidence (e.g., the examples of four peers) should matter relatively little, but when children are less familiar or capable at a task, their uncertainty may lead new evidence in the form of social comparison to have a measureable impact on children's persistence, as we saw in the case of 4- and 5-year-olds' counting in Experiments 1 and 2. In Experiment 3, we tested this hypothesis by choosing a task (sorting shapes) that is easy for four and five-year-olds, but near the zone of proximal development (Vygotsky, 1978) for three-year-olds. We predicted that four and five-year-olds would now be immune to social comparison evidence but that three-year-olds' persistence would be affected by it.

### 4.4.1 Method

#### 4.4.1.1 Participants

A total of 132 children were included in the final sample ( $m_{\text{age}} = 4.10$  years, range: 3.00-5.97 years, 48% girls). Nineteen additional children participated but were excluded from analysis due to experimenter error ( $n = 2$ ), technical difficulties with the tile dispenser ( $n = 2$ ), family interference ( $n = 7$ ), voluntarily withdrawing from the study ( $n = 7$ ), and a language barrier ( $n =$

1). Children were separated into two age groups, 3-year-olds (3.00-3.99) and 4- and 5-year-olds (4.00-5.97). Within each age group, children were randomly assigned to a *Peers Better*, *Peers Worse* and *No Peers* condition ( $n = 22$  per condition/age group); within each age group, ages were matched across conditions ( $p > .70$ ).

#### **4.4.1.2 Materials**

In the Initial Task and Evidence Phase, children saw pieces of paper with sets of shapes (circles, squares, triangles, and diamonds) outlined in a rectangle. The same line drawings of four animals used in Experiment 1 were shown to children in the *No Peers* condition. We again modified the game *Zingo*, such that each tile had a single shape (square, triangle, or circle) printed on it. The 60 tiles were divided into 30 pairs, including one pair for practice, with a different shape on each tile; pairs were loaded into the dispenser in a fixed random order. For children to sort the tiles, three small boxes 10 x 10 x 2 cm were taped together side-by-side; each box had a different shape pasted on the front. The same water toy from Experiments 1 and 2 as used.

#### **4.4.1.3 Design and Procedure**

The Procedure was identical to that in Experiment 1 except as follows.

Initial task: The experimenter showed children a sheet of paper with a large circle, triangle, and square and said, “These are my favorite shapes. I really like, circles, triangles, and squares. Which is your favorite?” This portion was added to remind young children the names of the shapes. Then, children were given a sheet with three sets of 4, 2 and 7 shapes and were asked to name each shape within the set. They received three star stickers on their paper.

Evidence Phase: Four shape sheets identical to those used in the Initial Task were used in the *Peers Worse* condition except that star stickers and red frowning faces were used to indicate that only one or two of the sets shapes had been named correctly. In the *Peers Better* all sheets had three stars and instead of sets of 4, 2, and 7 shapes, larger sets of shapes (16, 13, and 22) were depicted. Animal line drawings were shown in the *No Peers* condition.

Persistence Task: Children were asked to sort the shape tiles into one of three boxes, each marked with the target shape for sorting. The remainder of the Experiment followed the Procedure in Experiment 1.

#### 4.4.2 Results and Discussion

Once again, each pair of tiles sorted counted as a trial. The data were coded as in Experiment 1. Reliability between coders was high ( $r=.99, p < .001$ ). As predicted, 3-year-olds were affected by the social comparison evidence in the domain of shape sorting (Kruskal-Wallis test  $\chi^2(2) = 14.83, p < .001$ , Mean *Peers Worse*: 8.73 trials, 95% CI [4.68, 12.32]; Mean *No Peers*: 12.64 trials, 95% CI [8.51, 16.36]; Mean *Peers Better*: 19.36 trials, 95% CI [15.73, 23.18]). Three-year-olds the *Peers Worse* condition sorted for fewer trials than children in *Peers Better* condition ( $p < .001$ ), and children in the *Peers Better* condition sorted for significantly more trials than children in the *No Peers* condition ( $p = .01$ ). The *Peers Worse* and *No Peers* conditions were marginally significantly different ( $p = .07$ ). By contrast, four-year-olds showed no effect of condition (Kruskal-Wallis test  $\chi^2(2) = 0.341, p = .843$ , Mean *Peers Worse*: 18.41 trials, 95% CI [14.05, 22.91]; Mean *No Peers*: 17.00 trials, 95% CI [12.64, 21.36]; Mean *Peers Better*: 17.00 trials, 95% CI [13.27, 20.73]). See Figure 4.3.

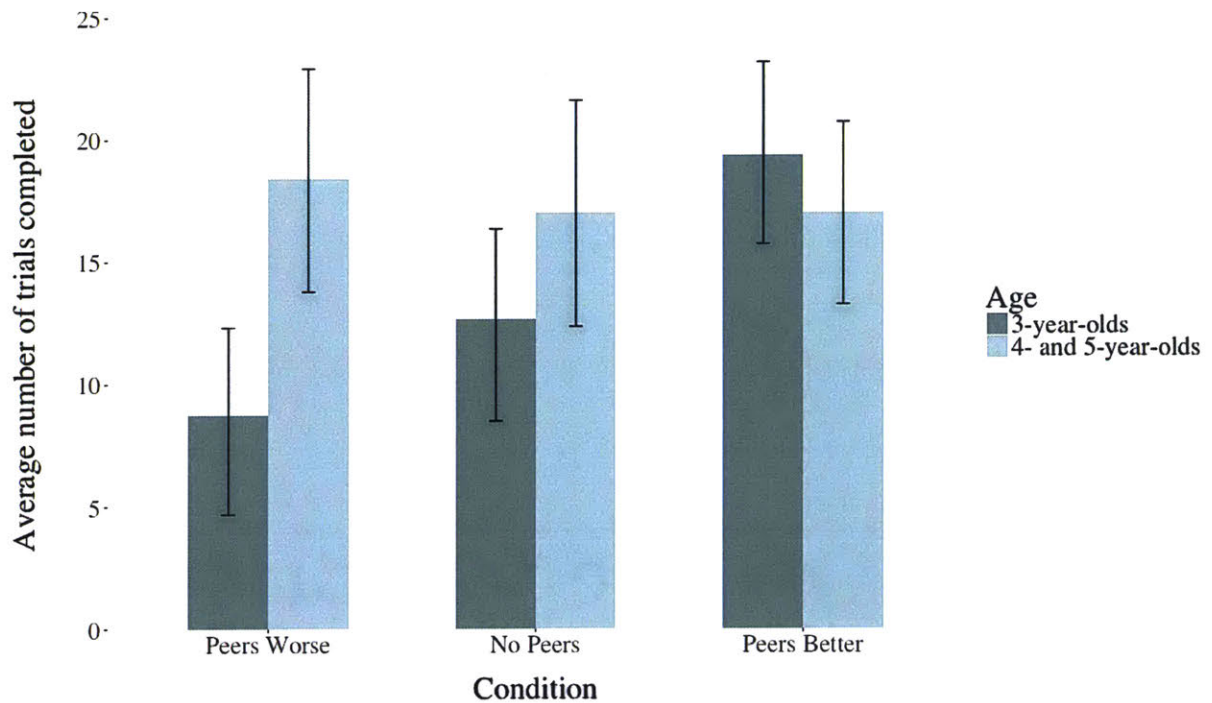


Figure 4.3. Results from Experiment 4. Error Bars represent 95% Confidence Intervals. Social comparison evidence affected persistence across conditions in 3-year-olds, but not 4- and 5-year-olds.

Consistent with the hypothesis that children use social comparison evidence to calibrate their persistence when their prior beliefs are more uncertain, 3-year-olds (for whom shape sorting is relatively challenging task) were affected by the social comparison evidence but 4- and 5-year-olds (who are confident shape-namers) were not. As the preceding studies suggest, this is not because four and five-year-olds are immune to social comparison, nor is it the case that they were either bored by the task or performing at ceiling; indeed the four-year-olds' persistence throughout (Means: 17-19 trials) was equivalent to the persistence of three-year-olds in the *Peers Better* condition (Mean: 19 trials). Rather the results suggest that children integrate social comparison evidence with their prior beliefs, so that evidence affects children most when they are uncertain about the abilities.



Like the four- and five-year-olds in Experiments 1 and 2, the three-year-olds in Experiment 3 were relatively more persistent given upward comparison (*Peers Better*) and only marginally significantly affected by downward comparison (*Peers Worse*).

#### **4.5 General Discussion**

In the current study, we found that preschoolers used social comparison evidence to calibrate their level of effort across two the domains. Across experiments, we found that when a task is in children's zone of proximal development, upwards social comparisons spur motivation. Compared to seeing evidence that peers performed better, when children see that their peers performed worse than they did, they persist relatively less. We found this pattern in children as young as 3 years, suggesting that the response to better performing peers is relatively early emerging and robust. These results leave open the question of whether children are insensitive to downward comparison compared to seeing no social comparison information or whether the effect is simply more graded: although the effects of seeing peers perform worse did not consistently reach significance in either Experiment 1 or Experiment 3, it is interesting that children in the *Peers Worse* condition persisted on numerically fewer trials than in the control condition. In sum, we find that rather than being insensitive to social comparison, we found that children ages 3 to 5 years interpreted their observations of better performing peers as evidence that they should allocate more effort to this task, given that improvement was possible.

Younger children's response of increasing effort upon learning it is possible to perform even better on a task accords with previous research showing that before entering formal schooling environments, children have relatively more mastery-oriented approaches to learning and their own abilities (Benenson & Dweck, 1986; Butler, 1989; Gelman, Heyman, & Legare,

2007; Nicholls, 1978), meaning they are primarily concerned with gaining new knowledge and skills rather than establishing their level of competence. That younger children are more focused on mastery compared to older children makes logical sense, given that, in general, older children's rates of learning are less rapid than young children's, and older children are more likely than younger children to believe that ability rather than effort accounts for success (Heyman et al., 2003). This notion is summarized well by the following statement, "In children's early environments, success is largely a function of paying attention and trying hard" (Cimpian, 2017, p. 401).

The results of the current study suggest that social comparison evidence supports children's self-directed learning even before they enter formal schooling, where explicit comparisons may be highlighted by adults. We argue that social comparison evidence in the domains under study provided children with a range of possible responses to educationally-relevant tasks and also exposed children to models who they may have emulated. However, if children were strictly modeling their own persistence after the amount of success they saw in the evidence phase, those in the *No Peers* conditions should have completed the fewest number of trials since they had no relevant models. This was not the case in any of the experiments. Rather children used social comparison information specifically to benchmark their own current competence and level of effort.

Of importance, this pattern likely depends on the domain at hand. In some cases, opportunity to allocate effort to remedy gaps in current skills or knowledge may not be present—especially for skills that are tied to essential traits. Children may not try as hard in response to upwards social comparison information for tasks where success is attributed to social category membership. This account is consistent with results showing that children as young as 4 years'

self-reported interest in a task as well as their performance can be affected by social comparison information (Butler, 1998; Rhodes & Brickman, 2008). Additionally, it is possible is that when social categories are highlighted at all, they prime a type of self-focus that overshadows more automatic considerations of learning or improvement (Cohn, 2004).

Another contribution of these empirical findings speaks to the importance of relevance for children's use of social comparison information, as has been shown in the adult literature (e.g., Buunk & Brenninkmeijer, 2001). When children observed evidence about their peers' performance in a task that they were already quite skilled at (4- and 5-year-olds in Experiment 3), we found no effect of social comparison. This finding is broadly consistent with the idea that learners integrate evidence with their prior beliefs (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Here we found that children were sensitive to social comparison evidence when they were relatively naïve about their own abilities and less skilled in a domain, but not when they were quite knowledgeable. Thus, less information for guiding learning could be gained from the behavior of others. The current findings point to the idea that children's current level of competency in a domain affects their sensitivity to social comparison evidence in that domain. Future research might also test the effect of upwards social comparison on self-directed learning on too difficult tasks (rather than too easy tasks) to see whether children will be less likely to integrate social comparison evidence when a task is beyond their zone of proximal development.

Although our findings suggest that upwards comparisons are motivating, we caution against the notion that all observations of better-performing peers will increase motivation. Sometimes being in a context where comparison is possible seems to shifts children towards easier tasks where failure is less likely (Dweck, 2000; Smiley & Dweck, 1994). Indeed if social comparison becomes competitive, young children as opposed to older children, show decreased

intrinsic motivation (Butler, 1989b; Deci, Betley, Kahle, Abrams, & Porac, 1981). Our work suggests that although social comparison may give children opportunities to calibrate their learning, in other settings or contexts it could trigger concerns about how they are perceived, whether by themselves or by others, leading them to shy away from tasks with which they would have otherwise gladly engaged. To better understand the range of possible responses, future work might look at how children's responses to social comparison evidence interact with children's mindsets (i.e., the belief that their abilities are fixed or malleable; Dweck, 2000; 2006) and potentially with essentialist beliefs about social categories (e.g., Cimpian & Dweck, 2007; Rhodes & Brickman, 2008).

