SEEKING SOCIAL INFORMATION DURING EARLY LANGUAGE COMPREHENSION AND WORD LEARNING

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kyle MacDonald
October 2018

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Michael C. Frank) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Hyowon Gweon)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(James McClelland)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Virginia A. Marchman)

Approved for the Stanford University Committee on Graduate Studies

Abstract

Children's rapid language development is one of the more remarkable features of human cognition. How do they learn so much despite ambiguous input and limited processing capabilities? Social learning theories argue for the importance of acquiring language from more knowledgeable adults who constrain the learning task. Statistical learning accounts emphasize the role of pattern detection and structure in children's language environment. Finally, active learning explanations focus on children's skill in seeking information to support their own learning. This thesis presents an integrative framework for unifying key ideas from these accounts, followed by a diverse set of case studies of eye movements during language processing and word learning. The empirical work highlights how children's gaze patterns can flexibly adjust to the demands of very different learning environments.

Chapter 1 uses the formalization of Optimal Experiment Design (OED) to provide an integrative account of information seeking within social contexts. Chapter 2 presents the first case study: children's eye movements during real-time processing of American Sign Language (ASL) where objects and language compete for visual attention. Chapter 3 provides a direct comparison of eye movements during signed and spoken language comprehension, proposing an information-seeking explanation of why ASL-learners are slower to shift gaze away from their social partners and to named objects. Chapters 4 and 5 generalize the information seeking explanation to the domain of novel word learning. Chapter 4 presents a series of large-scale word learning experiments, showing that the presence of social cues can change how adults distribute attention and memory during statistical learning. Finally, Chapter 5 presents an eye tracking study measuring how children's decisions to gather visual information about social partners vs. objects changes as a function of their knowledge of word-object links.

Dedication

This dissertation is dedicated to the memory of my Grammy, Sheila Paget. Thank you for being my greatest supporter and for always encouraging me to ask questions.

Acknowledgments

This dissertation would not have been possible without the support of the Deaf community. I am immensely grateful to the California School for the Deaf in Fremont, CA and to the families who participated in this research. Thanks to Karina Pedersen, Lisalee Egbert, Laura Petersen, Michele Berke, Sean Virnig, Pearlene Utley, and Rosa Lee Timm. I thank the developmental community, my labmates, and my friends Kara Weisman, MH Tessler, Erica Yoon. I thank Shane Blau, Kat Adams, Melanie Ashland for helpful discussion. I am grateful to the members of the Language and Cognition, Language Learning,and Social Cognition labs. I thank my parents, Morag and Richard, for their unconditional love and support and for their contributions to several key ideas in my disseration. I thank my sister, Kaitlin, for being there when I needed to someone to talk to. Finally, I thank my partner, Ellika, for everything – my days begin and end with you and for this I couldn't be more grateful.

Contents

Abstract **iv**

Dedication **v**

Acknowledgments **vi**

Introduction **1**

1 Active learning within social contexts **3**

1.1 Introduction 3

1.1.1 The scope of the integrative account 5

1.2 Part I: Learning from other people 7

1.2.1 Social contexts enhance attention and memory 7

1.2.2 Social contexts provide "good" information 9

1.2.3 Social contexts shape inferences and generalization 10

1.2.4 Social learning summary 12

1.3 Part II: Learning on your own 13

1.3.1 Active contexts enhances attention and memory 13

1.3.2 Active contexts provide "good" information 15

1.3.3 Can children select "good" information? 17

1.3.4 Active contexts shape inferences and generalization 19

1.3.5 Active learning summary 20

1.4 Part III: A formal account of active learning 21

1.4.1 Evidence of OED-like reasoning in human behavior 27

1.5 Part IV: Active learning within social contexts 30
1.5.1 Goals 32
1.5.2 Hypotheses 40
1.5.3 Queries 45
1.5.4 Answers 51
1.5.5 Stopping rules 56
1.6 Conclusions 58

2 Real-time American Sign Language comprehension 61
2.1 Introduction 61
2.1.1 ASL processing in adults 63
2.1.2 Lexical development in ASL 64
2.1.3 Research questions 65
2.2 Methods 65
2.2.1 Measures 67
2.2.2 Procedure 67
2.2.3 Stimuli 68
2.2.4 Trial Structure 69
2.2.5 Calculating measures of language processing efficiency 70
2.2.6 Analysis Plan 71
2.3 Results 73
2.3.1 Overview of looking behavior during real-time ASL comprehension 73
2.3.2 Evidence that eye movements during ASL processing index incremental sign comprehension 75
2.3.3 Real-time ASL comprehension in deaf and hearing children and deaf adults . 76
2.3.4 Links between children's age and efficiency in incremental sign comprehension 79
2.3.5 Links between children's incremental sign comprehension and productive vocabulary 80
2.4 Discussion 80
2.4.1 Limitations and open questions 82
2.5 Conclusion 84

3 Information seeking eye movements during language comprehension **85**

3.1 Introduction 85

4 Social information and word learning **86**

4.1 Introduction 86

4.2 Experiment 1 89

4.2.1 Method 89

4.2.2 Results and Discussion 91

4.3 Experiment 2 95

4.3.1 Method 96

4.3.2 Design and Procedure 96

4.3.3 Results and Discussion 96

4.4 Experiment 3 98

4.4.1 Method 99

4.4.2 Results and Discussion 100

4.5 Experiment 4 103

4.5.1 Method 104

4.5.2 Design and Procedure 104

4.5.3 Results and Discussion 105

4.6 General Discussion 106

4.6.1 Relationship to previous work 107

4.6.2 Limitations 109

4.7 Conclusions 111

5 Seeking social and statisical information during word learning **118**

Conclusion **119**

A Supplementary materials for Chapter 2 **120**

A.1 Model Specifications 120

A.1.1 Accuracy 121

A.1.2 Reaction Time 122

A.2 Sensitivity Analysis: Prior Distribution and Window Selection 123
A.3 Parallel set of non-Bayesian analyses 124
A.4 Analyses of phonological overlap and iconicity 125

B Supplementary materials for Chapter 4 **129**

References **136**

List of Tables

1.1 High level summary of three different paths through which active and social contexts influence learning. See Part I (social learning) and Part 2 (active learning) for more details and reviews of the behavioral evidence. 12

2.1 Age of ASL-learning children 66

2.2 Iconicity scores and phonological overlap for ASL stimuli 68

2.3 Summary of the four linear models using children's age and vocabulary size to predict accuracy and reaction time 79

4.1 Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 4.1. 93

4.2 Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 4.2. 97

A.1 Results for sensitivity analysis for Experiment 1.1 125

A.2 Results for MLE models 126

List of Figures

1 Schematic overview of the disseration content. 2

1.1 Schematic of an active causal learning context. 23

1.2 Schematic of an active word learning context. 28

1.3 Schematic of active learning within a social context. 33

2.1 Stimuli in the Visual Language Processing Task used in Experiment 1.1 69

2.2 Time course looking behavior for ASL-proficient adults and young ASL-learners . . . 74

2.3 The time course of looking behavior for young deaf and hearing ASL-learners 76

2.4 Scatterplots of relations between children's age and vocabulary and ASL processing . 78

4.1 Examples of stimuli for exposure and test trials from Experiments 4.1-4.4. 112

4.2 Experiment 4.1 results. 113

4.3 Experiment 4.2 results 114

4.4 Primary analyses of test trial performance in Experiment 3.3 115

4.5 Secondary analyses of test trial performance in Experiment 3.3 116

4.6 Experiment 4.4 results 117

A.1 Graphical representation of the accuracy model. 121

A.2 Graphical representation of the RT model. 122

A.3 Results of sensitivity analysis for Experiment 1.1. 124

A.4 association between degree of phonological overlap and RT/Accuracy. 127

A.5 association between degree of iconicity and RT/Accuracy. 128

Introduction

Overview of the argument here.

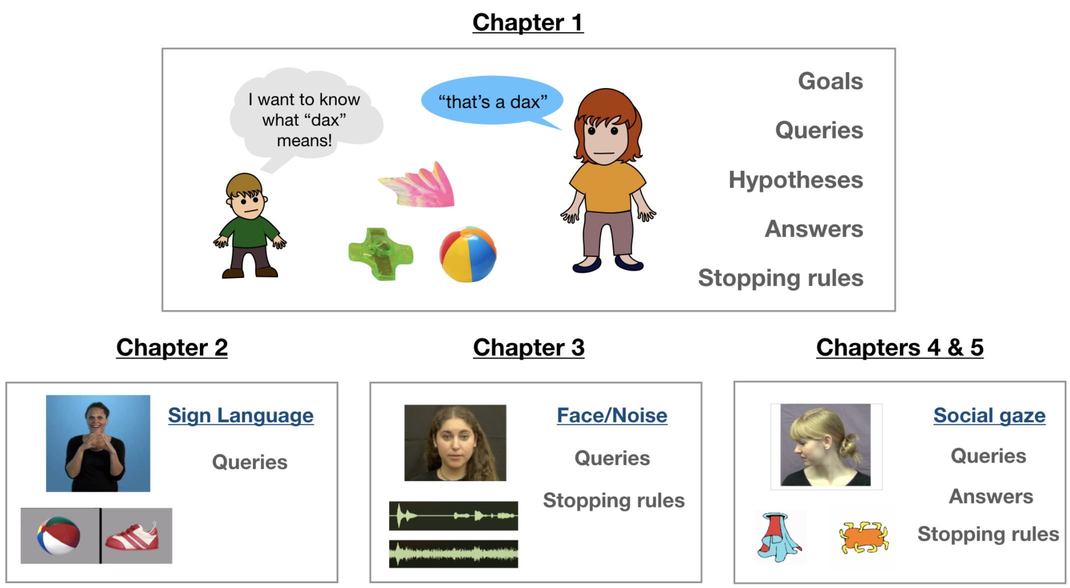


Figure 1: The upper panel shows a schematic overview of the components of an integrative framemwork of active learning within social contexts. The lower panels show the different case studies and pieces of the general model that correspond to each chapter of the disseration.

Chapter 1

Active learning within social contexts

1.1 Introduction

Human language learning is remarkable. Consider that children, despite limitations on general processing capabilities, acquire new lexical concepts at a high rate, eventually reaching an adult vocabulary ranging from 50,000 to 100,000 words (Bloom, 2002). And they accomplish this all while also developing motor skills, learning social norms, and building causal knowledge. How do we explain children's prodigious learning abilities?

Social learning accounts point out that children do not solve these problems on their own. Instead, children are typically surrounded by parents, other knowledgeable adults, or older peers – all of whom are likely ot know more than they do and are want to facilitate their learning. Social learning accounts also emphasize how social contexts can bootstrap children's learning via several mechanisms. For example, work on early language acquisition shows that social partners provide input that is tuned to children's cognitive abilities (Eaves Jr, Feldman, Griffiths, & Shafto, 2016; Fernald & Kuhl, 1987), that guides children's attention to important features in the world (C. Yu & Ballard, 2007), and that increases levels of sustained attention, which results in better learning (Kuhl, 2007; Yu & Smith, 2016).

Social contexts also change the computations (i.e., inferences) that support children's learning from evidence. Recent work in the fields of concept learning and causal intervention suggests that the presence of another person engages a set of psychological processes where the learner reasons about *why* other people performed specific actions. The critical insight comes from knowing that another person intentionally selected examples, allowing children to make stronger inferences that speed learning (Bonawitz & Shafto, 2016; Frank, Goodman, & Tenenbaum, 2009; Shafto, Goodman, & Griffiths, 2014). For example, children learn at different rates after observing the same evidence depending on whether they thought the behavior was accidental (less informative) or intentional (more informative). Moreover, adults and children will make even stronger inferences if they believe that another person selected their actions with the goal of helping them learn (i.e., teaching) (Shafto et al., 2012b).

However, children are not passive recipients of information – from people or the world. Instead, children actively select behaviors – for example, asking questions or choosing where to allocate visual attention – that change the content, pacing, and sequence of their learning experience. Recent theories of cognitive development have proposed the metaphor of "child as an intuitive scientist" and characterized early learning as a process of exploration and hypothesis testing following principles of the scientific method (Gopnik, Meltzoff, & Kuhl, 1999; Schulz, 2012). Moreover, recent empirical work across a variety of domains – education (Grabinger & Dunlap, 1995), machine learning (Castro et al., 2009; Settles, 2012), and cognitive science (D. B. Markant & Gureckis, 2014) – has directly compared learning trajectories in contexts marked by self-directed choice (active learning) as compared to settings where the learner has less control (passive learning). One conclusion from this work is that active contexts often lead to faster learning rates by enhancing attention and arousal or by providing learners with better information that is linked to their current goals, beliefs, and interests (see Gureckis & Markant (2012) for a review).

Thus, children are capable of guiding their own learning and social contexts provide particularly good learning environments. But how can we integrate these two proposals? Answering this question represents a significant step towards a complete theory of early learning since children's cognitive development often unfolds within social contexts and learners must integrate this social information when making decisions about what to learn.

In this paper, I propose an integrative account of active learning within social contexts. I use the framework of Optimal Experiment Design (Emery & Nenarokomov, 1998; Lindley, 1956) as a conceptual tool to bring social learning processes into contact with the underlying decision making that supports active learning. The key insight is that learning in the presence of other people plays a direct role in determining the *usefulness* of different actions. I organized the paper as follows. First, I define "active" and "social" learning to provide limits on the scope of phenomena that the integrative account aims to address. Then, I review the behavioral evidence showing that social (Part I) and active (Part II) contexts change how learning unfolds. From there I present the theory of Optimal Experiment Design (OED) (Part III) as a formal framework for understanding the decision-making process that supports active learning. Finally, I conclude by highlighting a series of novel links between the social learning account and self-directed choice, taking a step towards understanding how active learning unfolds within social contexts (Part IV).

1.1.1 The scope of the integrative account

Many theories of cognitive development have considered the relative contributions of "active" learning and "social" input to children's cognitive development. As a result, the terms are semantically "overloaded." So before reviewing the empirical evidence, it is worth defining "active learning" and "social contexts." I also want to scope the behaviors that the integrative account attempts to explain and highlight several distinctions that come up throughout the paper, including the different *timescales* through which social contexts affect active learning and the importance of others' *goals* for explaining social learning phenomena.

Learning can be "active" in a variety of ways. First, a child could be physically active, and this movement could change what information they extract from the experience. There is a large body of research that explores the effects of action experience on infants' learning (see Kontra, Goldin-Meadow, & Beilock (2012) for a review of work on embodied cognition). One classic example from Needham, Barrett, & Peterman (2002) shows that infants who physically hold and manipulate objects will outperform a control group on measures of object attention and exploration. Second, active learning could refer to children's contribution to processing incoming information. For example, young learners do not just passively accept other people's claims and will reject answers that conflict with their knowledge (Pea, 1982). Third, children might engage in self-generated "active" explanations. Lombrozo (2006) review evidence that 'self-explanation'can lead to better learning. Finally, active learning could refer to a decision-making process where children select, sequence, and pace their own learning experiences.

Here, I focus on active learning effects that arise via decision making. The key assumption is that active learners are trying to maximize the usefulness of their actions when choosing to gather information. By scoping the account to information seeking *decisions*, I do not aim to ignore the importance of other forms of active learning; instead, my goal is to constrain the space of possible connections between the active and social learning theories. Moreover, decisions during active learning still capture a rich set of behaviors, including pointing, eye movements, verbal question asking, and causal interventions.

Learning can also be "social" in a variety of ways. First, children could learn with another person present but without attending to or directly interacting with them. Research in social psychology shows that the mere presence of other people can facilitate performance of simple tasks and impair the performance of complex tasks (Cottrell, Wack, Sekerak, & Rittle, 1968; Uziel, 2007). Second, children could learn by looking to others as a guide, observing or imitating their behavior. In fact, children's capacity for faithful imitation is a critical feature separating human from non-human learning (Call, Carpenter, & Tomasello, 2005). Finally, children could both attend to the person and directly interact with them, entering a communicative learning context that engages powerful psychological reasoning processes that change learning.

In this paper, I define a "social context" as a learning environment where another agent is present. This definition includes all of the social learning behaviors – observation, imitation, and learning from direct interaction – discussed above. I use this broad definition to highlight the diverse ways that social contexts could shape children's decisions during learning. It is important to point out that the effects of social input could operate on at least three timescales: (1) in-the-moment (e.g., imitating another person), (2) over development (e.g., prior social input shaping current decisions), and (3) over cultural/evolutionary history (e.g., learning a conventionalized language system). This paper does not focus on timescales, but in particular sections, I highlight when they are relevant to the discussion.

In sum, the goal of this paper is to propose an integrative account of active learning within social contexts. I chose a specific definition of active learning – choices to seek information – and I will show that a range of social learning phenomena could shape these decisions. Before presenting the integrative account, I review evidence that both social and active learning (a) modulate processes such as attention and memory, (b) provide information that is particularly "good" for learning, and (c) change the strength of children's inferences and generalization.

1.2 Part I: Learning from other people

Social learning theories argue that children's rapid conceptual development is facilitated by the uniquely human capacity to transmit and acquire information from other people. A primary benefit of learning from others is that children gain access to knowledge that has accumulated over many generations; information that would be far too complex for any individual to figure out on their own (Boyd, Richerson, & Henrich, 2011). In addition to the cumulative effects, social contexts facilitate in-the-moment learning since more knowledgeable others can select input that is most useful for children's learning (Kline, 2015; Shafto et al., 2012b) and information that is likely to generalize beyond the current context (Csibra & Gergely, 2009).

There is a large body of empirical work on social learning across a variety of domains, e.g., language acquisition, causal learning, and concept learning. Importantly, these social learning effects operate through different pathways such as guiding attention, providing better information, and changing the strength of children's inferences. In this section, I briefly review the evidence for each pathway, with the goal of providing a high-level taxonomy of social learning effects.

1.2.1 Social contexts enhance attention and memory

From infancy humans preferentially attend to social information. For example, newborn infants prefer to look at face-like patterns compared to other abstract configurations (Johnson, Dziurawiec, Ellis, & Morton, 1991) and even show a preference for faces that make direct eye contact compared to faces that avert gaze (Farroni, Csibra, Simion, & Johnson, 2002). In the auditory domain, newborns prefer to listen to speech over non-speech (Vouloumanos & Werker, 2007), their mother's voice over strangers' voices (DeCasper, Fifer, Oates, & Sheldon, 1987), and infant-directed speech over adult-directed speech (Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Pegg, Werker, & McLeod, 1992). Moreover, recent work by Yu & Smith (2016), using head-mounted eye trackers to record parent-child interactions, shows that one-year-olds will sustain visual attention to an object longer when their parents had previously looked at that object.

These early attentional biases lead to differential learning in the presence of another person. For example, 4-month-olds show better memory for faces if that face gazed directly at them as compared to memory for a face with averted gaze (Farroni, Massaccesi, Menon, & Johnson, 2007). They also show enhanced memory for objects if an adult gazed at that object during learning (Cleveland, Schug, & Striano, 2007; Reid & Striano, 2005). Converging evidence comes from Thiessen, Hill, & Saffran (2005)'s work, showing that 7-month-olds perform better at word segmentation from infant-directed speech compared to adult-directed speech.

Kuhl (2007) refer to these effects as "social gating" phenomena since the presence of another person activates or enhances children's underlying computational learning mechanisms. One particularly striking piece of evidence for the social gating hypothesis comes from Kuhl, Tsao, & Liu (2003) study of infants' foreign-language phonetic learning. In this experiment, 9- to 10-month-old English-learning infants listened to Mandarin speakers either via live interactions or audiovisual recordings. Only the infants who heard Mandarin within social interactions were able to discriminate Mandarin-specific phonemes. In contrast, infants in the audiovisual recording condition showed no evidence of learning the phonemes despite similar amounts of exposure. Kuhl et al. (2003) also found that infants in the social interaction condition had higher rates of visual attention to the speaker, suggesting that the social context enhanced learning by increasing children's attention to the input.

Additional evidence comes from studies showing that when adults interact with an avatar controlled by a person rather than a computer, they experience higher levels of arousal, learn more, and pay more attention (Okita, Bailenson, & Schwartz, 2008). And recent work by Roseberry, Hirsh-Pasek, & Golinkoff (2014) found that children learn equally well from interactions with a person in a video chat (e.g., Skype) if they established social contingency, but they did not learn from watching communication between an adult and another child.

The common thread across this work is that the presence of another person increases attention. And as a result, social input becomes more salient and more likely to come into contact with general learning mechanisms. These changes occur within the child (endogenous effects) and in-the-moment of learning. However, social contexts can also provide better information, leading to exogenous effects on learning. In fact, social learning theories often start from the premise that early environments are unique because children are surrounded by people who know more than they do. Moreover, these individuals are invested in children's learning. Together, these features lead to contexts where more knowledgeable others select learning experiences that are particularly beneficial.

1.2.2 Social contexts provide "good" information

The idea that children's input might be shaped to facilitate their learning is a fundamental aspect of several theories of cognitive development [e.g., Zone of Proximal Development (Vygotsky, 1987), Guided Participation (Rogoff et al., 1993), and Natural Pedagogy (Csibra & Gergely, 2009)]. But how do social environments provide useful information?

One compelling set of evidence is that caregivers alter their communication style when speaking to children. Empirical work shows that when adults talk to children, they exaggerate prosody, reduce speed, shorten utterances, and elevate both pitch and affect (for a review, see Fernald & Simon (1984)). Subsequent empirical work shows that these features can help infants solve a variety of language acquisition challenges, including vowel learning (Adriaans & Swingley, 2017; De Boer & Kuhl, 2003), word segmentation (Fernald & Mazzie, 1991; Thiessen et al., 2005), word recognition (Singh, Nestor, Parikh, & Yull, 2009), and word learning (Graf Estes & Hurley, 2013).

Additional evidence that social contexts provide information tuned to individual learners comes from work on infants' early vocal production. For example, Goldstein & Schwade (2008) measured whether infants modified their babbling to produce more speech-like sounds after interacting with caregivers who provided either contingent or non-contingent responses to their babbling. Only infants in the contingent feedback condition changed their vocalization behavior to produce more adult-like language forms. Goldstein & Schwade (2008) hypothesized that the contingent input was particularly useful because it occurred soon after infants' vocalizations, making it easier to compare any discrepancies.

Converging support comes from research on children's early word learning. Social-pragmatic theories of language acquisition have long emphasized the importance of social cues for reducing referential uncertainty (Bloom, 2002; Clark, 2009; Hollich et al., 2000). Empirical work by C. Yu & Smith (2012a) shows that young learners tend to retain words that are accompanied by clear referential cues (e.g., adults' pointing and gaze direction), which serve to make a single object dominant in the visual field (Yu, Ballard, & Aslin, 2005; Yu & Smith, 2013). Moreover, correlational studies show positive links between early vocabulary development and parents' tendency to refer to objects that children are already attending to (i.e., "follow-in" labeling) (Tomasello & Farrar, 1986).

Together, these findings provide evidence that social contexts are likely to contain "useful" information. Similar to the attention/memory effects, these effects occur in-the-moment of learning. However, they are properties of the input, and their usefulness is derived from processes external to the learner, but internal to the learner's social partner (e.g., an adult reading the child's direction of gaze and providing a label). Social contexts also shape learning by engaging a sophisticated set of social reasoning processes that change how much children learn from new evidence.

1.2.3 Social contexts shape inferences and generalization

One defining feature of social learning is that people's actions are not random. Instead, people select behaviors with respect to some goal (e.g., to communicate a concept). If children are sensitive to others' goal-directed behavior, then they can reason about *why* someone chose an action. And this reasoning process can change how people interpret superficially similar behaviors.

Recent empirical and modeling work has formalized this social reasoning process within the framework of Bayesian models of cognition (Frank & Goodman, 2014; Goodman & Frank, 2016; Shafto et al., 2012b). Under this account, social learning is a process of belief updating that depends on two factors: the learner 's beliefs before seeing any evidence and what the learner thinks about the process that generated the evidence. If the learner assumes that someone selects an action with the intention to communicate, they can make "stronger" inferences.[1]

For example, Goodman, Baker, & Tenenbaum (2009) presented adults with written descriptions of the following causal learning scenarios. Someone generates a causal effect (e.g., growing flowers) by performing two actions at the same time (e.g., pouring a yellow liquid and a blue liquid). The person who generated the effect was either the participant or another person who already knew the causal structure. The participants' task was to identify the correct causal structure. When participants thought the other person was knowledgeable, they were more likely to think that performing *both* actions was necessary. In contrast, when the participant-generated the causal effect on their own (not knowing the causal structure), adults were less sure that both actions were necessary. Shafto et al. (2012b) interpreted these results as a psychological reasoning process such as: "if the other person were knowledgeable and wanted to generate the effect, then he would perform both actions." This finding also suggests that learners assume that others' goal-directed behaviors will be efficient

[1] Formally, these models change the likelihood term in Bayes theorem to capture a person's theory of how data are generated. I will discuss links between this formalization and the active learning account in Part IV.

and they should avoid performing unnecessary actions (e.g., pressing two buttons when pressing one would have been sufficient).

Similar effects of psychological reasoning on inference occur in word learning (Frank & Goodman, 2014; Xu & Tenenbaum, 2007b), selective trust in testimony (Shafto et al., 2012a), tool use (Sage & Baldwin, 2011), and concept learning (Shafto et al., 2014). Moreover, there is evidence that even young learners' inferences are sensitive to the presence of goal-directed behaviors. For example, Yoon, Johnson, & Csibra (2008) showed that 8-month-olds encode an object's identity if attention was directed by a communicative point, but they will encode an object's spatial location if attention was directed by a non-communicative reach. And Senju & Csibra (2008) found that infants will follow another person's gaze only if the gaze event was preceded by relevant, communicative cues (e.g., infant-directed speech or direct eye contact).

In addition to being easier to learn, information from other people is more likely to generalize beyond the current learning context. Csibra & Gergely (2009) argue that an assumption of *generalizability* is fundamental to "Natural Pedagogy" – a uniquely human communication system that allows adults to pass cultural knowledge to children. In these contexts, adults generate ostensive signals such as direct gaze, infant-directed speech, and infant-directed actions. These signals, in turn, direct infants' attention towards the adult, and bias infants to expect generalizable information.

Experimental work testing predictions of Natural Pedagogy shows that children tend to think that information presented in communicative contexts is generalizable (Butler & Markman, 2012; Yoon et al., 2008). For example, Butler & Markman (2012) showed that preschoolers were more likely to expect a novel causal property (e.g., magnetism) to generalize to new objects with the same shape if the causal property was demonstrated with pedagogical cues. Moreover, corpus analyses show that generic language (e.g., "birds fly") is common in everyday adult-child conversations (Gelman, Goetz, Sarnecka, & Flukes, 2008), suggesting that this generalizable information is prevalent in children's daily experience.

Across these studies, learners interpreted similar information in different ways depending on their assumptions about others' goals. These effects are different from the attentional (internal to the learner) and informational (external to the learner) explanations reviewed above in that the inferences based on social information are part of the underlying computations that support learning. However, parallel to Kuhl's Social Gating account, Natural Pedagogy argues that the presence of

Table 1.1: High level summary of three different paths through which active and social contexts influence learning. See Part I (social learning) and Part 2 (active learning) for more details and reviews of the behavioral evidence.

Learning component	**Active learning**	**Social learning**
Attention and memory	Engages attention, coordinates attention to the learning moment, and enhances uncertainty monitoring	Increases arousal and enhances attention/memory processes
Quality of the input	Generates high quality input that is linked to prior knowledge, hypotheses, and ability	Generates high quality input shaped to the learner's current cognitive abilities and goals
Inferences and generalization	Behaviors selected without knowledge of the target concept, leading to weaker inferences and generalization	Behaviors selected with knowledge of the target concept, leading to stronger inferences and generalization

pedagogical cues enhances processes internal to the learner such as attention. These theories and empirical work receive additional support from evolutionary models that emphasize the importance of pedagogy for the accumulation of human cultural knowledge (Boyd et al., 2011; Kline, 2015).

1.2.4 Social learning summary

The work on social learning reviewed in this section highlight several points. First, from an early age, children are surrounded by other people who know more than they do. Moreover, these more knowledgeable agents are invested in children's development and motivated to provide good learning opportunities. Second, children are driven to interact with other people, and these interactions are engaging and social partners guide attention to relevant information. Finally, social learning triggers a set of psychological reasoning mechanisms that build off children's capacity for detecting goal-directed behavior. Critically, the output of this reasoning is stronger inferences, allowing children to get more information out of the same amount of input.

However, it is clear that social input cannot account for all of children's rapid conceptual development, and that children are not just passive recipients of input from the world or other people. Instead, they actively process information and select behaviors that change what they learn. A parallel body of research on this topic – under the umbrella term of "active learning" – has developed alongside social learning theories. In the next section, I present the active learning account and the empirical work with the goal of clarifying the variety of ways that children shape their learning.

1.3 Part II: Learning on your own

The idea that children are "active" learners has also been an influential aspect of classic theories of cognitive development (e.g., Bruner, 1961; Berlyne, 1960). Recent theorizing has characterized cognitive development as a process of active hypothesis testing and theory revision following the principles of scientific inquiry (Gopnik et al., 1999; Schulz, 2012). Under this account, children drive their conceptual development by selecting actions to test theories about how the world works.

The effects of active learning have also been the focus of much empirical work in education (Grabinger & Dunlap, 1995; Prince, 2004), machine learning (Ramirez-Loaiza, Sharma, Kumar, & Bilgic, 2017; Settles, 2012), and cognitive psychology (Castro et al., 2009; Chi, 2009). The common thread across these diverse bodies of work is that active contexts – where people have control over their experience – lead to different (and often more rapid) learning outcomes when compared to passive contexts where people do not have control over the flow of information.

But how does active control change the way people learn? In this section, I present evidence for three mechanisms. These pathways parallel the social learning effects reviewed in Part I. First, active learning contexts enhance attention and memory, leading to stronger learning. Second, active learners use their uncertainty to select experiences that cater to their goals, beliefs, and capabilities. Third, active learning results in weaker inferences and generalization since there is no guarantee that self-generated examples will be informative. I conclude Part II with a discussion of why active learning might be particularly challenging to study. These challenges motivate the formalization of human inquiry as Optimal Experiment Design that I present in Part III.

1.3.1 Active contexts enhances attention and memory

A growing body of research shows that active control changes processes such as attention and arousal that support learning. In these experiments, researchers have compared learning outcomes for active and passive contexts across a variety of tasks such as episodic memory, casual intervention, and concept learning. The common finding is that active contexts result in faster and more robust learning. In a review of this literature, Markant, Ruggeri, Gureckis, & Xu (2016) propose that increased attention and memory drives the active learning advantage, with the precise effect determined by the type of control given to the learner (e.g., timing, content, timing & content).[2]

One compelling demonstration of the effects of different types of control comes from D. Markant et al. (2014) where they "deconstructed" the active learning process. Participants memorized the identities and locations of objects hidden in a grid. Markant et al. (2014) varied the *level* of control that participants had over the learning experience. Participants could control either: (1) the next location in the grid, (2) which item to reveal next, (3) the duration of each learning trial, and/or (4) the time between learning trials (i.e., inter-stimulus-interval or ISI). There was also a "yoked" control group of participants who saw the training data generated by the active learners. The yoked condition is critical for equating the informational content across learners while varying the level of control. There was an advantage in spatial memory for all levels of control, including the lowest amount in the ISI-only condition. These results suggest that learners benefitted from being able to coordinate the timing of information with their attentional processes, starting the next trial after they finished processing the previous one and once they were ready to learn something new.

Developmental studies have also used the deconstructing approach and found parallel active learning advantages for 6- to 8-year-olds in a spatial memory task (Ruggeri, Markant, Gureckis, & Xu, 2016). Other work has found similar benefits in word learning (Partridge, McGovern, Yung, & Kidd (2015); see also Kachergis, Yu, & Shiffrin (2013) for evidence in adults) and understanding causal structures (Schulz, 2012). For example, Sobel & Kushnir (2006) showed that preschoolers who generated interventions on a causal system learned more compared to yoked participants who either passively observed the same sequence of actions or re-created the same choices made by others (but see McCormack, Bramley, Frosch, Patrick, & Lagnado (2016)). Moreover, even infants benefit from active engagement with the learning environment. For example, Begus, Gliga, & Southgate (2014) found that 16-month-olds show better memory for information provided about an object they had demonstrated an interest in via pointing compared to an object that infants had previously ignored.

Additional evidence that active control enhances attention and memory comes from research on children's engagement with educational technology (for a review, see Hirsh-Pasek et al. (2015)). For example, Calvert, Strong, & Gallagher (2005) exposed preschool-aged children to two sessions of reading a computer storybook with an adult and manipulated whether the adult or the child controlled the mouse and could advance the story. Children in the adult-control condition showed

[2] For example, consider a child who is playing with a set of toys and turns to ask their parent, "What's this?" Here, the child is in control of both the selection (which toy gets labeled) and the timing (when the labeling event occurs) of information.

a decrease in attention to the storybook materials in the second session. In contrast, children who were given control maintained a constant level of engagement across both sessions, suggesting that active control provided a buffer against children losing interest in a repetitive task.

These results parallel the findings on attention/memory effects in social learning reviewed in Part I. Both active and social processes can modulate processes internal to the learner to facilitate in-the-moment learning. However, the effects of active control go beyond changing lower-level cognitive processes and modulate the *quality of information* that learners get from the world.

1.3.2 Active contexts provide "good" information

One defining feature of active learning is that people gather information that is useful for their development. This benefit arises because the learner has privileged access to their prior knowledge and current hypotheses, which they leverage to create more helpful learning contexts (e.g., asking a question about something that is particularly confusing). Research on this aspect of active learning focuses on how learners select actions to generate useful information, often comparing learning outcomes to passive contexts where the learner does not have control over the environment.

For example, Castro et al. (2009) directly compared adults' category learning in active vs. passive contexts to predictions from statistical learning theory to quantify any difference in human performance relative to the optimal model predictions. Participants saw a sequence of 3-D objects on a computer screen that varied along a single, continuous dimension (spiky to smooth) and were given feedback as to which category the stimulus belonged to. The participants' task was to learn the correct category boundary. In the active condition, learners could select which object they wanted to be labeled; whereas, in the passive condition, participants saw stimuli generated randomly from the correct categories. Active learning was always superior to passive learning with participants learning the category structure in less time and achieving higher levels of accuracy. However, human learners did not reach the performance of the optimal model, and the advantage for active over passive learning decreased in the more difficult (i.e., noisier) learning tasks.