Overall, these findings indicate impressive sophistication in young children learning from others' outcomes. Children's over-optimism in self-evaluations does not mean that they cannot take into account evidence about how their peers are performing. Rather, they use this information to calibrate their persistence while potentially remaining positive about the potential for growth and eventual mastery. At ages when they are still working to sort circles and squares and count sets of objects, children integrate social comparison evidence to calibrate their effort, suggesting this form of information may have cascading effects on learning starting in early childhood.

## Chapter 5

**This chapter is based on Magid, DePascale, & Schulz, (2017 *CogSci Proceedings*; Under Review). Four and five-year-olds infer differences in relative ability and appropriately allocate roles to achieve cooperative, competitive and prosocial goals**

Preschoolers are sensitive to differences in individuals' access to external resources (e.g., tools) in division of labor tasks. However, little is known about whether children consider differences in others' internal resources (e.g., ability) in allocating roles. Critically, these differences are relevant in collaborative contexts but may be irrelevant in competitive and prosocial ones. In three pre-registered experiments, we found that four and five-year-olds (mean: 54 months; range: 42-66 months;  $N = 132$ ) used age differences to infer relative ability and appropriately allocate the harder and easier of two tasks in a dyadic cooperative interaction (Experiment 1); they appropriately ignored differences in relative ability in competitive (Experiment 2) and prosocial (Experiment 3) contexts, instead assigning others the harder and easier roles, respectively. Thus three-and-a-half to five-year-olds evaluate their own abilities relative to others and effectively allocate roles to achieve diverse goals.

## 5.1 Introduction

Cooperation is a foundation of human culture and cognition, observed in activities as diverse as governing, hunting, fishing, building, and playing (Brownell, Ramani, & Zerwas, 2006; Rogoff, 1990; Tomasello, 1999). Young children begin cooperating in problem solving and social games by their first birthday, and the sophistication of their cooperative interactions increases over the first few years of life (e.g., Brownell & Carriger, 1990; Warneken, Chen, & Tomasello, 2010; for review see Warneken, 2017). Children cooperate by sharing food and toys (Brownell, Nichols & Svetlova, 2009; Hay, 1979), pointing to inform others (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello 2008), and assisting in goal-directed actions (Warneken & Tomasello, 2007). Children also appear to expect cooperation: when adults disengage from cooperative interactions, they protest (Ross & Lollis, 1987).

Across species, the most sophisticated forms of cooperation involve collaboration: cases in which individuals adjust their behavior to accomplish a goal (Boesch & Boesch, 1989). Children as young as three-and-a-half flexibly divide labor by coordinating on tasks involving different sub-goals (Ashley & Tomasello, 1998; Fletcher, Warneken, & Tomasello, 2012). Moreover, older preschoolers divide labor with respect to available resources: when the participant has both the tools needed to achieve a joint goal while their partner has only one, five-year-olds (though not three-year-olds) appropriately delegate to their partner the task corresponding to their partner's tool (Warneken, Steinwender, Hamann, & Tomasello, 2014).

In such cases, both partners are, in principle, equally capable of performing either role. However, people differ not just with respect to the availability of external resources, but also with respect to their internal resources, including motivation, physical ability, knowledge, and

intelligence.

Preschoolers are sensitive to such differences, selectively choosing to associate with and learn from more trustworthy, competent, and knowledgeable agents (e.g., Jara-Ettinger, Tenenbaum, & Schulz, 2015; Koenig & Jaswal, 2011; Koenig, Clément, & Harris, 2004; Kushnir, Vredenburg, & Schneider, 2013) and choosing to help and instruct more naive peers (Johnson, Pynn & Nisbet, 2002; Ziv & Frye, 2004; see Corriveau, Ronfard & Cui, 2017 for review). Three to five-year-olds also understand cognitive division of labor, recognizing that people specialize in different areas and may have more expertise in some areas than others (Danovitch & Keil, 2004; Lutz & Keil, 2002).

The degree to which children accurately represent their own strengths and weaknesses is somewhat more controversial. Work suggests that preschoolers are sometimes (excessively) optimistic about their abilities (Burhans & Dweck, 1995; Schneider, 1998). However, other research suggests that children as young as three engage in social comparison, evaluating their own performance relative to their peers (Butler, 1988; Magid & Schulz, 2015; Rhodes & Brickman, 2008) and can accurately assess whether they are good or bad at familiar tasks (e.g., Cimpian, Mu, & Erickson, 2012; Heyman, Dweck, & Cain, 1992). To the degree that children are sensitive to relative differences in ability (even if they are poor judges of their abilities in an absolute sense), they might recognize that they should take the easier task if they believe their partner is more capable, and the harder task if they believe their partner is less capable.

Critically however, allocating roles to “each according to their abilities” is only rational in collaborative tasks where both parties must succeed in order to achieve the goal. There are a number of other goal contexts in which partners engage in role allocation. Given that children are sensitive to goals across a range of cases, in that they track individuals’ goals and respond to

similar evidence differently depending on the overarching goal context (Buresh & Woodward, 2007; Carpenter, Call, & Tomasello, 2002; DiYanni, Nini, & Rheel, 2011; Meltzoff, 1995), we investigate role allocation across two additional contexts. Children will behave differently depending on the goals of a game, cooperating when necessary and competing when appropriate (see Green & Rechis, 2006 for a review). Additionally, when someone is in need, even very young children take steps to help, demonstrating a flexibility to take on prosocial goals when called for (see Warneken, 2015 for review).

Based on these findings, we expect unique patterns of role allocation when children compete and help. If the goal of a task is to optimize one's own chances of success, it makes sense to disregard relative differences in abilities and always assign the easier role to oneself and the harder task to one's opponent. Conversely, if the goal is prosocial – optimizing the other person's probability of success – it makes sense to assign the other person the easier of two tasks regardless of their abilities, especially if the decision-maker can do so at no cost to themselves.

Thus an adept social agent who has the power to allocate roles should do so differently in different goal contexts, considering the participants relative abilities in collaborative contexts when maximizing each participant's chances of success is beneficial but disregarding them in competitive and prosocial contexts where role assignment should be governed only by the relative difficulty of the tasks themselves.

Here we test children's aptitude at role assignment in these different contexts by introducing participants to two carnival-style games: a ring toss and ball toss. Each game had an easy and a hard version. (See Figure 5.1.) Individual participants received the easy version of one game and the hard version of the other. Children were not told that one game was "easy" and the other was "hard" but had an opportunity to try each game and judge for themselves.



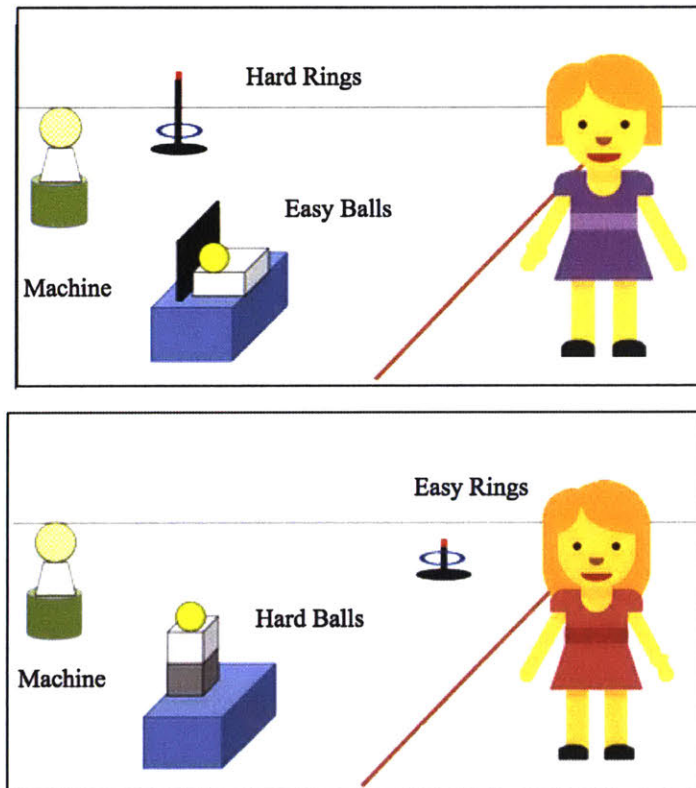
Participants were then told that another child was going to come to play with them. They were told that they had to choose one game for their partner, and one for themselves, and that if they both succeeded (i.e., getting a ring on a pole and a ball in the box) a special machine would activate.

In general, preschoolers expect that children their own age will know more than infants and less than adults (Taylor, Cartwright, & Bowden, 1991; see also Jaswal & Neely, 2006; Vanderborgh & Jaswal, 2010). We do not know of any research looking at whether children can use smaller age differences to infer relative abilities but given the anecdotal prevalence of age/ability attributions in everyday peer and sibling interactions we assumed that they would. Thus in one condition, children were told that the partner would be younger than themselves (*Younger Other condition*); in the other condition they were told that their partner would be older (*Older Other condition*). Consistent with previous work on collaborative problem solving (e.g., Warneken et al., 2014), the children recruited as participants for the study were three and-a-half to five-and-a-half years-old; thus children were told that the fictitious partner was two in the *Younger Other* condition and six in the *Older Other* condition.

There are a number of possible results. If children are poor judges of task difficulty, they should choose at chance. If children judge the tasks accurately, but try only to maximize their own chances of success (and ignore the collaborative nature of the task) they should choose the easy task for themselves and the hard task for their partner in both conditions. Conversely, if children tend to overestimate themselves (or underestimate their partner's ability) they should choose the hard task for themselves and the easy task for their partner in both conditions. However, if children's role allocation in cooperative tasks is sensitive to relative ability (as indexed by age), they should choose the easier game for their partner if their partner is younger,

and the easier game to themselves if their partner is older. In Experiment 1, we investigate how children allocate roles when working with another person. In Experiment 2, we ask how children allocate roles when working against another person. Finally, In Experiment 3, we evaluate whether children's ability to flexibly allocate roles extends to contexts where their goal is a prosocial one.

We chose to test children ages 3.5 to 5.5 years because previous work has shown that starting at 3.5 years, children can switch from one role in a collaborative interaction to another role (Ashley & Tomasello, 1998). Five-year-olds, but not three-year-olds can flexibly reason about what role to adopt in a cooperative interaction based on individuals' differential resources. We expected the task in the current study would be easier for younger children given that they do not need to remember or reason about the tools that their partner has available to them in the specific context of the study (c.f., Warneken et al., 2014). Thus children were told that the fictitious partner was two in the *Younger Other* condition and six in the *Older Other* condition so that there would be at least a ½ year gap between the age of the participant and the oldest possible age of the other participant.



*Figure 5.1.* Each participant saw only one setup (top or bottom). Participants practiced each game before allocating roles. Designations in the figure are those made by the experimenters, but were not shared with children.

## 5.2 Experiment 1

### 5.2.1 Method

#### 5.2.1.1 Participants

Procedures and our analysis plan for all experiments were pre-registered on the Open Science Framework ([osf.io/aq246](https://osf.io/aq246)). We assumed a large effect size (Cramer's  $V = .50$ ), and a power analysis indicated that 44 participants were required to reach a power of .90. All participants were recruited from an urban children's museum and randomly assigned to one of two conditions: Younger Other or Older Other. Forty-four children (mean age = 54 months;

range 43-66 months) were included in the final sample ( $n = 22$  per condition). Ten additional children did not pass the inclusion criteria. (See Procedure for details). An additional five children were tested but excluded due to parental interference ( $n = 3$ ) or failing to provide a response to the test question ( $n = 2$ ).

**5.2.1.2 Materials.** A felt mat (132x94 cm) was placed on the floor for game play. The mat was marked with three tape Xs and a line (16cm in front of the Xs) to indicate where participants should stand. Participants stood on the left and right Xs to play games and the center X to answer questions. Children played two games: a ring toss and a ball toss. Each game had two versions—one easier (Easy Rings, Easy Balls) and one harder (Hard Rings, Hard Balls). The ring toss used a plastic pole on a black circular plastic base. The easier version used a shorter pole (22cm with a 5cm red tip) and was closer to the tape line (13 cm away); the harder version used a taller pole (40cm with a 5cm red tip) and was farther from the tape line (65cm away). The ring toss game was played with blue rings (16cm diameter). Each ball toss game used a gray fabric box placed on top of a blue plastic crate (24x24x41cm) and was played with yellow plastic balls (26cm circumference). The easier version used a larger box (29x14x10cm) with a cardboard backboard (17x28cm) and was placed at the front of the crate, closer to the tape line (53cm). The harder version used a smaller box (14x14x10cm) elevated on a black box of the same size, and was placed at the back of the crate, farther from the tape line (77cm). The games were pre-tested with a separate group of children to establish that the easy games were in fact easier for children to score on than the harder games.

Half the participants played the Easy Rings and Hard Balls, half the Hard Rings and Easy Balls. Laminated cards (23x6cm) showed photographs of Older Other or Younger Other children. Children depicted in the photographs were either two-year-olds (10cm tall) or six-year-

olds (15cm tall), based on the condition. A laminated card of the same size had the word “YOU” printed in the center and was used to represent the participant. A remote-controlled LED light machine (12x13x12cm) was used for the joint task.