Using a similar approach, D. B. Markant & Gureckis (2014) investigated the effects of active vs. passive hypothesis testing on the rate of adults' category learning. They varied the difficulty of the learning task by testing two different types of category structures: a rule-based category, which differed along a single dimension (easier to learn), and an information-integration category, which varied along two dimensions (harder to learn). In the active condition, the learner could choose specific observations to test their beliefs; whereas, in the passive version, the sequence of data was generated randomly by the experiment. D. B. Markant & Gureckis (2014) also included a “yoked” condition where participants saw sequences of observations created by active learners but did not have control over the sequence. Similar to the Castro et al. (2009) findings, active participants learned the category structure faster and achieved a higher overall accuracy rate compared to the passive learners and the yoked-passive learners. D. B. Markant & Gureckis (2014) suggest that active learning was better because self-generated observations are linked to the learner’s hypothesis. Importantly, the advantage for active learners over the yoked participants suggests that the information value was linked to the individual learner. Also, the active learning advantage only held for the less complex, rule-based category.

The effects of active engagement also show up in studies of language use. For example, Schober & Clark (1989) asked pairs of adults to complete a task where one person (the director) used natural language to inform another person (the matcher) what order to arrange tangrams (geometric objects, called tans, which are put together to form shapes) in a 4x4 matrix. The matcher had a different level of control depending on condition assignment: (1) could actively participate (i.e., talk to the director), (2) could listen to the recorded conversation of director-matcher and pause the tape, and (3) could listen to the entire discussion but not pause the tape. Results showed that active participation led to a marked advantage over passive listening. Interestingly, the ability to stop the tape did not improve accuracy, suggesting that the active advantage was caused by informational value and not by control over the timing/ pacing of information. Also, participants in the passive conditions reported frustration (e.g., ”I don’t know what ‘this one’ means!) with being unable to correct early failures of understanding the novel conventions developed by the director. Thus, no amount of control over timing could make up for the lack of control over the information *content* in the overhearing conditions. Schober & Clark (1989) propose that conversations are a form of active information gathering about others’ intentions, and when people are unable to participate, they lose access to critical information that supports later comprehension.

These findings illustrate several points. First, the quality of active exploration was fundamentally linked to the learner’s understanding of the task: if the representation was weak, then self-directed learning was less useful. Second, the benefits of active control were tied to the individual learner’s prior knowledge and current hypotheses such that the same sequence of data did not provide "good" information for another person. And third, the benefits of active learning diminished with increased task difficulty because learners struggled to generate helpful examples. These challenges will come up again in Part III when I outline the formalization of human inquiry.

1.3.3 Can children select "good" information?

Research on children's pointing shows that young learners, who are not yet capable of more sophisticated information seeking behaviors such as verbal questions, can use nonverbal actions to facilitate learning. For example, Wu & Gros-Louis (2015) found that adults generate a higher number of object labels for objects that their 12-month-olds pointed to, suggesting that the infants' pointing elicited information that was especially useful for concrete word learning (for converging evidence, see Kishimoto, Shizawa, Yasuda, Hinobayashi, & Minami (2007); Goldin-Meadow, Goodrich, Sauer, & Iverson (2007); and Olson & Masur (2011)). Moreover, work by Begus & Southgate (2012) found that infants point more in the presence of a knowledgeable person compared to an incompetent person, suggesting that points signal a desire to learn as opposed to sharing attention. Finally, infants who produce more pointing gestures have larger vocabularies later in development (Rowe & Goldin-Meadow, 2009), providing additional evidence that even young infants are capable of generating actions that elicit information to support learning.

Later in development, when children being to acquire the requisite productive language skills, they start asking verbal questions to gather information. For example, in a corpus analysis of four children's parent-child conversations, Chouinard, Harris, & Maratsos (2007) found that children begin asking questions early in development (18 months) and at an impressive rate, ranging from 70-198 questions per hour of conversation. Chouinard et al. (2007) also coded the children's intent, finding that 71% of questions were to gather information, as opposed to seeking attention or clarification. Other corpus analyses provide converging evidence that questions are common in parent-child conversations (Davis, 1932) and that children use them to gather knowledge (Bova & Arcidiacono, 2013), persisting when they do not receive a satisfactory explanation (Frazier, Gelman, & Wellman, 2009).

Perhaps the most robust evidence that children are capable of efficient self-directed learning comes from research on causal learning. In these studies, children see a novel toy with an unknown causal structure. Then they play (design experiments) to figure out how the toy works. A nice feature of these tasks is that the space of possible actions and hypotheses are constrained such that it becomes possible to quantify the usefulness of children's causal interventions and compare them to the predictions of formal models of optimal information seeking (see Part III). Specifically, empirical work has found that preschoolers integrate prior beliefs and evidence to alter how they explore a causal system (e.g., testing a toy to learn the concept of balance-relations) (Bonawitz, Schijndel, Friel, & Schulz, 2012). Children spend more time exploring an object for which they saw confounded evidence for its causal structure, i.e., where there was more to be learned (Schulz & Bonawitz, 2007). And children become more efficient in producing causal interventions, as measured by informativeness, as they get older (6-8 years of age) (McCormack et al., 2016).

Moreover, even 8-month-old infants selectively explore objects that violate their prior expectations. Stahl & Feigenson (2015) showed infants events that violated an expectation about objects, either solidity, continuity, or support. They coded the types of actions that infants chose to perform on the objects during subsequent free play. Infants explored differently depending on the type of violation (e.g., banging the object after seeing a violation of solidity). The infants also spent more time playing with objects that violated expectations, and as a result, learned more about them. These results demonstrate an early sensitivity to uncertainty with infants using actions to test specific hypotheses that were linked to their prior experience.

Research on infants' selective visual and auditory attention provides another example of children's effective active learning. These studies start from two assumptions: that children possess limited cognitive resources and that children would benefit by attending to information that is likely to be learned. For example, Kidd, Piantadosi, & Aslin (2012) measured 7- and 8-month-olds' visual attention to a monitor that displayed a sequence of familiar objects (e.g., a toy truck). Within each series, infants saw trials that varied along a continuum from low to high complexity. Complexity was a measure of how surprising the current object was relative to the previous objects in that sequence. For example, if there were two objects (truck and ball) in a set, and the child saw [Truck-Truck-Truck-Truck] and the next trial was Truck, then this would be low surprisal/complexity. In contrast, if the child had seen the sequence [Truck-Ball-Ball-Ball] and the subsequent trial was a Truck, then this trial would be high surprisal/complexity. Infants looked longest on trials of intermediate complexity, choosing to disengage sooner when the object was either highly predictable or highly surprising. Kidd, Piantadosi, & Aslin (2014) extended these results to the auditory domain, showing a similar pattern of increased attention to sequences of intermediate complexity for nonsocial sounds such as a door closing or a train whistle. Kidd and her colleagues interpret these results as infants using prior experience to guide selective attention to find learnable information that is neither too simple (nothing to be gained) nor too complicated (too much to process).

There is also evidence that infants avoid spending time on information that is unlearnable. For example, Gerken, Balcomb, & Minton (2011) tested whether 17-month-olds would increase attention to a stream of input that consisted of a learnable structure (i.e., Russian feminine words take the endings oj and u, and masculine words take the endings ya and yem) as opposed to a random stream of input without any information to extract (i.e., word endings that are not diagnostic of category structure). Infants dishabituated sooner when listening to the unlearnable stream, suggesting that they were tracking the pace of learning and choosing to stop information gathering if progress was low.

1.3.4 Active contexts shape inferences and generalization

Active learning contexts also change the strength of learners' inferences and generalization. In contrast to the social learning effects discussed in Part I, active learners assume "weak" sampling, believing that examples are not generated from the true concept. This assumption, in turn, leads learners to make less-restrictive inferences and generalize more broadly. That is, active learners are aware that there the evidence they generate is not guaranteed to be informative, and they update their beliefs accordingly.

For example, Xu & Tenenbaum (2007a) tested 4-year-olds and adults' word generalization. In the task, participants learned the correct extension of a novel word after seeing three examples drawn from a set of 30 unfamiliar objects. The set of objects consisted of two basic-level categories (15 objects) and within each of the basic categories, there were three smaller, subordinate categories (5 objects) that varied in color, texture, and orientation. Participants saw three labeled examples from the subordinate category. The critical manipulation was whether the examples were selected by a teacher who knew the correct word extension (social) or by the learner (active) who did not know the appropriate extension. Both adults and children made stronger inferences in the teacher-driven condition, saying that the new word referred to the subordinate category; whereas, in the learner-driven case, children and adults tended to generalize the word to the basic-level.

Xu & Tenenbaum (2007a) explain these results as a sensitivity to the process that generated examples. When a knowledgeable teacher produced the labels, there was the good reason to think that the examples were linked to word meaning. Thus it would be surprising to see three examples from the smaller subordinate category. However, the active learners had no reason to think they generated examples from the true category, and therefore they extended the novel word more broadly to the basic-level. The takeaway from work is that even young children appear sensitive to the informativeness of the process that generates examples, and use this information to change how much they update beliefs based on new evidence.

1.3.5 Active learning summary

The work on active learning reviewed in this section highlights several points. First, from an early age, children are capable of efficient information seeking. Second, active learning is a complex area of research, covering a wide range of actions (e.g., pointing, visual attention, causal interventions, and verbal questions) and supported by a variety of cognitive processes. Moreover, active learning is by definition linked to the idiosyncrasy of individuals, and thus, does not function similarly across individuals, contexts, and learning domains. Finally, there is a multitude of factors that could influence the quality of children's active learning, creating a broad space of possibilities for researchers to test.

One way to constrain the space of hypotheses is to use formal models of the information seeking process. One useful approach is to perform an ideal observer analysis (Geisler, 2003) where a model is created to solve a task using all of the available information in the environment. This model does not include processing constraints such as limited attention or memory, and as a result, only produces errors based on the complexity or uncertainty present in the environment. By ignoring the processing level, these models allow researchers to focus on the structure of the learning task and the information available in the world. Moreover, the "ideal" model performance can then be compared to human behavior to see how much of the available information people use to solve the problem. This approach does not aim to make claims about human optimality; instead, the ideal-observer can be used as a method for scientific progress because it forces researchers to explicitly define the process of active learning as separable, underlying components, which can become the target of research.

Over the past three decades, researchers in both developmental and cognitive psychology have used the ideal-observer approach to understand human inquiry. Researchers have leveraged formal models of scientific reasoning that fall under the umbrella term Optimal Experiment Design (OED). The original purpose of the OED models was to help scientists select the best experiment from a set of possible experiments, where "best" is defined as the experiment that leads to the highest amount of information gained about the scientific phenomenon. Using this formalization, researchers have asked whether people's information seeking behaviors look qualitatively similar to predictions from OED models. Moreover, this approach has the benefit of using the same mathematical formalization as recent models of social learning: Bayesian ideal observer models (Frank et al., 2009; Goodman & Frank, 2016; Shafto et al., 2012a). These parallel formalizations suggest a way forward for integrating the active and social learning theories. This point is a focus of Part IV.

1.4 Part III: A formal account of active learning

Optimal Experiment Design (OED) (Emery & Nenarokomov, 1998; Lindley, 1956; Nelson, 2005) is a statistical framework for quantifying the "usefulness" of experiments. Lindley (1956) described the approach as a transition from viewing statistics as binary decision making to a practice of gathering information about the "state of nature" (p. 987). The concrete proposal is to design studies that maximize expected information gain (a measure borrowed from Information Theory and discussed in more detail below) and continue to collect data until the information gained reaches a pre-determined threshold.

The OED approach allows scientists to make design choices that maximize the effectiveness of their experiments, reducing inefficiency and cost. Consider the following toy example borrowed from Ouyang, Tessler, Ly, & Goodman (2016) where a researcher is interested in designing the best experiment to figure out whether people think a coin is fair or biased (i.e., a trick coin). Here the researcher's hypotheses correspond to different models of the coin $[M_{fair} : Bernoulli(p = 0.5)]$ and $[M_{bias} : Bernoulli(p)$ where $p \sim Uniform(0, 1)]$ and the experiments correspond to different sequences of coin flips that she could select as stimuli. Imagine that the researcher has limited time or resources and can only show a sequence of four coin flips, creating a space of 16 possible experiments. An OED model allows the researcher to select the best experiment that maximally differentiates the two hypotheses. For example, OED provides an answer to the question: how much better would it be to use $[HHHH]$ versus $[HTHT]$? Here, $[HHHH]$ is more informative because both the bias and the fair coin models make the same predictions for the $[HTHT]$ experiment, meaning we would not learn much from this test.

An applied example comes from Nelson, McKenzie, Cottrell, & Sejnowski (2010) where they used an OED model to differentiate competing theories of information seeking during adults' category learning. They created an OED model of their task, which included the design choices (what combination of features to show participants) and the relevant behavioral hypotheses (the different theories of category learning). They used the model to simulate the outcomes of using different stimulus sets, allowing them to choose stimuli for which the competing theories made very different predictions, speeding the rate of discovery.

Coenen, Nelson, & Gureckis (2017) provide a thorough review of the OED framework and its links to human information seeking. They outline the four critical parts of an OED model: (1) a set of hypotheses, (2) a set of queries (i.e., actions) to learn about the hypotheses, (3) a way to model the types of answers that each query could elicit, and (4) a way to score each answer with respect to the learning goal. They also highlight the importance of understanding learners' inquiry goals (what do people want to learn?) for engaging OED-like reasoning. The critical point is that without a clear learning goal, it becomes challenging to instantiate the hypotheses, questions, and answers that a learner should consider. In the rest of this section, I provide the mathematical details of the OED approach as described in Coenen et al. (2017). The goal is to provide a concrete foundation for the conceptual analysis of how social learning contexts can influence different components of active learning.

The OED model quantifies the *expected utility* of different information seeking actions. Formally, the set of queries is defined as $Q_1, Q_2, ..., Q_n = \{Q\}$. The expected utility of each query ($EU(Q)$) is a function of two factors: (1) the probability of obtaining a specific answer $P(a)$ weighted by (2) the usefulness of that answer for achieving the learning goal $U(a)$.

$$EU(Q) = \sum_{a \in q} P(a)U(a)$$

There are a variety of ways to define the usefulness function to score each answer. An exhaustive review is beyond the scope of this paper (for a detailed analysis of different approaches, see Nelson

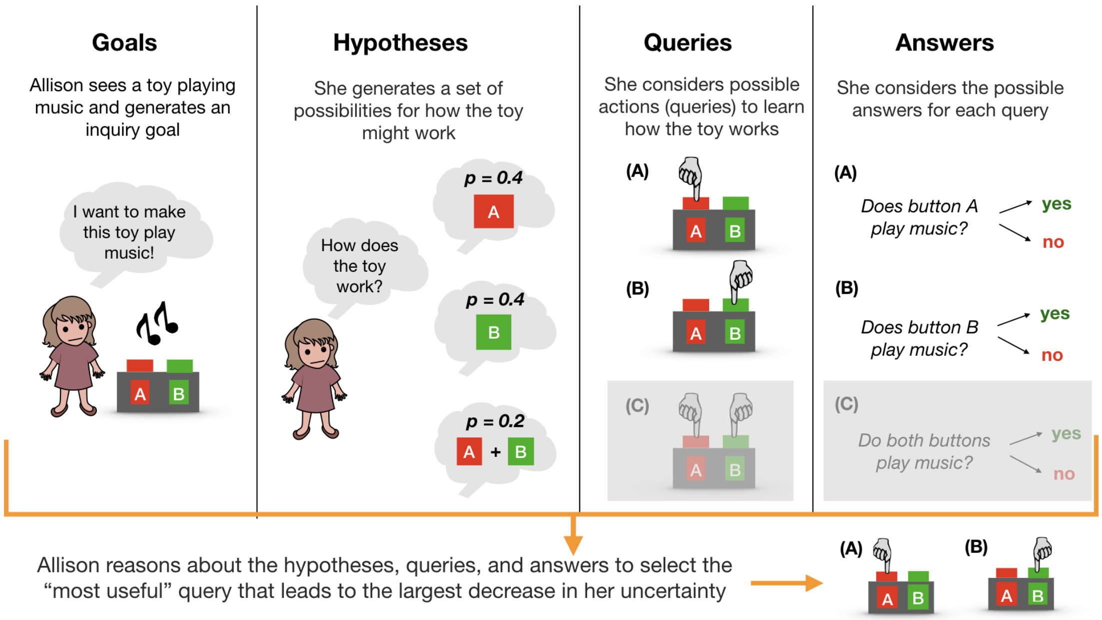


Figure 1.1: Schematic of an active causal learning context using the decomposition of Optimal Experiment Design. The learner generates an inquiry goal to learn how the toy works. She then considers hypotheses, including her subjective belief in each, placing stronger belief in the simpler, disjunctive hypotheses: only Button A or Button B. Next, she considers her possible queries (actions) and the potential outcomes if she took those actions. Together, these components quantify the expected usefulness of each action. If the learner chooses optimally, she picks the action that maximizes this expected utility. In this case, she chooses to press either Button A or B, but does not press both buttons since this action would produce confounded evidence and fail to reduce her uncertainty. See Part III in the text for mathematical details of the OED model.

(2005)). One standard method is to use *information gain*, which is defined as the change in the learner's overall uncertainty (difference in entoropy) before and after receiving an answer.

$$U(a) = ent(H) - ent(H|a)$$

Where $ent(H)$ is defined using Shannon entropy[3] (MacKay, 2003), which provides a measure of the overall amount of uncertainty in the learner's beliefs about the candidate hypotheses.

$$ent(H) = -\sum_{a \in A} P(h) log_2 P(h)$$

The conditional entropy computation is the same, but takes into account the change in the learner's beliefs after seeing an answer.

$$ent(H|a) = -\sum_{h \in H} P(h|a) log P(h|a)$$

To calculate the change in the learner's belief in a hypothesis $P(h|a)$, we use Bayes rule.

$$P(h|a) = \frac{P(h)P(a|h)}{P(a)}$$

If the researcher defines all these parts of the OED model (hypotheses, questions, answers, and the usefulness function), then selecting the optimal query is straightforward. The learner performs the expected utility computation for each query in the set of possible queries and picks the one that maximizes utility. In practice, the learner considers each possible answer, scores the answer with the usefulness function, and weights the score using the probability of getting that answer.

Before reviewing the behavioral evidence for OED-like reasoning in adults and children, I will present a worked example of how to compute the expected utility of a single query. The goal is to provide simple calculations that illustrate how reasoning about hypotheses, questions, and answers can lead to selecting useful actions. This example is slightly modified from Nelson (2005).[4]

Imagine that you are a biologist, and you come across a new animal that you think belongs to one

[3] Shannon entropy is a measure of unpredictability or amount of uncertainty in the learner's probability distribution over hypotheses. Intuitively, higher entropy distributions are more uncertain and harder to predict. For example, if the learner believes that all hypotheses are equally likely, then they are in a state of high uncertainty/entropy. In contrast, if the learner firmly believes in one hypothesis, then uncertainty/entropy is low.

[4] In the appendix, I include example code for instantiating the OED calculations as functions in the R programming language.

of two species: "glom" or "fizo." You cannot directly query the category identity, but you can gather information about the presence or absence of two features (eats meat? or is nocturnal?) that you know from prior research are more or less likely for each of the species. The following probabilities summarise this prior knowledge:

- $P(eatsMeat \mid glom) = 0.1$
- $P(eatsMeat \mid fizo) = 0.9$
- $P(nocturnal \mid glom) = 0.3$
- $P(nocturnal \mid fizo) = 0.5$

You also know from previous research that the probability of seeing a glom or a fizo in the wild is:

- $P(glom) = 0.7$
- $P(fizo) = 0.3$

Which feature should you test: eats meat? or sleeps at night? Intuitively, it seems better to test whether the creature eats meat because an answer to this question provides good evidence about whether the animal is a fizo since $P(eatsMeat \mid fizo) = 0.9$. However, the OED computation allows the biologist to go beyond this intuition and compute precisely how much better it is to ask the "eats meat?" question. All the scientist has to do is pass her knowledge about the hypotheses and features through the expected utility computation.

Here are the steps of the OED computation for calculating the utility of the "eats meat?" question. First, we use Bayes rule to calculate how much our beliefs would change if we received a "yes" or a "no" answer.[5]

$$P(glom \mid eatsMeat) = \frac{P(eatsMeat \mid glom) \times P(glom)}{P(eatsMeat)} = \frac{0.1 \times 0.7}{0.34} = 0.21$$

Next, we calculate the uncertainty over the Species hypothesis before doing any experiment. We do this by computing the prior entropy.

[5] Note that the $P(eatsMeat)$ term is computed by taking $P(eatsMeat) = [P(eatsMeat \mid glom)\ timesP(glom)] + [P(eatsMeat \mid fizo) \times P(fizo)] = (0.1 \times 0.7) + (0.9 \times 0.3) = 0.34$

$$\begin{aligned}
ent(Species) &= -\sum_{h \in H} P(h) \times log_2 P(h) \\
&= [-P(glom) \times log_2 P(glom)] + [-P(fizo) \times log_2 P(fizo)] \\
&= [-(0.7 \times log_2(0.7)] + [-(0.3 \times log_2(0.3)] \\
&= 0.88
\end{aligned}$$

To calculate information gain, we also need to compute our uncertainty over hypotheses conditional on seeing each answer, or the posterior entropy. First, for the "yes" answer:

$$\begin{aligned}
ent(Species|eatsMeat = yes) &= -\sum_{a \in A} P(Species \mid eatsMeat = yes) \times log_2 P(species \mid eatsMeat = yes) \\
&= [0.21 \times log_2(0.21)] + [0.79 \times log_2(0.79)] \\
&= 0.73
\end{aligned}$$

We use the difference between the prior and posterior entropy to compute the utility of the "yes" answer.

$$\begin{aligned}
U(a = yes) &= ent(Species) - ent(Species \mid eatsMeat = yes) \\
&= 0.88 - 0.73 \\
&= 0.15
\end{aligned}$$

Next, we do the same process for the "no" answer. First, we calculate the posterior entropy.

$$\begin{aligned}
ent(Species|eatsMeat = no) &= -\sum_{a \in A} P(Species \mid eatsMeat = no) \times log_2 P(species \mid eatsMeat = no) \\
&= [0.95 \times log_2(0.95)] + [0.05 \times log_2(0.05)] \\
&= 0.27
\end{aligned}$$

Again, we use the difference between the prior and posterior entropy to compute the utility of the "no" answer.

$$\begin{aligned} U(a = no) &= ent(Species) - ent(Species \mid eatsMeat = no) \\ &= 0.88 - 0.27 \\ &= 0.61 \end{aligned}$$

Note that the $U(a = no) > U(a = yes)$. This captures the intuition that learning that the animal does not eat meat would provide strong evidence against the "fizo" hypothesis since $P(eatsMeat \mid fizo) = 0.9$. Finally, to compute the overall expected information gain for the **"eats meat?" question**, we weight the utility of each answer by its probability:

$$\begin{aligned} EU(Q = eatsMeat) &= \sum_{a \in A} P(a)U(a) \\ &= [P(eatsMeat = yes) \times U(eatsMeat = yes)]+ \\ &\qquad [P(eatsMeat = no) \times U(eatsMeat = no)] \\ &= [0.34 \times 0.15] + [0.66 \times 0.61] \\ &= 0.46 \end{aligned}$$

If we performed the same steps to calculate the expected utility of the "sleeps at night?" question, we get $EU(Q = sleepsNight) = 0.026$. So if the biologist wants to maximize the chance of gaining useful information, she should select the "eats meat?" experiment since $EU(Q = eatsMeat) > EU(Q = sleepsNight)$.

1.4.1 Evidence of OED-like reasoning in human behavior

A growing body of psychological research has used the OED framework as a metaphor for active learning. The idea is that when people make decisions, they engage in a similar process of evaluating the "usefulness" of different actions relative to their learning goals. And they select behaviors that maximize the potential for gaining information. A success of the OED account is that it can capture a wide range of information seeking behaviors, including verbal question asking (Ruggeri & Lombrozo, 2015), planning interventions in causal learning tasks (Cook, Goodman, & Schulz, 2011), and decisions about where to look during scene understanding (Najemnik & Geisler, 2005). Figures 1 and 2 present schematic overviews of how OED principles could shape the learning process for two of these domains – causal learning (Figure 1) and word learning (Figure 2).

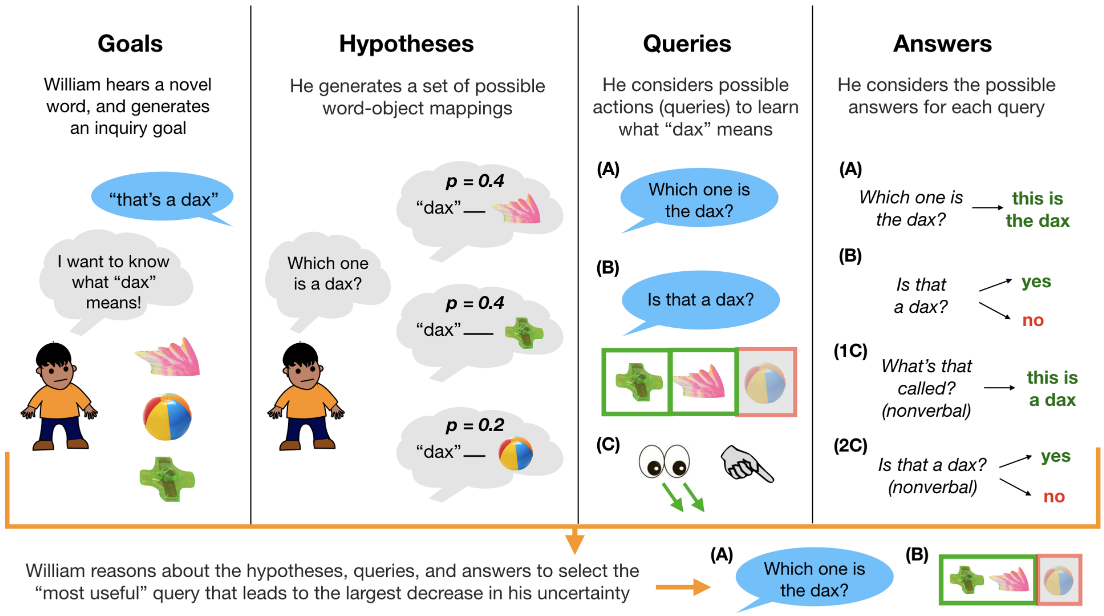


Figure 1.2: Schematic of an active word learning context using the decomposition of Optimal Experiment Design. Social input (hearing a novel word) triggers an inquiry goal. Then the learner considers potential hypotheses for the candidate word-object links, weighting each hypothesis by its prior probability. In this case, the learner thinks that the new word is less likely to refer to the familiar object BALL. Next, he considers possible queries (actions) and the potential outcomes of those actions. Note that in the word learning context, the child must direct queries towards a social partner, which provides the learner with more possible queries: both verbal (questions) and nonverbal (eye gaze; pointing). Note that the social partner must interpret the goal of the child's nonverbal queries. If the learner selects the action to maximize expected utility, then he would ask the most informative question, which removes all uncertainty for the meaning of 'dax' – 'What's that called?' If he does select the relatively less informative action of asking about a single object, he would be unlikely to ask about the familiar object BALL since there is less information to be gained from this query based on his prior beliefs.

One compelling use case of OED metaphor as a model of human behavior comes from Nelson (2005) study of eye movements during novel concept learning. Their model combined Bayesian probabilistic learning, which represents current knowledge as a probability distribution over concepts, with an OED model that calculated the usefulness of different patterns of eye movements. Here, eye movements were modeled as a type of question-asking behavior that gathered visual information about the target concept. Nelson (2005) found that participants' eye movements aligned with predictions from the OED model. Specifically, participants changed the dynamics of eye movements depending on how well they learned the target concepts. Early in learning, when the concepts were unfamiliar, the model generated a broader, less efficient distribution of fixations to explore all candidate features that could be used to categorize the stimulus. However, after the model began to learn the target concepts, eye movement patterns shifted to become more efficient and focused on a single stimulus dimension to maximize accuracy. This shift from exploratory to efficient eye movements matched adult performance on the task, suggesting that people's behavior was sensible given the structure of the learning problem and the uncertainty in the context.

Recent developmental work has used OED models to ask whether children are capable of selecting efficient behaviors that maximize learning goals. For example, Legare, Mills, Souza, Plummer, & Yasskin (2013) used a modified question asking game where 4- to 6-year-old children saw 16 cards with a drawing of an animal on them. The animals varied along several dimensions, including type, size, and pattern. Children could ask the experimenter yes/no questions to figure out which animal card was hidden in a box. Legare et al. (2013) coded questions as constraint-seeking (narrowing the set of possible cards by gathering information about a particular dimension (e.g., "Is it red?"), confirmatory (questions that provided redundant information), or ineffective that did not provide any useful information (e.g., "Does it have a tail?"). All children produced a high proportion of the more useful, constraint-seeking questions. Moreover, the number of constraint-seeking questions was correlated with accuracy in guessing the identity of the hidden card. Legare et al. (2013) interpreted these results as evidence for OED-like reasoning in children's question asking. Other work using the question-game finds that children prefer to direct questions to someone who is knowledgeable compared to someone who is inaccurate or ignorant, providing additional support for the OED hypothesis (Mills, Legare, Bills, & Mejias, 2010; Mills, Legare, Grant, & Landrum, 2011).

Another developmental example comes from Cook et al. (2011) study of causal learning.[6] They

[6] See Figure 1 for a schematic overview of the active causal learning context according the OED decomposition.

showed preschoolers a device that played music when beads were placed on top of it and manipulated the usefulness of different actions that children could take to test the device. Half of the children saw evidence that all types of beads could make the machine work, while the other half learned that only specific types of beads (defined by color) could make it go. Next, children could choose one of two sets of beads to test the machine. In one set the beads were stuck together, in the other, they could be separated. Children who learned that only some of the beads worked were twice as likely to select the separable beads. This finding suggests that children were reasoning about the amount of information they could gain by choosing the detachable beads since this choice would allow them to test each bead independently. In contrast, the children who believed that all beads worked had less to gain by picking the separable set. This result is compelling because it provides evidence that children's were reasoning about a decision (separable vs. stuck together beads) that would influence a future opportunity to generate useful information.

Although the OED approach has provided a formal account of seemingly unconstrained information seeking behaviors, there are several ways in which it falls short as an explanation of human self-directed learning. Coenen et al. (2017) argue that OED models make several critical assumptions about the learner and the learning task, including (1) the hypotheses/questions/answers under consideration, (2) that people are actually engaging in some expected utility computation in order to maximize the goal of knowledge acquisition, and (3) that the learner has sufficient cognitive capacities to carry out the calculations.

In the next section, I argue that limitations of the OED approach can be productively reconstrued as opportunities for understanding how learning from other people could scaffold active learning. I focus on integrating research and theory on social learning with five key components of the OED model: inquiry goals, hypotheses, questions, answers, and stopping rules (see Figures 1 and 2 for an overview). The proposal is that learning from more knowledgeable others provides the building blocks for children to engage in effective self-directed learning.

1.5 Part IV: Active learning within social contexts

Why is it important to integrate social contexts with active learning? First, children do not reinvent knowledge of the world, and while they can learn a tremendous amount from their behaviors, much of their generalization and abstraction is shaped by input from other people. Also, social learning can sometimes be the only way to learn something. Finally, children are often surrounded by parents, other adults, and older peers – all of whom may know more about the world than they do, creating contexts where the opportunity for social learning is ubiquitous.

Second, there is a body of empirical work showing that active learning can be biased and ineffective in systematic ways. For example, work by Klahr & Nigam (2004) showed that elementary school-aged children were less effective at discovering the principles of well-controlled experiments from their self-directed learning, but were capable of learning these principles from direct instruction. D. B. Markant & Gureckis (2014) showed that active exploration provided no benefit over passive input in category learning when there was a mismatch between the target concept and adults' prior hypotheses going into the learning task. And McCormack et al. (2016) found that 6-7 year-olds showed no benefit from active interventions on a causal system compared to observing another person perform the interventions.

In a comprehensive review of the self-directed learning literature, Gureckis & Markant (2012) point out that the quality of active exploration is linked to aspects of the learner's understanding of the task: if the representation is weak, then self-directed learning will be biased and ineffective. Coenen et al. (2017) go one step further and propose specific challenges for research on active learning. Here is a sampling of those open questions that are most relevant to this paper:

- What triggers inquiry behaviors in the first place?
- How do people construct a set of hypotheses?
- How do people generate a set of queries?
- What makes a "good" answer?
- How do people generate and weight possible answers to their queries?
- How does learning from answers affect query selection and belief change?

In the next section, I propose that theories of social learning accounts can start to address these challenges. I start from the OED model outlined in Coenen et al. (2017), and use it to integrate social and active learning. The benefit of this formalization is that it makes the different components of active learning explicit, highlighting the aspects that might be particularly challenging for young learners. Moreover, I think that we can reconstrue these "challenges" to the OED account as opportunities for understanding the role of other people in children's active learning. In each subsection, I define the challenge of active learning, discuss how social contexts could address each challenge, and highlight prior research that connects the active and social learning accounts. See Figure 3 for a schematic overview of the proposal.

1.5.1 Goals

An inquiry goal refers to the underlying motivation for people's information seeking behaviors. Often this is defined as a search for the correct hypothesis amongst a set of candidate hypotheses. Some examples of plausible inquiry goals that children might hold are:

- Is this person a reliable source of information? (selective learning)
- What is this speaker referring to? (word learning)
- What types of objects are called "daxes"? (category learning)
- How does this toy work? (causal learning)
- Where should I look next? (allocation of visual attention)

The importance of an explicit inquiry goal is that without a goal, it becomes difficult for the active learner to compare the utilities of different behaviors. That is, the learner cannot evaluate whether an action will lead to learning progress. Coenen et al. (2017) illustrate this point,

> The importance of such goals is made clear by the fact that in experiments designed to evaluate OED principles, participants are usually instructed on the goal of a task and are often incentivized by some monetary reward tied to achieving participants that goal. Similarly, in developmental studies, children are often explicitly asked to answer certain questions, solve a particular problem, or choose between a set of actions. (p. 32-33)

Thus, a prerequisite for understanding children's self-directed learning is a way to characterize children's goals. However, this is not trivial since children could consider a range of goals at any moment and there is no guarantee that learning progress be one of them. In fact, one line of theorizing about the OED hypothesis as a model of human inquiry argues that we should only expect to see effective information seeking in contexts where there are precise tasks and learning goals. For example, when a parent gives their child a new toy with several buttons on it and says, "Let's figure out how this toy works!" In this case, it becomes possible to ask whether the child approaches the task efficiently by selecting actions that provide useful information about the toy's causal structure.