**5.2.1.3 Procedure.** This study was approved by the institutional IRB. All children were tested individually in a quiet room at a children’s museum. Children were shown two games (either Easy Rings and Hard Balls, or Hard Rings and Easy Balls) and given the chance to practice each game four times. The game played first (rings or balls), the location of each game (right or left), and the version of each game (easy or hard) were counterbalanced across participants. After children practiced, the experimenter introduced the light machine and explained that players of the two games could work together to achieve a single joint goal: if the ball went in the box and a ring went on the pole at the same time, then the machine would light up. The experimenter introduced the participant to the fictional other child, named Jamie, by explaining that she had talked with the other child earlier that day and that s/he wanted to come play the games together with the participant. The experimenter then showed children a card with a picture of the other child and said that they were either a toddler (Younger Other) or a first-grader (Older Other). The experimenter then asked children their own age, so that they could specify that the other child was younger or older, by condition. The other child was matched by gender to the participant. For each category (Younger boy, Older boy, Younger girl, Older girl) two pictures were used to reduce the possibility that ancillary features of any picture might influence children’s choices or perceptions of the other child’s abilities. The photographs represented a diversity of races and ethnicities.

The experimenter then asked children to allocate roles by choosing which game the other child should play, placing the other child’s picture next to the game chosen for them and a card

with “You” written on it next to the game the participant chose for themselves. One game was designed to be easier than the other, however differences in motor skills or experience might lead different children to different conclusions, thus to ensure that the role allocation matched children’s judgment of the relative difficulty of the two games, we asked children “Which game was easier?” As a follow-up, children were asked why they chose the game they picked for the other child. Finally, we asked children if the other child was older or younger to ensure that they remembered the age of the intended partner. This last question was used as an inclusion criterion: children who did not answer correctly were not included in the analysis. Following these questions, the experimenter left the room briefly (15-30 seconds) and returned saying that she couldn’t find the other child. The experimenter then played the games with the child to turn on the light machine.

### **5.2.2 Pre-Registered Analyses and Results**

All data were coded from videotape by the second author. A naïve coder blind to condition and hypotheses recoded a randomly selected sample of 25% of the data for the three outcome measures: the question about role allocation, the question about game difficulty, and the question about Jamie’s age. Coders agreed 100% of the time (Cohen’s Kappa = 1).

In response to, “Which game was easier?” 37 of the 44 children (84%) responded that the game designed to be easier was easier for them. Children’s self-reported judgment was used in all analyses (consistent with the pre-registered design). This is because we did not expect the games designed to be easier would be perceived as such by every child due to individual children’s motor skill development and their prior experience with throwing balls or tossing rings. We did, however, expect that children would use their own notions of difficulty as proxies for other children’s notions of difficulty (Gweon, Asaba, & Bennett-Pierre, 2017).

All analyses to follow were pre-registered. As predicted, children's role assignments differed by condition ( $\chi^2(1) = 7.615, p = .006, V = .462$ ). In the Younger Other condition, 14 children (63%) children assigned their partner the Easy Game. By contrast, in the Older Other condition, only 4 children (18%) assigned their partner the Easy Game. See Figure 5.2. Collapsing across conditions, 72.72% of children assigned roles in a way corresponding to the difficulty of fulfilling each role in the joint task,  $p = .004$  by binomial test (two-sided). Thus children allocated roles in a way most likely to lead to their joint success and there was no significant difference in children's ability to allocate roles effectively in each condition ( $\chi^2(1) = 1.031, p = .310, V = .204$ ). Although we did not predict an effect of age on children's role allocation, given previous work showing that five-year-olds, but not three-year-olds allocate roles based on available resources (Warneken et al., 2014), we test whether the likelihood of participants allocating roles based on ability increased with age. As predicted, there was no effect of age in a logistic regression model using chronological age to predict corresponding role assignment,  $\beta = -.004, p = .995$ . This suggests that children ages three-and-a-half to five-and-a-half years can allocate roles in a cooperative interaction given inferred differences in ability.

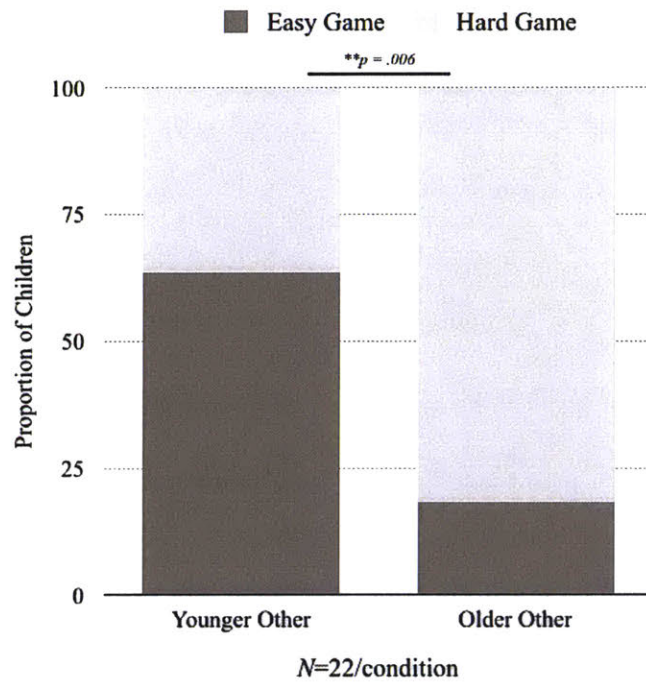


Figure 5.2. Proportion of children who chose the Easy Game or Hard Game for their partner by condition in Experiment 1: Cooperative Goal Context.

### 5.2.3 Discussion

These results suggest that children consider their own and their partner's relative abilities when allocating roles in a cooperative interaction. However, they leave open the question about the extent to which preschoolers simply assign harder games to older children and easier games to younger children without regard for the context. One explanation for this pattern that does not necessarily take into account the goal context of the task is a preference for procedural justice, or the equitable distribution of opportunity. Children as young as 5 years resist the unfair distribution of opportunity, and change the rules of an unfair game to provide equal opportunity to participants (Grocke, Rossano & Tomasello, 2015). While children are usually averse to inequitable distribution of resources, when there is a need or a reason to allocate opportunities



differently to different individuals, young children do so (Schmidt, Svetlova, Johe, & Tomasello, 2016).

Because children are sensitive to goals across a range of cases, we predicted that when the goal of a task is competitive, children's role allocation will reflect the demands of a competitive task. As noted, in competitive contexts, preschoolers should ignore relative ability and assign roles based on the relative difficulty of the tasks: assigning the easier game to themselves and the harder one to their opponents. We test this in Experiment 2.

### **5.3 Experiment 2**

#### **5.3.1 Method**

As in Experiment 1, all participants were recruited from an urban children's museum and randomly assigned to one of two conditions: Younger Other or Older Other. Forty-four children (mean age = 54 months; range 42-65 months) were included in the final sample ( $n = 22$  per condition). Seven additional children did not pass the inclusion criteria. (See Experiment 1 for details). Two additional children were tested but excluded due to parental interference. The materials were the same as those used in Experiment 1.

Children were introduced to and practiced the two games as in Experiment 1. After children practiced each game, the experimenter introduced the light machine and explained that the person who got a ball in the box or a ring on the pole first would win and get to turn on the machine. All other procedures were identical to those in Experiment 1.

#### **5.3.2 Results**

Data were coded as in Experiment 1. Inter-coder agreement was 100% (Cohen's Kappa =

1). In response to, “Which game was easier?” 33 of the 44 children (75%) responded that the game we had designed to be easier was easier for them.

All analyses to follow were pre-registered. As in Experiment 1, children’s own judgments were used for all analyses. In the Younger Other condition, 13 children (59%) assigned their partner the Hard Game. See Figure 5.3. In the Older Other condition, 18 children (82%) assigned their partner the Hard Game. As predicted, in Experiment 2 children’s role assignments did not differ by condition  $\chi^2(1) = 1.747, p = .186, V = .249$ . Collapsing across conditions, 70.45% of children assigned the harder game to the Other child,  $p = .010$  by binomial test (two-sided).

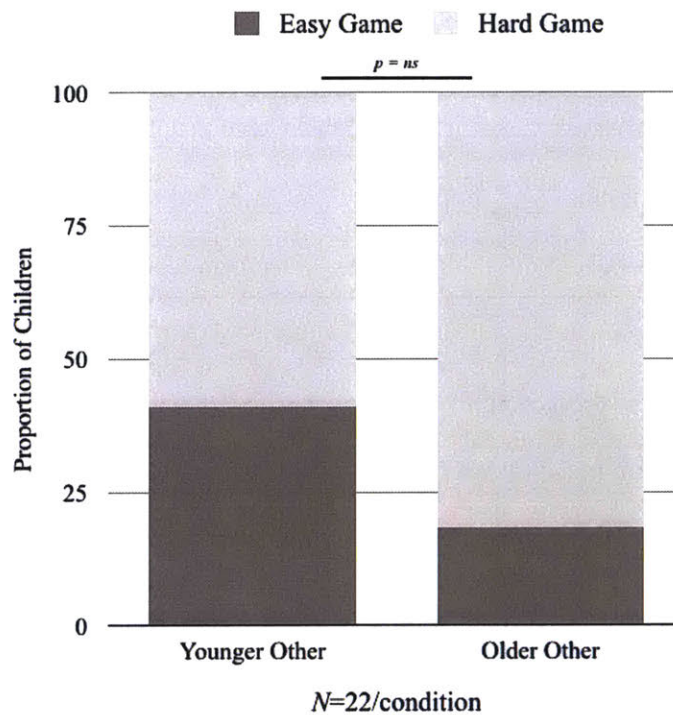


Figure 5.3. Proportion of children who chose the Easy Game or Hard Game for their partner by condition in Experiment 2—Competitive Context.

### 5.3.3 Discussion

These results suggest that preschoolers' flexible allocation of roles in Experiment 1 was not due merely to a tendency to assign younger children easier games and older children harder games. In the competitive context of Experiment 2, preschoolers assigned harder games to their opponent regardless of inferred ability. In Experiment 3, we look at whether children assign all children the easier task a prosocial context.

## **5.4 Experiment 3**

### **5.4.1 Method**

As in Experiments 1 and 2, participants were recruited from an urban children's museum and randomly assigned to one of two conditions: Younger Other or Older Other. Forty-four children (mean age = 55 months; range 43-66 months) were included in the final sample ( $n = 22$  per condition). Ten additional children did not pass the inclusion criteria. (See Experiment 1 for details.) Five additional children were tested but excluded due to parental interference. Materials were the same as the previous experiments.

The procedure was identical to Experiments 1 and 2 except that the experimenter explained that the children would get a sticker for the number of rings and balls they scored at the end of the game. After children practiced the games and were given the stickers, they were introduced to Jamie, who would come play after the participant. Children were told that Jamie would only have time to play one of the two games. Children were asked which game Jamie should play.

### **5.4.2 Results**

Data were coded as in Experiment 1. Inter-coder agreement was 97.22% (Cohen's Kappa

= 1). In response to, “Which game was easier?” 32 of the 44 children (73%) responded that the game we had designed to be easier was easier for them.

All analyses to follow were pre-registered. As previously, children’s own judgments were used for all analyses. In the Younger Other condition, 17 children (77%) assigned their partner the Easy Game. In the Older Other condition, 13 children (59%) assigned their partner the Easy Game. See Figure 5.4. As predicted, children’s role assignments did not differ by condition  $\chi^2(1) = .943, p = .332, V = .195$ . Collapsing across conditions, 30 of the 44 children (68%) assigned the Easy game to the Other child ( $p = .023$  by binomial test).

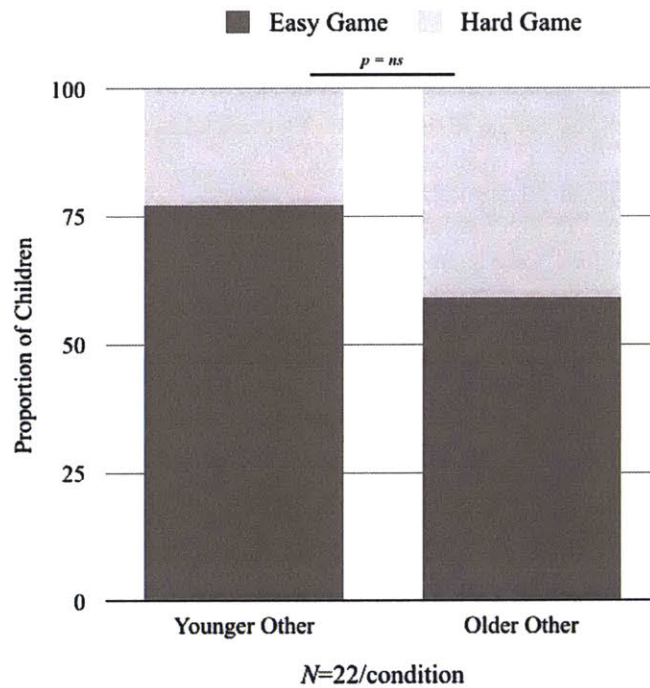


Figure 5.4. Proportion of children who chose the Easy Game or Hard Game for their partner by condition in Experiment 3—Prosocial Context.

### **5.4.3 Discussion**

The results from Experiment 3 provide further evidence that children's role assignment is sensitive to the goal context. Consistent with previous work (e.g., Buttleman, Carpenter, & Tomasello, 2009; Knudsen & Liszkowski, 2012; Martin & Olson, 2013), this suggests that young children can go beyond others' explicit requests in providing assistance to others. Preschoolers are sensitive to the difficulty of components of a task, and their helping behavior is sensitive to these dimensions of difficulty (see e.g., Bridgers, Gweon, & Jara-Ettinger, 2016 for similar results in selective teaching).

### **5.5. General Discussion**

In the current study, children effectively allocated roles in cooperative, competitive, and prosocial interactions. Although in principle preschoolers might simply choose which game to play based on level of enjoyment or choose a single strategy for role allocation independent of contexts (e.g., always assigning hard games to older children and easier games to others), instead we found patterns of role allocation that aligned with the way most likely to achieve the desired outcome. Preschoolers were able to consider differences in others' relative ability to effectively allocate roles to achieve a collaborative goal; they were also able to ignore differences in relative ability to allocate roles in competitive and prosocial contexts.

These findings support a large literature showing that children are adept at cooperating with others, in that they can take on complementary as well as parallel roles (Warneken et al., 2006), and they can divide labor and flexibly change roles within a task (Fletcher et al., 2012). The current study extends these findings to demonstrate that children are able to cooperate in a way that is based on something intrinsic about their collaborative partner. In addition, we found

that children are able to use themselves as a reference point when collaborating. Rather than always taking on the easier or harder task in Experiment 1, children's role allocation differed based on the age of the partner relative to the participant. Our results show that self-cognition is not only correlated with engagement and success in cooperative endeavors between 18 and 30 months (Brownell & Carriger, 1990; Brownell et al., 2006), but that self-cognition may play a role in more sophisticated forms of joint interactions.

Our study considered three goal contexts, but there are a few other goal contexts that would be interesting to investigate. In our version of a competitive context, the goal was to score first. However, in competitive sports, the goal is usually to see who prevails in a fair match, and sometimes handicaps are introduced on more able players (e.g., as in horse racing) to make the game more compelling. In a competitive context where the goal is not winning at all costs but winning at a fair match, previous work on children's preference procedural justice (e.g., Grocke et al., 2015) would support the idea that children would allocate roles in a way that would mirror the role allocation in Experiment 1: assigning the easier task to a younger player and the harder task to an older one. Indeed, across experiments, children were more likely to assign the Younger Other the easier game and the Older Other the harder game ( $p < .001$ ), suggesting they integrate the notion of procedural justice or appropriateness with goal context when assigning roles. Future work would be necessary to understand the precise way in which children combine notions of fairness with other goals.

If one has a general desire to be helpful, that goal could manifest itself differently depending on the particular way in which one wants to help. For example, one type of prosocial goal might be for the other individual to succeed on the task, in which case the child should assign the individual the easier game. However, another goal that is also compatible with the

prosocial context would be for the individual to develop her skills; in this case, the optimal decision might be to assign the less competent individual might be asked to do a harder part of a task. Future research might look at children's sensitivity to these more nuanced goal contexts, and also look at how children's ability to allocate roles extends beyond dyadic interactions to third party contexts and larger social groups.

The current study also suggests that children can use relatively fine-grained differences in age (two years) to infer relative ability. However, age is a coarse proxy for ability: younger individuals are sometimes more skilled, or skilled in different respects, than older ones. Research suggests that preschoolers are sensitive to these factors (Jaswal & Neely, 2006; Koenig & Jaswal, 2011; Kushnir et al., 2013; Lutz & Keil, 2002; Vanderborgh & Jaswal, 2010). Future work might investigate whether children can use more nuanced indices of ability differences to allocate roles.

Additionally, the current study focused on individuals' relative competence to complete a task. However, other factors both internal to the individual (e.g., motivation, trustworthiness, energy level, informational access, etc.) and external (access to tools, relative proximity, etc.) affect someone's ability to complete a goal, influencing the costs of goal-directed actions and the probability of success. Previous research suggests that children evaluate the costs and rewards of others' goal-directed actions (Jara-Ettinger, et al., 2015; Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015; Liu, Ullman, Tenenbaum, & Spelke, 2017); future research might look quantitatively at whether children's role assignment varies with calculations of expected utility (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). For the present, this study suggests that well before children have much, if any, experience delegating responsibility they appropriately invoke and ignore ability differences in assigning roles to achieve diverse goals.

# Chapter 6

## 6.1 Summary of Chapters 2-5

The findings presented in previous research and in this thesis show that across many different circumstances, young children are able to reason about their own and others' knowledge, skills, and abilities. This reasoning motivates and directs their learning and decision-making. Thus, children are rational not only in how they learn about the world but also in how they use information from the world and other people to learn themselves.

Chapter 2 investigated whether children expect another agent's beliefs to change after the agent observes evidence and found that 4.5- to 6-year-olds expect third parties to learn rationally. This work contributes to our understanding of the sophistication of children's reasoning *about* learning. While many other studies have demonstrated that children themselves learn rationally from data (e.g., Gopnik & Wellman, 2012; Schulz, 2012; Tenenbaum et al., 2011; Xu & Garcia, 2008) and represent others beliefs (e.g., Wellman et al., 2001), this work suggests children infer what others will believe from more nuanced forms of evidence, even when those resulting beliefs contradict children's own knowledge.

In Chapter 3, we found that children's own exploration was guided by the difficulty of the problem they were tasked with solving. Our findings suggest that the ability to monitor one's uncertainty, an aspect of metacognition that is often considered to be late developing (Lipowski et al., 2013), may be available to young children after all. We observed this to be the case in a situation where the task was in an intuitive domain, namely shaking boxes to guess how many marbles are inside, and when the degree of uncertainty could be readily represented and



alternative hypotheses disambiguated via perception. This work also connects to the findings from Chapter 2 demonstrating that children use representations of another's internal mental state and the discriminability of two alternative hypotheses to guide their inferences and exploration. Together, these studies show that young children represent not only evidence they observe through seeing or hearing—a sample of balls being drawn from a box or the number of marbles being shaken in a box—but they also represent unobservable variables, such as alternative hypotheses, in order to make inferences. This result has implications for our understanding of uncertainty monitoring and control processes, and provides additional support for a growing trend towards the idea that even young children are metacognitively sophisticated (Ghetti, Hembacher & Coughlin, 2013; Lyons & Ghetti, 2010)

Chapter 4 probed children's understanding of social comparison, and asked how information about a peer's performance might influence their motivation. We found that children calibrate their effort based on what they observed their peers do, but only for tasks relevant to their current learning goals. As young as age 3, children seem to represent aspects of their own current skill level relative to others and use this information to determine when more effort may be required. This work has implications not only for our scientific understanding of when children begin to attend to peers to learn about themselves which had remained unclear from past work (e.g., Butler, 1998; Ruble et al., 1994), but also points to future directions in applied research looking at how motivation is influenced by peers in real-world settings. Chapter 5 demonstrated that children ages 3-and-a-half to 5-and-a-half use relative ability assessments to decide who should be responsible for a component of a task in solving a problem. This work suggests that beyond being highly motivated to cooperate with others (e.g., Brownell et al., 2006), young children understand how to optimally allocate roles across contexts. While one

could argue that children do not *need* to represent their own abilities or update their theories of themselves with evidence since their abilities are changing rapidly, Chapters 4 and 5 suggest that children's reasoning about their abilities plays a key role in their self-directed learning. Thus, we do not have to wait until age 6 or 7 to use information about the self and others to make decisions that benefit our learning.

## **6.2 Limitations and open questions**

As with many paradigms in developmental science, our investigations are limited by the laboratory context. To investigate our core questions, we structured children's observations of the evidence to ensure they attended to key features of the problem. In addition, we elicited responses directly after putting forth a circumscribed set of evidence. Thus, we do not yet know how children would reason about changes in others' beliefs or how they might use social comparison information in more naturalistic and noisy settings. However, the results presented in this proposal advance our understanding of what is possible when children are given the opportunity to consider their own and others' processes for learning and problem solving.

A second limitation of the current work, and direction for future investigation, concerns the lack of study at the level of individuals. From previous studies, we know that young children's beliefs about the relationship between ability and effort vary across individuals by age 7 (e.g., Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2006; Dweck & Leggett, 1988; Gunderson et al., 2013), and there is variation in metacognitive abilities as early as age 5 (Vo et al., 2014). While we found evidence of sophisticated reasoning in children as a group, as our studies in Chapters 4 and 5 were not designed to assess individual differences, we did not have the measures to assess whether some children are more influenced by social comparison

information or more motivated to attempt challenging tasks, perhaps as a function of their achievement mindsets (Dweck, 2006). We found weak effects of age on children's performance in the tasks in Chapters 2 and 3, suggesting there may be development of metacognition in the domains we studied, but we did not have subtle enough measures to evaluate the strength of this relationship. While our findings do not take on questions about individual differences, the development of new data collection methods such as Lookit (Scott & Schulz, 2017) provide researchers the opportunity to design studies to test hypotheses about individual differences.

Additionally, while we have shown sophisticated reasoning across each of the four papers, we know that when trying to understand others' perspectives in daily life, children (and even adults) may struggle with analogous problems, for example understanding others' interpretation of evidence in the political arena (see Shapiro & Bloch-Elkon, 2008). For example, although we found in Chapter 2 that young children can reason about how others' prior beliefs influence interpretation of evidence, there are still many examples in everyday life of individuals failing to integrate their knowledge of another's prior beliefs with mutually observed evidence to infer the reason for a resulting difference in opinion. Thus, there must be other cognitive and non-cognitive components that drive our ability (and willingness) to understand other people's thought processes.

Moreover, we might also want to know what abilities continue to develop across one's life. There are a number of different aspects of metacognition that have been studied by both developmental cognitive scientists, including person metacognition, metacognitive monitoring, and metacognitive control (see Flavell, 1979 for a discussion of the types of metacognition). While we find evidence of metacognitive sophistication in this thesis, there are clearly domains and tasks, such as making confidence judgments about one's own answers, that develop into

middle childhood (Destan et al., 2014). Additionally, there are aspects of our self-representations that likely become more sharp and differentiated, for instance when we have to represent aspects of our own costs that did not exist before (e.g., limited time and money) and when we encounter new experiences, like forging close personal relationships. Going forward, theories of the development of metacognition should consider both children's early successes alongside their failures.

One outcome variable of interest in Chapters 4 and 5 is children's motivation and how it is influenced by metacognition. Even for adults, the link between metacognition and motivation can be challenging to articulate. The behavioral methods of the studies reported here aim to constrain the space of goals that children are trying to achieve by rewarding particular outcomes (e.g., identifying the correct number of marbles inside of a box; scoring before a partner). However, in real life our goals are not constrained in this way. Children who participated in the studies in Chapter 3 may simultaneously have wanted to know how many marbles there are inside the box and shake the box to generate sounds to listen to. Those who participated in Chapter 5, left to their own devices, may have wanted not only to maximize the possibility of attaining the joint goal but also to play the game they needed more practice at. Thus even the process of tracking one's own goals and acting in accordance with them can present a substantial challenge for a human learner as well as other intelligent systems.

### **6.3 The relationship between reasoning about the self and others**

#### **6.3.1 Structure and dynamics**

One question often raised in the broader discussion of metacognition is whether, and if so to what degree, reasoning about the self and reasoning about others reflect the same processes.

A parallel question concerning theory of mind has also received a good deal of attention from scholars in philosophy, psychology, and cognitive neuroscience. Researchers have posited different mechanisms underlying theory of mind, two of which I have discussed: (1) the proposal that people use an understanding of their *own* minds to model or simulate other people's minds (e.g., Perner, 1996), and (2) the argument that people have an abstract, representational theory of how people think that is deployed to generally reason about the self as well as others (e.g., Saxe, 2006).

An argument in line with the simulation account is the “like me” hypothesis, where infants come to understand others by leveraging their understanding of their own internal desires and their relationship to behavior to infer others' desires from their behaviors (Meltzoff, 2007, Sommerville, Woodward, & Needham, 2005; Woodward, Sommerville, & Guajardo, 2001). While not an explanation for the development of reasoning about others, evidence from speeded response tasks with adults shows that people seem to impute their own mental states to others (Bradford, Jentsch, & Gomez, 2015; Nickerson, 1999), and that inhibiting one's own mental state to conjure another's requires top-down control (Decety & Sommerville, 2003). On the basis of these and other findings, some have argued that modeling others after the self plays a primary and enduring role in supporting our rich social cognitive abilities (e.g., Decety & Sommerville, 2003; Meltzoff, 2007).