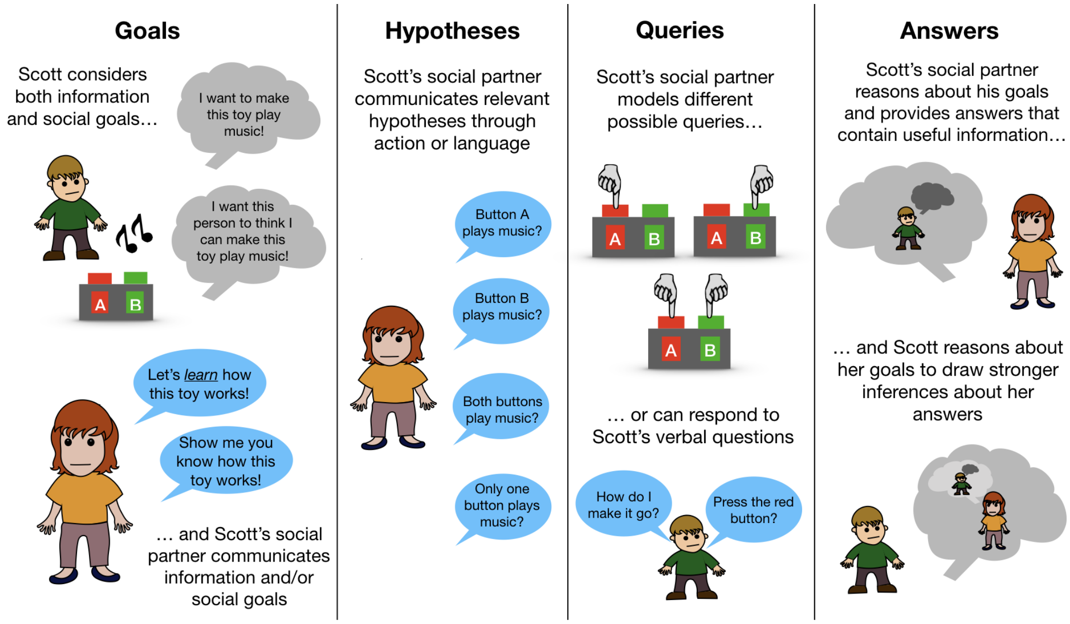


Figure 1.3: Schematic of active learning within a social context. Each panel shows how social information could influence a different component of the active learning process. These social effects occur in-the-moment of learning or over developmental time. Also, the cause of the social effect varies from the mere presence of another person to triggering a sophisticated psychological reasoning process about others' goal-directed behaviors. Note that the panels correspond to the different sub-sections in Part IV of the text.

This example illustrates how children's interactions with other people could play a role in triggering inquiry behavior. Both adults and older peers can construct contexts with clear learning goals to support children's information seeking. This connection draws on influential ideas in cognitive development that frame social learning as a form of scaffolding (Zone of Proximal Development (Vygotsky, 1987), Guided Participation (Rogoff et al., 1993), and Guided Play (Weisberg, Hirsh-Pasek, & Golinkoff, 2013)). Under these accounts, adults place children in contexts that present something new to be learned but importantly contain learning goals that are achievable given children's current capabilities.

One example comes from Weisberg et al. (2013) "guided play" proposal. They define guided play as an intermediate learning context, falling between unstructured free play and constrained direct instruction. The precise boundaries between these contexts are challenging to define, but the critical dimension is the level of control that the adult has over the interaction. Weisberg et al. (2013) present the following example to highlight the difference between guided play and direct instruction:

> For example, a teacher with the goal of teaching new vocabulary words could take a direct instruction approach, by telling children the meanings of the new words they encounter in a storybook or by showing examples: "This is a helmet. A helmet goes on your head to stop your head from getting hurt if you fall off your bike." Or, she could take a guided-play approach, introducing the new words in the context of a child's play episode while encouraging children to think broadly about the word's meaning: "She's got a helmet on while riding her bike. What do you think would happen if she fell off her bike and wasn't wearing her helmet? (p. 106)

While these contexts appear to be quite similar, the key difference is whether the child initiated the activity. In free play, the child drives the interaction; whereas, in direct instruction, the more knowledgeable person explicitly tells the learner what to do, asks questions, and demonstrates new concepts. Weisberg et al. (2013) propose the guided play provides the right combination of structure and opportunity for exploration, leading to better learning.

One less-emphasized feature of the guided play proposal is the adult's role in establishing a clear learning goal. In the previous example, in the absence of a social partner, it is unclear what goals the child would pursue when playing with the storybook. However, if there is an adult who has knowledge about the book and communicates that she wants to talk about it, then the child can begin to seek information. In the language of the OED framework, the adult creates a context that triggers inquiry goals, which scaffolds children's active learning. See the "goals" panel in Figure 3 for another example of how a social partner could communicate an inquiry goal in a causal learning context.

In addition to communicating goals, the mere presence of other people changes children's tendency to detect whether they understand something. This capacity for monitoring one's uncertainty is a sub-component of metacognition (thinking about thinking) and has been the focus of much developmental research (see Lyons & Ghetti (2010) for a review). The upshot of this research is that while children might be capable of showing behaviors that are sensitive to uncertainty (e.g., selectively exploring an object with ambiguous causal structure), the ability to explicitly access the underlying representation of uncertainty is slower develop.

The primary evidence for this slow developmental trajectory is children's failures to monitor and verbally report their uncertainty, even when it should be obvious. For example, Markman (1979) had elementary school aged children read paragraphs containing inconsistent information (e.g., "fish can't see without light, and there's no light at the bottom of the ocean, but some fish at the bottom of the ocean only know their food by its color."). After reading the paragraph, children answered a series of ten questions that gave them the opportunity to ask for clarification about the inconsistencies in the text. Overall, children performed poorly, as Markman (1979) put it, "even highly motivated children" were unable to detect inconsistencies in the incoming information. However, if the experimenter gave children a warning or a challenge to find a problem with the essay, then they were able to generate more questions, suggesting that they more likely to monitor uncertainty while reading.

While not discussed in these terms, the "warning" manipulation can be reconstrued as an intervention of the social context on children's expectations about incoming information. Perhaps children approached Markman (1979)'s task with the default expectation that the adult would provide complete and helpful information, making it less important to allocate resources to uncertainty monitoring. In fact, the assumption that other people will be helpful is a critical feature of social learning theories, and the empirical work on children's pedagogical sampling assumptions reviewed in Part I provides empirical support. However, when the social context shifts these expections, then children might be more likely to monitor the input for inconsistencies, and in turn, be more likely to generate inquiry goals.

Converging evidence comes a study by Kim, Paulus, Sodian, & Proust (2016) where they measured 3- to 4-year-olds' uncertainty monitoring in contexts where children either did or did not expect to communicate with another person. In the task, children were either knowledgeable or ignorant about which toys an experimenter hid in a box. In the "informing" condition, children had to tell another person about the contents of the box, but they were given the opportunity to "opt-out" of responding if they were unsure ("Max wants to know what's inside the box. Can you help him? If you do not want to tell him, it's okay. I can tell him."). In the "explicit" condition, children only reported on their uncertainty. Interestingly, the children who had the potential to inform another person showed higher rates of uncertainty monitoring, opting-out more often when they did not know what was inside the box. Similar to the Markman (1979) study, the social context shifted children's expectations, making them more likely to detect lack of knowledge. These results also connect with work showing that the process of generating explanations for other people can reveal inconsistencies in understanding (Lombrozo, 2006).

Another interesting effect of social contexts on children's goals comes from work exploring how social input can shape children's implicit theories of intelligence (Dweck & Leggett, 1988). Implicit theories of intelligence refer to children's internal working models of the world and provide general frameworks for processing information and generating predictions about behavior. Dweck & Leggett (1988) propose a cognitive model where holding different implicit theories of intelligence leads children to adopt different goal orientations, which manifest in different choice behavior during learning. For example, if a child holds the belief that intelligence is malleable (an incremental theory), they will want to increase competence (select a learning goal) and therefore be more likely to choose tasks that facilitate learning.

Empirical work shows that social partners can directly intervene on children's learning vs. performance goals. For example, Elliott & Dweck (1988) manipulated elementary school-aged children's goals by presenting them with a choice between one of two tasks described in the following ways:

- *Performance task.* In this box we have problems of different levels. Some are hard, some are easier. If you pick this box, although you won't learn new things, it will really show me what kids can do.

- *Learning task*. If you pick the task in this box, you'll probably learn a lot of new things. But you'll probably make a bunch of mistakes, get a little confused, maybe feel a little dumb at times — but eventually you'll learn some useful things.

Elliott & Dweck (1988) found that when the social partener oriented children's towards learning goals, they tended to choose the more difficult task despite being morewere likely to make mistakes and appear incompetent. In another study, Dweck & Leggett (1988) showed that children who already held performance goals viewed effort on a task as an index of ability, whereas children with learning goals viewed effort as a means for improvement. Moreover, both lab-based experiments and observational work provide evidence that the language adults use when praising children can influence whether children adopt an incremental theory, which lead to more inquiry goals (Cimpian, Arce, Markman, & Dweck, 2007; Gunderson et al., 2013).

Taken together, the research on implicit theories suggests that social contexts can also communicate the *value* of learning goals relative to other goals that children may pursue. The Elliott & Dweck (1988) finding – that children oriented toward learning goals selected more challenging tasks – maps onto behaviors that the OED framework aims to characterize. That is, the goal manipulation encouraged children to select actions to make progress on learning.

Interactions with other people can also introduce other "social" goals that influence children's information seeking. The OED framework only includes informational goals such that the utility of a behavior is determined by whether it results in a reduction of uncertainty. However, empirical and modeling work in other domains has extended the basic information-seeking account to include *situation-specific utility functions* that include goals such as saving time, money, or cognitive resources. For example, Meder & Nelson (2012) designed a series of experiments where "pure" information seeking goals (e.g., maximizing accuracy) were placed at odds with the reward structure of the task. The critical manipulation was the addition of asymmetric rewards for correct and incorrect categorization decisions in a binary classification task. Asymmetric rewards forced the learner to choose the less likely category to earn the highest reward (i.e., they must be willing to forego the goal of being accurate and make mistakes to get the highest score). When these goals were put in conflict, participants' behavior was mixed. Participants only reduced their preference for information seeking and improving accuracy when the asymmetric reward structure was made explicit via task instructions. However, Meder & Nelson (2012) take their results as evidence that people were considered goals other than pure information gain or reward maximization, suggesting the need to consider additional factors when trying to understand what behaviors people think are useful.

One important feature of social contexts is that they engage a process of psychological reasoning about others' mental lives. And once the learner starts thinking about the other person, they could begin to consider an additional set of "social" goals that may conflict with or support their learning goals. Consider the "goals" panel of the schematic active-social learning context shown in Figure 3. If the learner is worried about whether his social partner thinks he is smart, then he might prioritize actions that minimize the chance of making a mistake and perhaps not even attempt to make the toy work. Alternatively, the learner might seek out easy tasks to demonstrate his competence at the expense of choosing actions that help him learn about the world. On the other hand, the child's social partner could explicitly communicate the learning goal using natural language, e.g., "Let's learn how this toy works!"

Recent advances in modeling pragmatic communication[7] provide an interesting connection with the situation-specific-utility functions in the OED framework. For example, Yoon & Frank (2017) model speakers' decisions to use polite speech as opposed to direct speech (e.g., indirect language such as "I don't think that dress looks phenomenal on you" as opposed to "It looks terrible") as a tradeoff between maximizing informational and social goals. In their model, speakers reason about whether their utterance will communicate information faithfully with as little effort as possible (information goal) and/or cause harm to one's own or another's self-image (social goal). Yoon & Frank (2017)'s empirical results suggest that speakers are in fact balancing social and epistemic goals, suggesting that both are necessary to account for polite speech.

We can make a direct connection between the politeness model and the OED account. Both assume that people produce actions to maximize utility, but the politeness model expands the type of information that people use to compute the usefulness of an action. An intriguing possibility for future research is to adapt the utility-theoretic approach used by Yoon & Frank (2017) to model children's information seeking behaviors such as asking questions in different social contexts. These ideas also connect to the effects of task framing (performance vs. learning oriented) on children's decisions to attempt more challenging tasks (Dweck & Leggett, 1988). The presence of another person could be modeled as an increase in the weight that children place on maximizing social goals,

[7]Rational Speech Act (RSA) framework for pragmatic reasoning. The RSA approach models language comprehension and production as, "a process of recursive reasoning about what speakers would have said, given a set of communicative goals" (p.819) (Goodman & Frank, 2016)

leading children to select easier tasks where they can appear competent. This is a reconstrual of the goal-orienting account reviewed above since it characterizes the behavior as an output from a mixture of goals as opposed to the social context triggering either a performance or a learning goal.

One significant gap in research on children's inquiry goals is an estimate of how often children experience contexts with clear learning goals in their daily lives. That is, we do not yet have a theory of the kinds of environments that would lead children to generate learning goals. However, research on *guided participation* across cultures provides an interesting counter-example (Rogoff et al., 1993). Rogoff et al. (1993) provided parents and their toddlers with a set of novel objects and coded the amount of "caregiver orienting" behavior. The study was conducted in four different cultural communities (a Mayan Indian town in Guatemala, a middle-class urban group in the United States, a tribal village in India, and a middle-class urban neighborhood in Turkey) that varied in how separated children were from adult activities and whether formal schooling was emphasized. Caregiver orienting behavior was defined as,

> Caregiver orients child involved introducing new information or structure to the child (at any point in the episode) regarding the overall goals or a key part of the event or what was expected in the situation. Orienting framed a major goal, not just specific little directives for particular actions. (p. 43)

Parents in all four communities produced high rates of structuring and orienting behaviors (with the lowest rate of structuring being 81% of play episodes). Thus, when placed in a structured activity, adults make sure children are aware of the goal (e.g., learning the function of the novel toy). However, the communities differed in how often children were directly involved with adult activities in day-to-day life, with the children raised in rural villages often having early access to adult economic and social events. An open question is whether older peers and adults need to be directly engaged with the child to trigger inquiry goals. Perhaps increased access to observing adult goal-directed behaviors could facilitate children to generate learning goals since children would see lots of adult activities that they do not understand but are motivated to learn about (e.g., cooking, shopping, working).

In sum, understanding children's goals represent a critical step in characterizing the relative contribution active learning to cognitive development. Social contexts have the power to trigger inquiry goals, but they also introduce social goals that might be at odds with children's information seeking. One direction for future research is to quantify the prevelance of different goals in children's everyday experience. It would be useful to know how much of children's daily activities involve settings where there is a clear learning goal and whether the goal was generated independently or communicated by older peers and adults. It would also be useful to know how the distribution of these goals change as a function of development, especially as children enter school and across different cultural contexts where children have differential access to structured (e.g., lessons and sports) vs. unstructured activies (e.g., free play). Moving beyond the lab-based studies of OED-like reasoning is not trivial and will require leveraging recent advances in large-scale observational data collection about children's daily experiences. However, it is important to ask whether the efficient information-seeking behaviors that children produce in the lab would "scale up" to the complexity of real world learning environments, especially if we want to use the OED model as an account of cognitive development.

1.5.2 Hypotheses

Once a learing goal is established, the next step of inquiry is to decide what hypotheses to consider. Intuitively, a hypothesis is a candidate explanation about how the world works. For example, consider the schematic learning contexts shown in Figures 1 and 2. In the casual context, the hypotheses for how the toy works could include: (1) Button A, (2) Button B, (3) Buttons A and B simultaneously. In the word learning context, the child might think the new word "dax" means: (1) dax = object A, (2) dax = object B, or (3) dax = object C.[8]

The set of hypotheses under consideration is critical for quantifying effective self-directed learning in the OED account. The usefulness function of expected information gain outlined in Part III works by comparing the learner's uncertainty over hypotheses before and after she sees an answer ($U(a) = ent(H) - ent(H|a)$). Without knowing the conents of the hypothesis space, it becomes challenging to select the best choice for reducing uncertainty. Put another way; the OED framework does not readily deal with situations where learners might have to consider a large space of hypotheses, might hold the wrong hypotheses, or might perform actions without considering any hypotheses at all. This is an especially important challenge for developmental accounts that draw on OED principles since these scenarios seem quite plausible for young learners.

[8]Note that this hypothesis space is simplified since it only considers the possibility of one-to-one word-object mappings.

However, one useful function of social learning contexts is that they can provide a clear set of possible explanations for the actual state of the world. In practice, adults and older peers who have access to the correct hypothesis can constrain the space of children's hypotheses to facilitate information seeking. This effect of social context parallels the discussion of the role of social partners in constraining goals in the previous section.

One relevant case study comes from work on children's early word learning. The challenge for a young word learner is that even the simplest of words, concrete nouns, are often used in complex contexts with multiple possible referents, which in turn have many conceptually natural properties that a speaker could talk about. This ambiguity creates the potential for an (in principle) unlimited amount of hypotheses that children could consider when trying to figure out the meaning a novel word. Remarkably, word learning proceeds despite this massive uncertainty, with estimates of adult vocabularies ranging from 50,000 to 100,000 distinct lexical concepts (Bloom, 2002).

It does not seem plausible for children to entertain all hypotheses about possible word-object links. But which ones should they consider? One proposed solution is that word learners only consider a single word-object link at a time (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013). Under this account, the child makes an initial guess about the word meaning, and then only stores that word-object link until she receives sufficient evidence that her initial hypothesis was incorrect. If she does see enough counter-evidence, then she will switch to a new, single hypothesis that better matches the statistics in the input.[9] However, another influential account of early word learning, inspired by basic associative learning principles, argues that word learners store more than a single hypothesis. Under this theory, children's hypothesis spaces are reduced gradually via the aggregation of word-object co-occurrence statistics across multiple labeling events (Siskind, 1996; C. Yu & Smith, 2012b). Support for this account comes from experimental work showing that both adults and young infants can use word-object co-occurrence statistics to learn word meaning from individually ambiguous naming events (Smith & Yu, 2008). Moreover, adults show evidence of being able to recall multiple word-object links from an initial naming event (Yurovsky & Frank, 2015).

The critical difference between these proposals is how much information learners store in the

[9]This "propose-but-verify" account parallels work by Bonawitz, Denison, Gopnik, & Griffiths (2014) in the domain of causal learning, which suggests that a "Win-Stay, Lose-Sample" algorithm (inspired by efficient sampling procedures in computer science) provides a better explanation of children's hypothesis testing behaviors compared to an algorithm that enumerates the entire hypothesis space.

hypothesis space. Some of our own work provides evidence that the social context can modulate the content of the learner's hypothesis space (MacDonald et al., 2017b). Inspired by ideas from Social-pragmatic theories of language acquisition that emphasize the importance of social cues for word learning (Bloom, 2002; Clark, 2009; Hollich et al., 2000), we showed adults a series of word learning contexts that varied in ambiguity depending on whether there was a useful social cue to reference available (a speaker's gaze). We then measured learners' memory for alternative word-object links. People flexibly responded to the amount of ambiguity in the input, and as uncertainty increased, they tended to store more word-object hypotheses. Moreover, we found that learners stored representations with different levels of fidelity as a function of the reliability of the social cue. When the speaker was a less reliable source of information, learners distributed attention and memory broadly, storing more hypotheses.

These results provide evidence that the content of learners' hypothesis spaces changed as a function of social information. Further suppport for this idea comes from experimental work showing that even children as young as 16 months prefer to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin, 1993), and analyses of naturalistic parent-child labeling events shows that young learners tended to retain labels accompanied by clear referential cues, which served to make a single object dominant in the visual field (C. Yu & Smith, 2012a). One important direction for future research is to measure the full causal pathway from variation in social information through children's hypothesis spaces to their information seeking behaviors. For example, it would be interesting to know whether learners' subsequent questions or decisions about where to allocate attention would be affected by the social context in which they were first exposed to a new word.

A second case study that illustrates the importance of considering the learner's hypothesis space comes from work by Lucas, Bridgers, Griffiths, & Gopnik (2014) where they compared children and adult's capacity for learning different kinds of causal structures. In the task, participants saw a series of events that consisted of a training phase where an experimenter placed objects on a box that either played or did not play music. The participant's goal was to learn which objects, or combination of objects, made the box work. In the disjunctive condition, only single objects (A or C, but not B) made the toy play music. In the conjunctive condition, only the combination of two different objects (A **and** C) would make the toy play music. After seeing several demonstrations, both children and adults were tested on ambiguous events where they could either infer a disjunctive or conjunctive causal relationship. Only children learned the conjunctive relationship, even though the evidence favored this interpretation. The authors speculate that adults were biased towards the disjunctive hypothesis since it is a more common causal relationship in everyday experience; whereas children have accumulated less experience and have diffuse prior beliefs, which caused them to update their beliefs faster when they saw evidence in favor of the conjunctive hypothesis.[10]

The striking difference between adults and children in the Lucas et al. (2014) study suggests that it is not trivial to know what hypotheses children bring to the learning task. Moreover, features of the immediate social context could bias the content of children's hypothesis space. In fact, Lucas et al. (2014) suggest that,

> Most of the time, adults do not need to dramatically change their beliefs or abandon their hypotheses for dramatically different ones. Indeed, doing so would be a liability: adults are expected to make accurate predictions and good decisions, not bold inductive leaps. Adults are also unlikely to have caregivers to correct their errors and save them from poor choices. (p. 295)

This proposal makes a testable prediction: that contexts in which learners "feel" safe to make mistakes would lead children to consider and test a broader range of hypotheses. Another effect of the social context highlighted in Lucas et al. (2014)'s study is that the experimenters scaffolded children and adults consideration of the disjunctive and conjunctive hypotheses. That is, their actions constrained the hypothesis space to isolate learning and information seeking effects. The "Hypotheses" panel of Figure 3 presents a schematic illustration of how social partners can constrain hypotheses in the moment of learning. Similar to the research on goals, the majority of the work on hypotheses has focused on lab-based studies. Thus, it is an open question as to how much of role adults play in communicating relevant hypotheses during children's daily experience.

Social information can also shape children's hypothesis space by providing the necessary input to revise children's incorrect intuitive theories about how the world works. Research on conceptual change provides evidence that children integrate social input with their current hypotheses (Gelman, 2009). Conceptual change refers to a "radical" reconstruction of an intuitive theory about how the

[10] Figure 1 illustrates a version of the active causal learning scenario. In this case, the learner believes that pressing both buttons to activate the toy is a less likely hypothesis. As a result, the query to test the conjunctive hypothesis "press both buttons" becomes less useful for gaining information since it would produce confounded evidence concerning her other, disjunctive hypotheses.

world works. For example, elementary school-aged children tend to hold a mixture of beliefs about the shape of the earth (Vosniadou & Brewer, 1992). These theories range from a flat earth theory that matches children's everyday perceptual experiences (i.e., walking on flat ground) to the adult-like, sphere model, which reflects the actual state of the world. Interestingly, some children hold intermediate beliefs such as a dual-earth theory where there are two earths: "a round one which is up in the sky and a flat one where people live" (p. 550). The fact that some children entertain a dual-earth theory suggests that they are actively integrating aspects of their initial theory with the information they get from other people who already hold the correct, sphere theory. Moreover, for children to hold the sphere model at all, they must have learned it from social input.

Additional evidence for the role of social input comes from work on children's reasoning about biological concepts. For example, the concept of "alive" takes years to fully develop, with younger children (under ten years of age) often claiming that only animals, and not plants, are alive. Opfer & Siegler (2004) tested the hypothesis that evidence of goal-directed movement is critical for children's extension of the "alive" concept. In their study, 5-year-olds' were trained in different ways to think about the concept of a "living thing" and then asked whether they believed that plants were alive. Children either learned that plants were capable of goal-directed movement (e.g., "The house plant is growing this way. It needs the sunlight over here."), that plants were capable of growth, or that plants need water to survive. Children in the goal-directed movement condition showed the most significant theory revision, saying that plants were also living things more consistently on a post-intervention categorization task. Converging evidence comes from research on the links between language experience and performance on various cognitive tasks. For example, empirical work shows that deaf children without access to a natural language perform worse on Theory of Mind tasks (Peterson & Siegal, 2000); that Korean speakers perform better than English speakers on tasks that require categorizing based on tight vs. loose distinctions, which are lexicalized in Korean (McDonough, Choi, & Mandler, 2003); and that exposure to a first language reduces infants' capacity to detect non-native phonetic contrasts (Maurer & Werker, 2014).

One conclusion from the work on children's conceptual change is that certain hypotheses are more primitive and children require social input to revise them (flat earth -> spherical earth). Taken together, the case studies reviewed in this section illustrate several points relevant to the integrative active-social learning account. First, the set of hypotheses that children consider are likely to be quite different from adults (and possibly different from what the experimenter thinks the child is considering). Second, children generate hypotheses using a mixture of prior knowledge, expectations about the task, and social input. Third, there are (at least[11]) two timescales through which social learning can shape hypotheses. An in-the-moment timescale where others' behavior constrains hypotheses for the current task – for example, referential gaze indicating candidate word-object mappings or an adult suggesting a disjunctive (one-block) vs. a conjunctive (two-block) theory for how to make a toy work. And a developmental timescale where prior interactions with other people and cultural learning modify the hypotheses that children bring to the learning task (e.g., the conceptual change and language effects reviewed above).

1.5.3 Queries

Queries in the OED framework refer to the experiments that a scientist could conduct to gather information about their hypotheses. Queries in human information seeking can refer to a variety of actions, including verbal questions, pushing a button to figure out how a toy works, and decisions about where to look. One of the strengths of the OED account is the fact that it provides general principles that explain such a broad range of behaviors.

However, the challenge for the young learner is to discover what behaviors are available and of those actions which might be particularly useful for gathering information. In this section, I illustrate how the social learning context provides valuable input to this learning process via demonstrations of useful actions that learners could take to gather information. I suggest that this process involves adults' modeling useful information seeking and children recognizing and imitating these behaviors.

It seems obvious that children would look to older peers or adults to learn what actions are useful. However, a large body of empirical work suggests that even young infants will not imitate every action that they see. Instead, children show evidence of "rational imitation" and will look for cues about others' goals and use this information to selectively imitate. For example, Gergely, Bekkering, & KirÃąly (2002) measured how often 14-month-old infants imitated an adult's actions – turning on a light with her head (less efficient) instead of her hands (more efficient) – as a function of whether there was a relevant explanation for selecting the less efficient action (whether the adult's

[11]We could make a further distinction between the developmental and the cultural timescales, where developmental refers to information acquired from interactions with others in the child's lifetime (e.g., disjunctive causal structures are more likely to occur in the world) and cultural refers to information that has accumulated throughout human evolutionary history (e.g., access to natural language or concepts such as the spherical earth.)

hands were occupied). They found a substantial difference in imitation rates across conditions (69% in the hands-free vs. 21% in the hands-occupied), suggesting that children recognized the reason for the inefficient action and chose to ignore the means and focus on the goal of turning the light on in the most efficient way possible.

The high rates of imitation in the hands-free condition highlight another component of learning from others' actions: that children tend to overimitate behaviors even when these actions are not directly relevant to the task. For example, Call et al. (2005) compared imitation behaviors of 2-year-old children after they watched someone demonstrate how to open a tube using only the necessary actions or using the actions plus a style component unrelated to opening the tube (e.g., removing the tube's cap with an exaggerated twisting motion). Almost all children (93%) imitated the causally irrelevant action, providing evidence that they were focused on reproducing each of the experimenter's actions and not just reproducing the outcome of opening the tube. Other empirical work shows that social factors matter for children's decisions to imitate. Carpenter et al. (1998a) showed that 14- and 18-month-olds were less likely to mimic an adult's action if the action was marked verbally as a mistake (e.g., "Whoops!"). Buchsbaum, Gopnik, Griffiths, & Shafto (2011) provide evidence that the children are more likely to overimitate when the adult is described as a "knowledgeable teacher" as opposed to "naive." And Carpenter, Call, & Tomasello (2002) showed that giving children explicit information about another person's goals before a causal demonstration leads to an increase in imitation and learning of the correct casual structure. Together, the work on imitation suggests that children could use others' goal-directed, information seeking behaviors as input to help them learn what queries are useful.

Research on verbal question asking provides insight into how social input guides children's developing information seeking. First, consider that the very act of asking a question in natural language requires that children have acquired a conventionalized symbolic system, which must have been learned from social input. Second, both experimental work and corpus analyses provide evidence that children's question-asking becomes more varied and productive over the first years of life as they get exposed to more complex language input (see Chouinard et al. (2007) and Legare et al. (2013) studies reviewed in Part II). Moreover, as children develop, they improve the timing of their turn-taking during question-answer exchanges, reducing the length of gaps between turns (Casillas, 2014). Interestingly, Casillas (2014) also showed that adults were sensitive to children's developing question-answering skills, waiting to ask more difficult questions until children were older, and modifying their questions if children were confused (e.g., "Who is this? -> What's he called? -> Who is he? -> What is his name?").

The majority of research on children's question asking has focused on aspects of the child's behavior, exploring how the type, content, and effectiveness of questions changes as children develop. However, several studies have measured children's exposure to questions in their input. For example, Yu & Shafto (2017) coded parent-child interactions from the CHILDES database to measure the amount of "pedagogical" questions in children's input. They differentiate "pedagogical" from "information seeking" questions by coding whether the adult already knew the answer. For example, "What's that called?" would be pedagogical; whereas, "What did you do at school?" would be information seeking. Approximately 30% of parents' questions were pedagogical, 60% were information seeking, and 10% were rhetorical (i.e., not intended to be answered verbally). Parents also directed a smaller proportion of pedagogical questions to older children. Yu & Shafto (2017) speculate that the function of pedagogical questions is to help children learn.

It is interesting to consider how children might internalize adults' question asking behaviors (modifications, pedagogical questions) and incorporate them as part of their own question asking repertoire. However, more work is needed to understand the link between adults' question-asking practices and children's behaviors. This research could be especially interesting since observational studies have found that that parents' use of wh-questions predicts children's later vocabulary and verbal reasoning outcomes (Rowe, Leech, & Cabrera, 2017) and children of parents who were trained to ask "good" questions during book reading episodes at home also asked better questions during book reading sessions at school (Birbili & Karagiorgou, 2009). One explanation for these associations is that wh-questions challenge children to produce more complex responses that build verbal abilities. However, another intriguing causal pathway is that the frequency and type of questions that parents shape children's information seeking skills by providing templates for useful questions.

After children generate a set of possible questions, they have to evaluate the relative usefulness (i.e., utility) of the different queries. But how do children learn the features of a good question? One solution is for children to observe other people's question asking behaviors, recognize which questions are useful, and leverage their imitation skills to model those behaviors. In fact, there is evidence from work with adults showing a substantial difference between people's question-generating (harder) and question-evaluation (easier) skills. For example, Rothe, Lake, & Gureckis (2015) asked a group of adults to play a modified “Battleship” game where they had to find the location of three ships that consisted of 2-4 tiles and could be oriented in either the vertical or horizontal direction on a 6x6 grid. Participants gathered information sequentially by uncovering one tile at a time. At different points in the task, the game would stop, and participants could ask any question using natural language. Rothe et al. (2015) used a formal OED model to measure the expected information gain of each question, and people rarely produced high information value questions. However, in a follow-up experiment Rothe et al. (2015) had a different group of adults play the Battleship game, but this time participants had access to the list of questions generated by participants the free-form version, In this contexts, adults were quite good at recognizing and selecting high information value questions.

Developmental work provides additional evidence of this production-comprehension asymmetry. For example, children younger than the age of three have difficulty generating appropriate verbal questions compared to their older peers in “Twenty Questions” style tasks that are designed to measure question-asking skills (Mills et al., 2010, 2011). However, when Mills, Danovitch, Grant, & Elashi (2012) tested 3- to 5-year-old’s capacity to learn from observing third-party question-answer exchanges, they found that even the youngest children were capable of using information elicited by others’ yes/no questions to identify the contents of a box. Interestingly, children recognized the value of others’ question-answer exchanges, paying more attention to them as compared to third-party exchanges that did not include question-answer exchanges. These results suggest that even at an age where generating questions “from scratch” might be difficult, children can observe and learn from questions that occur in their social environment. Mills et al. (2011) also explored this phenomenon by directly manipulating whether children were exposed to a training phase where adults modeled useful questions before playing the question asking game. They found that even though the youngest children were not successful at constructing good questions, they were able to ask useful questions at a much higher rate following exposure to explicit modeling.

Work with elementary-school-aged children in the domain of scientific inquiry also shows that generating a good question is a challenging aspect of inquiry skills. One relevant example comes from Kuhn & Pease (2008)’s intervention study comparing children trained on scientific inquiry skills (e.g., understanding the objectives of inquiry and identifying questions) to a group of slightly older students who had not participated in the direct instruction training. Children in the training group showed progress. In contrast, children in the comparison group failed to develop these formal scientific inquiry skills in the absence of a particular kind of input. kuhn2008needs summarizing the key results,

> Consistent with the findings of Kuhn and Dean (2005), identifying a question appears to play a key role in making the rest of the inquiry cycle productive . . . Like other components of the inquiry process, this skill is not one a student learns once and has mastered. (p. 555)

Another way that social contexts influence the set of possible queries is by adding a "social target" for information seeking. That is, the child now has an additional choice: to gather information from the non-social world or other people. The "Questions" panel in Figure 3 shows a child trying figure out how a toy works. If there is no social partner present, they could still try actions that test the system directly and seek information from the world. But if another person is present, the child now has the option to ask verbal questions or to seek help via nonverbal cues (e.g., pointing, facial expressions).

Recent empirical work has explored the factors that influence children's decisions to seek information from other people. For example, Fitneva, Lam, & Dunfield (2013) measured 4- to 6-year-old's decisions about how to learn the features of a novel social category: "moozles." Critically, the target concept was either visible (hair color) or invisible (knowledge of a foreign language). Children could choose to look directly at the moozle or ask a moozle expert. Children tended to "look" for the visible property and "ask" for the invisible property. These results suggest that children have some meta-understanding of the kinds of information that is particularly useful to learn from social partners.