On the other hand, in favor of arguments for an abstract representational theory of mind, are the observed commonalities between mental state reasoning for the self and others. Specifically, proponents of this position argue that if simulation were the way in which people construe mental states, then we should expect mental state understanding to be more precocious or elaborate for the self than for others (Leslie, 1987). Aside from the studies discussed above

showing some privileging of the self over others in reasoning about beliefs (e.g., Bradford et al., 2015), some argue that there is no additional evidence that people have a deeper understanding of their own mental states (Carruthers, 2011). Moreover, there is little evidence from the developmental literature to suggest that mental states are attributed to self *before* others. For example, in the unexpected contents task (Perner, Leekam, & Wimmer, 1987), children are just as likely to report having had a false belief about what was inside a Smarty's container when the belief is one's own from a moment ago or that of another person in the present (Happé, 2003). Additionally, 3-year-olds make the same pattern of errors on false belief tasks when reasoning about self as when reasoning about others (Gopnik & Astington, 1988; Moore, Pure, & Furrow, 1990; Gopnik & Slaughter, 1991), suggesting that while very young children's theory of mind appears qualitatively different from that of adults, it operates similarly over computations of their own and others' beliefs. Further evidence for a representational theory of mind comes from work showing that congenitally blind people successfully reason about the mental states of sighted others, who have vastly different perceptual experiences from their own (Bedny, Pascual-Leone, & Saxe, 2009). Together, these findings suggest that theory of mind reasoning does not come about through a direct simulation of what the self experiences. Because theory of mind is encompassed by metacognition, there is utility in extending this analysis to not only mental states but also to characteristics of individuals' skills and traits. The literature reviewed in Chapter 1.3 (e.g., Schneider, 1998, Stipek, 1981) shows that children's evaluations of others' traits and skills are more accurate than evaluations of their own. These findings suggests that in reasoning about the abilities of others, we do not appear to use a model of ourselves.

In sum, some researchers have proposed that we predict and explain other people's behavior by modeling or simulating our own minds, while others argue we have an abstract,

representational, theory of mind that allows us to interpret both our own and others' behavior.

This thesis does not directly address the issue of simulation versus representation, but the current studies can shed light on the degree to which we reason about characteristics of the self and others in tandem, rather than first modeling the self to understand features of others.

The results from Chapter 4 demonstrate that children integrate information about peers' performance to direct their own learning. It is unlikely that the children in this study had to depend on a mode of themselves to interpret the observed information about other children's effort and abilities. Because children adjusted their own behavior (e.g., counting more or less tiles) based on information about how their peers had performed, they could not have used a specific model of themselves to understand other children's performance. These findings provide circumstantial evidence that rather than using ourselves to model others, we use information about others to better understand ourselves. Indeed, integrating information about other people may be a crucial component of how we reason about ourselves because the outcomes of our actions are often situation dependent and thus may be subjective without a point of reference (Festinger, 1954). Thus, having access to what others do in the same situation may provide a lens of objectivity to evaluations of ourselves that then guide self-directed behavior.

Crucially, this is not to say that children's reasoning about others always precedes reasoning about the self. Instead, children appear to be able use the same principles at the same time to reasoning about ability and effort in the context of themselves and other people. Evidence for this shared system Chapter 5, where children reasoned about their own and another (hypothetical) individual's skills, and used those representations to allocate roles across several contexts. Individual levels of competency on these motor tasks (throwing a ball into a bucket), and the academic tasks (counting objects) in Chapter 4, seem to be construed along the same

dimension for the self and others. In this sense, there is an analogous relationship between ability and outcome, and effort and outcome when those constructs concern the self as when they concern other people. Finally, in Chapter 2 we found that children held consistent beliefs that another individual would learn from evidence in a rational manner. Although there is a lively debate in education about the degree to which learning processes vary across individuals (e.g., Jonassen & Grabowski, 2012), at least in this constrained case, children had consistent expectations about the metacognitive process of another person's learning from data.

While children do have expectations about learning and behavior that apply across individuals, the results presented in thesis, as well as a host of past literature, also make a strong case for children also having a nuanced theory of individual differences. Understanding that knowledge, abilities, and traits differ across people, including the self, support a variety of sophisticated interpersonal interactions, including knowing how you can best teach another person (e.g., Gweon & Schulz, in press; Shafto, Goodman & Frank, 2012), who you should ask for help (Kushnir et al., 2013), and how you can best comfort a particular individual. In addition, using evidence to place oneself and others in a common, abstract, space of characteristics is crucial for guiding role allocation for more complex tasks, as well as in helping to determine what you and others should specialize in later in life. This claim is currently speculative. More investigations directly testing how young children reason about themselves and others across a variety of contexts is necessary to ascertain the degree to which children have a rich, flexible theory of individuals.



### 6.3.2 The self as a special case

The previous section discussed the common structure and dynamics of reasoning about the minds and characteristics of the self and others. However, intuitively, there seems to be something unique about our considerations of our own cognition. Thus, it is useful to consider what might make the self a special case of reasoning about individuals at large. The maxim “know thyself” exalted by philosophers, both ancient and modern, suggests that a kind of enlightenment can be achieved through profound consideration of one’s self. Indeed, many scholars share the idea that thinking about ourselves serves a different function than thinking about others. Evidence from neuroimaging studies show that while there is substantial overlap between regions of the brain that are active when considering the both the self and other people, there is also a network of brain regions that is more active when thinking about the self than others, suggesting that self-relevant cognitive processing might be specialized through the evolution of the brain (Heatherton, Wyland, Macrae, Demos, Denny, & Kelley, 2006; Saxe, Moran, Scholz, & Gabrieli, 2006; Vanderwal, Hunyadi, Grupe, Connors, & Schultz, 2008). This specialization may exist because we guide our own behavior, and basing our actions on an understanding of our own utilities and abilities (though not necessarily explicitly or consciously). These representations of the self, including reflections of our preferences, skills, and values enable us maximize positive experiences and minimize negative ones (Epstein, 1973).

This line of thought suggests that notions of a unique and aware self are critical to what it means to be intelligent, and to make intelligent decisions. Indeed, current approaches to artificial intelligence are actively seeking ways to build machines that reflect on their own operations and algorithms (Cox, 2005; Lewis, Platzner, Rinner, Torresen, & Yao, 2016). Building this type of intelligence has proved to be a challenging task. In particular, it is challenging to know whether

our phenomenology of the self, and our experience of being conscious, is the source of certain aspects of our intelligence (Kricher et al., 2002).

The work presented here supports on direction for future inquiry: perhaps what it means to be self-aware is in great part about being “other-aware.” Thus, to design truly intelligent artificial intelligence, further study of how we reason about people as individuals, and ourselves in context of others, is necessary. These future investigations would ideally include investigations that ask, at an algorithmic level, how we integrate reasoning about the self and others. A better understanding of these processes would provide critical information for developing intelligent systems that learn as humans learn.

#### **6.4 Concluding remarks**

This thesis contributes to our understanding of metacognition by showing that while reasoning about the self may serve different functions than reasoning about others, these aspects of reasoning appear to share common substrates. Additionally, from early in life, we reason about others in order to better understand ourselves. Taken together, the results presented here show in a variety of ways that cognition about cognition appears to underlie sophisticated learning and problem solving in children as young as 3 years.

One triumph, and puzzle, of self-directed learning is that the entity directing the learning is constantly changing. Our abilities, especially in childhood, but even into late adulthood, continuously shift. What was once difficult is now easy, and what was once easy may now be difficult. As Madeleine L’Engle wrote (1972), “A self is not something static, tied up in a pretty parcel and handed to the child, finished and complete. A self is always becoming” (p. 10). As such, while the mechanisms behind the metacognitive abilities of young children queried in this

thesis may remain relatively constant over our lifetimes, because the self and others change, the application of these mechanisms must be especially flexible.

## References

- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. *The self in social judgment, 1*, 85-106.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics, 95*, 1082-1095.
- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126.
- Asaba, M., & Gweon, H. (2016). Who should I tell? Young children correct and maintain others' beliefs about the self. *Proceedings of the 38<sup>th</sup> Annual Conference of the Cognitive Science Society*.
- Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development, 7*, 143-163.
- Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology, 20*, 131-144.
- Ávila, L. T., Chiviawowsky, S., Wulf, G., & Lewthwaite, R. (2012). Positive social-comparative feedback enhances motor learning in children. *Psychology of Sport and Exercise, 13*, 849-853.
- Azmitia, M. (1988). Peer interaction and problem solving: When are two heads better than one? *Child Development, 59*, 87-96.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110-118.
- Baker, C.L., Saxe, R., & Tenenbaum, J.B. (2009). Action understanding as inverse planning. *Cognition, 113*, 329-349.

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bedny, M., Pascual-Leone, A., & Saxe, R. R. (2009). Growing up blind does not change the neural bases of Theory of Mind. *Proceedings of the National Academy of Sciences, 106*, 11312-11317.
- Bélangier, M. J., Atance, C. M., Varghese, A. L., Nguyen, V., & Vendetti, C. (2014). What will I like best when I'm all grown up? Preschoolers' understanding of future preferences. *Child Development, 85*, 2419-2431.
- Bell, J. A., & Livesey, P. J. (1985). Cue significance and response regulation in 3-to 6-year-old children's learning of multiple choice discrimination tasks. *Developmental Psychobiology, 18*, 229-245.
- Benenson, J. F., & Dweck, C. S. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development, 57*, 1179-1187.
- Berger, S. E., & Adolph, K. E. (2003). Infants use handrails as tools in a locomotor task. *Developmental Psychology, 39*, 594.
- Berger, S. E., Adolph, K. E., & Kavookjian, A. E. (2010). Bridging the gap: solving spatial means–ends relations in a locomotor task. *Child Development, 81*, 1367-1375.
- Bernard, S., Clément, F., & Mercier, H. (2016). Wishful thinking in preschoolers. *Journal of Experimental Child Psychology, 141*, 267-274.
- Bertenthal, B. I., & Fischer, K. W. (1978). Development of self-recognition in the infant. *Developmental Psychology, 14*, 44.
- Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018-1034.
- Bjork, R. A., Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*.

- Cambridge, MA: MIT Press.
- Bjorklund, D. F., & Green, B. L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist*, *47*, 46.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246-263.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*, 161-186.
- Boesch, C., & Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tai National Park. *American Journal of Physical Anthropology*, *78*, 547-573.
- Bonawitz, E. B., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*, 215-234.
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*, 322-330.
- Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom*. ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education.
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, *42*, 500-513.
- Boseovski, J. J., Chiu, K., & Marcovitch, S. (2013). Integration of behavioral frequency and intention information in young children's trait attributions. *Social Development*, *22*, 38-57.
- Bradford, E. E., Jentsch, I., & Gomez, J. C. (2015). From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition*, *138*, 21-34.

- Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2016). Children consider others' expected costs and rewards when deciding what to teach. *In Proceedings of the 38th annual conference of the cognitive science society* (pp. 559-564). Cognitive Science Society.
- Brim, O.G. (1976). Life-span development of the theory of oneself: Implications for child development. In H. W. Reese (Ed.) *Advances in Child Development and Behavior* (Vol. 11). New York: Academic Press.
- Brody, G. H., Stoneman, Z., & MacKinnon, C. E. (1982). Role asymmetries in interactions among school-aged children, their younger siblings, and their friends. *Child Development, 53*, 1364-1370.
- Brosseau-Liard, P., Cassels, T., & Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in accurate versus confident speakers. *PloS One, 9*, e108308.
- Brosseau-Liard, P. E., & Birch, S. A. (2010). 'I bet you know more and are nicer too!': what children infer from others' accuracy. *Developmental Science, 13*, 772-778.
- Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*, 353-376.
- Brownell, C. A., & Carriger, M. S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Development, 61*, 1164-1174.
- Brownell, C. A., Ramani, G. B., & Zerwas, S. (2006). Becoming a Social Partner With Peers: Cooperation and Social Understanding in One-and Two-Year-Olds. *Child Development, 77*, 803-821.
- Brownell, C., Nichols, S. R., & Svetlova, M. (2013). Converging developments in prosocial behavior and self-other understanding in the second year of life: the second social-cognitive

- revolution. *Navigating the Social World: What Infants, Children, and Other Species Teach Us*, 385-390.
- 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*, 331-340.
- Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*, 287-314.
- Burhans, K. K., & Dweck, C. S. (1995). Helplessness in early childhood: The role of contingent worth. *Child Development*, *66*, 1719-1738.
- Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development*, *10*, 67-74.
- Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, *7*, 3-15.
- Butler, R. (1989a). Mastery versus ability appraisal: A developmental study of children's observations of peers' work. *Child Development*, *60*, 1350-1361.
- Butler, R. (1989b). Interest in the task and interest in peers' work in competitive and noncompetitive conditions: A developmental study. *Child Development*, *60*, 562-570.
- Butler, R. (1998). Age trends in the use of social and temporal comparison for self-evaluation: Examination of a novel developmental hypothesis. *Child Development*, *69*, 1054-1073.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*, 337-342.
- Butterworth, G. (1992). Origins of self-perception in infancy. *Psychological Inquiry*, *3*, 103-111.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge: MIT Press.



- Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*, 235-251.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616.
- Carpendale, J.I. & Chandler, M.J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686-1706.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315-330.
- Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding “prior intentions” enables two-year-olds to imitatively learn a complex task. *Child Development, 73*, 1431-1441
- Carruthers, P. (2011). *The opacity of mind: an integrative theory of self-knowledge*. Oxford, UK: Oxford University Press..
- Chandler, M. J., & Carpendale, J. I. (1998). Inching toward a mature theory of mind. *Self-Awareness: Its nature and development*, 148-190.
- Cimpian, A. (2010). The impact of generic language about ability on children's achievement motivation. *Developmental Psychology, 46*, 1333-1340.
- Cimpian, A. (2017). Early reasoning about competence is not irrationally optimistic, nor does it stem from inadequate cognitive representations. *Handbook of competence and motivation (2nd edition): Theory and application*, 387-407.
- Cimpian, A., Arce, H. M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science, 18*, 314-316.
- Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to

- a social category undermines children's achievement. *Psychological Science*, 23, 533-541.
- Cohn, M.A. (2004). Rescuing our heroes: Positive perspectives on upward comparisons in relationships, education, and work. In P.A. Linley & S. Joseph (Eds.), *Positive Psychology in Practice* (pp. 218-237). Hoboken, NJ: John Wiley & Sons.
- Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, 120, 341-349.
- Cooper, C. R. (1980). Development of collaborative problem solving among preschool children. *Developmental Psychology*, 16, 433-440.
- Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow preschoolers prefer nondissenters as informants. *Psychological Science*, 20, 372-377.
- Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, 12, 426-437.
- Cox, M. T. (2005). Metacognition in computation: A selected research review. *Artificial Intelligence*, 169, 104-141.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13, 148-153.
- Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, 27, 111-133.
- Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling-of-knowing judgments. *Child Development*, 54, 1480-1486.
- D'Andrade, R. (1987). A folk model of the mind. In Holland D., Quinn, N. (Eds.), *Cultural Models in Language and Thought*. Cambridge, UK: Cambridge University Press.
- Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist?: Developmental shifts in ways of clustering knowledge. *Child Development*, 75, 918-931.

- De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*, 1037-1060.
- Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences, 7*, 527-533.
- Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Personality and Social Psychology Bulletin, 7*, 79-83.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*, 355-361.
- Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130*, 335-347.
- Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition, 126*, 285-300.
- Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5-to 7-year-old children. *Journal of Experimental Child Psychology, 126*, 213-228.
- Diamond, A. (2006). The early development of executive functions. *Lifespan Cognition: Mechanisms of change, 210*, 70-95.
- DiYanni, C., Nini, D., & Rheel, W. (2011). Looking good versus doing good: Which factors take precedence when children learn about new tools? *Journal of Experimental Child Psychology, 110*, 575-591.
- Donaldson, M. L. (1986). *Children's explanations: A psycholinguistic study*. Cambridge,

- England: Cambridge University Press.
- Droege, K. L., & Stipek, D. J. (1993). Children's use of dispositions to predict classmates' behavior. *Developmental Psychology, 29*, 646.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374-380.
- Dweck, C. (2006). *Mindset: The new psychology of success*. New York: Random House.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.
- Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review, 23*, 660-679.
- Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. Mineola, NY: Dover Publications.
- Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development, 64*, 830-847.
- Edwards, C. P., & Lewis, M. (1979). Young children's concepts of social relations: Social functions and social objects. In *The child and its family* (pp. 245-266). New York, NY: Springer US.
- Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist, 28*, 404.
- Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and

- metacognitive regulation. *Consciousness and Cognition*, 9, 288-307.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34, 906.
- Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1, 324-340.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, 1, 97-112.
- Fletcher, G. E., Warneken, F., & Tomasello, M. (2012). Differences in cognitive processes underlying the collaborative activities of children and chimpanzees. *Cognitive Development*, 27, 136-153.
- Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111, 8410-8415.
- Frey, K. S., & Ruble, D. N. (1985). What children say when the teacher is not around: Conflicting goals in social comparison and performance assessment in the classroom. *Journal of Personality and Social Psychology*, 48, 550-562.
- Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2016). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science*, 19, 275-283.
- Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate:

- Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology*, *110*, 561-574.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis (3rd edition)*. London, England: Chapman & Hall/CRC Press.
- Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, *78*, 757-774.
- Gentile, B., Twenge, J. M., & Campbell, W. K. (2010). Birth cohort differences in self-esteem, 1988–2008: A cross-temporal meta-analysis. *Review of General Psychology*, *14*, 261.
- Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7*, 287-292.
- Gergely, G., Nádasdy, Z., Csibra, G., & Biró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165-193.
- Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives*, *7*, 160-165.
- Giles, J. W., & Heyman, G. D. (2003). Preschoolers' beliefs about the stability of antisocial behavior: Implications for navigating social challenges. *Social Development*, *12*, 182-197.
- Giuri, P., Ploner, M., Rullani, F., & Torrìsi, S. (2010). Skills, division of labor and performance in collective inventions: Evidence from open source software. *International Journal of Industrial Organization*, *28*, 54-68.
- Goodman, N., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. (2008). Church: a language for generative models with non-parametric memoization and approximate inference. In *Uncertainty in Artificial Intelligence*.

- Goodman, N.D., & Tenenbaum, J.B. (electronic). Probabilistic Models of Cognition. Retrieved from <http://probmods.org>.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality, *Brain and Behavioral Sciences*, *16*, 1-14.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26-37.
- Gopnik, A., & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development*, *62*, 98-110.
- Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*, 145-171.
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085-1108.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620.
- Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, *113*, 3492-3496.
- Green, V. A., & Recheis, R. (2006). Children's cooperative and competitive interactions in limited resource situations: A literature review. *Journal of Applied Developmental Psychology*, *27*, 42-

59.

- Grocke, P., Rossano, F., & Tomasello, M. (2015). Procedural justice in children: Preschoolers accept unequal resource distributions if the procedure provides equal opportunities. *Journal of Experimental Child Psychology, 140*, 197-210.
- Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent praise to 1-to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development, 84*, 1526-1541.
- Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science, 7*, 464-481.
- Gweon & Schulz (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332*, 1524.
- Gweon, H., & Saxe, R. (2013). Developmental cognitive neuroscience of theory of mind. In *Neural Circuit Development and Function in the Brain* (pp. 367-377).
- Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition, 132*, 335-341.
- Gweon, H., Tenenbaum, J., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 9066-9071.
- Gweon, H., & Schulz, L.E. (in press). From exploration to instruction: Children learn from exploration and tailor their demonstrations to observers' goals and competence. *Child Development*.
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665-668



- Happé, F. (2003). Theory of mind and the self. *Annals of the New York Academy of Sciences, 1001*, 134-144.
- Harris, P. L. (1994). Thinking by children and scientists: False analogies and neglected similarities. *Mapping the mind: Domain specificity in cognition and culture, 294*.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.
- Harter, S. (1982). The perceived competence scale for children. *Child development, 87-97*.
- Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.
- Hay, D. F. (1979). Cooperative interactions and sharing between very young children and their parents. *Developmental Psychology, 15*, 647.
- Heatherston, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience, 1*, 18-25.
- Heckhausen, H. (1984). Emergent achievement behavior: Some early developments. *Advances in Motivation and Achievement, 3*, 1-32.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, N.Y.: Wiley.
- Hembacher, E., & Ghetti, S. (2014). Don't Look at My Answer Subjective Uncertainty Underlies Preschoolers' Exclusion of Their Least Accurate Memories. *Psychological Science, 25*, 1768-1776.
- Hermes, J., Behne, T., & Rakoczy, H. (2015). The role of trait reasoning in young children's selective trust. *Developmental Psychology, 51*, 1574.
- Hewlett, B. S., Fouts, H. N., Boyette, A. H., & Hewlett, B. L. (2011). Social learning among

- Congo Basin hunter–gatherers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366, 1168-1178.
- Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, 77, 1-19.
- Heyman, G. D., Dweck, C. S., & Cain, K. M. (1992). Young children's vulnerability to self-blame and helplessness: Relationship to beliefs about goodness. *Child Development*, 63, 401-415.
- Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development*, 74, 516-534.
- Jara-Ettinger, E., Baker, C.L., & Tenenbaum, J.B. (2012). Learning what is where from social observations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*, 515-520.
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, 20, 589-604.
- Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, 140, 14-23.
- Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science*, 26, 633-640.
- Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, 61, 248-272.
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, 17, 757-758.

- Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*, 1541-1547.
- Johnson-Pynn, J. S., & Nisbet, V. S. (2002). Preschoolers effectively tutor novice classmates in a block construction task. *Child Study Journal, 32*, 241-256.
- Johnson, S. C., Booth, A., & O'Hearn, K. (2001). Inferring the goals of a nonhuman agent. *Cognitive Development, 16*, 637-656.
- Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: implications for contingent self-worth and coping. *Developmental Psychology, 35*, 835-847.
- Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*, 449-460.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One, 7*, e36399.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*, 1795-1804.
- Kircher, T. T., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., ... & David, A. S. (2000). Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognitive Brain Research, 10*, 133-144.
- Knudsen, B., & Liskowski, U. (2012). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy, 17*, 672-691.
- Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition, 87*, 179-208.
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261-1277.

- Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development, 82*, 1634-1647.
- Koenig, M. A., & Sabbagh, M. A. (2013). Selective social learning: New perspectives on learning from others. *Developmental Psychology, 49*, 399.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694-698.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349.
- Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In *Metacognition* (pp. 1-17). New York, NY: Springer US.
- Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition, 94*, 95-103.
- Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?" The distinction between causal knowledge and word knowledge guides preschoolers' selective requests for information. *Developmental Psychology, 49*, 446-453.
- Kushnir, T., Xu, F. & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*, 1134-1140.
- L'Engle, M. (1972). *A Circle of Quiet: The Crosswicks Journal, Book One*. New York, NY: Farrar, Straus & Giroux.
- Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology, 20*, 163-198.
- Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development, 84*, 1253-

1268.

- Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, 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.
- Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science, 357*, 1290-1294.
- Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological review, 94*, 412.
- Lewis, P. R., Platzner, M., Rinner, B., Tørresen, J., & Yao, X. (Eds.). (2016). *Self-aware computing systems: An engineering approach*. New York, NY: Springer US.
- Lewthwaite, R., & Wulf, G. (2010). Social-comparative feedback affects motor skill learning. *The Quarterly Journal of Experimental Psychology, 1*, 1-12.
- Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology, 49*, 1505.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*, 732-739.
- Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of Young Children's Trait Understanding: Behavior-to-Trait Inferences and Trait-to-Behavior Predictions. *Child Development, 78*, 1543-1558.
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science, 358*, 1038-1041.
- Lockhart, K. L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of

- traits: Protective optimism? *Child Development*, 73, 1408-1430.
- Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, 73, 1073-1084.
- Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, 104, 19751-19756.
- Lyons, K. E., & Ghetti, S. (2010). Metacognitive development in early childhood: New questions about old assumptions. In *Trends and Prospects in Metacognition Research* (pp. 259-278). New York, NY: Springer US.
- Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, 82, 1778-1787.
- Magid & Schulz (2015). Quit while you're ahead: Preschoolers persistence and willingness to accept challenges are affected by social comparison. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*.
- Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language*, 51, 458-472.
- Markant, D. B., Ruggeri, A., Gureckis, T. M., & Xu, F. (2016). Enhanced memory as a common effect of active learning. *Mind, Brain, and Education*, 10, 142-152.
- Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology*, 83, 377.
- Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child Development*, 69, 1030-

1053.

- Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38*, 376.
- Martin, A., & Olson, K. R. (2013). When kids know better: paternalistic helping in 3-year-old children. *Developmental Psychology, 49*, 2071.
- Martin, A., Lin, K., & Olson, K. R. (2016). What You Want Versus What's Good for You: Paternalistic Motivation in Children's Helping Behavior. *Child Development, 87*, 1739-1746.
- Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122*, 47.
- Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1263.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18*, 196-204.
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838.
- Meltzoff, A. N. (2007). 'Like me': a foundation for social cognition. *Developmental Science, 10*, 126-134.
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*, 530.
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463-477.

- Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*, 225-229.
- Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*, 722-730.
- Morris, W. N., & Nemcek Jr, D. (1982). The development of social comparison motivation among preschoolers: Evidence of a stepwise progression. *Merrill-Palmer Quarterly*, 413-425.
- Morse, S., & Gergen K.J. (1970). Social comparison, self-consistency, and the concept of the self. *Journal of Personality and Social Psychology*, *16*, 148.
- Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2005). On the specificity of the relation between executive function and children's theories of mind.
- Mostache, H.S. & Bragonier, P. (1981). An observational study of social comparison in preschoolers. *Child Development*, *52*, 376-378.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*, 33.
- Myers, L. J., & Liben, L. S. (2012). Graphic symbols as “the mind on paper”: Links between children’s interpretive theory of mind and symbol understanding. *Child Development*, *83*, 186-202.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, *70*, 680.
- Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children’s sequential information search is sensitive to environmental probabilities. *Cognition*, *130*, 74-80.