Additional evidence comes from work by Lockhart, Goddu, Smith, & Keil (2016). They asked 5- to 11-year-old children whether a person "growing up on their own" could learn facts that varied along a dimension of learnability based on direct perceptual experience (that the sky is blue vs. that the earth is round). Children of all ages performed quite well, showing that could verbally articulate the kinds of information that would require interactions with other people. The items in this study did not test children's appreciation for more indirect forms of social learning: that is, to learn that the sky is blue relies on having a conventional language system for referring to concepts. Gelman

(2009) refers to this as “a hidden level of cultural input” (p. 2). The critical point is that even in learning contexts that appear entirely self-directed, children’s massive amounts of accumulated experience with other people shape the hypotheses and questions that children consider.

Another set of relevant examples comes from work on children’s help-seeking behaviors. Vredenburgh & Kushnir (2016) had children build toys that required multiple steps, and on each step, children were given the opportunity to ask for help from the experimenter. Each step varied in difficulty and children naturally varied in their toy building skill. Children asked for help when the step on more challenging steps, suggesting that preschoolers sought help systematically. Moreover, work by Gweon & Schulz (2011) found that 16-month-old infants are selective help-seekers, turning to a social target to request information or acting on the world depending on which information source was more likely to help them achieve their current goal. In this case, the infants’ goal was to make a malfunctioning toy produce music, and the critical manipulation was whether children saw evidence that explained the likely cause of failure being the toy versus their capacity for making the toy play music. When the toy was likely to be broken, they reached for a new object (queried the world), but in contrast, when the evidence suggested that the child was the issue, then they sought help from a nearby adult.

Some of our work has explored how the presence of another person changes the set of information seeking behaviors available (MacDonald et al., 2017a). Inspired by theories of natural vision that characterize eye movements as an information seeking mechanism, we asked whether children and adults would allocate more visual attention to a speaker when the linguistic signal was noisy to support the goal of rapid language understanding. We used an eye-tracking task to measure participants’ gaze patterns while they processed clear or degraded speech (speech with brown noise added). Both children and adults spent more time fixating on the speaker in the degraded speech context. Interestingly, children and adults were also more accurate in word recognition even though the speech was noisy and difficult to process. This result suggests that listeners were compensating for the uncertainty in the auditory channel by gathering visual information from the speaker. Critically, listeners would not have been able to gather this information if the speaker was not present (e.g., listening to a noisy recording) and in clear view.

In sum, queries provide the tools for information seeking. However, more research is needed to understand the link between children’s input and the set of questions they consider . One proposal is that children use their powerful imitative learning abilities to model the question-asking behaviors demonstrated by more knowledgeable others. Moreover, social contexts can fundamentally change the set of actions available to the learner by providing a social target for information seeking behaviors.

1.5.4 Answers

In the OED framework, answers refer to possible states of the world after the learner generates a query. An answer is useful if it results in a substantial decrease in uncertainty about the actual state of the world. The challenge for the active learner can be separated roughly into two parts: (1) figure out what which answers are likely and (2) decide how much you should learn from an answer after seeing it. The social context plays a role in each component and is the focus of the current section.

Defining the specific features of a "good" answer is challenging. Intuitively, a good answer gives the learner information that they did not already know, that they were interested in learning, and that is likely to be useful beyond the current context (i.e., to generalize). Even within the formal OED framework, there have been a variety of ways to instantiate the utility function (e.g., information gain, probability gain, and Kullback-Leibler divergence) to compute the value of an answer (see Nelson (2005)). All of these information-theoretic utility functions take into account the learner's prior beliefs represented as probability distributions over hypotheses and calculate the impact that an answer would have on the learner's beliefs represented as conditional probability distributions.

Several social learning accounts argue that a key function of social information is to provide useful answers. For example, evolutionary models of cultural learning argue that the human capacity for efficiently transferring knowledge between individuals allows for the gradual accumulation of small improvements that eventually lead to complex tools, beliefs, and practices that would be difficult, if not impossible, for any individual to discover on their own (Kline, 2015). Boyd et al. (2011) provide the following example,

> For example, a rare chance observation might allow a hunter to associate a particular spoor with a wounded polar bear, or to link the color and texture of ice with its stability on windy days just after a thaw. Such rare cues allow accurate low-cost inferences about the environment. However, most individuals will not observe these cues, and thus making the same inference will be much more difficult for them. Organisms that cannot imitate

> must rely on individual learning, even when it is difficult and error-prone. They are stuck with whatever information that nature offers. (p. 10921)

The fundamental idea is that communicating useful and difficult to acquire information enhanced evolutionary fitness. Thus, there is an a priori reason to expect that information from other people will be useful.

This concept is critical to Csibra & Gergely (2009)‘s theory of “Natural Pedagogy” reviewed in Part I. Specifically, they argue that an assumption of *generalizability* is a fundamental component of adults’ communication with children. Their account has three elements: adults transmit generic information, adults provide ostensive cues to signal generalizable knowledge, and children show sensitivity to these cues, treating information differently when they are present. Evidence for a bias towards generalizability comes from a set of empirical studies showing that infants will generalize more often when learning information accompanied by ostensive communicative cues such as eye gaze or child-directed speech. For example, infants are more likely to generalize the positive vs. negative valence associated with a specific object-person pairing to a new person if the valence was demonstrated with pedagogical cues (Gergely, Egyed, & KirÃąly, 2007). Also, infants are more likely to encode the stable features of an object, as opposed to its location in space, if a communicative signal such as a point guided their attention (Yoon et al., 2008).

Even if learning occurs in a social context with a default assumption of useful and generalizable information, not all answers are equally informative. Thus, the second challenge for information seekers is to evaluate possible responses to a query to figure out how much they would update beliefs. This challenge is perhaps one of the more developed connections between the formal social and active learning accounts. Researchers have made progress in modeling the influence of different assumptions that a learner could make about the generative process of answers. For example, Shafto et al. (2012b) lay out a continuum of sampling assumptions:

- *Weak sampling*: answers generated at random from the set of all possible answers (independent of target hypothesis)
- *Strong sampling*: answers generated at random from the set of answers that are true of the correct hypothesis (linked to target hypothesis)
- *Pedagogical sampling*: answers generated that maximize the learner’s belief in the correct hypothesis (linked to target hypothesis and consider alternative hypotheses)

Critically, if the learner assumes strong or pedagogical sampling, then they can make stronger inferences that speed learning. For example, if we see someone press two buttons to activate a device, we are more likely to think that both buttons were necessary if that person knew how the machine worked and wanted to communicate to us how it worked. Otherwise, if one of the buttons would have been sufficient, then it would not make sense for them to perform the less efficient action of pressing both buttons. The effects of these sampling assumptions are fundamentally psychological. They require the learner to reason about others' goals and to reason about whether other people are thinking about their goals. See the "Answers" panel of Figure 3 for a visual illustration of this recursive psychological reasoning process within the active learning context.

Empirical support for the pedagogical sampling account comes from a range of domains/tasks, including word learning (Frank et al., 2009), pragmatic inference (Frank & Goodman, 2012), and causal reasoning (Bonawitz et al., 2011) (see the section on inferences and generalization in social learning Part I). We can also revisit these findings and connect them directly to specific components of the OED model of human inquiry. Consider Xu & Tenenbaum (2007a) finding: that learners are "sensitive" to the sampling process that generated the examples. When a knowledgeable teacher selected the examples, learners assumed the examples indicated the true word meaning. And if this were the case, then it would be surprising to see three examples drawn from the smaller subordinate category. Formally, they modeled sensitivity to sampling assumptions by changing the learner's belief in the probability of hearing a particular label (l_i) given a specific object (o_i) and word meaning (m), modifying the likelihood function in their Bayesian cognitive model: $p(x_i \mid m) \propto p(l_i \mid o_i, m)$. In this case, the likelihood function for the teacher-driven condition was designed to capture the idea that learners should prefer "smaller" or more restrictive hypotheses if they are confident that the teacher generated labels based on the actual word meaning.

This formalization provides a direct connection with the OED model of human inquiry. Specifically, when a learner simulates the possible answers, she considers how much each answer will update her beliefs. This reasoning process is modeled by computing the difference between the learner's prior and posterior uncertainty (i.e., entropy): $U(a) = ent(H) - ent(H|a)$. The learners' sampling assumptions naturally enter the information seeking calculus through the posterior entropy term $ent(H|a) = -\sum_{h \in H} P(h|a) log P(h|a)$. Intuitively, this part of the model captures the idea that not all answers are equally useful, and answers that are generated to help us learn are more informative and should lead to a stronger change in the learner's beliefs.

The approach of building more sophisticated likelihood functions has also been used to capture another aspect of evaluating the utility of an answer: that not all social partners are equally reliable sources of information. When learning from the testimony of others, there is always a possibility that the information could be inaccurate or misleading. This reasoning process might be especially important for young learners who acquire much of their information via interactions with others. A growing body of evidence suggests that even very young infants are capable of *selective* learning, rejecting answers that conflict with their knowledge (Pea, 1982) and seeking information from people who tend to provide good answers in the past (Koenig, Clement, & Harris, 2004).

For example, empirical work shows that preschoolers track and integrate a speaker's prior instances of accuracy to figure out if they are trustworthy and will use this information to guide subsequent learning from that speaker's future claims (Koenig et al., 2004). In these studies, children evaluate a speaker's current testimony after the speaker establishes a record of reliability or unreliability by labeling or mislabeling familiar objects. Across these studies, preschoolers are consistently less likely to direct questions towards and learn from a previously unreliable person. Moreover, Chow, Poulin-Dubois, & Lewis (2008) found that 14-month-olds are less likely to follow the gaze of a person who had been unreliable in the past, i.e., someone who had consistently directed gaze towards an empty location in space. Finally, children's selective learning appears sensitive to external cues, preferring to learn familiar over unfamiliar teachers (Corriveau & Harris, 2009), adults over peers (Rakoczy, Hamann, Warneken, & Tomasello, 2010), and ingroup over outgroup members (MacDonald, Schug, Chase, & Barth, 2013).

Converging evidence comes from Gweon, Pelton, Konopka, & Schulz (2014) work on children's exploration behavior after seeing pedagogical demonstrations of varying quality. In this study, children played with toys that either had a single or multiple functions (e.g., spinning globe or flashing light) until they independently discovered the correct number of functions. Then, depending on condition assignment, they saw a puppet either teach all of the functions (informative teacher) or a subset of the functions (under-informative teacher). Finally, the puppet teacher introduced a new toy to the participant and demonstrated a single novel function. Critically, when the puppet teacher had been under-informative in their previous teaching, children spent more time exploring the object functions that the teacher did not demonstrate. That is, children did not make the stronger inference that there were no other functions to discover.

The upshot of the selective learning literature is that children are not entirely credulous when they encounter information. Instead, they actively reason about features such as expertise and helpfulness avoiding people that they think are unreliable. Interestingly, the typical outcome measures in studies of selective learning are children's information-gathering decisions: whom to direct questions towards and how long to explore a novel toy. These behaviors map directly onto the decisions in the OED model of human inquiry and suggest that children consider the expected utility of others' answers when deciding to gather information.

Similar to the pedagogical reasoning effects, the selective learning phenomena have also been modeled by modifying the likelihood function in a Bayesian cognitive model. For example, Shafto et al. (2012a) proposed that selective learning in object labeling scenarios can be explained as children reasoning about both the helpfulness and knowledgeability of speakers when they produce a given label, l. Here the child's goal is to select a speaker that increases the chance that they get good information that matches the true state of the world ($label = correct$). Formally, they specify this likelihood function as:

$$P(l \mid s, k, h) = \sum_{b} P(l \mid b, h) P(b \mid k, s)$$

This function captures the idea that the probability of a label depends on the true state of the world and features of the speaker: their knowledge (k) and helpfulness (h). This is decomposed into two parts: (1) the speaker's belief (b) about the label $P(b|k, s)$, which depend on their knowledge (k) and the true label (s), and (2) the speaker's probability of producing a label that matches their belief, which depends on their helpfulness (h). Using this model, Shafto et al. (2012a) captured several qualitative findings from the selective trust literature, including children's demonstrated preference for accurate over inaccurate speakers. While the precise mathematical details of the model are less important, the key takeaway is that the same modeling approach can account for children's behavior in different domains. This parallel suggests that children reason about the utility of answers before deciding to seek information from a social partner and after having received information from another person.

One consequence of the ideas discussed in this section is that features of the individuals who are present in a social learning context can change whether information seeking occurs at all. If a child is in a setting that is unlikely to provide useful answers, then generating an information-seeking behavior, even if the action has the potential to return useful information, becomes less valuable. Thus, a challenge for the self-directed learner is to figure out whether helpful answers are likely to occur. However, this is a less-developed area of research, and more work is needed to understand whether the expected usefulness of answers within a *context* might reduce information seeking because the costs in mental energy or time are too high. The dual consideration of costs and benefits in active learning has been the focus of recent advances in machine learning (Haertel, Seppi, Ringger, & Carroll, 2008) and children's developing social cognition (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). It would be interesting to merge these cost-based approaches with ideas from social learning theory to ask how social contexts modify the costs of different behaviors.

1.5.5 Stopping rules

A stopping rule describes a threshold that causes people to cease information seeking and generate a behavior. The concept draws on ideas from probability theory that have been used to model how random variables change as a function of time. Rules can be information or time-based. For example, when researching for a paper, a student might generate the time-based stopping rule – read for two hours – or a more information-based rule – read until I understand this concept. The learner's goal is to figure out the stopping rule that balances achieving the inquiry goal with reducing unnecessary time/effort put into the task.

Studies of children's information seeking have primarily focused on measuring whether children persist if their initial request is not satisfied. For example, Frazier et al. (2009) analyzed parent-child question-answer exchanges from the CHILDES database to see if children show evidence of seeking causal explanations when they ask *how* and *why* questions. To address this hypothesis, Frazier et al. (2009) measured the probability of children re-asking the same question and the likelihood of asking a different, follow-up question after receiving either an explanatory response (e.g., CHILD: "Why you put yogurt in there?" ADULT: "Yogurt's part of the ingredients") or a non-explanatory response (e.g., CHILD: "How do you get sick?" ADULT: "I don't know."). Children were more than twice as likely to re-ask a question after getting a non-explanatory response (24%) compared to an explanatory answer (9.4%), providing evidence that they continued to collect information until their inquiry goal was satisfied.

Converging evidence comes from Deborah, Louisa Chan, & Holt (2004)'s work exploring children's intended meaning when they ask "What is it?" about objects. Children's propensity for asking follow-up questions was measured after they were given either a name or a functional explanation in response to an ambiguous request ("What is it?" Or "What's this?"). Similar to Frazier et al. (2009)'s findings, 2- to 4-year-olds asked more follow-up questions when adults provided an object label, suggesting that they intended to ask about the object's function and persisted to get this information. Children were also more likely to change the form of their ambiguous questions to more specifically target functional explanations in the object label condition. Together, these studies suggest that young children are sensitive to when they have gathered sufficient information to address their questions. In the majority of this work, the social context influenced children's stopping decisions by giving them the information they desired.

However, social contexts can shape children's stopping decisions in other ways. A recent body of research shows that adults' demonstrations of pedagogy directly influence children's decisions about whether to persist in exploration. For example, Bonawitz et al. (2011) showed that preschoolers spend less time exploring an object and are less likely to discover alternative object-functions after an adult explicitly taught a single function (Bonawitz et al., 2011). The explanation for this effect is similar to the pedagogical inference work reviewed in the "Answers" section. A pedagogical demonstration by a knowledgeable teacher provides evidence for that function and against the existence of other functions; otherwise, the teacher would have demonstrated the other features. We can reconstrue this finding as an effect of social input on children's stopping rules. When the social context indicates that there is less to learn, children adopt a lower threshold for ending their search.

It is not the case that information acquired in social contexts always reduces information search. In fact, Butler & Markman (2012) study of children's inductive inferences provides evidence that a pedagogical demonstration leads to an increase in exploration. Preschoolers saw an unfamiliar object and learned a novel causal property (that the object could magnetically pick up paper clips). In one condition, the experimenter demonstrated the property with a pedagogical framing ("Look, watch this!"). In the other condition, it was an accidental demonstration ("Oops!"). After the pedagogical demonstration, children spent more than twice as much time exploring the objects and generated three times as many attempts to make the objects pick up the paper clips.

These are seemingly contradictory effects of pedagogy on children's stopping rules. However, they both arise because socially transmitted information is more informative and warrants stronger inferences. In the causal learning case, the stronger inference is about the lack of alternative object-functions, leading to less exploration. In the inductive inference case, the stronger inference is about the generalizability of the causal property, leading to more exploration. These findings also highlight the useful distinction between social information available in the moment and social input acquired from previous interactions.

Understanding why children decide to stop gathering information is a promising area for future research. Studies could focus on children's developing capacity to reason about the cost of actions for themselves and others. This cost term is critical to deciding when it is not worth gathering additional information. In fact, recent theorizing in the field of social cognition proposes that intuitive reasoning about the costs/benefits of others actions is the core of our social cognition. Jara-Ettinger, Gweon, Schulz, & Tenenbaum (2016) describe this idea as a Naive Utility Calculus where, "... human social cognition is structured around a basic understanding of ourselves and others as intuitive utility maximizers." (p. 589).

There has also been a growing interest in developing "cost-sensitive" active learning algorithms in the field of machine learning (Haertel et al., 2008), with researchers beginning to define costs in increasingly sophisticated ways. For example, Settles, Craven, & Friedland (2008) point out that the cost of information gathering should not be measured as a reduction in the number of training trials if those training trials vary in length because specific questions are more challenging to answer. Thus, an efficient active learner should want to ask questions that maximize information gain, but they should also take into account the cost incurred by others (e.g., time or mental effort) to provide that information. This idea is fundamentally psychological in that it expands the utility computation to include some measure of how our behavior affects others' actions and/or mental states. Another open question is how children develop skill in reasoning about more complex costs (e.g., reputation management), and how children integrate their costs with the costs incurred by others.

1.6 Conclusions

In this paper, I proposed a way forward for integrating active and social learning theory. I argued that the formal framework of Optimal Experiment Design (OED) is a useful way to characterize the effects of social contexts on children's active learning. The OED framework defines information seeking as a process of expected utility maximization for gaining information. There are four model components – goals, hypotheses, questions, and answers – and other factors that are important but exist outside the model such as stopping rules. I used the OED decomposition to bring social learning phenomena into contact with active learning theory, and I suggested that the social context can influence children's active learning by:

1. communicating and/or triggering a diverse set of goals both informational and/or social
2. shaping the content of children's hypothesis spaces both in-the-moment and/or over a developmental timescale
3. serving as a model of possible and useful queries for children to imitate
4. providing useful and generalizable answers to children's queries
5. modulating when children decide to stop collecting information

The heart of this proposal is that an integrative account of active learning within social contexts represents a valuable step towards understanding cognitive development. I have tried to argue that the social and active learning accounts have much to be gained by considering the other. Active learning accounts benefit by understanding how learners function within social contexts. This social information can allow researchers to create more sophisticated utility functions that take into account the social goals that people consider when deciding what to learn. This seems especially important for characterizing children's active learning since observational studies of learning environments suggest that opportunities to learn from social interaction are ubiquitous. On the other hand, researchers interested in social learning can benefit from advances in the field of active learning by connecting their ideas to the rich traditions of machine learning, decision theory, and statistics. Moreover, children's information seeking decisions are often used as dependent variables in studies of social learning phenomena. Thus, a secondary benefit would be a deeper understanding of the factors that influence the measurement of social learning effects.

Finally, I think research on active learning could move beyond studies that lack a social context and begin to document the presence of learning goals, constrained hypothesis spaces, and the quality of answers in children's everyday experiences. This line of inquriy falls out of the integrative approach emphasized in this paper. It will require a shift from relying on highly-controlled lab experiments of children's active learning skill to leveraging large-scale, observational datasets. And while this approach adds complexity to our experiments and models, I think that the benefit will be a far greater understanding of how children's active learning operates over fundamentally social input.

Chapter 2

Real-time American Sign Language comprehension

2.1 Introduction[1]

Finding meaning in a spoken or a signed language requires learning to establish reference during real-time interaction – relying on audition to interpret spoken words, or on vision to interpret manual signs. Starting in infancy, children learning spoken language make dramatic gains in their efficiency in linking acoustic signals representing lexical forms to objects in the visual world. Studies of spoken language comprehension using the looking-while-listening (LWL) procedure have tracked developmental gains in language processing efficiency by measuring the timing and accuracy of young children's gaze shifts as they look at familiar objects and listen to simple sentences (e.g., "Where's the ball?") naming one of the objects (Fernald, Zangl, Portillo, & Marchman, 2008; Law & Edwards, 2014; Venker, Eernisse, Saffran, & Ellis Weismer, 2013). Such research finds that eye movements to named objects occur soon after the auditory information is sufficient to enable referent identification, and often prior to the offset of the spoken word (Allopenna, Magnuson, & Tanenhaus, 1998). Moreover, individual differences in the speed and accuracy of eye movements in response to familiar words predict vocabulary growth and later language and cognitive outcomes (Fernald,

[1] This chapter is published in MacDonald et al. (2018a). Real-time lexical comprehension in young children learning American Sign Language. *Developmental science*, e12672.

Perfors & Marchman, 2006; Marchman & Fernald, 2008). Together, these results suggest that gaze shifts to objects in response to spoken language reflect a rapid integration of linguistic and visual information, and that variability in the timing of these gaze shifts provides researchers a way to measure the efficiency of the underlying integration process.

Much less is known about how language influences visual attention during sign language comprehension, especially in young learners. Given the many surface-level differences between signed and spoken languages, it is not immediately clear whether the findings from spoken language will generalize to signed languages or whether they are specific to mechanisms of language comprehension in the auditory modality. In particular, studies with children learning spoken languages find that these skills undergo dramatic developmental changes over the 2nd and 3rd years of life. Moreover, there are significant relations between variation in efficiency in online language processing, as indexed by language-driven eye movements, and measures of linguistic achievement, such as vocabulary size and scores on standardized tests (Fernald et al., 2006; Marchman & Fernald, 2008). Will individual variation in language processing among children learning a signed language also be related to their age and vocabulary outcomes, as observed in children learning a spoken language?

Here we address this question by developing precise measures of speed and accuracy in real-time sign language comprehension by children learning American Sign Language (ASL). First, we estimate the extent to which adults and children tend to shift visual attention to a referent and away from the language source prior to the offset of a sign naming an object in the visual scene. Will signers wait until the end of the signed utterance, perhaps to reduce the probability of missing upcoming linguistic information? Or will signers shift gaze incrementally as the signs unfold in time, initiating saccades soon after there is enough information in the signal to identify the referent, similar to children and adults processing spoken language? Another related possibility is that signers would produce incremental gaze shifts to the named objects while still monitoring the linguistic signal in the periphery. This analysis provides an important first step towards validating the linking hypothesis that eye movements generated in our task reflect efficiency of sign recognition, rather than some other process, such as attending to the objects after the process of sign comprehension is complete. If children and adults produce rapid gaze shifts prior to target sign offset, this would provide positive evidence of incremental ASL processing.

Next, we compare the time course of ASL processing in deaf and hearing native ASL-learners to ask whether having the potential to access auditory information in their day-to-day lives would change the dynamics of eye movements during ASL processing. Do deaf and hearing native signers show parallel patterns of looking behavior driven by their similar language background experiences and the in-the-moment constraints of interpreting a sign language (i.e., fixating on a speaker as a necessary requirement for gathering information about language)? Or would the massive experience deaf children have in relying on vision to monitor both the linguistic signal and the potential referents in the visual world result in a qualitatively different pattern of performance compared to hearing ASL learning, e.g., waiting until the end of the sentence to disengage from the signer? This analysis is motivated by prior work that has used comparisons between native hearing and deaf signers to dissociate the effects of learning a visual-manual language from the effects of lacking access to auditory information (e.g., Bavelier, Dye, & Hauser, 2006).

Finally, we compare timing and accuracy of the eye movements of young ASL-learners to those of adult signers, and ask whether there are age-related increases in processing efficiency that parallel those found in spoken languages. We also examine the links between variability in children's ASL processing skills and their expressive vocabulary development. A positive association between these two aspects of language proficiency, as previously shown in children learning spoken languages, provides important evidence that skill in lexical processing efficiency is a language-general phenomenon that develops rapidly in early childhood, regardless of language modality.

2.1.1 ASL processing in adults

Research with adults shows that language processing in signed and spoken languages is similar in many ways. As in spoken language, sign recognition is thought to unfold at both the lexical and sub-lexical levels. Moreover, sign processing is influenced by both lexicality and frequency; non-signs are identified more slowly than real signs (Corina & Emmorey, 1993) and high frequency signs are recognized faster than low frequency signs (Carreiras, GutiÃĪrrez-Sigut, Baquero, & Corina, 2008). Recent work using eye-tracking methods found that adult signers produce gaze shifts to phonological competitors, showing sensitivity to sub-lexical features, and that these shifts were initiated prior to the offset of the sign, showing evidence of incremental processing (Lieberman, Borovsky, Hatrak, & Mayberry, 2015). In addition, Caselli and Cohen-Goldberg (2014) adapted a computational model, developed for spoken language (Chen & Mirman, 2012), to explain patterns of lexical access in sign languages, suggesting that the languages share a common processing architecture.

However, differences between spoken and signed languages in both sub-lexical and surface features of lexical forms could affect the time course of sign recognition (for reviews, see Carreiras, 2010 and Corina & Knapp, 2006). For example, Emmorey and Corina (1990) showed deaf adults repeated video presentations of increasingly longer segments of signs in isolation and asked them to identify the signs in an open-ended response format. In the same study, English-speaking adults heard repeated presentations of increasingly longer segments of spoken words. Accurate identification of signs required seeing a smaller proportion of the total sign length compared to words (see also Morford & Carlsen, 2011), suggesting that features of visual-manual languages, such as simultaneous presentation of phonological information, might increase speed of sign recognition. Moreover, Gutierrez and colleagues (2012) used EEG measures to provide evidence that semantic and phonological information might be more tightly linked in the sign language lexicon than in the spoken language lexicon. Thus there is evidence for both similarities and dissimilarities in the processes underlying spoken-word and manual-sign recognition. However, with a few exceptions (e.g. Lieberman et al., 2015, 2017), most of this work has relied on offline methods that do not capture lexical processing as it unfolds in time during naturalistic language comprehension. In addition, no previous studies have characterized how young ASL-learners choose to divide visual attention between a language source and the nonlinguistic visual world during real-time language comprehension.

2.1.2 Lexical development in ASL

Diary studies show that ASL acquisition follows a similar developmental trajectory to that of spoken language (Lillo-Martin, 1999; Mayberry & Squires, 2006). For example, young signers typically produce recognizable signs before the end of the first year and two-sign sentences by their 2nd birthday (Newport & Meier, 1985). And as in many spoken languages (Waxman et al., 2013), young ASL-learners tend first to learn more nouns than verbs or other predicates (Anderson & Reilly, 2002). However, because children learning ASL must rely on vision to process linguistic information and to look at named objects, it is possible that basic learning processes, such as the coordination of joint visual attention, might differ in how they support lexical development (Harris & Mohay, 1997). For example, in a study of book reading in deaf and hearing dyads, Lieberman, Hatrak, and Mayberry (2015) found that deaf children frequently shifted gaze to caregivers in order to maintain contact with the signed signal. Hearing children, in contrast, tended to look continuously at the book, rarely shifting gaze while their caregiver was speaking. This finding suggests that the modality of the linguistic signal may affect how young language learners negotiate the demands of processing a visual language while simultaneously trying to fixate on the referents of that language.

This competition for visual attention in ASL could lead to qualitatively different looking behavior during real-time ASL comprehension, making the link between eye movements and efficiency of language comprehension in ASL less transparent. On the one hand, demands of relying on vision to monitor both the linguistic signal and the named referent might cause signers to delay gaze shifts to named objects in the world until the end of the target sign, or even the entire utterance. In this case, eye movements would be less likely to reflect the rapid, incremental influence of language on visual attention that is characteristic of spoken language processing. Another possibility is that ASL-learners, like spoken language learners, will shift visual attention as soon as they have enough linguistic information to do so, producing saccades prior to the offset of the target sign. Evidence for incremental language processing would further predict that eye movements during ASL processing could index individual differences in speed of incremental comprehension, as previously shown in spoken languages.

2.1.3 Research questions

Adapting the LWL procedure for ASL enables us to address four questions. First, to what extent do children and adult signers shift their gaze away from the language source and to a named referent prior to the offset of the target sign? Second, how do deaf and hearing ASL-learners compare in the time course of real-time lexical processing? Third, how do patterns of eye movements during real-time language comprehension in ASL-learners compare to those of adult signers? Finally, are individual differences in ASL-learners' processing skill related to age and to expressive vocabulary development?

2.2 Methods

Participants were 29 native, deaf and hearing ASL-learning children (17 females, 12 males) and 16 fluent adult signers (all deaf), as shown in Table 1. Since the goal of the current study was to document developmental changes in processing efficiency in native ASL-learners, we set strict inclusion criteria. The sample consisted of both deaf children of deaf adults and hearing Children of Deaf Adults (CODAs), across a similar age range. It is important to note that all children, regardless of hearing status, were exposed to ASL from birth through extensive interaction with at least one caregiver fluent in ASL and were reported to experience at least 80% ASL in their daily lives. Twenty-five of the 29 children lived in households with two deaf caregivers, both fluent in ASL. Although the hearing children could access linguistic information in the auditory signal, we selected only ASL-dominant learners who used ASL as their primary mode of communication both within and outside the home (10 out of 13 hearing children had two deaf caregivers). Adult participants were all deaf, fluent signers who reported using ASL as their primary method of communication on a daily basis. Thirteen of the 16 adults acquired ASL from their parents and three learned ASL while at school.

Our final sample size was determined by our success over a two-year funding period in recruiting and testing children who met our strict inclusion criteria – receiving primarily ASL language input. It is important to note that native ASL-learners are a small population. The incidence of deafness at birth in the US is less than .003%, and only 10% of the 2-3 per 1000 children born with hearing loss have a deaf parent who is likely to be fluent in ASL (Mitchell & Karchmer, 2004). In addition to the 29 child participants who met our inclusion criteria and contributed adequate data, we also recruited and tested 17 more ASL-learning children who were not included in the analyses, either because it was later determined that they did not meet our stringent criterion of exposure to ASL from birth (n = 12), or because they did not complete the real-time language assessment due to inattentiveness or parental interference (n = 5).

Table 2.1: Age (in months) of hearing and deaf ASL-learning participants

Hearing status	**n**	**Mean**	**SD**	**Min**	**Max**
deaf	16	28.0	7.5	16	42
hearing	13	29.4	11.2	18	53
all children	29	28.6	9.2	16	53

2.2.1 Measures

Expressive vocabulary size: Parents completed a 90-item vocabulary checklist, adapted from Anderson and Reilly (2002), and developed specifically for this project to be appropriate for children between 1Â · and 4 years of age. Vocabulary size was computed as the number of signs reported to be produced by the child.

ASL Processing: Efficiency in online comprehension was assessed using a version of the LWL procedure adapted for ASL learners, which we call the Visual Language Processing (VLP) task. The VLP task yields two measures of language processing efficiency, reaction time (RT) and accuracy. Since this was the first study to develop measures of online ASL processing efficiency in children of this age, several important modifications to the procedure were made, as described below.

2.2.2 Procedure

The VLP task was presented on a MacBook Pro laptop connected to a 27” monitor. The child sat on the caregiver’s lap approximately 60 cm from the screen, and the child’s gaze was recorded using a digital camcorder mounted behind the monitor. To minimize visual distractions, testing occurred in a 5’ x 5’ booth with cloth sides. On each trial, pictures of two familiar objects appeared on the screen, a target object corresponding to the target noun, and a distracter object. All picture pairs were matched for visual salience based on prior studies with spoken language (Fernald et al., 2008). Between the two pictures was a central video of an adult female signing the name of one of the pictures. Participants saw 32 test trials with five filler trials (e.g. “YOU LIKE PICTURES? MORE WANT?”) interspersed to maintain children’s interest.

Coding and Reliability. Participants’ gaze patterns were video recorded and later coded frame-by-frame at 33-ms resolution by highly-trained coders blind to target side. On each trial, coders indicated whether the eyes were fixated on the central signer, one of the images, shifting between pictures, or away (off), yielding a high-resolution record of eye movements aligned with target noun onset. Prior to coding, all trials were pre-screened to exclude those few trials on which the participant was inattentive or there was external interference. To assess inter-coder reliability, 25% of the videos were re-coded. Agreement was scored at the level of individual frames of video and averaged 98% on these reliability assessments.

Table 2.2: Iconicity scores (1 = not iconic at all; 7 = very iconic) and degree of phonological overlap (out of 5 features) for each sign item-pair. Values were taken from ASL-LEX, a database of lexical and phonological properties of signs in ASL.

Item Pair (iconicity score 1-7)	**Number of matched features**	**Matched features**
bear (3.0) – doll (1.2)	1	Movement
cat (4.6) – bird (4.5)	3	Selected Fingers, Major Location, Sign Type
car (6.2) – book (6.7)	4	Selected Fingers, Major Location, Movement, Sign Type
ball (5.7) – shoe (1.5)	4	Selected Fingers, Major Location, Movement, Sign Type

2.2.3 Stimuli

Linguistic stimuli. To allow for generalization beyond characteristics of a specific signer and sentence structure, we recorded two separate sets of ASL stimuli. These were recorded with two native ASL signers, using a different alternative grammatical ASL sentence structures for asking questions (see Petronio and Lillo-Martin, 1997):

- Sentence-initial wh-phrase: “HEY! WHERE [target noun]?”
- Sentence-final wh-phrase: “HEY! [target noun] WHERE?”

Each participant saw one stimulus set which consisted of one ASL question structure, with roughly an even distribution of children across the two stimulus sets (16 saw sentence-initial wh-phrase structure; 13 saw the sentence-final wh-phrase structure). To prepare the stimuli, two female native ASL users recorded several tokens of each sentence in a child-directed register. Before each sentence, the signer made a hand-wave gesture commonly used in ASL to gain an interlocutor’s attention before initiating an utterance. These candidate stimuli were digitized, analyzed, and edited using Final Cut Pro software, and two native signers selected the final tokens. The target nouns consisted of eight object names familiar to most children learning ASL at this age.

Visual stimuli. The visual stimuli consisted of colorful digitized pictures of objects corresponding to the target nouns presented in four fixed pairs (cat—bird, car—book, bear—doll, ball—shoe). See Table 2 for information about the degree of phonological overlap in each item-pair and the degree of iconicity for each sign (values were taken from ASL-LEX [Caselli et al., 2017]).[2] Images were digitized pictures presented in fixed pairs, matched for visual salience with 3–4 tokens of each object

[2]We did not find evidence that these features were related to the speed or accuracy of participants’ eye movements in our task. However, this study was not designed to vary these features systematically. See Appendix XX for the analysis.

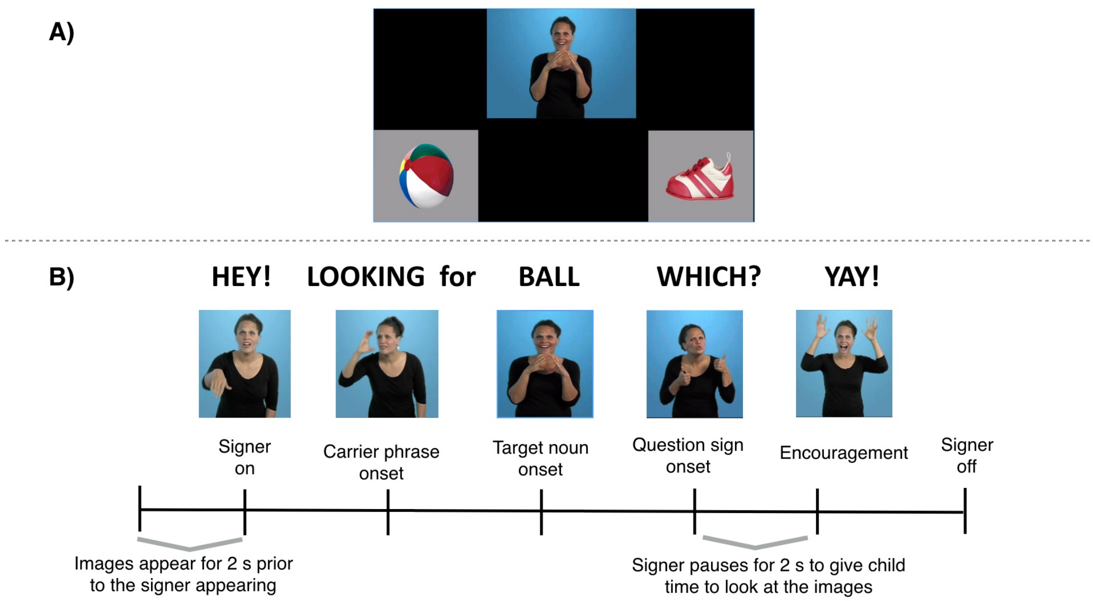


Figure 2.1: Configuration of visual stimuli (1A) and trial structure (1B) for one question type (sentence final wh-phrase) shown in the central video on the VLP task.

type. Each object served as target four times and as distracter four times for a total of 32 trials. Side of target picture was counterbalanced across trials.

2.2.4 Trial Structure

Figure 1 shows the structure of a trial with a sentence-final wh-phrase, one of the two question types in the VLP task. On each trial, children saw two images of familiar objects on the screen for 2 s before the signer appeared, allowing time for children to inspect both images. Next, children saw a still frame of the signer for one second, so they could orient to the signer prior to sentence onset. The target sentence was then presented, followed by a question and 2-s hold, followed by an exclamation to encourage attention to the task. This structure is nearly identical to the auditory LWL task, differing only in the addition of the 2-s hold. The hold was included to give participants additional time to shift gaze from the signer to the objects.

2.2.5 Calculating measures of language processing efficiency

Computing target sign onset and offset. In studies of spoken language processing, target word onset is typically identified as the first moment in the auditory signal when there is acoustic evidence of the target word. However, in signed languages like ASL, phonological information is present in several components of the visual signal simultaneously – for example, in one or both hands as well as in the face of the signer - making it difficult to determine precisely the beginning of the target sign. Because sign onset is critical to operationalizing speed of ASL comprehension in this task, we applied an empirical approach to defining target-sign onset. We used a gating task in which adult signers viewed short videos of randomly presented tokens that varied in length. Two native signers first selected a sequence of six candidate frames for each token, and then 10 fluent adult signers unfamiliar with the stimuli watched videos of the target signs in real-time while viewing the same picture pairs as in the VLP task. Participants indicated their response with a button press. For each sign token, the onset of the target noun was operationalized as the earliest video frame? at which adults selected the correct picture with 100% agreement. To determine sign offset, two native signers independently marked the final frame at which the handshape of each target sign was no longer identifiable. Agreements were resolved by discussion. Sign length was defined as sign offset minus sign onset (Median sign length was 1204 ms, ranging from 693-1980 ms).

Reaction Time. Reaction time (RT) corresponds to the latency to shift from the central signer to the target picture on all signer-to-target shifts, measured from target-noun onset. We chose cutoffs for the window of relevant responses based on the distribution of children's RTs in the VLP task, including the middle 90% (600-2500 ms) (see Ratcliff, 1993). Incorrect shifts (signer-to-distracter [19%], signer-to-away [14%], no shift [8%]) were not included in the computation of median RT. The RT measure was reliable within participants (Cronbach's $\alpha = 0.8$).

Target Accuracy. Accuracy was the mean proportion of time spent looking at the target picture out of the total time looking at either target or distracter picture over the 600 to 2500 ms window from target noun onset. We chose this window to be consistent with the choice of the RT analysis window. This measure of accuracy reflects the tendency both to shift quickly from the signer to the target picture in response to the target sign and to maintain fixation on the target picture. Mean proportion looking to target was calculated for each participant for all trials on which the participant was fixating on the center image at target-sign onset. To make accuracy proportion scores more suitable for modeling on a linear scale, all analyses were based on scores that were scaled in log space using a logistic transformation. The Accuracy measure was reliable within participants (Cronbach's $\alpha = 0.92$)

Proportion Sign Length Processed Prior to Shifting. As a measure of incremental processing, we used the mean proportion of the target sign that children and adults saw before generating an initial eye movement away from the central signer. Because target signs differed in length across trials, we divided each RT value by the length of the corresponding target sign. Previous research on spoken language suggests that at least 200 ms is required to program an eye-movement (Salverda, Kleinschmidt, & Tanenhaus, 2014), so we subtracted 200 ms from each RT to account for eye movements that were initiated during the end of the target sign (proportion target sign = $\frac{(RT - 200\text{ms})}{\text{Sign Length}}$). Mean proportion of sign processed was computed for each token of each target sign and then averaged over all target signs within participants, reflecting the amount of information signers processed before generating an eye movement, on average. A score of ≥ 1.0 indicates that a signer tended to initiate eye movements to the target pictures after sign offset. An average < 1.0 indicates eye-movements were planned during the target sign, reflecting the degree to which signers showed evidence of incremental language processing.

2.2.6 Analysis Plan

We used Bayesian methods to estimate the associations between hearing status, age, vocabulary, and RT and accuracy in the VLP task. Bayesian methods are desirable for two reasons: First, Bayesian methods allowed us to quantify support in favor of a null hypothesis of interest – in this case, the absence of a difference in real-time processing skills between age-matched deaf and hearing ASL learners. Second, since native ASL learners are rare, we wanted to use a statistical approach that allowed us to incorporate relevant prior knowledge to constrain our estimates of the strength of association between RT/accuracy on the VLP task and age/vocabulary.

Concretely, we used prior work on the development of real-time processing efficiency in children learning spoken language (Fernald et al., 2008) to consider only plausible linear associations between age/vocabulary and RT/accuracy, thus making our alternative hypotheses more precise. In studies with adults, the common use of eye movements as a processing measure is based on the assumption that the timing of the first shift reflects the speed of their word recognition (Tanenhaus, Magnuson, Dahan, & Chambers, 2000).[3] However, studies with children have shown that early shifts are more likely to be random than later shifts (Fernald et al., 2008), suggesting that some children's shifting behavior may be unrelated to real-time ASL comprehension. We use a mixture-model to quantify the probability that each child participant's response is unrelated to their real-time sign recognition (i.e., that the participant is responding randomly, or is "guessing"), creating an analysis model where participants who were more likely to be guessers have less influence on the estimated relations between RT and age/vocabulary. Note that we use this approach only in the analysis of RT, since "guessing behavior" is integral to our measure of children's mean accuracy in the VLP task, but not to our measure of mean RT. The Supplemental Material available online provides more details about the analysis model, as well two additional sensitivity analyses, which provide evidence that our results are robust to different specifications of prior distributions and to different analysis windows. We also provide a parallel set of analyses using a non-Bayesian approach, which resulted in comparable findings.

To provide evidence of developmental change, we report the strength of evidence for a linear model with an intercept and slope, compared to an intercept-only model in the form of a Bayes Factor (BF) computed via the Savage-Dickey method (Wagenmakers et al., 2010). To estimate the uncertainty around our estimates of the linear associations, we report the 95% Highest Density Interval (HDI) of the posterior distribution of the intercept and slope. The HDI provides a range of plausible values and gives information about the uncertainty of our point estimate of the linear association. Models with categorical predictors were implemented in STAN (Stan Development Team, 2016), and models with continuous predictors were implemented in JAGS (Plummer, 2003). Finally, we chose the linear model because it a simple model of developmental change with only two parameters to estimate, and the outcome measures – mean RT and Accuracy for each participant – were normally distributed. All of the linear regressions include only children's data and take the form: *processing measure* *age* and *processing measure* *vocabulary*.

[3]The assumption that first shifts reflects speed of incremental word recognition depends on the visual display containing candidate objects with minimal initial phonological overlap. If there are phonological competitors present (e.g., candy vs. candle), then participants' early shifting behavior could reflect consideration of alternative lexical hypotheses for the incoming linguistic information.

2.3 Results

The results are presented in five sections addressing the following central questions in this research. First, where do ASL users look while processing sign language in real-time? Here we provide an overview of the time course of looking behavior in our task for both adults and children. Second, would young ASL-learners and adult signers show evidence of rapid gaze shifts that reflect lexical processing, despite the apparent competition for visual attention between the language source and the nonlinguistic visual world? In this section, we estimate the degree to which children and adults tended to initiate eye-movements prior to target sign offset, providing evidence that these gaze shifts occur prior to sign offset and index speed of incremental ASL comprehension. Third, do deaf and hearing native signers show a similar time course of eye movements, despite having differential access to auditory information in their daily lives? Or would deaf children's daily experience relying on vision to monitor both the linguistic signal and the potential referents in the visual world result in a qualitatively different pattern of performance, e.g., their waiting longer to disengage from the signer to seek the named object? Fourth, do young ASL-learners show age-related increases in processing efficiency that parallel those found in spoken languages? Here we compare ASL-learners' processing skills to those of adult signers and exploring relations to age among the children. Finally, is individual variation in children's ASL processing efficiency related to the size of their productive ASL vocabularies?

2.3.1 Overview of looking behavior during real-time ASL comprehension

The first question of interest was where do ASL users look while processing sign language in real-time? Figure 2 presents an overview of adults (2A) and children's (2B) looking behavior in the VLP task. This plot shows changes in the mean proportion of trials on which participants fixated the signer, the target image, or the distracter image at every 33-ms interval of the stimulus sentence. At target-sign onset, all participants were looking at the signer on all trials. As the target sign unfolded, the mean proportion looking to the signer decreased rapidly as participants shifted their gaze to the target or the distracter image. Proportion looking to the target increased sooner and reached a higher asymptote, compared to proportion looking to the distracter, for both adults and children. After looking to the target image, participants tended to shift their gaze rapidly back to the signer, shown by the increase in proportion looking to the signer around 2000 ms after target-noun onset.

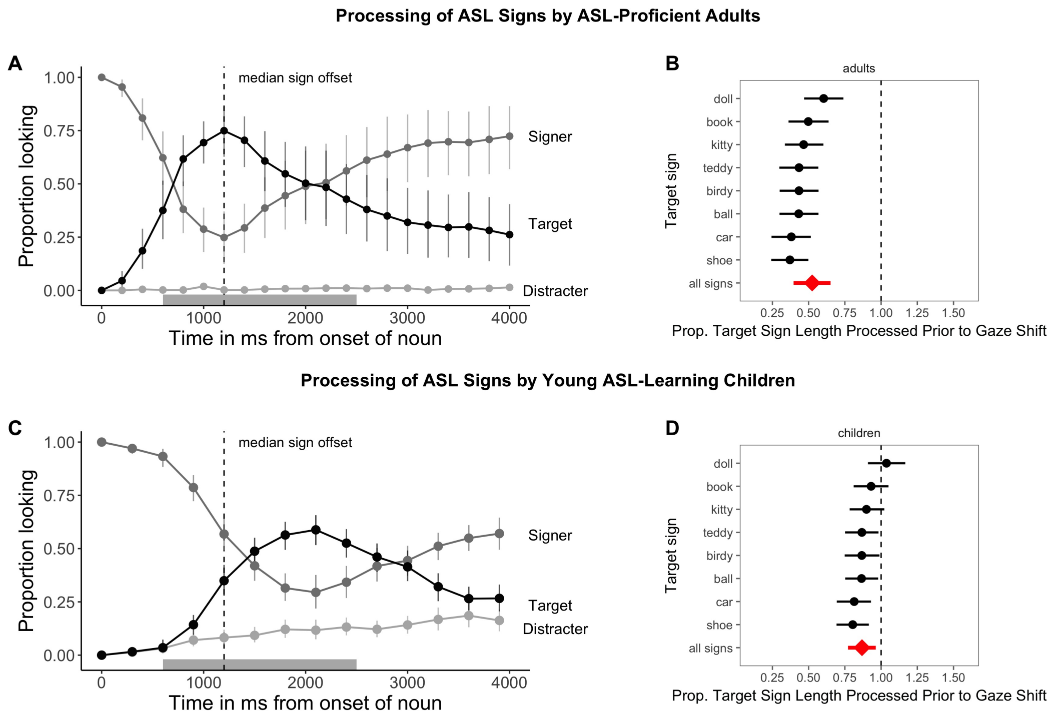


Figure 2.2: The time course of looking behavior for ASL-proficient adults (2A) and young ASL-learners (2C). The curves show mean proportion looking to the signer (dark grey), the target image (black), and the distracter image (light grey). The grey shaded region marks the analysis window (600-2500ms); error bars represent +/- 95% CI computed by non-parametric bootstrap. The mean proportion of each target sign length (see the Methods section for details on how sign length was defined) processed prior to shifting visual attention away from the language source to a named object for adults (2B) and children (2D). The diamond indicates the mean estimate for all signs. The dashed vertical line corresponds to a median proportion of 1.0. Error bars represent 95% Highest Density Intervals.

Adults tended to shift to the target picture sooner in the sentence than did children, and well before the average offset of the target sign. Moreover, adults rarely looked to the distractor image at any point in the trial. This systematic pattern of behavior – participants reliably shifting attention from the signer to the named object and back to the signer – provides qualitative evidence that the VLP task is able to capture interpretable eye movement behavior during ASL comprehension.

2.3.2 Evidence that eye movements during ASL processing index incremental sign comprehension

One of the behavioral signatures of proficient spoken language processing is the rapid influence of language on visual attention, with eye movements occurring soon after listeners have enough information to identify the named object. Our second question of interest was whether young ASL-learners and adult signers would also show evidence of rapid gaze shifts in response to signed language, despite the apparent competition for visual attention between the language source and the nonlinguistic visual world. Or would signers delay their shifts until the very end of the target sign, or even until the end of the utterance, perhaps because they did not want to miss subsequent linguistic information?

To answer these questions, we conducted an exploratory analysis, computing the proportion of each target sign that participants processed before generating an eye movement to the named object. Figure 2 shows this measure for each target sign for both adults (2B) and children (2D). Adults shifted prior to the offset of the target sign for all items and processed on average 51% of the target sign before generating a response (M = 0.51, 95% HDI [0.35, 0.66]). Children processed 88% of the target sign on average, requiring more information before shifting their gaze compared to adults. Children reliably initiated saccades prior to the offset of the target sign overall (M = 0.88, 95% HDI [0.79, 0.98]) and for five out of the eight signed stimuli.

These results suggest that young signers as well as adults process signs incrementally as they unfold in time (for converging evidence see Lieberman et al., 2015, 2017). It is important to point out that we would not interpret signers waiting until the end of the sign or the end of the sentence as evidence against an incremental processing account since there could be other explanations for that pattern of results such as social norms of looking at a person until they finish speaking. However, this result provides positive evidence that eye movements in the VLP task provide an index of speed of incremental ASL comprehension, allowing us to perform the subsequent analyses that estimate (a) group differences in looking behavior and (b) links between individual variation in speed and accuracy of eye movements during ASL processing and variation in productive vocabulary.

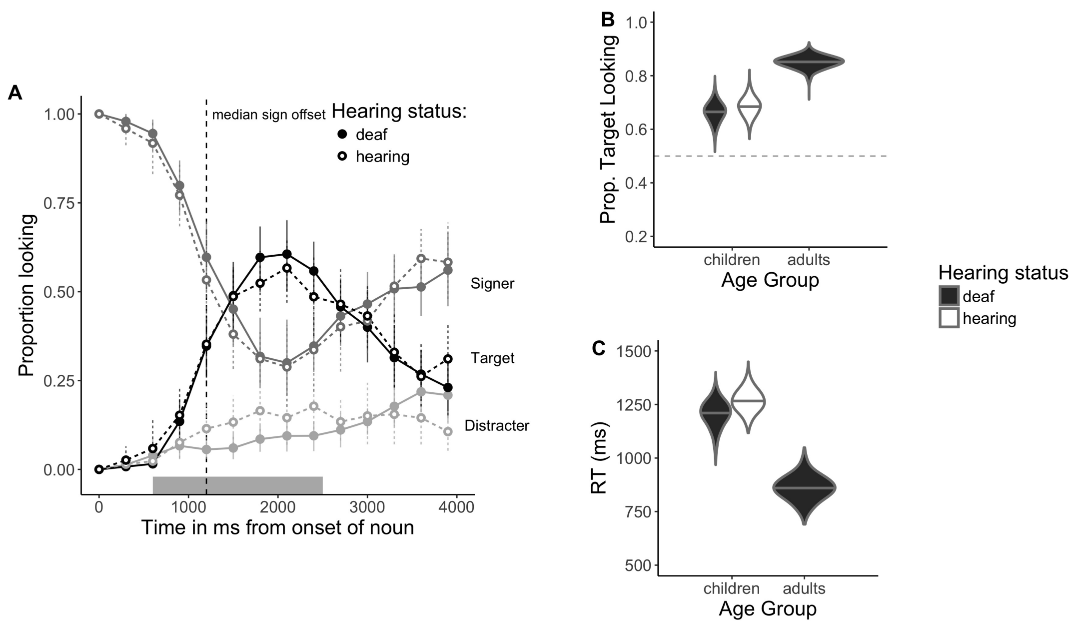


Figure 2.3: The time course of looking behavior for young deaf and hearing ASL-learners (3A). Filled circles represent deaf signers, while open circles represent hearing signers; All other plotting conventions are the same as in Figure 2. Panels B and C show full posterior distributions over model estimates for mean Accuracy (3B) and Reaction Time (3C) for children and adults. Fill (white/black) represents children's hearing status. (Note that there were no hearing adult signers in our sample).

2.3.3 Real-time ASL comprehension in deaf and hearing children and deaf adults

The third question of interest was whether deaf and hearing native signers show a similar time course of lexical processing, driven by their similar language experiences and the in-the-moment constraints of interpreting a sign language in real time? Or would deaf children's daily experience relying on vision to monitor both the linguistic signal and the potential referents in the visual world result in a qualitatively different pattern of performance, e.g., their waiting longer to disengage from the signer to seek the named object?

Figure 3A presents the overview of looking behavior for deaf and hearing children. At target-sign onset, all children were looking at the signer on all trials. Overall, deaf and hearing children showed a remarkably similar time course of looking behavior: shifting away from the signer, increasing looks to the target, and shifting back to the signer at similar time points as the sign unfolded. To quantify any differences, we compared the posterior distributions for mean accuracy (Figure 3B) and mean RT (Figure 3C) across the deaf and hearing groups. We did not find evidence for a difference in mean accuracy (M_(hearing) ãĂŰ=0.68,ãĂŰ MãĂŮ*(deaf=) 0.65; ÎšãĂŮ*diff= 0.03, 95% HDI [-0.07, 0.13]) or RT ãĂŰ(MãĂŮ*(hearing) ãĂŰ=1265.62 ms,ãĂŰ MãĂŮ*(deaf=) 1185.05 ms; ÎšãĂŮ_diff= 78.32 ms, 95% HDI [-86.01 ms, 247.04 ms]), with the 95% HDI including zero for both models. These parallel results provide evidence that same-aged hearing and deaf native ASL-learners showed qualitatively similar looking behavior during real-time sentence processing, suggesting that decisions about where to allocate visual attention are not modulated by differential access to auditory information, but rather are shaped by learning ASL as a first language (see Bavelier et al., 2006 for a review of the differential effects of deafness compared to learning a visual language on perception and higher-order cognitive skills). Moreover, these results provide additional justification (over and above children's highly similar language background experience) for analyzing all the native ASL-learning children together, regardless of hearing status, in the subsequent analyses.

Returning to the overview of looking behavior shown in Figure 2, we see that adults tended to shift to the target picture sooner in the sentence than did children, and well before the average offset of the target sign. Moreover, adults rarely looked to the distractor image at any point in the trial. To quantify these age-related differences we computed the full posterior distribution for children and adults' mean Accuracy (Figure 3B) and RT (Figure 3C). Overall, adults were more accurate (M_adults= 0.85, M_(children)= 0.68, Îš_diff = 0.17, 95% HDI for the difference in means [0.11, 0.24]) and faster to shift to the target image compared to children (M_adults= 861.98 ms, M_(children)= 1229.95 ms; Îš_diff = -367.76 ms, 95% HDI for the difference in means [-503.42 ms, -223.85 ms]). This age-related difference parallels findings in spoken language (Fernald et. al., 2006) and shows that young ASL learners are still making progress towards adult-levels of ASL processing efficiency. Next, we compared real-time processing efficiency in ASL-learners and adult signers. Returning to the overview of looking behavior shown in Figure 2, we see that adults tended to shift to the target picture sooner in the sentence than did children, and well before the average offset of the target sign. Moreover, adults rarely looked to the distractor image at any point in the trial. To quantify these differences we computed the full posterior distribution for children and adults' mean

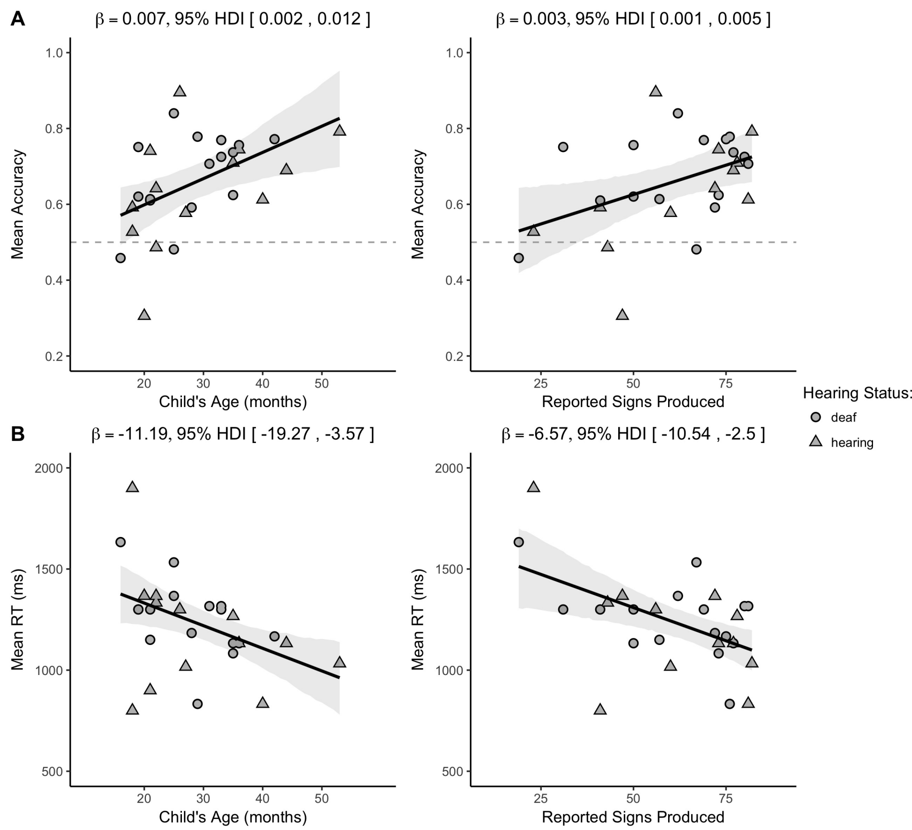


Figure 2.4: Scatterplots of relations between children's age and vocabulary and measures of their mean accuracy (4A) and mean RT (4B) in the VLP procedure. Shape represents children's hearing status. The solid black line is the maximum a posteriori model estimate for the mean accuracy at each age point. The shaded gray regions represent the 95% Highest Density Interval (range of plausible values) around the regression line.

Accuracy (Figure 3B) and RT (Figure 3C). Overall, adults were more accurate (M_adults= 0.85, M_(children)= 0.68, Îš_diff = 0.17, 95% HDI for the difference in means [0.11, 0.24]) and faster to shift to the target image compared to children (M_adults= 861.98 ms, M_(children)= 1229.95 ms; Îš_diff = -367.76 ms, 95% HDI for the difference in means [-503.42 ms, -223.85 ms]). This age-related difference parallels findings in spoken language (Fernald et. al., 2006) and shows that young ASL learners are still making progress towards adult-levels of ASL processing efficiency.

2.3.4 Links between children's age and efficiency in incremental sign comprehension

The fourth question of interest was whether young ASL-learners show age-related increases in processing efficiency that parallel those found in spoken languages. To answer this question, we estimated relations between young ASL learners' age-related increases in the speed and accuracy with which they interpreted familiar signs (see Table 3 for point and interval estimates). Mean accuracy was positively associated with age (Figure 4A), indicating that older ASL learners were more accurate than younger children in fixating the target picture. The Bayes Factor (BF) indicated that a model including a linear association was 12.8 times more likely than an intercept-only model, providing strong evidence for developmental change. The Îš estimate indicates that, for each month of age, children increased their accuracy score by 0.007, i.e., an increase of ˜1% point, meaning that over the course of one year the model estimates a ˜12% point gain in accuracy when establishing reference in the VLP task. Mean RTs were negatively associated with age (Figure 4A), indicating that older children shifted to the target picture more quickly than did younger children. The BF was ˜14, providing strong evidence for a linear association. The model estimates a ˜11 ms gain in RT for each month, leading to a ˜132 ms gain in speed of incremental ASL comprehension over one year of development.

Together, the accuracy and RT analyses showed that young ASL learners reliably looked away from the central signer to shift to the named target image in the VLP task. Importantly, children varied in their response times and accuracy, and this variation was meaningfully linked to age. Thus, like children learning spoken language, ASL learners improve their real-time language processing skills over the second and third years of life as they make progress towards adult levels of language fluency.

Table 2.3: Summary of the four linear models using children's age and vocabulary size to predict accuracy (proportion looking to target) and reaction time (latency to first shift in ms). BF is the Bayes Factor comparing the evidence in favor of linear model to an intercept-only (null) model; Mean Beta is the mean of the posterior distribution for the slope parameter for each model (i.e., the linear association); and the Highest Density Interval (HDI) shows the interval containing 95% of the plausible slope values given the model and the data.

Model specification	**Bayes Factor**	**Mean Beta**	**95% HDI**

Accuracy ~ Age	12.8	0.007	[0.002, 0.012]
Accuracy ~ Vocab	6.8	0.003	[0.001, 0.005]
RT ~ Age	14.4	-11.2 ms	[-19.3 ms, -3.6 ms]
RT ~ Vocab	18.7	-6.6 ms	[-10.5 ms, -2.5 ms]

2.3.5 Links between children's incremental sign comprehension and productive vocabulary

The final question of interest was whether individual differences in processing skills were related to the size of children's ASL vocabularies. As shown in Figure 4B, children with higher accuracy scores also had larger productive vocabularies (BF = 6.8), with the model estimating a 0.003 increase for each additional sign known. Moreover, children who were faster to recognize ASL signs were those with larger sign vocabularies (BF = 18.7), with each additional sign resulting in a ~7 ms decrease in estimated RT. Taken together, older children and children with larger expressive vocabularies were more accurate and efficient in identifying the referents of familiar signs. It is important to point out that the independent effect of vocabulary size on ASL processing could not be assessed here given the correlation between age and vocabulary (r = 0.76) in our sample of children ages one to four years. However, these findings parallel results in the substantial body of previous research with monolingual children learning spoken languages, such as English (Fernald et al., 2006) and Spanish (Hurtado, Marchman, & Fernald, 2007).

2.4 Discussion

Efficiency in establishing reference in real-time lexical processing is a fundamental component of language learning. Here, we developed the first measures of young ASL learners' real-time language comprehension skills. There are five main findings from this research.

First, both adults and children showed a similar qualitative pattern of looking behavior as signs unfolded in time. They began by looking at the signer to gather information about the signed sentence, before shifting gaze to the named object, followed by a return in looking to the signer. All signers allocated very few fixations to the distractor image at any point during the signed sentence.

Second, children and adults tended to shift their gaze away from the signer and to the named referent prior to sign offset, providing evidence of incremental ASL processing. This rapid influence of language on visual attention in ASL is perhaps even more striking since premature gaze shifts could result in a degraded the linguistic signal processed in the periphery or in missing subsequent linguistic information altogether. Furthermore, evidence of incremental gaze shifts suggests that eye movements during ASL processing index efficiency of lexical comprehension, as previously shown in spoken languages, which is important for future work on the psycholinguistics of early sign language acquisition.

Third, deaf and hearing native signers, despite having differential access to auditory information, showed remarkably similar looking behavior during real-time ASL comprehension. Even though the deaf and hearing children had differential access to auditory information in their daily lives, this experience did not change their overall looking behavior or the timing of their gaze shifts during ASL comprehension. Instead, both groups showed parallel sensitivity to the in-the-moment constraints of processing ASL in real time. That is, both deaf and hearing children allocated similar amounts of visual attention to the signer, presumably because this was the only fixation point in the visual scene that also provided information with respect to their goal of language comprehension. This is in stark contrast to what hearing children could potentially do in a similar grounded language comprehension task where a speaker was a potential visual target. In that case, the hearing listener could choose to look at the speaker or to look elsewhere, without losing access to the incoming language via the auditory channel. Thus, they can look while they listen.

Fourth, like children learning spoken language, young ASL-learners were less efficient than adults in their real-time language processing, but they showed significant improvement with age over the first four years. Moreover, although all target signs were familiar to children, older children identified the named referents more quickly and accurately than younger children. This result suggests that the real-time comprehension skills of children who are learning ASL in native contexts follow a similar developmental path to that of spoken language learners, as has been shown in previous work on ASL production (Lillo-Martin, 1999; Mayberry & Squires, 2006). By developing precise measures of real-time ASL comprehension, we were able to study children's language skills earlier in development as compared to other methods.

Fifth, we found a link between ASL processing skills and children's productive vocabularies. ASL-learning children who knew more signs were also faster and more accurate to identify the correct referent than those who were lexically less advanced. These results are consistent with studies of English- and Spanish-learning children, which find strong relations between efficiency in online language comprehension and measures of linguistic achievement (Fernald et al., 2006; Marchman & Fernald, 2008).

2.4.1 Limitations and open questions

This study has several limitations. First, while the sample size is larger than in most previous studies of ASL development, it is still relatively small compared to many studies of spoken language acquisition - an unsurprising limitation, given that native ASL-learners are a rare population. Thus more data are needed to characterize more precisely the developmental trajectories of sign language processing skills. Second, testing children within a narrower age range might have revealed independent effects of vocabulary size on ASL processing, which could not be assessed here given the correlation between age and vocabulary size in our broad sample of children from one to four years. To facilitate replication and extension of our results, we have made all of our stimuli, data, and analysis code publicly available (`https://github.com/kemacdonald/SOL`).

Third, we did not collect measures of age-related gains in children's general cognitive abilities. Thus, it is possible that our estimates of age-related changes in lexical processing are influenced by children's developing efficiency in other aspects of cognition, e.g., increased control of visual attention. Work on the development of visual attention from adolescence to early adulthood shows that different components of visual attention (the ability to distribute attention across the visual field, attentional recovery from distraction, and multiple object processing) develop at different rates (Dye and Bavelier, 2009). Moreover, work by Elsabbagh et. al., (2013) shows that infants become more efficient in their ability to disengage from a central stimulus to attend to a stimulus in the periphery between the ages 7 months and 14 months. However, there is a large body of work showing that features of language use and structure (e.g., the frequency of a word, a word's neighborhood density, and the amount of language input a child experiences) affect the speed and accuracy of eye movements in the Looking-While-Listening style tasks (see Tanenhaus et al., 2000 for a review). Thus, while it possible that age-related improvements in general cognitive abilities are a factor in our results, we think that the strength of the prior evidence suggests that more efficient gaze shifts in the VLP task are indexing improvements in the efficiency of incremental ASL comprehension.

A fourth limitation is that characteristics of our task make it difficult to directly compare our findings with previous work on ASL processing by adults. For example, in contrast to prior gating studies (e.g., Emmorey & Corina, 1990; Morford & Carlsen, 2011), our stimuli consisted of full sentences in a child-directed register, not isolated signs, and we used a temporal response measure rather than an open-ended untimed response. However, it is interesting to note that the mean reaction time of the adults in our task (M = 862 ms) is strikingly close to the average performance of native adult signers in Lieberman et al.'s (2015) "unrelated" condition (M = 844 ms). In addition, we did not select stimuli that parametrically varied features of signs that may influence speed of incremental ASL comprehension, including iconicity and degree of phonological overlap. However, we were able to use a recently created database of lexical and phonological properties of 1000 signs (Caselli et. al., 2017) to explore this possibility. We did not see evidence that iconicity or degree of phonological overlap influenced speed or accuracy of eye movements in children or adults in our sample of eight target signs (see Figures S4 and S5 in the online supplement).