- Nelson, T. O. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, *122*, 269–273.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of experimental psychology: Learning, Memory, and Cognition*, *14*, 676.
- Nelson, T. O., Dunlosky, K. J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*, 207–213
- Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development*, *49*, 800-814.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*, 328.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737.
- O'Neill, D. K. (1996). Two-Year-Old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659-677.
- Onishi, K.H. & Baillargeon, R. (2005). Do 15-month-old-infants understand false beliefs? *Science*, *308*, 255-258.
- Parsons, J. E., & Ruble, D. N. (1977). The development of achievement-related expectancies. *Child Development*, *48*, 1075-1079.
- Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*, 79-85.

- Pepitone, E. A. B. (Ed.). (1980). *Children in cooperation and competition: Toward a developmental social psychology*. Lanham, MD: Lexington books.
- Perner, J., 1993. *Understanding the Representational Mind*. The MIT Press, Cambridge, MA
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125-137.
- Pillow, B. H., & Mash, C. (1999). Young children's understanding of interpretation, expectation and direct perception as sources of false belief. *British Journal of Developmental Psychology*, 17, 263-276.
- Pomerantz, E. M., Ruble, D. N., Frey, K. S., & Greulich, F. (1995). Meeting goals and confronting conflict: Children's changing perceptions of social comparison. *Child Development*, 66, 723-738.
- Pratt, C. & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, 61, 973-982.
- Rai, R., & Mitchell, P. (2006). Children's ability to impute inferentially based knowledge. *Child Development*, 77, 1081-1093.
- Reyna, V. F. (1996). Conceptions of memory development with implications for reasoning and decision making. *Annals of Child Development*, 12, 87-118.
- Rhodes, M., & Brickman, D. (2008). Preschoolers' responses to social comparisons involving relative failure. *Psychological Science*, 19, 968-972.
- Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, 55, 550-560.
- Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, 9, 1027.

- Rogoff, B. (1990). Apprenticeships in thinking. *Cognitive development in social context*. Oxford, England: Oxford University Press.
- Ross, H. S., & Lollis, S. P. (1987). Communication within infant social games. *Developmental Psychology, 23*, 241.
- Ross, H. S., Recchia, H. E., & Carpendale, J. I. (2005). Making sense of divergent interpretations of conflict and developing an interpretive understanding of mind. *Journal of Cognition and Development, 6*, 571-592.
- Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process I. In *Advances in experimental social psychology*, Vol. 10, pp. 173-220). Academic Press.
- Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebel, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology, 16*, 105-115.
- Ruble, D. N., Eisenberg, R., & Higgins, E. T. (1994). Developmental Changes in Achievement Evaluation: Motivational Implications of Self-Other Differences. *Child Development, 65*, 1095-1110.
- Ruble, D. N., Parsons, J. E., & Ross, J. (1976). Self-evaluative responses of children in an achievement setting. *Child Development, 47*, 990-997.
- Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201-224.
- Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on Scientific Thinking: Children's Understanding of the Hypothesis-Evidence Relation. *Child Development, 64*, 1617-1636.

- Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology, 52*, 2159.
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054-1070.
- Saxe, R. (2009). The neural evidence for simulation is weaker than I think you think it is. *Philosophical Studies, 144*, 447-456.
- Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience, 1*, 229-234.
- Schmidt, M. F., Svetlova, M., Johe, J., & Tomasello, M. (2016). Children's developing understanding of legitimate reasons for allocating resources unequally. *Cognitive Development, 37*, 42-52.
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science, 1*, 291-297.
- Schulz, L. E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences, 16*, 382-389.
- Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*, 1045.
- Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology, 44*, 1266.
- Schwanenflugel, P. J., Fabricius, W. V., & Noyes, C. R. (1996). Developing organization of

- mental verbs: Evidence for the development of a constructivist theory of mind in middle childhood. *Cognitive Development*, 11, 265-294.
- Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002).
- Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, 1, 4-14.
- Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, 22, 878-880.
- 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.
- Shapiro, R. Y., & Bloch-Elkon, Y. (2008). Do the facts speak for themselves? Partisan disagreement as a challenge to democratic competence. *Critical Review*, 20, 115-139.
- Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, 450, 102-105.
- Siegel, M. H., Magid, R.W., Tenenbaum, J. B., & Schulz, L. E. (2014; in prep). Black boxes: Hypothesis testing via indirect perceptual evidence. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*.
- Siegel, M. H., Tenenbaum, J.B., & McDermott, J. (in prep). Inferring the number of objects from collision sounds.
- Siegel, M.H. & McDermott, J.H. (2014). Inferring the Number of Objects from Collision Sounds. Poster presented at the Association for Research in Otolaryngology. Baltimore, MD.
- Skerry, A. E., Carey, S.E., & Spelke, E.S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of*

- Sciences, 110, 18728-18733.*
- Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. *Child Development, 65, 1723-1743.*
- Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58, 424-433.*
- Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62, 468-483.*
- Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two-to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology, 3, 222-237.*
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96, B1-B11.*
- Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science, 30, 759-774.*
- Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology, 21, 195.*
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10, 89-96.*
- Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348, 91-94.*
- Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development, 11, 176-190.*
- Stipek, D. J. (1981). Children's perceptions of their own and their classmates' ability. *Journal of Educational Psychology, 73, 404-410.*

- Stipek, D. J. (1984). The development of achievement motivation. *Research on Motivation in Education*, 1.
- Stipek, D. J., & Hoffman, J. M. (1980). Development of children's performance-related judgments. *Child Development*, 51, 912-914.
- Stipek, D. J., Gralinski, J. H., & Kopp, C. B. (1990). Self-concept development in the toddler years. *Developmental Psychology*, 26, 972.
- Stipek, D., Recchia, S., McClintic, S., & Lewis, M. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development*, i-95.
- Sussan, D., & Son, L. K. (2007). The training of metacognitive monitoring in children. *Columbia Undergraduate Science Journal*, 2, 98-109.
- Swallow, S. R., & Kuiper, N. A. (1992). Mild depression and frequency of social comparison behavior. *Journal of Social and Clinical Psychology*, 11, 167-180.
- Tamis-LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental Psychology*, 44, 734.
- Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development*, 62, 1334-1351.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331, 1279-1285.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of experimental psychology: Learning, Memory, and Cognition*, 25, 1024.

- Thorndike, E. L. (1911). *Individuality*. Boston, MA: Houghton Mifflin.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*, 509-529.
- van Schijndel, T. J., Visser, I., van Bers, B. M., & Raijmakers, M. E. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104-119.
- VanderBorgh, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development*, *18*, 61-71.
- Vanderwal, T., Hunyadi, E., Grupe, D. W., Connors, C. M., & Schultz, R. T. (2008). Self, mother and abstract other: an fMRI study of reflective social processing. *Neuroimage*, *41*, 1437-1446.
- Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, *15*, 159-176.
- Veroff, J. (1969). Social comparison and the development of achievement motivation. In C. P. Smith (Ed.), *Achievement related motives in children* (pp.46-101). New York: Russell Sage.
- Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills metacognition in the numerical domain. *Psychological Science*, *25*, 1712-1721.
- Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599-637.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*, 34-41.
- Walton G.E., Bower N.J., Bower T.G. (1992). Recognition of familiar faces by newborns. *Infant*



*Behavior and Development* 15, 265–269.

Warneken, F. (2015). Precocious prosociality: Why do young children help? *Child Development Perspectives*, 9, 1-6.

Warneken, F. (2017). How children solve the two challenges of cooperation. *Annual Review of Psychology*.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, 11, 271-294.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, 77, 640-663.

Warneken, F., Steinwender, J., Hamann, K., & Tomasello, M. (2014). Young children's planning in a collaborative problem-solving task. *Cognitive Development*, 31, 48-58.

Watson, J. S., & Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly of Behavior and Development*, 18(3), 219-227.

Wellman, H. M. (1983). Metamemory revisited. In *Trends in Memory Development research* (Vol. 9, pp. 31-51). Basel, Switzerland: Karger Publishers.

Wellman, H. M. (1985). The origins of metacognition. Pp. 1-32 in *Metacognition, cognition, and human performance*, Vol. 1: Theoretical perspectives, edited by D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller. London: Academic Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72, 655-684.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *A Vol. in the educational psychology series. Development of achievement*

- motivation* (pp. 91-120). San Diego, CA, US: Academic Press.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451-469.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition 13*, 103-128.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-regulation* (pp. 531-566).
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.
- Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. *Intentions and intentionality: Foundations of Social Cognition*, 149-169.
- Xu, F. & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-olds infants, *Cognition, 112*, 97-104.
- Xu, F. & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 5012-5015.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.
- Zaitchik, D. (1991). Is only seeing really believing?: Sources of the true belief in the false belief task. *Cognitive Development, 6*, 91-103.
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15*, 41-44.

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*, 307-313.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics, 85*, 9-23.

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development, 19*, 457-477.

## Appendix I. Chapter 2 Supporting Information

### A1.1 Computational Model

To help clarify our proposal and specify what counts as “rational inference” in these contexts, we developed a computational model that provides quantitative predictions for each experimental condition. The model specifies how a rational agent would behave when presented with the same task that we gave our participants. Although many studies have used Bayesian models to assess children’s ability to update their *own* prior beliefs from data (see Gopnik & Wellman, 2012; Schulz, 2012; and Tenenbaum et al., 2011 for reviews) to our knowledge, this is the first attempt to consider children’s ability to predict when *another* agent will (or will not) change his mind by considering both that agent’s access to the data and his prior beliefs. Finally, note that in suggesting that children’s rational inferences on behalf of a third party can be captured by a Bayesian inference model, we do not mean to suggest that children have conscious, meta-cognitive access to these computations; rather, we suggest that such sophisticated computations may underlie the many implicit, rapid, accurate judgments that support everyday social cognition. Figure 4 in the Main Text displays the predictions of our model for each of the candidate hypotheses of Table 1 in the Main Text.

The model is specified at two levels. First, we built a model of the Frog as a rational learner, given the information that he has available to him. Then, we modeled children’s rational inferences about the Frog. Thus two levels of rational inference are represented: the Frog’s beliefs about the location of the box, and the child’s beliefs about the Frog’s beliefs.

We adopt a Bayesian framework for modeling both these levels of rational inference. Bayesian inference models a learning event as an interaction of two factors: the agent’s prior beliefs about a hypothesis, before seeing new data:  $p(h)$ , and the probability that the hypothesis is

true given the newly observed data, the likelihood  $p(D | h)$ . These combine to yield the agent's updated posterior belief  $p(h | D)$ . Given new data bearing on a hypothesis, Bayes' rule specifies how a rational agent should update her beliefs as:

$$p(h | D) \propto p(D | h)p(h).$$

We now turn to the model of the Frog's inference, from the perspective of an ideal observer (which we can consider the child as approximating). On each experimental trial, the experimenter draws ducks from the Duck box, either randomly or selectively, and the boxes may or may not have been switched. At that point, both the child and the Frog know whether the sample is drawn randomly or selectively but only the child knows whether the boxes have been switched. However, the Frog has some prior belief  $p_{switch}$  about whether the boxes were switched in his absence (given the demonstration that they can be switched), which is equivalent to having a prior belief about which box the ducks are being drawn from. We can specify these as  $p(h_{duck}) = p_{switch}$  for the Duck box and  $p(h_{ball}) = 1 - p_{switch}$  for the Ball box. The Frog must integrate this prior belief with his observation of three (or five) ducks being drawn from the box. Under random sampling, the probability of drawing  $n$  ducks and zero balls, with replacement<sup>6</sup>, from the duck box is  $h_{duck} = \left(\frac{45}{60}\right)^n = \left(\frac{3}{4}\right)^n$ ; similarly the probability of drawing  $n$  ducks from the ball box is  $h_{ball} = \left(\frac{1}{4}\right)^n$  (these quantities specify the likelihood of the data given each hypothesis). Under selective sampling, the experimenter explicitly reached into the box to pull out a duck and thus

---

<sup>6</sup> While the experiment used sampling without replacement, our model used sampling with replacement because the analysis is conceptually simpler and for large populations (i.e., the 60 objects in the box here) the difference between the distributions underlying sampling with and without replacement is negligible.

the probability is 1 from each box. The posterior beliefs  $p(h_{duck} | D)$  and  $p(h_{ball} | D)$  are then given by Bayes' theorem above:

$$p(h_{duck} | D) \propto p_{switch} * \left(\frac{3}{4}\right)^n \quad \text{and} \quad p(h_{ball} | D) \propto p_{switch} * \left(\frac{1}{4}\right)^n \quad (1)$$

in the case of random sampling, and

$$p(h_{duck} | D) \propto p_{switch} \quad \text{and} \quad p(h_{ball} | D) \propto p_{switch} \quad (2)$$

in the case of selective sampling.