We also cannot yet make strong claims about processing in signed vs. spoken languages in absolute terms because the VLP task included the signer as a central fixation, resulting in different task demands compared to the two-alternative procedure used to study children's spoken language processing (e.g., Fernald et al. 1998). However, a direct comparison of the timecourse of eye movements during signed and spoken language processing is a focus of our ongoing work (MacDonald et al., 2017). Nevertheless, the current results reveal parallels with previous findings showing incremental processing during real-time spoken language comprehension (see Tanenhaus et al., 2000) and sign language comprehension in adults (Lieberman et al., 2015). Moreover, we established links between early processing efficiency and measures of vocabulary in young ASL-learners, suggesting that parallel mechanisms drive language development, regardless of the language modality.

Finally, our sample is not representative of most children learning ASL in the United States. Since most deaf children are born to hearing parents unfamiliar with ASL, many are exposed quite inconsistently to sign language, if at all. We took care to include only children exposed to ASL from birth. The development of real-time ASL processing may look different in children who have inconsistent or late exposure to ASL (Mayberry, 2007). An important step is to explore how variation in ASL processing is influenced by early experience with signed languages. Since children's efficiency in interpreting spoken language is linked to the quantity and quality of the speech that they hear

(Hurtado, Marchman, & Fernald, 2008; Weisleder & Fernald, 2013), we would expect similar relations between language input and outcomes in ASL-learners. We hope that the VLP task will provide a useful method to track precisely the developmental trajectories of a variety of ASL-learners.

2.5 Conclusion

This study provides evidence that both child and adult signers rapidly shift visual attention as signs unfold in time and prior to sign offset during real-time sign comprehension. In addition, individual variation in speed of lexical processing in child signers is meaningfully linked to age and vocabulary. These results contribute to a growing literature that highlights parallels between signed and spoken language development when children are exposed to native sign input, suggesting that it is the quality of children's input and not features of modality (auditory vs. visual) that facilitate language development. Moreover, similar results for deaf and hearing ASL-learners suggest that both groups, despite large differences in their access to auditory information in their daily lives, allocated attention in similar ways while processing sign language from moment to moment. Finally, these findings indicate that eye movements during ASL comprehension are linked to efficiency of incremental sign recognition, suggesting that increased efficiency in real-time language processing is a language-general phenomenon that develops rapidly in early childhood, regardless of language modality.

Chapter 3

Information seeking eye movements during language comprehension

3.1 Introduction[1]

[1]Parts of this chapter were published as MacDonald et al. (2017a) An information-seeking account of eye movements during spoken and signed language comprehension and MacDonald et al. (2018b) Adults and preschoolers seek visual information to support language comprehension in noisy environments. Proceedings of the 39th and 40th Annual Meetings of the Cognitive Science Society.

Chapter 4

Social information and word learning

4.1 Introduction[1]

Learning the meaning of a new word should be hard. Consider that even concrete nouns are often used in complex contexts with multiple possible referents, which in turn have many conceptually natural properties that a speaker could talk about. This ambiguity creates the potential for an (in principle) unlimited amount of referential uncertainty in the learning task.[2] Remarkably, word learning proceeds despite this uncertainty, with estimates of adult vocabularies ranging between 50,000 to 100,000 distinct words (Bloom, 2002). How do learners infer and retain such a large variety of word meanings from data with this kind of ambiguity?

Statistical learning theories offer a solution to this problem by aggregating cross-situational statistics across labeling events to identify underlying word meanings (Siskind, 1996; C. Yu & Smith, 2007). Recent experimental work has shown that both adults and young infants can use word-object co-occurrence statistics to learn words from individually ambiguous naming events (Smith & Yu,

[1]This chapter is published in MacDonald et al. (2017b). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology* , 94, 67-84.

[2]This problem is a simplified version of Quine's *indeterminacy of reference* (Quine, 1960): That there are many possible meanings for a word ("Gavigai") that include the referent ("Rabbit") in their extension, e.g., "white," "rabbit," "dinner." Quine's broader philosophical point was that different meanings ("rabbit" and "undetached rabbit parts") could actually be extensionally identical and thus impossible to tease apart.

2008; Vouloumanos, 2008). For example, Smith and Yu (2008) taught 12-month-olds three novel words simply by repeating consistent novel word-object pairings across 10 ambiguous exposure trials. Moreover, computational models suggest that cross-situational learning can scale up to learn adult-sized lexicons, even under conditions of considerable referential uncertainty (Smith, Smith, & Blythe, 2011).

Although all cross-situational learning models agree that the input is the co-occurrence between words and objects and the output is stable word-object mappings, they disagree about how closely learners approximate the input distribution (for review, see Smith, Suanda, & Yu 2014). One approach has been to model learning as a process of updating connection strengths between multiple word-object links (McMurray, Horst, & Samuelson, 2012), while other approaches have argued that learners store only a single word-object hypothesis (Trueswell et al., 2013). In recent experimental and modeling work Yurovsky and Frank (2015) suggest an integrative explanation: learners allocate a fixed amount of attention to a single hypothesis and distribute the rest evenly among the remaining alternatives. As the set of alternatives grows, the amount of attention allocated to each object approaches zero.

In addition to the debate about representation, researchers have disagreed about how to characterize the ambiguity of the input to cross-situational learning mechanisms. One way to quantify the uncertainty in a naming event is to show adults video clips of caregiver-child interactions and measure their accuracy at guessing the meaning of an intended referent (Human Simulation Paradigm: HSP [Gillette, Gleitman, Gleitman, and Lederer, 1999]). Using the HSP, Medina, Snedeker, Trueswell, and Gleitman (2011) found that approximately 90% of learning episodes were ambiguous ($< 33\%$ accuracy) and only 7% were relatively unambiguous ($> 50\%$ accuracy). In contrast, Yurovsky, Smith, and Yu (2013) found a higher proportion of clear naming events, with approximately 30% being unambiguous ($> 90\%$ accuracy). Consistent with this finding, Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, and Trueswell (2013) showed that the proportion of unambiguous naming episodes varies across parent-child dyads, with some parents rarely providing highly informative contexts and others' doing so relatively more often.[3]

Thus, representations in cross-situational word learning can appear distributional or discrete, and

[3]The differences in the estimates of referential uncertainty in these studies could be driven by the different sampling procedures used to select naming events for the HSP. Yurovsky, Smith, and Yu (2013) sampled utterances for which the parent labeled a co-present object, whereas Medina, Snedeker, Trueswell, et al. (2011) randomly sampled any utterances containing concrete nouns. Regardless of these differences, the key point here is that variability in referential uncertainty across naming events exists and thus could alter the representations underlying cross-situational learning.

the input to statistical learning mechanisms can vary along a continuum from low to high ambiguity. These results raise an interesting question: could learners be sensitive to the ambiguity of the input and use this information to alter the representations they store in memory? In the current line of work, we investigated how the presence of referential cues in the social context might alter the ambiguity of the input to statistical word learning mechanisms.

Social-pragmatic theories of language acquisition emphasize the importance of social cues for word learning (Bloom, 2002; Clark, 2009; Hollich et al., 2000). Experimental work has shown that even children as young as 16 months prefer to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin, 1993). In an analysis of naturalistic parent-child labeling events, Yu and Smith (2012) found that young learners tended to retain labels that were accompanied by clear referential cues, which served to make a single object dominant in the visual field. And correlational studies have demonstrated strong links between early intention-reading skills (e.g., gaze following) and later vocabulary growth (Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998b). Moreover, studies outside the domain of language acquisition have shown that the presence of social cues: (a) produce better spatial learning of audiovisual events (Wu, Gopnik, Richardson, & Kirkham, 2011), (b) boost recognition of a cued object (Cleveland et al., 2007), and (c) lead to preferential encoding of an object's featural information (Yoon et al., 2008). Together, the evidence suggests that social cues could alter the representations stored during cross-situational word learning by modulating how people allocate attention to the relevant statistics in the input.

The goal of our current investigation was to ask whether the presence of a valid social cue – a speaker's gaze – could change the representations underlying cross-situational word learning. We used a modified version of Yurovsky and Frank (2015)'s paradigm to provide a direct measure of memory for alternative word-object links during cross-situational learning. In Experiment 1, we manipulated the presence of a referential cue at different levels of attention and memory demands. At all levels of difficulty, learners tracked a strong single hypothesis but were less likely to track multiple word-object links when a social cue was present. In Experiment 2, we replicated the findings from Experiment 1 using a more ecologically valid social cue. In Experiment 3, we moved to a parametric manipulation of referential uncertainty by varying the reliability of the speaker's gaze. Learners were sensitive to graded changes in reliability and retained more word-object links as uncertainty in the input increased. Finally, in Experiment 4, we equated the length of the initial naming events with and without the referential cue. Learners stored less information in the presence of gaze even when they had visually inspected the objects for the same amount of time. In sum, our data suggest that cross-situational word learners are quite flexible, storing representations with different levels of fidelity depending on the amount of ambiguity present during learning.

4.2 Experiment 1

We set out to test the effect of a referential cue on the representations underlying cross-situational word learning. We used a version of Yurovsky and Frank (2015)'s paradigm where we manipulated the ambiguity of the learning context by including a gaze cue from a schematic, female interlocutor. Participants saw a series of ambiguous exposure trials where they heard one novel word that was either paired with a gaze cue or not and selected the object they thought went with each word. In subsequent test trials, participants heard the novel word again, this time paired with a new set of novel objects. One of the objects in this set was either the participant's initial guess (Same test trials) or one of the objects was *not* their initial guess (Switch test trials). Performance on Switch trials provided a direct measure of whether referential cues influenced the number of alternative word-object links that learners stored in memory. If learners performed worse on Switch trials after an exposure trial with gaze, this would suggest that they stored fewer additional objects from the initial learning context.

4.2.1 Method

Participants

We posted a set of Human Intelligence Tasks (HITs) to Amazon Mechanical Turk. Only participants with US IP addresses and a task approval rate above 95% were allowed to participate, and each HIT paid 30 cents. 50-100 HITs were posted for each of the 32 between-subjects conditions. Data were excluded if participants completed the task more than once or if participants did not respond correctly on familiar object trials (131 HITs). The final sample consisted of 1438 participants.

Stimuli

Figure 1 shows screenshots taken from Experiment 1. Visual stimuli were black and white pictures of familiar and novel objects taken from Kanwisher, Woods, Iacoboni, and Mazziotta (1997). Auditory stimuli were recordings of familiar and novel words by an AT&T Natural Voices ™(voice: Crystal) speech synthesizer. Novel words were 1-3 syllable pseudowords that obeyed all rules of English phonotactics. A schematic drawing of a human speaker was chosen for ease of manipulating the direction of gaze, the referential cue of interest in this study. All experiments can be viewed and downloaded at the project page: `https://kemacdonald.github.io/soc_xsit/`.

Design and Procedure

Participants saw a total of 16 trials: eight exposure trials and eight test trials. On each trial, they heard one novel word, saw a set of novel objects, and were asked to guess which object went with the word. Before seeing exposure and test trials, participants completed four practice trials with familiar words and objects. These trials familiarized participants to the task and allowed us to exclude participants who were unlikely to perform the task as directed, either because of inattention or because their computer audio was turned off.

After the practice trials, participants were told that they would now hear novel words and see novel objects and that their task was to select the referent that "goes with each word." Over the course of the experiment, participants heard eight novel words two times, with one exposure trial and one test trial for each word. Four of the test trials were *Same* trials in which the object that participants selected on the exposure trial was shown with a set of new novel objects. The other four test trials were *Switch* trials in which one of the objects was chosen at random from the set of objects that the participant did not select on exposure.

Participants were randomly assigned to one of the 32 between-subjects conditions (4 Referents X 4 Intervals X 2 Gaze conditions). Participants either saw 2, 4, 6, or 8 referents on the screen and test trials occurred at different intervals after exposure trials: either 0, 1, 3, or 7 trials from the initial exposure to a word. For example, in the 0-interval condition, the test trial for that word would occur immediately following the exposure trial, but in the 3-interval condition, participants would see three additional exposure trials for other novel words before seeing the test trial for the initial word. The interval conditions modulated the time delay and the number of intervening trials between learning and test, and the number of referents conditions modulated the attention demands present during learning.

Participants were assigned to either the Gaze or No-Gaze condition. In the Gaze condition, gaze was directed towards one of the objects on exposure trials; in the No-Gaze condition, gaze was always directed straight ahead (see Figure 1 for examples). At test, gaze was always directed straight ahead. To show participants that their response had been recorded, a red box appeared around the selected object for one second. This box always appeared around the selected object, even if participants' selections were incorrect.

4.2.2 Results and Discussion

Analysis plan

The structure of our analysis plan is parallel across all four experiments. First, we examined accuracy on exposure trials in the Gaze condition and then we compared response times on exposure trials across the Gaze and No-Gaze conditions. These analyses tested whether learners were (a) sensitive to our experimental manipulation and (b) altered their allocation of attention in response to the presence of a social cue. Accuracy on exposure trials was defined as selecting the referent that was the target of gaze in the Gaze condition. (Note that there was no "correct" behavior for exposure trials in the No-Gaze condition.) Next, we examined accuracy on test trials to test whether learners' memory for alternative word-object links changed depending on the ambiguity of the learning context. Accuracy on test trials (both Same and Switch) was defined as selecting the referent that was present during the exposure trial for that word.

The key behavioral prediction of our hypothesis was that the presence of gaze would result in reduced memory for multiple word-object links, operationalized as a decrease in accuracy on Switch test trials after seeing exposure trials with a gaze cue. To quantify participants' behavior, we used mixed-effects regression models with the maximal random effects structure justified by our experimental design: by-subject intercepts and slopes for each trial type (Barr, 2013). We limited all models to include only two-way interactions because the critical test of our hypothesis was the interaction between gaze condition and trial type, and we did not have theoretical predictions for any possible three-way or four-way interactions.

In the main text, we only report effects that achieved statistical significance at the $\alpha = .05$

threshold. In the Appendix, we report the full model specification and output for each of the models in the paper. All models were fit using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2013), and all of our data and our processing/analysis code can be viewed in the version control repository for this paper at `https://github.com/kemacdonald/soc_xsit`.

Exposure trials

To ensure that our referential cue manipulation was effective, we compared participants' accuracies on exposure trials in the Gaze condition to a model of random behavior defined as a Binomial distribution with a probability of success $\frac{1}{NumReferents}$. Correct performance was defined as selecting the object that was the target of the speaker's gaze. Following Yurovsky and Frank (2015), we fit logistic regressions for each gaze, referent, and interval combination specified as `Gaze Target` $\sim$ `1 + offset(logit(1/Referents))`. The offset encoded the chance probability of success given the number of referents, and the coefficient for the intercept term shows on a log-odds scale how much more likely participants were to select the gaze target than would be expected if participants were selecting randomly. In all conditions, participants used gaze to select referents on exposure trials more often than expected by chance (smallest $\beta = 1.4$, z = 9.38, $p < .001$). However, the mean proportion of gaze following varied across conditions (overall $M = 0.84$, range: 0.77–0.93).

We were also interested in differences in participants' response times across the experimental conditions. Since these trials were self-paced, participants could choose how much time to spend inspecting the referents on the screen, thus providing an index of participants' attention. To quantify the effects of gaze, interval, and number of referents, we fit a linear mixed-effects model that predicted participants' inspection times as follows: `Log(Inspection time)` $\sim$ `(Gaze * Log(Interval) + Log(Referents))`^{2} `+ (1 | subject)`. We found a significant main effect of the number of referents ($\beta = 0.34$, p < .001) with longer inspection times as the number of referents increased, a significant interaction between gaze condition and the number of referents ($\beta =$ -0.27, p < .001) with longer inspection times in the No-Gaze condition, especially as the number of referents increased, and a significant interaction between gaze condition and interval ($\beta =$ -0.08, $p = 0.004$) with longer inspection times in the No-Gaze condition, especially as the number of intervening trials increased (see the top row of Figure 2). Shorter inspection times on exposure trials with gaze provide evidence that the presence of a referential cue focused participants' attention on a single referent and away

Predictor	Estimate	Std. Error	z value	p value	
Intercept	3.01	0.29	10.35	$< .001$	***
Switch Trial	-1.36	0.24	-5.63	$< .001$	***
Gaze Condition	0.12	0.26	0.47	0.64	
Log(Interval)	-0.45	0.11	-4.08	$< .001$	***
Log(Referents)	0.23	0.11	2.02	0.04	*
Switch Trial*Gaze Condition	-1.09	0.12	-9.07	$< .001$	***
Switch Trial*Log(Interval)	0.52	0.05	9.50	$< .001$	***
Switch Trial*Log(Referent)	-0.59	0.09	-6.49	$< .001$	***
Gaze Condition*Log(Interval)	0.06	0.06	1.00	0.32	
Gaze Condition*Log(Referent)	0.20	0.09	2.15	0.03	*
Log(Interval)*Log(Referent)	-0.04	0.04	-1.02	0.31	

Table 4.1: Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 4.1.

from alternative word-object links.

Test trials

Next, we explored participants' accuracy in identifying the referent for each word in all conditions for both kinds of test trials (see the bottom row of Figure 2). We first compared the distribution of correct responses made by each participant to the distribution expected if participants were selecting randomly defined as a Binomial distribution with a probability of success $\frac{1}{NumReferents}$. Correct performance was defined as selecting the object that was present on the exposure trial for that word. We fit the same logistic regressions as we did for exposure trials: `Correct ~ 1 + offset(logit(1/Referents))`. In 31 out of the 32 conditions for both Same and Switch trials, participants chose the correct object more often than would be expected by chance (smallest $\beta =$ 0.36, $z = 2.44$, $p = 0.01$). On Switch trials in the 8-referent, 3-interval condition, participants' responses were not significantly different from chance ($\beta = 0.06$, z = 0.33, $p = 0.74$). Participants' success on Switch trials replicates the findings from Yurovsky and Frank (2015) and provides direct evidence that learners encoded more than a single hypothesis in ambiguous word learning situations even under high attentional and memory demands and in the presence of a referential cue. To quantify the effects of gaze, interval, and number of referents on the probability of a correct response, we fit the following mixed-effects logistic regression model to a filtered dataset where we removed participants who did not reliably select the object that was the target of gaze on exposure trials:[4] `Correct` $\sim$ `(Trial Type + Gaze + Log(Interval) + Log(Referents))`^{2} `+ offset(logit(`$^1/_{Referents}$`)) + (TrialType | subject)`. We coded interval and number of referents as continuous predictors and transformed these variables to the log scale.[5]

Table 1 shows the output of the logistic regression. We found significant main effects of the number of referents ($\beta = 0.23$, $p < .001$) and interval ($\beta = -0.45$, $p < .001$), such that as each of these factors increased, accuracy on test trials decreased. We also found a significant main effect of trial type ($\beta = -1.36$, $p < .001$), with worse performance on Switch trials. There were significant interactions between trial type and interval ($\beta = 0.52$, $p < .001$), trial type and referents ($\beta = -0.59$, $p < .001$), and gaze condition and referents ($\beta = 0.2$, $p < .05$). These interactions can be interpreted as meaning: (a) the interval between exposure and test affected Same trials more than Switch trials, (b) the number of referents affected Switch trials more than Same trials, and (c) participants performed slightly better at the higher number of referents in the Gaze condition. The interactions between gaze condition and referents and between referents and interval were not significant. Importantly, we found the predicted interaction between trial type and gaze condition ($\beta = -1.09$, $p < .001$), with participants in the Gaze condition performing worse on Switch trials. This interaction provides direct evidence that the presence of a referential cue reduces participants' memory for alternative word-object links.

We were also interested in how the length of inspection times on exposure trials would affect participants' accuracy at test. So we fit an additional model where participants' inspection times were included as a predictor. We found a significant interaction between inspection time and gaze condition ($\beta = -0.17$, $p = 0.01$) such that longer inspection times provided a larger boost to accuracy in the No-Gaze condition. Importantly, the key test of our hypothesis, the interaction between gaze condition and trial type, remained significant in this alternative version of the model (β = -1.02, p = p < .001).

Taken together, the inspection time and accuracy analyses provide evidence that the presence of a referential cue modulated learners' attention during learning, and in turn made them less likely to track multiple word-object links. We saw some evidence for a boost to performance on Same

[4]We did not predict that there would be a subset of participants who would not follow the gaze cue, thus this filtering criterion was developed posthoc. However, we think that the filter is theoretically motivated because we would only expect to see an effect of gaze if participants actually used the gaze cue. The filter removed 94 participants (6% of the sample). The key inferences from the data do not depend on this filtering criterion.

[5]If we allowed for three-way interactions in the model, the key interaction between gaze condition and trial type remained significant (β = -1.3, p = 0.006).

trials in the Gaze condition at the higher number of referent and interval conditions, but reduced tracking of alternatives did not always result in better memory for learners' candidate hypothesis. This finding suggests that the limitations on Same trials may be different than those regulating the distribution of attention on Switch trials.

There was relatively large variation in performance across conditions in the group-level accuracy scores and in participants' tendency to *use* the referential cue on exposure trials. Moreover, we found a subset of participants who did not reliably use the gaze cue at all. It is possible that the effect of gaze was reduced because the referential cue that we used – a static schematic drawing of a speaker – was relatively weak compared to the cues present in real-world learning environments. Thus we do not yet know how learners' memory for alternatives during cross-situational learning would change in the presence of a stronger and more ecologically valid referential cue. We designed Experiment 2 to address this question.

4.3 Experiment 2

In Experiment 2, we set out to replicate the findings from Experiment 1 using a more ecologically valid stimulus set. We replaced the static, schematic drawing with a video of an actress. While these stimuli were still far from actual learning contexts, they included a real person who provided both a gaze cue and a head turn towards the target object. To reduce the across-conditions variability that we found in Experiment 1, we introduced a within-subjects design where each participant saw both Gaze and No-Gaze exposure trials in a blocked design. We selected a subset of the conditions from Experiment 1 and tested only the 4-referent display with 0 and 3 intervening trials as between-subjects manipulations. Our goals were to replicate the reduction in learners' tracking of alternative word-object links in the presence of a referential cue and to test whether increasing the ecological validity of the cue would result in a boost to the strength of learners' recall of their candidate hypothesis.

4.3.1 Method

Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiment 1. 100 HITs were posted for each condition (1 Referent X 2 Intervals X 2 Gaze conditions) for a total of 400 paid HITs (33 HITs excluded).

Stimuli

Audio and picture stimuli were identical to Experiment 1. The referential cue in the Gaze condition was a video (see Figure 1). On each exposure trial, the actress looked out at the participant with a neutral expression, smiled, and then turned to look at one of the four images on the screen. She maintained her gaze for 3 seconds before returning to the center. On test trials, she looked straight ahead for the duration of the trial.

4.3.2 Design and Procedure

Procedures were identical to those of Experiment 1. The major design change was a within-subjects manipulation of the gaze cue where each participant saw exposure trials with and without gaze. The experiment consisted of 32 trials split into 2 blocks of 16 trials. Each block consisted of 8 exposure trials and 8 test trials (4 Same trials and 4 Switch trials) and contained only Gaze or No-gaze exposure trials. The order of block was counterbalanced across participants.

4.3.3 Results and Discussion

We followed the same analysis plan as in Experiment 1. We first analyzed inspection times and accuracy on exposure trials and then analyzed accuracy on test trials.

Exposure trials

Similar to Experiment 1, participants' responses on exposure trials differed from those expected by chance (smallest β = 3.39, z = 31.99, p < .001), suggesting that gaze was effective in directing participants' attention. Participants in Experiment 2 were more consistent in their use of gaze with the video stimuli compared to the schematic stimuli used in Experiment 1 ($M_{Exp1} = 0.8$, $M_{Exp2} =$

Predictor	Estimate	Std. Error	z value	p value	
Intercept	4.04	0.18	21.97	$< .001$	***
Switch Trial	-2.99	0.19	-16.11	$< .001$	***
Gaze Condition	-0.10	0.16	-0.63	0.53	
Log(Interval)	-0.93	0.10	-9.23	$< .001$	***
Switch Trial*Gaze Condition	-0.71	0.16	-4.49	$< .001$	***
Switch Trial*Log(Interval)	0.79	0.10	8.03	$< .001$	***
Gaze Condition*Log(Interval)	0.15	0.08	2.05	0.04	*

Table 4.2: Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 4.2.

0.91), suggesting that using a real person increased participants' willingness to follow the gaze cue.

We replicated the findings from Experiment 1. Inspection times were shorter when gaze was present (β = -1.1, $p < .001$) and in the 3-interval condition (β = -0.48, $p < .001$). The interaction between gaze and interval was not significant, meaning that gaze had the same effect on participants' inspection times at both intervals (see Panel A of Figure 3).

Test trials

Across all conditions for both trial types, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.58$, z = 9.32, $p < .001$). We replicated the critical finding from Experiment 1: after seeing exposure trials with gaze, participants performed worse on Switch trials, meaning they stored fewer word-object links ($\beta = -0.71$, $p < .001$).[6] Participants were also less accurate as the interval between exposure and test increased (β = -0.93, $p < .001$) and on the Switch trials overall ($\beta = -2.99$, $p < .001$).

In addition, there was a significant interaction between trial type and interval ($\beta = 0.79$, $p < .001$), with worse performance on Switch trials in the 3-interval condition. The interaction between gaze condition and interval was also significant ($\beta = 0.15$, $p = 0.041$), such that participants in the gaze condition were less affected by the increase in interval. Similar to Experiment 1, we did not see evidence of a boost to performance on Same trials in the gaze condition.

Next, we added inspection times on exposure trials to the model. Similar to Experiment 1, the key interaction between gaze and trial type remained significant in this version of the model (β = -0.54, $p < .001$). We also found an interaction between inspection time and trial type ($\beta = 0.21$, $p = 0.05$),

[6] As in Experiment 1, we fit this model to a filtered dataset removing participants who did not reliably use the gaze cue.

with longer inspection times providing a larger boost to performance on Switch trials (i.e., stronger memory for alternative word-object links). This result differs slightly from Experiment 1 where we found an interaction between trial type and inspection time, with longer inspection times providing a larger boost to accuracy in the No-Gaze condition. Despite this subtle difference, we speculate that inspection times likely played a similar role in both experiments, with longer inspection times leading to better performance on Switch trials since these trials depended on encoding multiple word-object links. It is also possible that the interaction between gaze condition and inspection time that we found in Experiment 1 was influenced by the different number of referents and interval conditions.

The results of Experiment 2 provide converging evidence for our primary hypothesis that the presence of a referential cue reliably focuses learners' attention away from alternative word-object links and shifts them towards single hypothesis tracking. Moving to the video stimulus led to higher rates of selecting the target of gaze on exposure trials, but did not result in a boost to performance on Same trials. This finding suggests that the level of attention and memory demand present in the learning context might modulate the effect of gaze on the fidelity of learners' single hypothesis.

Thus far we have shown that people store different amounts of information in response to a categorical manipulation of referential uncertainty. In both Experiments 1 and 2, the learning context was either entirely ambiguous (No-Gaze) or entirely unambiguous (Gaze). But not all real-world learning contexts fall at the extremes of this continuum. Could learners be sensitive to more subtle changes in the quality of the input? In our next experiment, we tested a prediction of our account: whether learners would store more word-object links in response to graded changes in referential uncertainty during learning.

4.4 Experiment 3

In Experiment 3, we explored whether learners would allocate attention and memory flexibly in response to *graded* changes in the referential uncertainty that was present during learning. To test this hypothesis, we moved beyond a categorical manipulation of the presence/absence of gaze, and we parametrically varied the reliability of the referential cue. We manipulated cue reliability by adding a block of familiarization trials where we varied the proportion of Same and Switch trials. If participants saw more Switch trials, this provided direct evidence that the speaker's gaze was a less reliable cue to reference because the gaze target on exposure trials would not appear at test. This design was inspired by a growing body of experimental work showing that even young children are sensitive to the prior reliability of speakers and will use this information to decide whom to learn novel words from (e.g., Koenig, Clement, & Harris, 2004).

4.4.1 Method

Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiment 1 and 2 (27 HITs excluded). 100 HITs were posted for each reliability level (0%, 25%, 50%, 75%, and 100%) for total of 500 paid HITs.

Design and Procedure

Procedures were identical to those of Experiments 1 and 2. We modified the design of our cross-situational learning paradigm to include a block of 16 familiarization trials (8 exposure trials and 8 test trials) at the beginning of the experiment. These trials served to establish the reliability of the speaker's gaze. To establish reliability, we varied the proportion of Same/Switch trials that occurred during the familiarization block. Recall that on Switch trials the gaze target did not show up at test, which provided evidence that the speaker's gaze was not a reliable cue to reference. Reliability was a between-subjects manipulation such that participants either saw 8, 6, 4, 2, or 0 Switch trials during familiarization, which created the 0%, 25%, 50%, 75%, and 100% reliability conditions. After the familiarization block, participants completed another block of 16 trials (8 exposure trials and 8 test trials). Since we were no longer testing the effect of the presence or absence of a referential cue, all exposure trials throughout the experiment included a gaze cue. Finally, at the end of the task, we asked participants to assess the reliability of the speaker on a continuous scale from "completely unreliable" to "completely reliable."

4.4.2 Results and Discussion

Exposure trials

Participants reliably chose the referent that was the target of gaze at rates greater than chance (smallest $\beta = 2.62$, z $= 31.99$, $p < .001$). We fit a mixed effects logistic regression model predicting the probability of selecting the gaze target as follows: `Correct-Exposure ~ Reliability Condition * Subjective Reliability + (1 | subject)`. We found an effect of reliability condition ($\beta = 3.28$, $p = 0.03$) such that when the gaze cue was more reliable, participants were more likely to use it ($M_{0\%} = 0.83$, $M_{25\%} = 0.82$, $M_{50\%} = 0.87$, $M_{75\%} = 0.9$, $M_{100\%} = 0.94$). We also found an effect of subjective reliability ($\beta = 7.26$, $p < .001$) such that when participants thought the gaze cue was reliable, they were more likely to use it. This analysis provides evidence that participants were sensitive to the reliability manipulation both in how often they used the gaze cue and in how they rated the reliability of the speaker at the end of the task.

Test trials

Next, we tested whether the reliability manipulation altered the strength of participants' memory for alternative word-object links in the second block of test trials that followed the initial familiarization phase. Across all conditions, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.42$, z $= 3.69$, $p < .001$). Our primary prediction was an interaction between reliability and test trial type, with higher levels of reliability leading to worse performance on Switch trials (i.e., less memory allocated to alternative word-object links). To explore this prediction, we performed four complementary analyses: our primary analysis, which tested the effect of the reliability manipulation, and three secondary analyses, which explored the effects of participants' (a) use of the gaze cue, (b) subjective reliability assessments, and (c) inspection time on exposure trials.

Reliability condition analysis

To test the effect of reliability, we fit a model predicting accuracy at test using reliability condition and test trial type as predictors. We found a significant main effect of trial type ($\beta = -3.95$, $p < .001$), with lower accuracy on Switch trials. We also found the key interaction between reliability condition and trial type ($\beta =$ -0.76, $p = 0.044$), such that when gaze was more reliable, participants performed worse on Switch trials (see Panel A of Figure 4). This interaction suggests that people store more word-object links as the learning context becomes more ambiguous. However, the interaction between reliability and trial type was not particularly strong, and – similar to Experiment 1 – performance varied across conditions (see the 50% reliable condition in Panel A of Figure 4). So to provide additional support for our hypothesis, we conducted three follow-up analyses.

Gaze use analyses

We would only expect to see a strong interaction between reliability and trial type if learners chose to use the gaze cue during exposure trials. To test this hypothesis, we fit two additional models that included two different measures of participants' use of the gaze cue. First, we added the number of exposure trials on which participants chose to use the gaze cue as a predictor in our model. We found a significant interaction between use of the gaze cue on exposure trials and trial type ($\beta = -1.43$, $p < .001$) with worse performance on Switch test trials when participants used gaze on exposure trials (see Panel B of Figure 4). We also found an interaction between gaze use and reliability ($\beta = 0.97$, $p = 0.004$) such that when gaze was more reliable, participants were more likely to use it. The β value for the interaction between trial type and reliability changed from -0.76 to -0.62, ($p = 0.086$). This reduction suggests that participants' tendency to use the gaze cue is a stronger predictor of learners' memory for alternative word-object links compared to our reliability manipulation.[7]

We also hypothesized that the reliability manipulation might change how often individual participants chose to use the gaze cue throughout the task. To explore this possibility, we fit a model with the same specifications, but we included a predictor that we created by binning participants based on the number of exposure trials on which they chose to follow gaze (i.e., a gaze following score). We found a significant interaction between how often participants chose to follow gaze on exposure trials and trial type ($\beta = -0.26$, $p < .001$), such that participants who were more likely to use the gaze cue performed worse on Switch trials, but not Same trials (see Panel A of Figure 5).[8] Taken together, the two analyses of participants' use of the gaze cue provide converging evidence that when the speaker's gaze was reliable participants were more likely to use the cue, and when they followed gaze, they tended to store less information from the initial naming event.

[7] We are grateful to an anonymous reviewer for suggesting this analysis, but we would like to note that it is exploratory.

[8] We found this interaction while performing exploratory data analyses on a previous version of this study with an independent sample (N = 250, $\beta = -0.24$, $p < .001$). The results reported here are from a follow-up study where testing this interaction was a planned analysis.

Subjective reliability analysis

The strong interaction between use of the gaze cue and memory for alternative word-object links suggests that participants' subjective experience of reliability in the experiment mattered. Thus, we fit the same model but substituted subjective reliability for the frequency of gaze use as a predictor of test trial performance. We found a significant interaction between trial type and participants' subjective reliability assessments ($\beta = -1.63$, $p = 0.01$): when participants thought the speaker was more reliable, they performed worse on Switch trials, but not Same trials (see Panel B of Figure 5).

Inspection time analyses

Finally, we analyzed the effect of inspection times on exposure trials, fitting a model using inspection time, trial type, and reliability condition to predict accuracy at test. We found a main effect of inspection time ($\beta = 0.31$, $p = 0.001$), with longer inspection times leading to better performance for both Same and Switch trials. The interaction between inspection time and reliability condition was not significant. The key interaction between reliability condition and trial type remained significant in this version of the model (β = -0.58, p = 0.048).

Next, we explored the factors that influenced inspection time on exposure trials by fitting a model to predict inspection times as a function of reliability condition and participants' use of the gaze cue. We found a main effect of participants' use of the gaze cue (-0.32, $p < .001$) with shorter inspection times when participants followed gaze. The main effect of reliability condition and the interaction between reliability and use of gaze were not significant. These analyses provide evidence that inspection times were similar across the different reliability conditions and that use of the gaze cue was the primary factor affecting how long participants explored the objects during learning.