Thus we have a posterior distribution over the two hypotheses, where the posterior probability that the sample is drawn from the Duck box increases as the number of randomly sampled ducks increases, and remains equal to the prior under selective sampling. This reflects our intuition that the evidence is stronger with each new randomly sampled duck and unchanged with each selectively sampled duck.

Having specified a rational model of the Frog's inference, we now describe our model of the experimental participants. We propose that children can approximately simulate the above inference, and when asked to say where they think the Frog thinks the duck box is, they report the output of this computation, subject to two approximations. As is standard practice when modeling behavioral responses (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Gweon, Tenenbaum, & Schulz, 2010; Xu & Tenenbaum, 2007), we assume that children probability match; that is, the frequency with which they select responses is proportional to the posterior probability of each hypothesis. In a population of participants, this rule gives a distribution of responses that mimics the distribution of posterior beliefs, and it is an efficient scheme for approximating probabilistic inference (Vul, Goodman, Griffiths, & Tenenbaum, 2014). We also consider the possibility that on each trial, there is a nonzero probability that children may have been inattentive or confused. We therefore include a noisy response parameter,  $p_{error}$ , estimating

the probability that a participant gives a box choice selected uniformly at random, instead of the response predicted by the model. Thus our model at this point has two parameters: the Frog's prior belief,  $p_{switch}$ , about whether the boxes were switched, and the noisy response parameter,  $p_{error}$ . To estimate  $p_{switch}$ , we used the ratio of children's responses on the initial belief question:

$$p_{switch} = \frac{28}{178} = .157$$

We have no analogous way to derive a plausible independent and numerically precise estimate of  $p_{error}$ . For the results displayed in Figure 4, we set  $p_{error} = .25$ ; as children had to pass two inclusion checks, at most 25% of included children could have been answering at chance.

While we have described our model mathematically, it is possible to implement this model implicitly by simple sampling operations, without making any explicit statistical calculations. We describe one such implementation written in the probabilistic programming language Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008; Goodman & Tenenbaum, 2014).

We implemented the Rational Learning model, and all of the alternative models presented in Table 1 in the probabilistic programming language Church (Goodman et al., 2008; Goodman & Tenenbaum, 2014). We used the webchurch implementation, available at <https://github.com/probmods/webchurch> or interactively at <https://probmods.org/play-space.html>. To evaluate the following Church code, copy and paste the code text into the environment available in the latter link.

The following code block is sufficient to reproduce all of the model predictions described here; to obtain the predictions for individual conditions given a specified model, modify the variables num-draws, actual-switch, and sampling-manner as described in the text. To obtain the

predictions of different (alternative) models, replace the uncommented code with the appropriate commented code below.

```
;; UTILITIES
(define (count-satisfying x l)
  (length (filter (lambda (r) (equal? r x)) l)))

(define (percentage-satisfying x l)
  (/ (count-satisfying x l) (length l)))

(define (aggregate-samples lst)
  (list '(left right) (list (percentage-satisfying 'left lst) (percentage-satisfying 'right lst))))

;; BEGIN MODEL SPECIFICATION

(define duck-box-num-ducks 45)
(define duck-box-num-balls 15)

(define ball-box-num-ducks 15)
(define ball-box-num-balls 45)

(define (make-number-container dd db bd bb)
  (lambda (box object)
    (case box
      (('duck) (case object
                  (('duck) dd)
                  (('ball) db)))
      (('ball) (case object
                  (('duck) bd)
                  (('ball) bb))))))

(define nums
  (make-number-container
   duck-box-num-ducks
   duck-box-num-balls
   ball-box-num-ducks
   ball-box-num-balls))
```



```
(define (random-sample-from-box box)
  (if (flip (/ (nums box 'duck) (+ (nums box 'duck) (nums box 'ball))))
      'duck
      'ball))
```

```
(define (sample-from-box box sampling-manner likes)
  (case sampling-manner
    (('random)
     (random-sample-from-box box)
    )
    (('selective)
     likes)))
```

```
(define (sample-n-from-box box sampling-manner likes n)
  (let ((f (lambda () (sample-from-box box sampling-manner likes))))
    (repeat n f)))
```

```
(define (make-boxes l r)
  '((left ,l) (right ,r)))
```

```
(define (left boxes)
  (assoc 'left boxes))
```

```
(define (right boxes)
  (assoc 'right boxes))
```

```
(define (box-majority-object box) (last box))
(define (left-box-primary-object boxes) (last (left boxes)))
(define (right-box-primary-object boxes) (last (right boxes)))
```

```
(define (which-box object boxes)
  (cond
    ((equal? (left-box-primary-object boxes) object) (left boxes))
    ((equal? (right-box-primary-object boxes) object) (right boxes))))
```

```
(define (which-side box) (first box))
```

```
(define (where-is ob boxes)
  (which-side (which-box ob boxes)))
```

```

(define (box-on-side left-or-right boxes)
  (case left-or-right
    (('left) (left boxes))
    (('right) (right boxes))))

(define (switch-boxes boxes)
  (make-boxes (right-box-primary-object boxes) (left-box-primary-object boxes)))

(define (other-box side)
  (if (equal? side 'left) 'right 'left))

;; model of the child and the experiment (external to the frog)
(define (child-sample)

  (rejection-query
   (define pre-boxes (make-boxes 'duck 'ball)) ;; arrangement of boxes before
   ;; frog leaves (observed by frog)

   (define actual-switch? true) ;; whether the boxes are switched or kept in place

   (define num-draws 3) ;; how many objects are sampled (3 or 5)
   (define sampling-manner 'random) ;; whether sampling is random or selective
   (define experimenter-preference 'duck) ;; experimenter's target during selective sampling

   (define post-boxes (if actual-switch? (switch-boxes pre-boxes) pre-boxes)) ;; box locations
   when frog returns (unknown to frog)
   (define which-box-drawn-from (where-is 'duck post-boxes)) ;; which box is being sampled
   from

   (define observed-sample (repeat num-draws (lambda () 'duck))) ;; the set of objects sampled
   (observed by frog)

   ;; child's model of the frog

   ;; BEGIN MODEL-SPECIFIC CODE – set which-model to 'a-'f to enable each alternative
   model
   (define which-model 'a)
   (define (frog-sample)
     .....
```

```

;; (a) full model
.....
(case which-model
  ('a)
  (rejection-query
    (define frog-switch-prior .159) ;; how much frog expects the 'switch' trick
    (define frog-thinks-switch? (flip frog-switch-prior)) ;; whether the frog guesses the boxes
were switched, before sample

    (define frog-belief-post-boxes (if frog-thinks-switch? (switch-boxes pre-boxes) pre-boxes))
;; where the frog thinks the boxes are, before observing sample

    (define imagined-sample (sample-n-from-box
      (box-majority-object (box-on-side which-box-drawn-from frog-belief-post-
boxes)) ;; which-box-drawn-from is directly observed
      sampling-manner ;; directly observed by frog
      experimenter-preference ;; directly observed by frog
      (length observed-sample))) ;; sample frog imagines might be seen,
conditioned on his beliefs about the boxes

    (define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes)) ;; frog's updated
belief on box locations after seeing sample

    ;; variable of interest
    frog-belief-where-duck-is

    ;; condition
    (equal? imagined-sample observed-sample)))
.....
;; (b) PRIOR BELIEF MODEL
.....
('b)
(rejection-query
  (define frog-belief-post-boxes post-boxes)

  (define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes))

  ;; variable of interest
  frog-belief-where-duck-is))

```

```

.....
;; (c) LOCATION MODEL
.....
('c)
(rejection-query
 (define frog-switch-prior .159)
 (define frog-thinks-switch? (flip frog-switch-prior))
 (define frog-belief-post-boxes (if frog-thinks-switch? (switch-boxes pre-boxes) pre-boxes))

 (define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes))
 (define imagined-sample (sample-n-from-box
   (box-majority-object (box-on-side which-box-drawn-from frog-belief-post-
boxes)) ;; which-box-drawn-from is directly observed
   sampling-manner      ;; directly observed by frog
   experimenter-preference ;; directly observed by frog
   0)) ;; does not learn from sampling, modeled as didn't see samples.
 (define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes))

;; variable of interest
frog-belief-where-duck-is))
.....
;; (d) SAMPLED DATA MODEL
.....
('d)
(case sampling-manner
 ('random)
 (rejection-query
 (define frog-switch-prior .5)
 (define frog-thinks-switch? (flip frog-switch-prior))

 (define frog-belief-post-boxes (if frog-thinks-switch? (switch-boxes pre-boxes) pre-
boxes))

 (define imagined-sample (sample-n-from-box
   (box-majority-object (box-on-side which-box-drawn-from frog-belief-
post-boxes))
   sampling-manner
   experimenter-preference
   (length observed-sample))))

```

```

(define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes))

;; variable of interest
frog-belief-where-duck-is

;; condition
(equal? imagined-sample observed-sample)))

(('selective)
(rejection-query
(define frog-switch-prior .5)
(define frog-thinks-switch? (flip frog-switch-prior))

(define frog-belief-post-boxes (if frog-thinks-switch? (switch-boxes pre-boxes) pre-
boxes))

(define imagined-sample (repeat (length observed-sample)
(lambda ()
(if (flip) 'duck 'ball))))
(define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes))
;; variable of interest
frog-belief-where-duck-is)))

.....
;; (e) RANDOM STAY, SELECTIVE SHIFT MODEL
.....
(('e)
(cond ((equal? sampling-manner 'random) which-box-drawn-from)
((equal? sampling-manner 'selective) (other-box which-box-drawn-from))))

.....
;; (f) CHANCE RESPONSE MODEL
.....
(('f)
(rejection-query
(define frog-switch-prior .5)
(define frog-thinks-switch? (flip frog-switch-prior))

(define frog-belief-post-boxes (if frog-thinks-switch? (switch-boxes pre-boxes) pre-boxes))

```

```

(define imagined-sample (repeat (length observed-sample)
                               (lambda ()
                                 (if (flip) 'duck 'ball))))
(define frog-belief-where-duck-is (where-is 'duck frog-belief-post-boxes)
 frog-belief-where-duck-is)))

;; END MODEL-SPECIFIC CODE

(define noise .25) ;; noisy response model: probability of child making a random response
(define response (if (flip noise)
                    (uniform-draw '(left right))
                    (frog-sample))) ;; child's response (possibly noisy)

response
true))

(define child-distribution (aggregate-samples (repeat 1000 child-sample)))
child-distribution

```

**Appendix II. Chapter 3 Supporting Information**

Table S1. Experiment 1 results of linear mixed effects models predicting proportion exploration time, Models 1-4.

Model	Predictors	$\beta$ Estimate	Error	<i>t</i> value
1	Discriminability	0.192	0.043	4.460
2	Trial order	-0.111	0.041	-2.724
3	Discriminability	0.180	0.042	4.255
	Trial order	-0.092	0.038	-2.437
4	Discriminability	0.118	0.042	4.227
	Trial order	-0.090	0.038	-2.381
	Marbles in box	0.011	0.022	0.503

Table S2. Experiment 2 results of linear mixed effects models predicting proportion exploration time, Models 1-4.

Model	Predictors	$\beta$ Estimate	Error	<i>t</i> value
1	Discriminability	0.119	0.026	4.561
2	Trial order	-0.017	0.040	-0.428
3	Discriminability	0.119	0.026	4.527
	Trial order	-0.012	0.036	-0.320
4	Discriminability	0.116	0.027	4.317
	Trial order	-0.013	0.037	-0.345
	Marbles in box	0.007	0.013	0.524



Table S3. Experiment 3 results of linear mixed effects models predicting proportion exploration time, Models 1-4.

Model	Predictors	$\beta$ Estimate	Error	<i>t</i> value
1	Discriminability	0.129	0.031	4.129
2	Trial order	-0.119	0.028	-4.269
3	Discriminability	0.108	0.030	3.640
	Trial order	-0.102	0.027	-3.792
4	Discriminability	0.104	0.030	3.448
	Trial order	-0.102	0.027	-3.823
	Marbles in box	-0.013	0.012	-1.082

Table S4. Experiment 4 results of linear mixed effects models predicting proportion exploration time, Models 1-4.

Model	Predictors	$\beta$ Estimate	Error	<i>t</i> value
1	Discriminability	0.130	0.032	4.135
2	Trial order	0.009	0.037	0.252
3	Discriminability	0.130	0.032	4.107
	Trial order	0.004	0.035	0.129
4	Discriminability	0.130	0.031	4.080
	Trial order	0.006	0.035	0.161
	Marbles in box	0.008	0.017	0.448

Table S5. Joint analysis from Experiments 1-4 results of linear mixed effects models predicting proportion exploration time, Models 1-4.

Model	Predictors	$\beta$ Estimate	Error	<i>t</i> value
1	Discriminability	0.130	0.016	8.380
2	Trial order	-0.059	0.018	-3.215
3	Discriminability	0.127	0.015	8.260
	Trial order	-0.051	0.017	-2.961
4	Discriminability	0.127	0.015	8.218
	Trial order	-0.050	0.017	-2.949
	Marbles in box	0.004	0.007	0.571