Together, these four analyses show that when the speaker's gaze was more reliable, participants were more likely to: (a) use the gaze cue, (b) rate the speaker as more reliable, and (c) store fewer word-object links, showing behavior more consistent with single hypothesis tracking. These findings support and extend the results of Experiments 1 and 2 in several important ways. First, similar to Experiment 2, participants' performance on Same trials was relatively unaffected by changes in performance on Switch trials. The selective effect of gaze on Switch trials provides converging evidence that the limitations on Same trials may be different than those regulating the distribution of attention on Switch trials. Second, learners' use of a referential cue was a stronger predictor of reduced memory for alternative word-object links compared to our reliability manipulation. Although we found a significant effect of reliability on participants' use of the gaze cue, participants' tendency to use the cue remained high. Consider that even in the 0% reliability condition the mean proportion of gaze following was still 0.82. It is reasonable that participants would continue to use the gaze cue in our experiment since it was the only cue available and participants did not have a strong reason to think that the speaker would be deceptive.

The critical contribution of Experiment 3 is to show that learners respond to a graded manipulation of referential uncertainty, with the amount of information stored from the initial exposure tracking with the reliability of the cue. This graded accuracy performance shows that learners stored alternative word-object links with different levels of fidelity depending on the amount of referential uncertainty present during learning.

Across Experiments 1-3, learners tended to store fewer word-object links in unambiguous learning contexts when a clear referential cue was present. However, in all three experiments, participants' responses on exposure trials controlled the length of the trial, meaning that when participants used the gaze cue, they also spent less time visually inspecting the objects. Thus, we do not know whether there is an independent effect of referential cues on the representations underlying cross-situational learning, or if the effects found in Experiments 1-3 are entirely mediated by a reduction in inspection time. In Experiment 4, we addressed this possibility by removing participants' control over the length of exposure trials, which made the inspection times equivalent across the Gaze and No-Gaze conditions.

4.5 Experiment 4

In Experiment 4, we asked whether a reduction in visual inspection time in the gaze condition could completely explain the effect of social cues on learners' reduced memory for alternative word-object links. To answer this question, we modified our paradigm and made the length of exposure trials equivalent across the Gaze and No-Gaze conditions. In this version of the task, participants were shown the objects for a fixed amount of time regardless of whether gaze was present. We also included two different exposure trial lengths in order to test whether gaze would have a differential effect at shorter vs. longer inspection times. If the presence of gaze reduces learners' memory for multiple word-object links, then this provides evidence that referential cues affected the underlying representations over and above a reduction in inspection time.

4.5.1 Method

Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiments 1, 2, and 3. 100 HITs were posted for each condition (1 Referent X 2 Intervals X 2 Inspection Time conditions) for a total of 400 paid HITs (37 HITs excluded).

Stimuli

Audio, picture, and video stimuli were identical to Experiments 2 and 3. Since inspection times were fixed across conditions, we wanted to ensure that participants were aware of the time remaining on each exposure trial. So we included a circular countdown timer located above the center video. The timer remained on the screen during test trials but did not count down since participants could take as much time as they wanted to respond on test trials.

4.5.2 Design and Procedure

Procedures were identical to those of Experiment 1-3. The design was identical to that of Experiment 2 and consisted of 32 trials split into 2 blocks of 16 trials. Each block consisted of 8 exposure trials and 8 test trials (4 Same trials and 4 Switch trials) and contained only Gaze or No-Gaze exposure trials. The order of block was counterbalanced across participants.

The major design change was to make the length of exposure trials equivalent across the Gaze and No-Gaze conditions. We randomly assigned participants to one of two inspection time conditions: Short or Long. Initially, the length of the inspection times was based on participants' self-paced inspection times in the Gaze and No-Gaze conditions in Experiment 2 (Short = 3 seconds; Long = 6 seconds). However, after pilot testing, we added three seconds to each condition to ensure that participants had enough time to respond before the experiment advanced (Short = 6 seconds; Long = 9 seconds). If participants did not respond in the allotted time, an error message appeared informing participants that time had run out and encouraged them to respond within the time window on subsequent trials.

4.5.3 Results and Discussion

We did not see strong evidence of an effect of the different inspection times. Thus, all of the results reported here collapse across the short and long inspection time conditions. For all analyses, we removed the trials on which participants did not respond within the fixed inspection time on exposure trials (0.05% of trials).

Exposure Trials

Participants' responses on exposure trials differed from those expected by chance (smallest $\beta = 2.95$, z = 38.08, $p < .001$), suggesting that gaze was again effective in directing participants' attention. Similar to Experiment 2, participants were quite likely to use the gaze cue when it was a video of an actress ($M_{0-interval} = 0.93$, $M_{3-interval} = 0.95$).

Test Trials

Figure 6 shows performance on test trials in Experiment 4. In the majority of conditions, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.2$, z = 2.2, $p < .05$). However, participants' responses were not different from chance on Switch trials after exposure trials with gaze in the 3-interval condition ($\beta = 0.17$, $p = 0.06$).

We replicate the key finding from Experiments 1-3: after seeing exposure trials with gaze, participants were less accurate on Switch trials ($\beta = 0.9$, $p < .001$). Since inspection times were fixed across the Gaze and No-Gaze conditions, this finding provides evidence that the presence of a referential cue did more than just reduce the amount of time participants' spent inspecting the potential word-object links. In contrast to Experiments 2 and 3, visual inspection of Figure 6 suggested that the referential cue provided a boost to accuracy on Same trials. To assess the simple effect of gaze on trial type, we computed pairwise contrasts using the *lsmeans* package in R with a Bonferroni correction for multiple comparisons (Lenth, 2016). Accuracy was higher for Same trials in the Gaze condition ($\beta = 0.49$, $p < .001$), but lower for Switch trials ($\beta = -0.41$, $p < .001$). The boost in accuracy on Same trials differs from Experiments 2 and 3 and suggests that making inspection times equivalent across conditions allowed the social cue to affect the strength of learners' memory for their candidate hypothesis.

The results of Experiment 4 help to clarify the effect of gaze on memory in our task, providing evidence that the presence of a referential cue did more than just reduce participants' visual inspection time. Instead, gaze reduced memory for alternative word-object links even when people had the same opportunity to visually inspect and encode them. We also found evidence of a boost for learners' memory of their candidate hypothesis in the gaze condition, an effect that we saw at the higher number of referents and the longer intervals in Experiment 1, but that we did not see in Experiments 2 or 3. One explanation for this difference is that in Experiment 4, since participants' use of gaze was independent of the length of exposure trials, inspection times in the gaze condition were longer compared to those in Experiments 1-3. Thus, it could be that the combination of a gaze cue coupled with the opportunity to continue attending to the gaze target led to a boost in performance on Same trials relative to trials without gaze.

4.6 General Discussion

Tracking cross-situational word-object statistics allows word learning to proceed despite the presence of individually ambiguous naming events. But models of cross-situational learning disagree about how much information is actually stored in memory, and the input to statistical learning mechanisms can vary along a continuum of referential uncertainty from unambiguous naming instances to highly ambiguous situations. In the current line of work, we explore the hypothesis that these two factors are fundamentally linked to one another and to the social context in which word learning occurs. Specifically, we ask how cross-situational learning operates over social input that varies the amount of ambiguity in the learning context.

Our results suggest that the representations underlying cross-situational learning are quite flexible. In the absence of a referential cue to word meaning, learners tended to store more alternative word-object links. In contrast, when gaze was present learners stored less information, showing behavior consistent with tracking a single hypothesis (Experiments 1 and 2). Learners were also sensitive to a parametric manipulation of the strength of the referential cue, showing a graded increase in the tendency to use the cue as reliability increased, which in turn resulted in a graded decrease in memory for alternative word-object links (Experiment 3). Finally, learners stored less information in the presence of gaze even when they were shown the objects for the same amount of time (Experiment 4).

In Experiments 2 and 3 reduced memory for alternative hypotheses did not result in a boost to memory for learners' candidate hypothesis. This pattern of data suggests that the presence of a referential cue selectively affected one component of the underlying representation: the number of alternative word-object links, and not the strength of the learners' candidate hypothesis. However, in Experiments 1 and 4, we did see some evidence of stronger memory for learners' initial hypothesis in the presence of gaze: at the higher number of referents and interval conditions (Experiment 1), and when the length of exposure trials was equivalent across the Gaze and No-Gaze conditions (Experiment 4). We speculate that the relationship between the presence of a referential cue and the strength of learners' candidate hypothesis is modulated by how the cue interacts with attention. In Experiment 1, gaze may have provided a boost because, in the absence of gaze, attention would have been distributed across a larger number of alternatives. And, in Experiment 4, gaze may have led to better memory because it was coupled with the opportunity for sustained attention to the gaze target. More work is needed in order to understand precisely when the presence of gaze affects this particular component of the representations underlying cross-situational learning.

In Experiments 1-3, longer inspection times (i.e., more time spent encoding the word-object links during learning) led to better memory at test. We did, however, find slightly different interaction effects across our studies. In Experiment 1, longer inspection times led to higher accuracy in the No-Gaze condition for both Same and Switch trials. In Experiment 2, longer inspection times provided a larger boost to performance on Switch trials compared to Same trials, regardless of gaze condition. Despite these differences, we speculate that inspection time played a similar role across these studies: When a social cue was present, learners' attention was focused and inspection times tended to be shorter, which led to worse performance on Switch trials (i.e., reduced memory for alternative word-object links). Interestingly, in Experiment 4, we found an effect of social cues on memory for alternatives even when participants were given the same opportunity to visually inspect the objects, suggesting that gaze does more than just modulate visual attention during learning.

4.6.1 Relationship to previous work

Why might a decrease in memory for alternatives fail to increase the strength of learners' memory for their candidate hypothesis? One possibility is that participants did not shift their cognitive resources from the set of alternatives to their single hypothesis, but instead chose to use the gaze information to reduce inspection time, thus conserving their resources for future use. Griffiths, Lieder, and Goodman (2015) formalize this behavior by pushing the rationality of computational-level models down to the psychological process level. In their framework, cognitive systems are thought to be adaptive in that they optimize the use of their limited resources, taking the cost of computation (e.g., the opportunity cost of time or mental energy) into account. For example, Vul, Goodman, Griffiths, and Tenenbaum (2014) showed that as time pressure increased in a decision-making task, participants were more likely to show behavior consistent with a less cognitively challenging strategy of matching, rather than with the globally optimal strategy. In the current work, we found that learners showed evidence of altering how they allocated cognitive resources based on the amount of referential uncertainty present during learning, spending less time inspecting alternative word-object links and reducing the number of links stored in memory when uncertainty was low.

Our results fit well with recent experimental work that investigates how attention and memory can constrain infants' statistical word learning. For example, Smith and Yu (2013) used a modified cross-situational learning task to show that only infants who disengaged from a novel object to look at both potential referents were able to learn the correct word–object mappings. Moreover, Vlach and Johnson (2013) showed that 16-month-olds were only able to learn from adjacent cross-situational co-occurrence statistics, and unable to learn from co-occurrences that were separated in time. Both of these findings make the important point that only the information that comes into contact with the learning system can be used for cross-situational word learning, and this information is directly influenced by the attention and memory constraints of the learner. These results also add to a large literature showing the importance of social information for word learning (Bloom, 2002; Clark, 2009) and to recent work exploring the interaction between statistical learning mechanisms and other types of information (Frank et al., 2009; Koehne & Crocker, 2014; C. Yu & Ballard, 2007). Our findings suggest that referential cues affect statistical learning by modulating the amount of information that learners store in the underlying representations that support learning over time.

Is gaze a privileged cue, or could other, less-social cues (e.g., an arrow) also affect the representations underlying cross-situational learning? On the one hand, previous research has shown that gaze cues lead to more reflexive attentional responses compared to arrows (Friesen, Ristic, & Kingstone, 2004), that gaze-triggered attention results in better learning compared to salience-triggered attention (Wu & Kirkham, 2010), and that even toddlers readily use gaze to infer novel word meanings

(Baldwin, 1993). Thus, it could be that gaze is an especially effective cue for constraining word learning since it communicates a speaker's referential intent and is a particularly good way to guide attention. On the other hand, the generative process of the cue – whether it is more or less social in nature – might be less important; instead, the critical factor might be whether the cue effectively reduces uncertainty in the naming event. Under this account, gaze is placed amongst a set of many cues that could produce similar effects as those reported here. Future work could explore a wider range of cues to see if they modulate the representations underlying cross-situational learning in a similar way.

How should we characterize the effect of gaze on attention and memory in our task? One possibility is that the referential cue acts as a filter, only allowing likely referents to contact statistical learning mechanisms (C. Yu & Ballard, 2007). This 'filtering account' separates the effect of social cues from the underlying computation that aggregates cross-situational information. Another possibility is that referential cues provide evidence about a speaker's communicative intent (Frank et al., 2009). In this model, the learner is reasoning about the speaker and word meanings simultaneously, which places inferences based on social information as part of the underlying computation. A third possibility is that participants thought of the referential cue as pedagogical. In this context, learners assume that the speaker will choose an action that is most likely to increase the learner's belief in the true state of the world (Shafto et al., 2012b), making it unnecessary to allocate resources to alternative hypotheses. Experiments show that children spend less time exploring an object and are less likely to discover alternative object-functions if a single function is demonstrated in a pedagogical context (Bonawitz et al., 2011). However, because the results from the current study cannot distinguish between these explanations, these questions remain topics for future studies specifically designed to tease apart these possibilities.

4.6.2 Limitations

There are several limitations to the current study that are worth noting. First, the social context that we used was relatively impoverished. Although we moved beyond a simple manipulation of the presence or absence of social information in Experiment 3, we nevertheless isolated just a single cue to reference, gaze. But real-world learning contexts are much more complex, providing learners access to multiple cues such as gaze, pointing, and previous discourse. In fact, Frank, Tenenbaum, and Fernald (2013) analyzed a corpus of parent-child interactions and concluded that learners would do better to aggregate noisy social information from multiple cues, rather than monitor a single cue since no single cue was a consistent predictor of reference. In our data, we did see a more reliable effect of referential cues when we used a video of an actress, which included both gaze and head turn as opposed to the static, schematic stimuli, which only included gaze. It is still an open and interesting question as to how our results would generalize to learning environments that contain a rich combination of social cues.

Second, we do not yet know how variations in referential uncertainty during learning would affect the representations of young word learners, the age at which cross-situational word learning might be particularly important. Recent research using a similar paradigm as our own did not find evidence that 2- or 3-year-olds stored multiple word-object links; instead, children only retained a single candidate hypothesis (Woodard, Gleitman, & Trueswell, 2016). However, performance limitations on children's developing attention and memory systems (Colombo, 2001; Ross-sheehy, Oakes, & Luck, 2003) could make success on these explicit response tasks more difficult. Moreover, our work suggests that different levels of referential uncertainty in naturalistic learning contexts (see Medina et al., 2011; Yurovsky & Frank, 2015) might evoke different strategies for information storage, with learners retaining more information as ambiguity in the input increases. Thus, we think that it will be important to test a variety of outcome measures and learning contexts to see if younger learners show evidence of storing multiple word meanings during learning.

In addition, previous work with infants has shown that their attention is often stimulus-driven and sticky (Oakes, 2011), suggesting that very young word learners might not effectively explore the visual scene in order to extract the necessary statistics for storing multiple alternatives. It could be that referential cues play an even more important role for young learners by filtering the input to cross-situational word learning mechanisms and guiding children to the relevant statistics in the input. In fact, recent work has shown that the precise timing of features such as increased parent attention and gesturing towards a named object and away from non-target objects were strong predictors of referential clarity in a naming event (Trueswell et al., 2016). It could be that the statistics available in these particularly unambiguous naming events are the most useful for cross-situational learning.

Finally, the current experiments used a restricted cross-situational word learning scenario, which differs from real-world language learning contexts in several important ways. One, we only tested a single exposure for each novel word-object pairing; whereas, real-world naming events are best characterized by discourse where an object is likely to be named repeatedly in a short amount of time (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014). Two, the restricted visual world of 2-8 objects on a screen combined with the forced-choice response format may have biased people to assume that all words in the task must have referred to one of the objects. But, in actual language use, people can refer to things that are not physically co-present (e.g., Gleitman, 1990), creating a scenario where learners would not benefit from storing additional word-object links in the absence of clear referential cues. Finally, we presented novel words in isolation, removing any sentential cues to word meaning (e.g., verb-argument relations). In fact, previous work with adults has shown that cross-situational learning mechanisms only operate in contexts where sentence-level constraints do not completely disambiguate meaning (Koehne & Crocker, 2014). Thus, we need more evidence to understand how the representations underlying cross-situational learning change in response to referential uncertainty at different timescales and in richer language contexts that more accurately reflect real-world learning environments.

4.7 Conclusions

Word learning proceeds despite the potential for high levels of referential uncertainty and despite learners' limited cognitive resources. Our work shows that cross-situational learners flexibly respond to the amount of ambiguity in the input, and as referential uncertainty increases, learners tend to store more word-object links. Overall, these results bring together aspects of social and statistical accounts of word learning to increase our understanding of how statistical learning mechanisms operate over fundamentally social input.

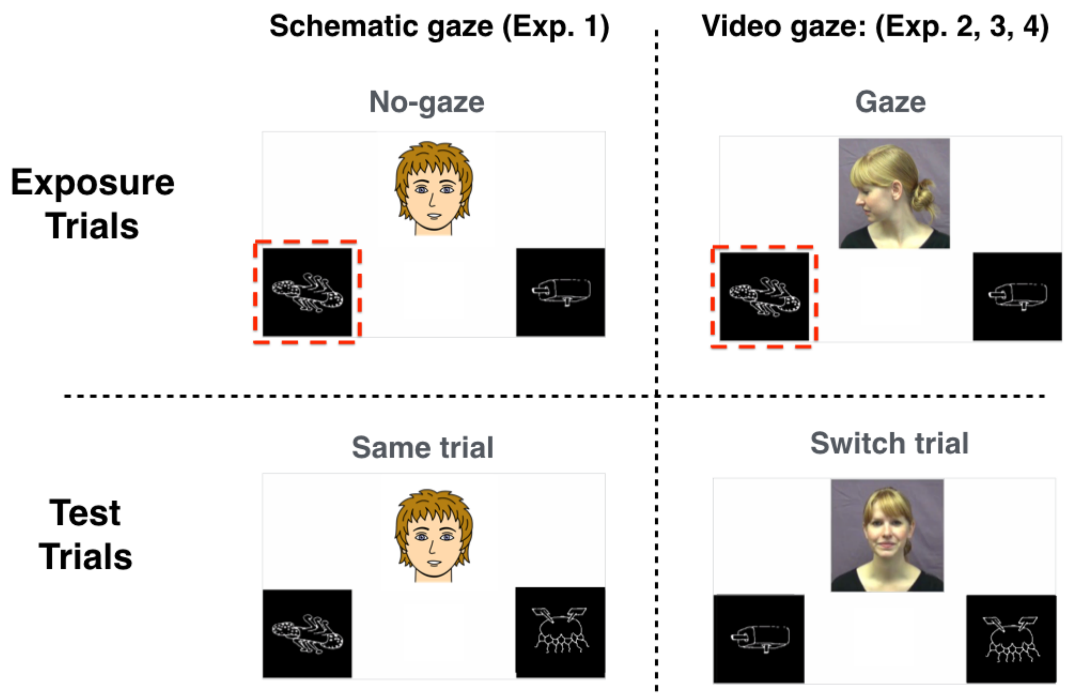


Figure 4.1: Screenshots of exposure and test trials from Experiments 1-4. The top left panel shows an exposure trial in the No-gaze condition using the schematic gaze cue (Experiment 4.1). The top right panel shows an exposure trial in the Gaze condition using the video gaze cue (Experiments 4.2-4.4). Participants saw either Gaze or No-gaze exposure trials depending on condition assignment, and participants saw both types of test trials: Same (bottom left panel) and Switch (bottom right panel). On Same trials, the object that participants chose during exposure appeared with a new novel object. On Switch trials the object that participants did not choose appeared with a new novel object. Participants either saw 2, 4, 6, or 8 referents on the screen depending on condition assignment.

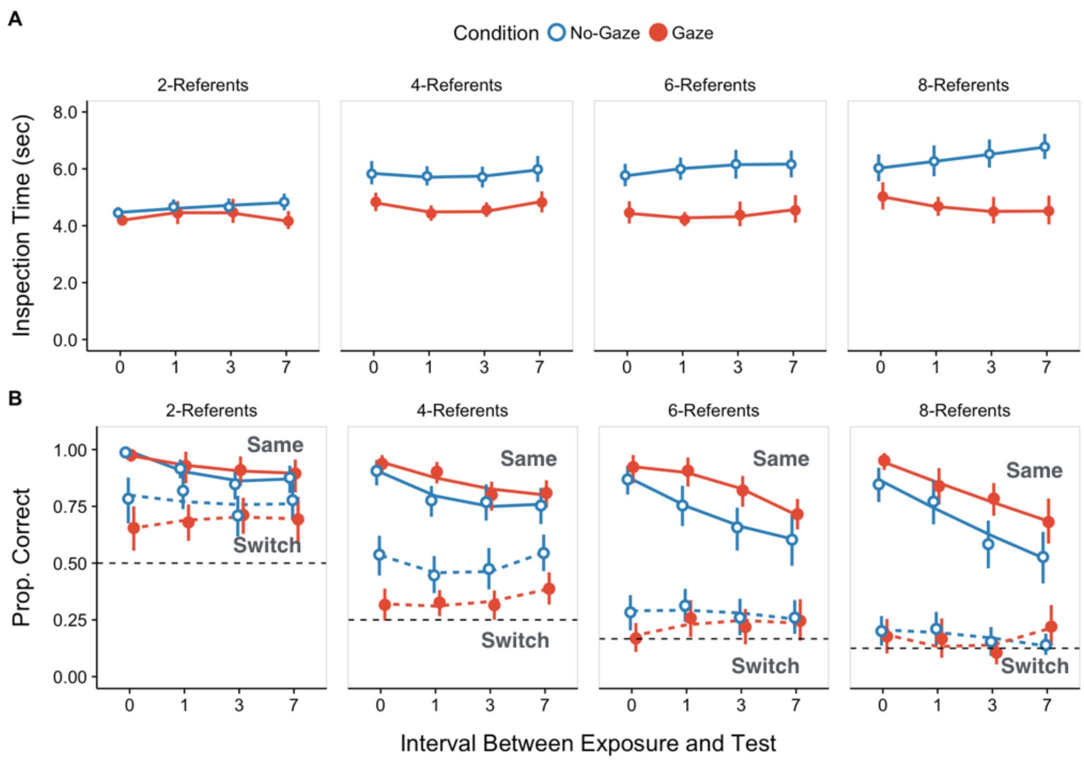


Figure 4.2: Experiment 4.1 results. The top row shows average inspection times on exposure trials for all experimental conditions as a function of the number of trials that occurred between exposure and test. Each panel represents a different number of referents, and line color represents the Gaze and No-Gaze conditions. The bottom row shows accuracy on test trials for all conditions as a function of the number of intervening trials. The horizontal dashed lines represent chance performance for each number of referents, and the type of line (solid vs. dashed) represents the different test trial types (Same vs. Switch). Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

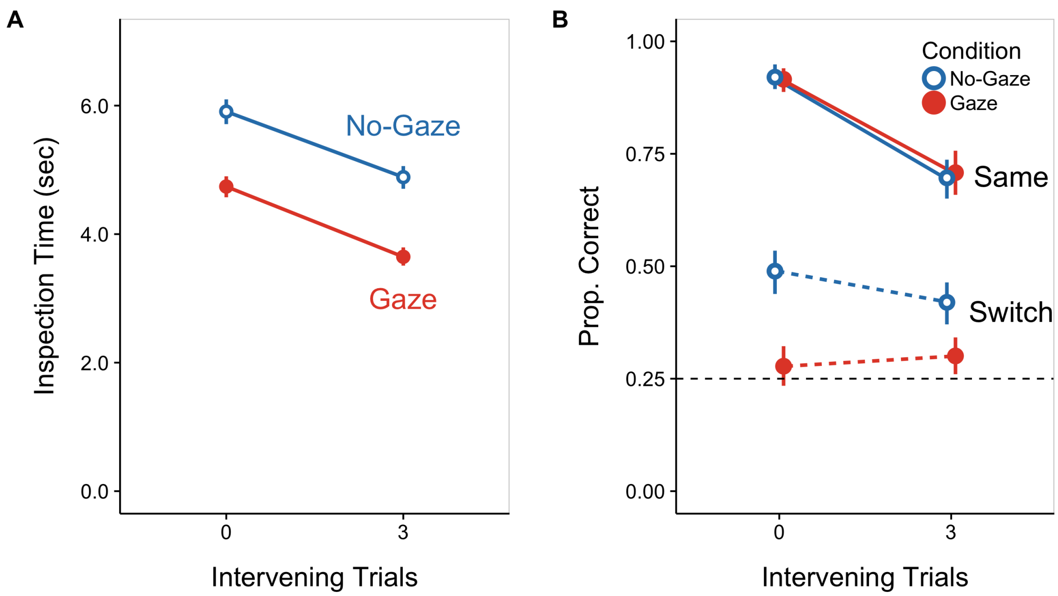


Figure 4.3: Experiment 2 results. Panel A shows inspection times on exposure trials with and without gaze. Panel B shows accuracy on Same and Switch test trials. All plotting conventions are the same as in Figure 2. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

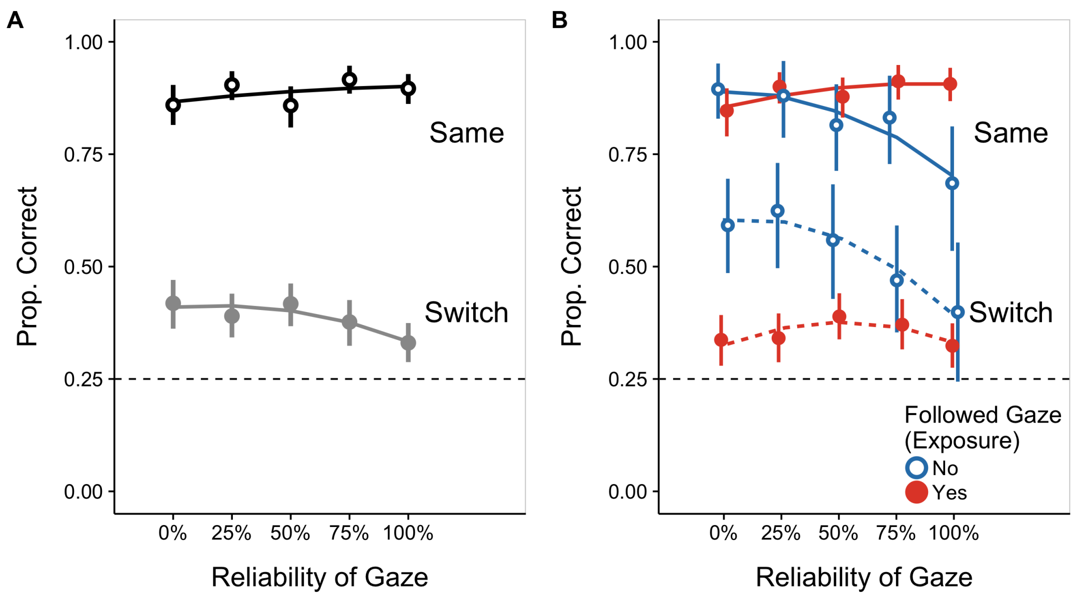


Figure 4.4: Primary analyses of test trial performance in Experiment 3. Panel A shows performance as a function of reliability condition. Panel B shows performance as a function of reliability condition and whether participants chose to follow gaze on exposure trials. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

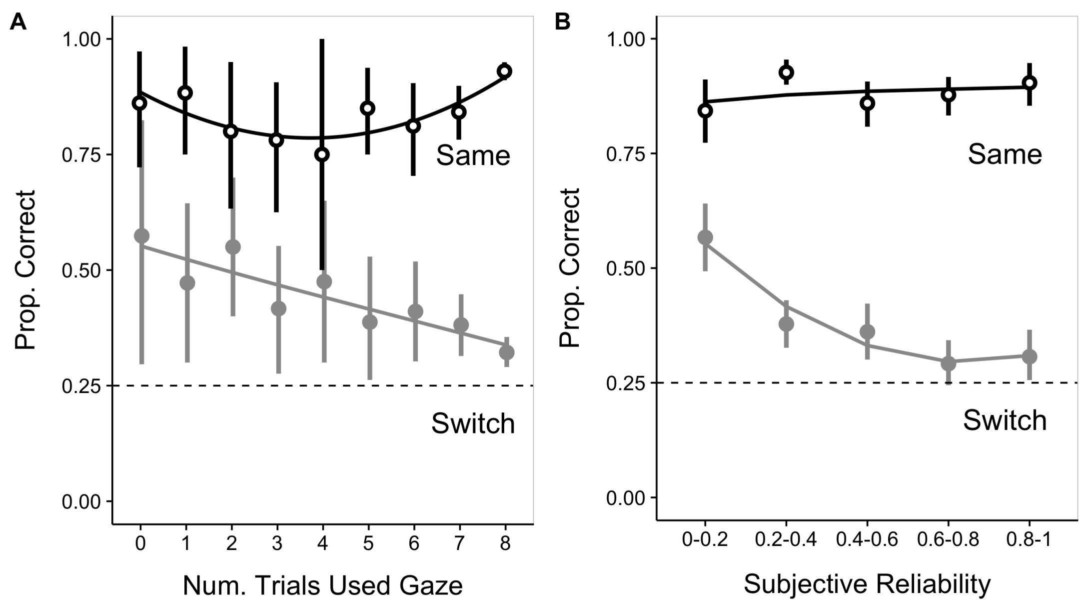


Figure 4.5: Secondary analyses of test trial performance in Experiment 3. Panel A shows accuracy as a function of the number of exposure trials on which participants chose to use the gaze cue. Panel B shows accuracy as a function of participants' subjective reliability judgments. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

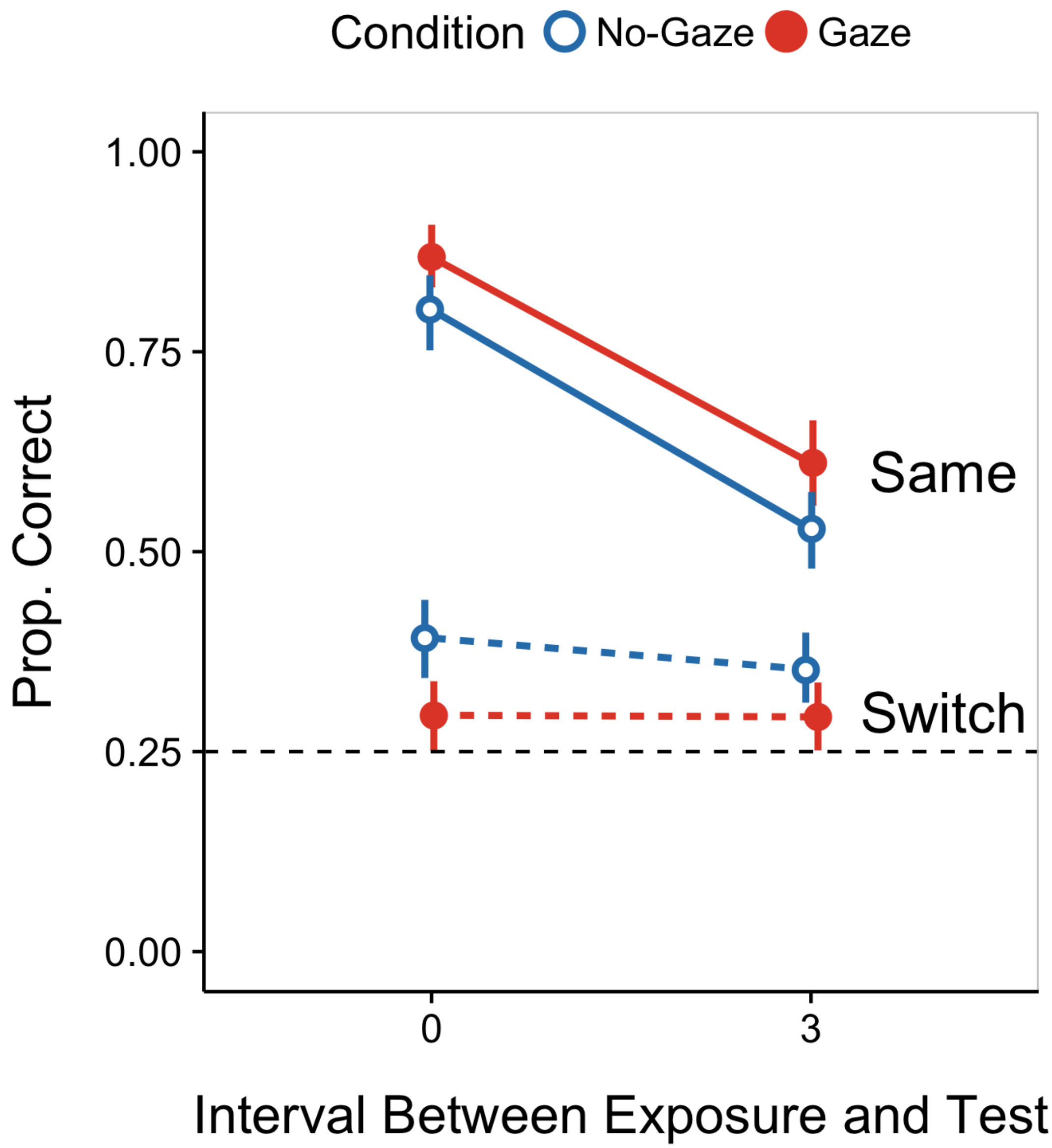


Figure 4.6: Experiment 4.4 results. Accuracy on test trials in Experiment 4 collapsed across the Long and Short inspection time conditions. The dashed line represents chance performance. Color and line type indicate whether there was gaze present on exposure trials. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

Chapter 5

Seeking social and statisical information during word learning

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

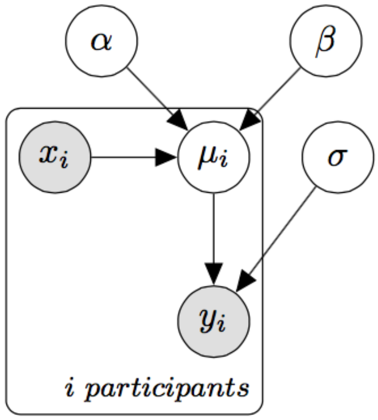
Appendix A

Supplementary materials for Chapter 2

In this appendix, we present four pieces of supplemental information. First, we provide details about the Bayesian models used to analyze the data. Second, we present a sensitivity analysis that provides evidence that the estimates of the associations between age/vocabulary and accuracy/reaction time (RT) are robust to different parameterizations of the prior distribution and different cutoffs for the analysis window. Third, we present the results of a parallel set of analyses using a non-Bayesian approach to show that these results are consistent regardless of choice of analytic framework. And fourth, we present two exploratory analyses measuring the effects of phonological overlap and iconicity on RT and accuracy. In both analyses, we did not see evidence that these factors changed the dynamics of eye movements during ASL processing

A.1 Model Specifications

Our key analyses use Bayesian linear models to test our hypotheses of interest and to estimate the associations between age/vocabulary and RT/accuracy. Figure S1 (Accuracy) and S2 (RT) present graphical models that represent all of the data, parameters, and other variables of interest, and their dependencies. Latent parameters are shown as unshaded nodes while observed parameters and data are shown as shaded nodes. All models were fit using JAGS software (Plummer, 2003) and adapted



$$\alpha \sim \text{Gaussian}(0,\, 10)$$
$$\beta \sim \text{TruncatedGaussian}_{+}(0,\, 1)$$
$$\sigma \sim \text{Uniform}(1/1000,\, 1000)$$
$$\mu_i \leftarrow \alpha + \beta * x_i$$
$$y_i \sim \text{Gaussian}(\mu_i,\, \sigma)$$

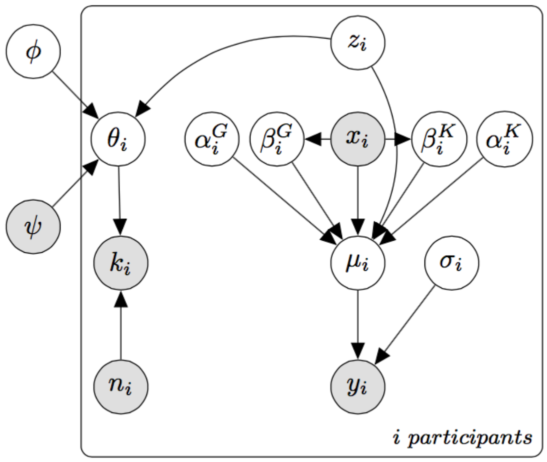
Figure A.1: Graphical model representation of the linear regression used to predict accuracy. The shaded nodes represent observed data (i.e., each participant's age, vocabulary, and mean accuracy). Unshaded nodes represent latent parameters (i.e., the intercept and slope of the linear model).

from models in Kruschke (2014) and Lee and Wagenmakers (2014).

A.1.1 Accuracy

To test the association between age/vocabulary and accuracy we assume each participant's mean accuracy is drawn from a Gaussian distribution with a mean, μ, and a standard deviation, σ. The mean is a linear function of the intercept, α, which encodes the expected value of the outcome variable when the predictor is zero, and the slope, β, which encodes the expected change in the outcome with each unit change in the predictor (i.e., the strength of association).

For α and σ, we use vague priors on a standardized scale, allowing the model to consider a wide range of plausible values. Since the slope parameter β is critical to our hypothesis of a linear association, we chose to use an informed prior: that is, a truncated Gaussian distribution with a mean of zero and a standard deviation of one on a standardized scale. Centering the distribution at zero is conservative and places the highest prior probability on a null association, to reduce the chance that our model overfits the data. Truncating the prior encodes our directional hypothesis that accuracy should increase with age and larger vocabulary size. And using a standard deviation of one constrains the plausible slope values, thus making our alternative hypothesis more precise. We constrained the slope values based on previous research with children learning spoken language



$$\alpha \sim \text{Gaussian}(0,\ 10)$$
$$\beta \sim \text{TruncatedGaussian}_{-}(0,\ 1)$$
$$\sigma_i \sim \text{Uniform}(1/1000,\ 1000)$$
$$z_i \sim \text{Bernoulli}(0.5)$$
$$\phi \sim \text{Uniform}(0.5,\ 1)$$
$$\psi \leftarrow 0.5$$
$$\theta_i \leftarrow \begin{cases} \phi & \text{if } z_i = 1 \\ \psi & \text{if } z_i = 0 \end{cases}$$
$$k_i \sim \text{Binomial}(\theta_i,\ n_i)$$
$$\mu_i \leftarrow \begin{cases} \alpha_i^K + \beta_i^K * x_i & \text{if } z_i = 1 \\ \alpha_i^G + \beta_i^G * x_i & \text{if } z_i = 0 \end{cases}$$
$$y_i \sim \text{Gaussian}(\mu_i,\ \sigma_i)$$

Figure A.2: Graphical model representation of the linear regression plus latent mixture model (i.e., guessing model). The model assumes that each individual participant's first shift is either the result of guessing or knowledge. And the latent indicator z_i determines whether that participant is included in the linear regression estimating the association between age/vocabulary and RT.

showing that the average gain in accuracy for one month of development between 18-24 months to be ~1.5% (Fernald, Zangl, Portillo, & Marchman, 2008).

A.1.2 Reaction Time

The use of RT as a processing measure is based on the assumption that the timing of a child's first shift reflects the speed of their incremental language comprehension. Yet, some children have a first shift that seems to be unassociated with this construct: their first shift behavior appears random. We quantify this possibility for each participant explicitly (i.e., the probability that the participant is a "guesser") and we create an analysis model where participants who were more likely to be guessers have less of an influence on the estimated relations between RT and age/vocabulary.

To quantify each participant's probability of guessing, we computed the proportion of signer-to-target (correct) and signer-to-distracter (incorrect) shifts for each child. We then used a latent mixture model in which we assumed that the observed data, k_i, were generated by two processes (guessing and knowledge) that had different overall probabilities of success, with the "guessing group" having a probability of 50%, ÏĹ, and the "knowledge" group having a probability greater than 50%, ϕ. The group membership of each participant is a latent indicator variable, z_i, inferred based on that participant's proportion of correct signer-to-target shifts relative to the overall proportion of correct shifts across all participants (see Lee & Wagenmakers (2014) for a detailed discussion of this modeling approach). We then used each participant's inferred group membership to determine whether they were included in the linear regression. In sum, the model allows participants to contribute to the estimated associations between age/vocabulary and RT proportional to our belief that they were guessing.

As in the Accuracy model, we use vague priors for α and σ on a standardized scale. We again use an informed prior for β, making our alternative hypothesis more precise. That is, we constrained the plausible slope values based on previous research with children learning spoken language showing that the average gain in RT for one month of development between 18-24 months to be ˜30 ms (Fernald, Zangl, Portillo, & Marchman, 2008).

A.2 Sensitivity Analysis: Prior Distribution and Window Selection

We conducted a sensitivity analysis to show that our parameter estimates for the associations between accuracy/RT and age/vocabulary are robust to decisions about (a) the analysis window and (b) the specification of the prior distribution on the slope parameter. Specifically, we varied the parameterization of the standard deviation on the slope, allowing the model to consider a wider or narrower range of values to be plausible a priori. We also fit these different models to two additional analysis windows +/- 300 ms from the final analysis window: 600-2500 ms (the middle 90% of the RT distribution in our experiment). Figure S3 shows the results of the sensitivity analysis, plotting the coefficient for the β parameter in each model for the three different analysis windows for each specification of the prior. All models show similar coefficient values, suggesting that inferences about the parameters are not sensitive to the exact form of the priors. Table S1 shows the Bayes Factors for all models across three analysis windows and fit using four different vales for the slope prior. The Bayes Factor only drops below 3 when the prior distribution is quite broad (standard deviation of 3.2) and only for the longest analysis window (600-2800 ms). In sum, the strength of evidence for a

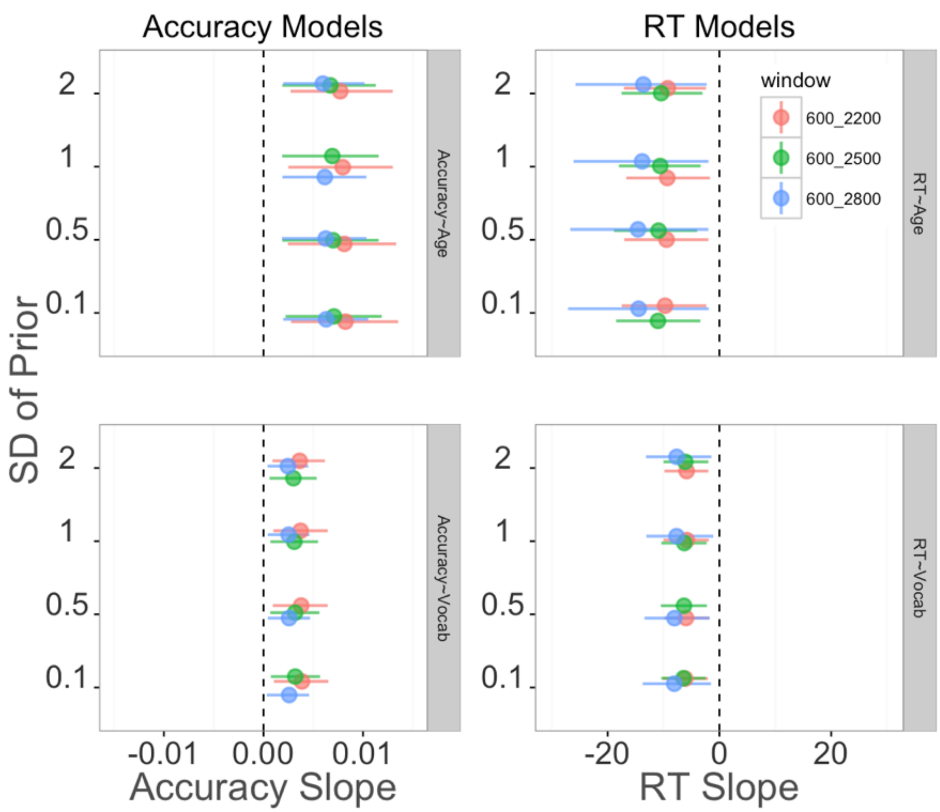


Figure A.3: Coefficient plot for the slope parameter for four different parameterizations of the prior and for three different analysis windows. Each panel shows a different model. Each point represents a Îš coefficient measuring the strength of association between the two variables. Error bars are 95% HDIs around the coefficient. Color represents the three different analysis windows.

linear association is robust to the choice of analysis window and prior specification.

A.3 Parallel set of non-Bayesian analyses

First, we compare Accuracy and RT of native hearing and deaf signers using a Welch Two Sample t-test and do not find evidence that these groups are different (Accuracy: $t(28) = 0.75$, $p = 0.45$, 95% CI on the difference in means [-0.07, 0.14]; RT: $t(28) = 0.75$, $p = 0.46$, 95% CI on the difference in means [-125.47 ms, 264.99 ms].

Second, we test whether children and adults tend to generate saccades away from the central signer prior to the offset of the target sign. To do this, we use a One Sample t-test with a null

Table A.1: Bayes Factors for all four linear models fit to three different analysis windows using four different parameterizations of the prior distribution for the slope parameter.

Analysis window	**SD Slope**	**Acc˜Age**	**Acc˜Vocab**	**RT˜Age**	**RT˜Vocab**
600 âĂŞ 2200 ms	3.2	6.2	3.7	2.4	4.1
NA	1.4	14.1	5.5	3.5	8.6
NA	1.0	19.4	8.9	5.0	9.2
NA	0.7	22.7	11.6	7.8	17.0
600 âĂŞ 2500 ms	3.2	11.0	2.3	5.6	6.1
NA	1.4	9.7	4.0	13.8	10.5
NA	1.0	12.8	6.8	12.5	18.2
NA	0.7	15.6	6.8	17.9	20.7
600 âĂŞ 2800 ms	3.2	6.0	1.1	1.2	1.4
NA	1.4	10.7	2.6	3.5	4.7
NA	1.0	13.5	4.0	3.7	4.0
NA	0.7	15.2	4.6	5.5	5.6

hypothesis that the true mean is not equal to 1, and we find evidence against this null (Children: M = 0.88, t(28) = -2.92, p = 0.007, 95% CI [0.79, 0.96]; Adults: M = 0.51, t(15) = -6.87, p < 0.001, 95% CI [0.35, 0.65])

Third, we fit the four linear models using MLE to estimate the relations between the processing measures on the VLP task (Accuracy/RT) and age/vocabulary. We follow recommendations from Barr (2008) and use a logistic transform to convert the proportion accuracy scores to a scale more suitable for the linear model. Table XXX shows the results.

A.4 Analyses of phonological overlap and iconicity

First, we analyzed whether phonological overlap of our item-pairs might have influenced adults and children's RTs and accuracy. Signs that are higher in phonological overlap might have been more difficult to process because they are more confusable. Here, phonological overlap is quantified as the number of features (e.g., Selected Fingers, Major Location, Movement, Sign Type) that both signs shared. Values were taken from a recently created database (ASL-LEX) of lexical and phonological properties of nearly 1,000 signs of American Sign Language (Caselli et al., 2017). Our item-pairs varied in degree of overlap from 1-4 features. We did not see evidence that degree of phonological overlap influenced either processing measure in the VLP task. Next, we performed

Table A.2: Results for the four linear models fit using Maxiumum Likelihood Estimation. All p-values are one-sided to reflect our directional hypotheses about the VLP measures improving over development.

Model specification	**Mean Beta value**	**std. error**	**t-statistic**	**p-value**
logit(accuracy) ~ age + hearing status	0.003	0.012	2.59	0.008
logit(accuracy) ~ vocabulary + hearing status	0.002	0.006	2.27	0.015
RT ~ age + hearing status	-10.050	4.620	-2.17	0.019
RT ~ vocabulary + hearing status	-6.340	2.180	-2.91	0.003

a parallel analysis, exploring whether the iconicity of our signs might have influenced adults and children's RT and accuracy. It is possible that highly iconic signs might be easier to process because of the visual similarity to the target object. Again, we used ASL-LEX to quantify the iconicity of our signs. To generate these values, native signers were asked to explicitly rate the iconicity of each sign on a scale of 1-7, with 1 being not iconic at all and 7 being very iconic. Similar to the phonological overlap analysis, we did see evidence that degree of iconicity influenced either processing measure for either age group in the VLP task.

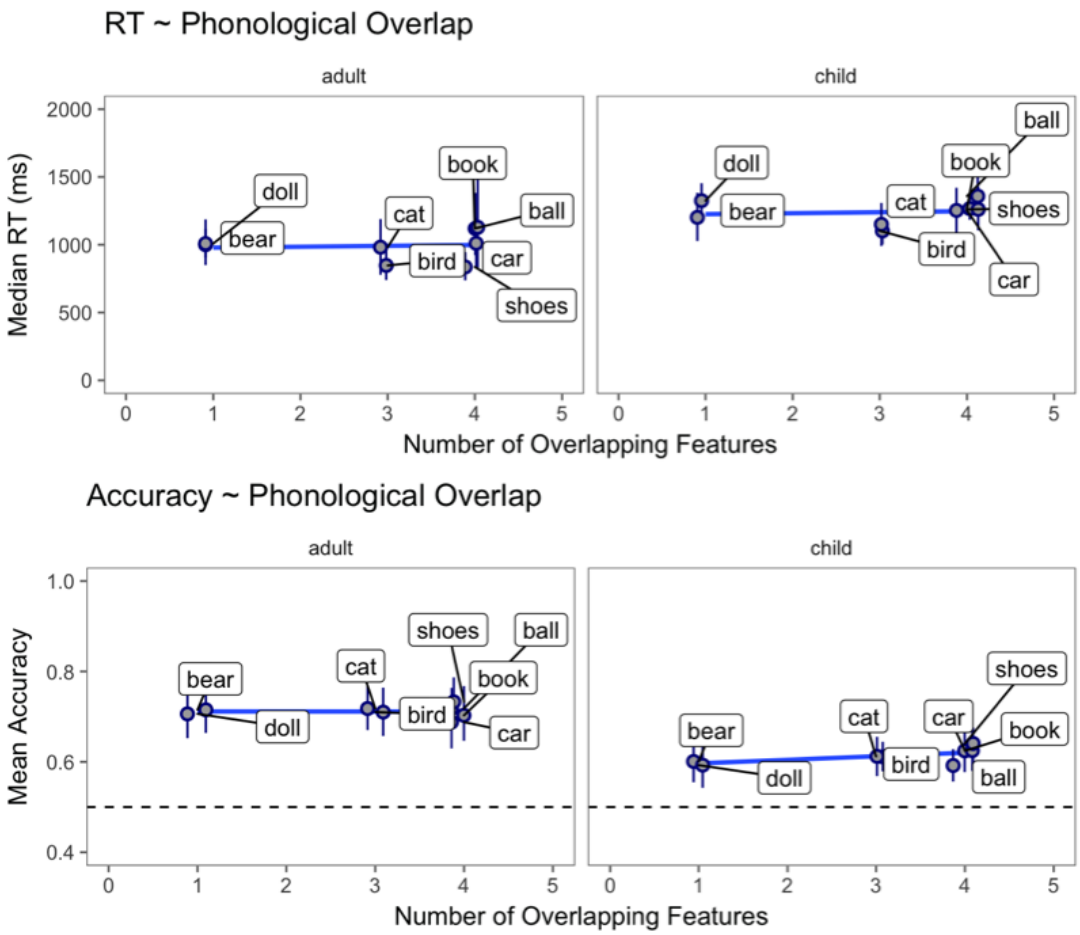


Figure A.4: Scatterplot of the association between degree of phonological overlap and RT (top row) and accuracy (bottom row) for both adults (left column) and children (right column). The blue line represents a linear model fit.

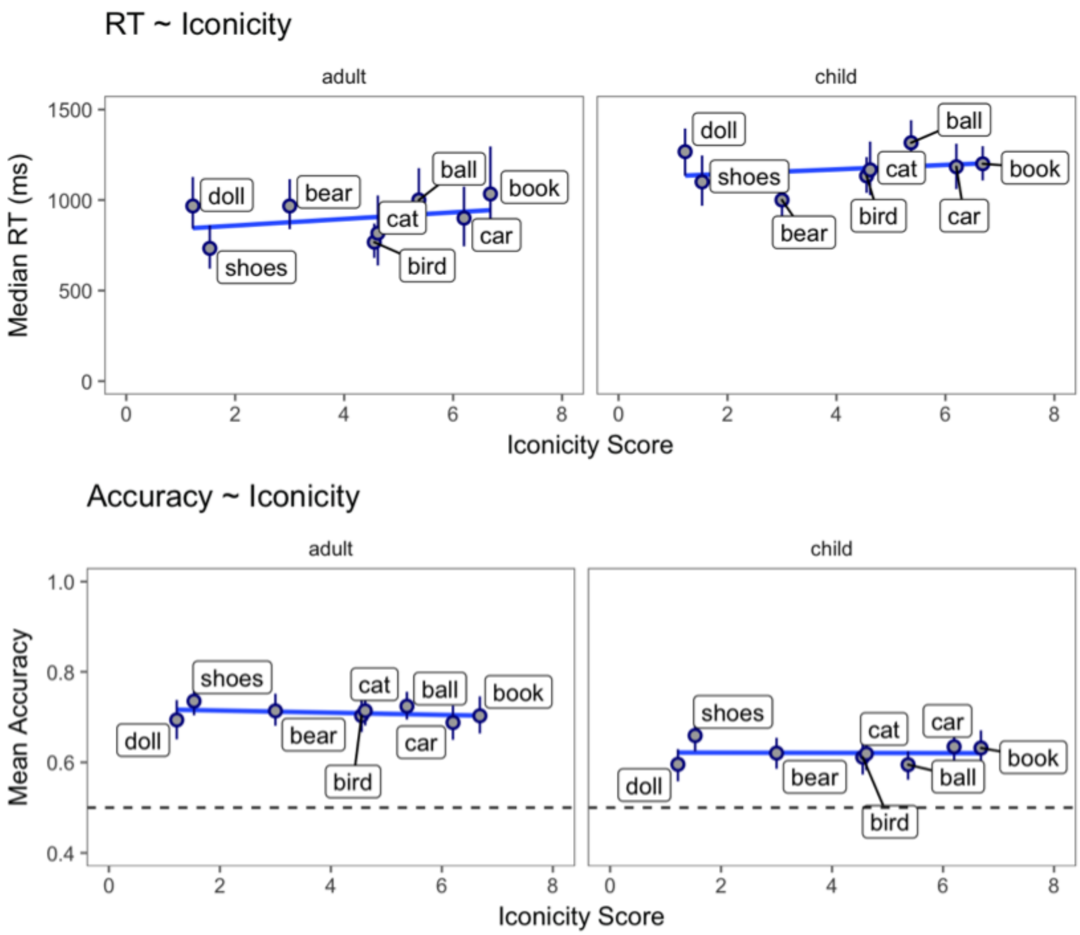


Figure A.5: Scatterplot of the association between degree of iconicity and RT (top row) and accuracy (bottom row) for both adults (left column) and children (right column). The blue line represents a linear model fit.

Appendix B

Supplementary materials for Chapter 4

Table A1. Length of inspection times on exposure trials in Experiment 1 as a function of gaze, interval, and number of referents

`Log(Inspection time)` $\sim$ `(Gaze + Log(Interval) + Log(Referents))`^{2} `+ (1 | subject)`

term	estimate	std.error	t.value	p.value	
Intercept	0.83	0.10	8.19	< .001	***
Gaze Condition	0.16	0.11	1.48	0.138	
Log(Interval)	0.06	0.05	1.33	0.184	
Log(Referents)	0.34	0.04	7.91	< .001	***
Gaze Condition*Log(Interval)	-0.08	0.03	-2.86	0.004	**
Gaze Condition*Log(Referent)	-0.27	0.04	-6.01	< .001	***
Log(Interval)*Log(Referent)	-0.00	0.02	-0.19	0.849	

Table A2. Accuracy on test trials in Experiment 1 with inspection times on exposure trials included as a predictor

`Correct` $\sim$ `(Trial Type + Gaze + Log(Interval) + Log(Referents) +`

`Log(Inspection Time))`^{2} `+ offset(logit(`$^1/_{Referents}$`)) + (TrialType | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	2.89	0.34	8.49	< .001	***
Switch Trial	-1.45	0.25	-5.76	< .001	***
Gaze Condition	0.12	0.27	0.43	0.669	
Log(Interval)	-0.47	0.11	-4.15	< .001	***
Log(Referents)	0.05	0.14	0.39	0.693	
Log(Inspection Time)	0.20	0.15	1.38	0.169	
Switch Trial*Gaze Condition	-1.02	0.13	-7.86	< .001	***
Switch Trial*Log(Interval)	0.52	0.06	9.39	< .001	***
Switch Trial*Log(Referent)	-0.62	0.09	-6.67	< .001	***
Switch Trial*Log(Inspection Time)	0.09	0.07	1.36	0.174	
Gaze Condition*Log(Interval)	0.09	0.06	1.61	0.107	
Gaze Condition*Log(Referent)	0.36	0.10	3.68	< .001	***
Gaze Condition*Log(Inspection Time)	-0.17	0.07	-2.55	0.011	*
Log(Interval)*Log(Referent)	-0.05	0.04	-1.26	0.207	
Log(Interval)*Log(Inspection Time)	0.02	0.03	0.54	0.589	
Log(Referents)*Log(Inspection Time)	0.05	0.05	0.94	0.345	

Table A3. Length of inspection times on exposure trials in Experiment 2 as a function of gaze and interval

`Log(Inspection time) ~ Gaze * Log(Interval) + (1 | subject)`

term	estimate	std.error	t.value	p.value	
Intercept	3.90	0.08	50.69	$< .001$	***
Gaze Condition	-1.10	0.05	-20.90	$< .001$	***
Log(Interval)	-0.48	0.05	-8.77	$< .001$	***
Gaze Condition*Log(Interval)	-0.02	0.04	-0.60	0.549	

Table A4. Accuracy on test trials in Experiment 2 with inspection times on exposure trials included as a predictor

`Correct ~ (Trial Type + Gaze + Log(Interval) + Log(Inspection Time))`^{2} `+`

`offset(logit(`$^1/_{Referents}$`)) + (TrialType | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	3.51	0.29	12.13	$< .001$	***
Gaze Condition	0.13	0.23	0.58	0.559	
Switch Trial	-3.12	0.26	-12.21	$< .001$	***
Log(Interval)	-0.88	0.14	-6.34	$< .001$	***
Log(Inspection Time)	0.15	0.13	1.14	0.255	
Switch Trial*Gaze Condition	-0.54	0.17	-3.21	0.001	**
Gaze Condition*Log(Interval)	0.16	0.09	1.85	0.064	.
Gaze Condition*Log(Inspection Time)	-0.14	0.10	-1.37	0.172	
Switch Trial*Log(Interval)	0.77	0.10	8.00	$< .001$	***
Switch Trial*Log(Inspection Time)	0.21	0.11	1.96	0.05	.
Log(Interval)*Log(Inspection Time)	0.04	0.06	0.77	0.44	

Table A5. Accuracy on exposure trials in Experiment 3 as a function of reliability condition and participants' subjective reliability judgments

`Correct-Exposure` $\sim$ `Reliability Condition * Subjective Reliability +`
`offset(logit(`$^1/_{Referents}$`)) + (1 | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	3.07	0.98	3.13	0.002	**
Reliability Condition	3.28	1.50	2.19	0.029	*
Subjective Reliability	7.26	1.73	4.21	$< .001$	***
Reliability Condition*Subjective Reliability	-4.58	2.72	-1.68	0.093	.

Table A6. Accuracy on test trials in Experiment 3 as a function of reliability condition

`Correct` $\sim$ `Trial Type * Reliability Condition + offset(logit(`$^1/_{Referents}$`)) +`
`(Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	4.70	0.36	13.10	$< .001$	***
Trial Type	-3.95	0.36	-10.92	$< .001$	***
Reliability Condition	0.38	0.37	1.03	0.302	
Reliability Condition*Trial Type	-0.76	0.38	-2.01	0.044	*

Table A7. Accuracy on test trials in Experiment 3 as a function of reliability condition and participants' use of gaze on exposure trials

`Correct` $\sim$ `(Trial Type + Reliability Condition + Correct-Exposure)`^{2}

`+ offset(logit(`$^1/_{Referents}$`)) + (Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	4.50	0.39	11.59	$< .001$	***
Correct Exposure	0.07	0.29	0.26	0.796	
Trial Type	-2.70	0.38	-7.07	$< .001$	***
Reliability Condition	-0.43	0.44	-0.98	0.325	
Correct Exposure*Trial Type	-1.43	0.26	-5.41	$< .001$	***
Correct Exposure*Reliability	0.97	0.33	2.92	0.004	**
Reliability Condition*Trial Type	-0.62	0.36	-1.72	0.086	.

Table A8. Accuracy on test trials in Experiment 3 as a function of each participants' accuracy on exposure trials

`Correct` $\sim$ `Trial Type * Total Correct Exposure + offset(logit(`$^1/_{Referents}$`)) +`

`(Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	2.73	0.39	7.01	$< .001$	***
Total Exposure Correct	0.14	0.06	2.49	0.013	*
Trial Type	-1.39	0.39	-3.55	$< .001$	***
Total Exposure Correct*Trial Type	-0.26	0.06	-4.66	$< .001$	***

Table A9. Accuracy on test trials in Experiment 3 as a function of each participants' subjective reliability judgment

`Correct` $\sim$ `Trial Type * Subjective Reliability + offset(logit(`$^1/_{Referents}$`)) + (Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	4.54	0.44	10.33	$< .001$	***
Subjective Reliability	0.40	0.58	0.69	0.493	
Trial Type	-3.44	0.44	-7.81	$< .001$	***
Subjective Reliability*Trial Type	-1.63	0.59	-2.78	0.005	**

Table A10. Accuracy on test trials in Experiment 3 as a function of reliability condition and inspection time on exposure trials

`Correct` $\sim$ `(Trial Type + Reliability condition + Trial Type + Log(Inspection Time))`^{2} `+ offset(logit(`$^1/_{Referents}$`)) + (Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	3.11	0.20	15.94	$< .001$	***
Log(Inspection Time)	0.31	0.09	3.31	0.001	**
Trial Type	-2.75	0.20	-13.64	$< .001$	***
Reliability Condition	0.50	0.30	1.66	0.097	.
Log(Inspection Time)*Trial Type	0.03	0.09	0.34	0.736	
Log(Inspection Time)*Reliability Condition	-0.20	0.11	-1.83	0.067	.
Trial Type*Reliability Condition	-0.58	0.29	-1.97	0.048	*

Table A11. Accuracy on test trials in Experiment 4 as a function of gaze and interval

`Correct` $\sim$ `(Trial Type + Gaze + Log(Interval))`^{2} `+ offset(logit(`$^1/_{Referents}$`)) +`
`(Trial Type | subject)`

term	estimate	std.error	z.value	p.value	
Intercept	3.37	0.16	21.32	$< .001$	***
Trial Type	-3.18	0.16	-19.93	$< .001$	***
Gaze Condition	-0.48	0.14	-3.52	$< .001$	***
Log(Interval)	-0.84	0.10	-8.59	$< .001$	***
Trial Type*Gaze Condition	0.90	0.14	6.63	$< .001$	***
Trial Type*Log(Interval)	0.80	0.09	8.71	$< .001$	***
Gaze Condition*Log(Interval)	-0.01	0.07	-0.10	0.917	

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(02), 395–418.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). Lme4: Linear mixed-effects models using eigen and s4. *R Package Version*, *1*(4).

Begus, K., Gliga, T., & Southgate, V. (2014). Infants learn what they want to learn: Responding to infant pointing leads to superior learning.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

Berlyne, D. E. (1960). Conflict, arousal, and curiosity.

Birbili, M., & Karagiorgou, I. (2009). Helping children and their parents ask better questions: An intervention study. *Journal of Research in Childhood Education*, *24*(1), 18–31.

Bloom, P. (2002). *How children learn the meaning of words.* The MIT Press.

Bonawitz, E. B., Schijndel, T. J. van, Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234.

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

Bonawitz, E., & Shafto, P. (2016). Computational models of development, social influences. *Current Opinion in Behavioral Sciences*, *7*, 95–100.

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*(3), 322–330.

Bova, A., & Arcidiacono, F. (2013). Investigating children's why-questions: A study comparing argumentative and explanatory function. *Discourse Studies*, *15*(6), 713–734.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(Supplement 2), 10918–10925.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(01), 207–220.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*.

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

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416–1428.

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (pan troglodytes) and human children (homo sapiens). *Animal*

Cognition, *8*(3), 151–163.

Calvert, S. L., Strong, B. L., & Gallagher, L. (2005). Control as an engagement feature for young children's attention to and learning of computer content. *American Behavioral Scientist*, *48*(5), 578–589.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998a). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 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*(5), 1431–1441.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998b). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, i–174.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Casillas, M. (2014). Turn-taking. *Pragmatic Development in First Language Acquisition*, 53–70.

Castro, R. M., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2009). Human active learning. In *Advances in neural information processing systems* (pp. 241–248).

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.

Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, i–129.

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science*, *11*(5), 761–770.

Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science*, *18*(4), 314–316.

Clark, E. V. (2009). *First language acquisition.* Cambridge University Press.

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5-and 7-month-old infants. *Infant and Child Development*, *16*(3), 295–306.

Coenen, A., Nelson, J. D., & Gureckis, T. (2017). Asking the right questions about human inquiry.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595.

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*(3), 426–437.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, *9*(3), 245.

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

Davis, E. A. (1932). The form and function of children's questions. *Child Development*, *3*(1), 57–74.

De Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129–134.

Deborah, G. K. N., Louisa Chan, E., & Holt, M. B. (2004). When children ask,"What is it?" What do they want to know about artifacts? *Psychological Science*, *15*(6), 384–389.

DeCasper, A. J., Fifer, W. P., Oates, J., & Sheldon, S. (1987). Of human bonding: Newborns prefer their mothers' voices. *Cognitive Development in Infancy*, 111–118.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256.

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, *123*(6), 758.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5.

Emery, A., & Nenarokomov, A. V. (1998). Optimal experiment design. *Measurement Science and Technology*, *9*(6), 864.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605.

Farroni, T., Massaccesi, S., Menon, E., & Johnson, M. H. (2007). Direct gaze modulates face recognition in young infants. *Cognition*, *102*(3), 396–404.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104.

Fitneva, S. A., Lam, N. H., & Dunfield, K. A. (2013). The development of children's information gathering: To look or to ask? *Developmental Psychology*, *49*(3), 533.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24.

Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult–child conversation. *Child Development*, *80*(6), 1592–1611.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319.

Geisler, W. S. (2003). Ideal observer analysis. *The Visual Neurosciences*, *10*(7), 12–12.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, *60*, 115–140.

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, *4*(1), 1–31.

Gergely, G., Bekkering, H., & KirÃąly, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature*, *415*(6873), 755–755.

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

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid 'labouring in vain'by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–979.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, *10*(6), 778–785.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Goodman, N. D., Baker, C. L., & Tenenbaum, J. B. (2009). Cause and intent: Social reasoning in causal learning. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2759–2764).

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn.* William Morrow & Co.

Grabinger, R. S., & Dunlap, J. C. (1995). Rich environments for active learning: A definition. *Research in Learning Technology*, *3*(2).

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

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*(5), 1526–1541.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning a cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*(3), 335–341.

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*(6037), 1524–1524.

Haertel, R. A., Seppi, K. D., Ringger, E. K., & Carroll, J. L. (2008). Return on investment for active learning. In *Proceedings of the nips workshop on cost-sensitive learning* (Vol. 72).

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015).

Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., . . . Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naÃŕve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 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.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*(1), 200–213.

Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, *9*(1), 133–142.

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*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Kim, S., Paulus, M., Sodian, B., & Proust, J. (2016). Young children's sensitivity to their own ignorance in informing others. *PloS One*, *11*(3), e0152595.

Kishimoto, T., Shizawa, Y., Yasuda, J., Hinobayashi, T., & Minami, T. (2007). Do pointing gestures by infants provoke comments from adults? *Infant Behavior and Development*, *30*(4), 562–567.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, *38*.

Koehne, J., & Crocker, M. W. (2014). The interplay of cross-situational word learning and sentence-level constraints. *Cognitive Science*.

Koenig, M. A., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science*, *4*(4), 731–739.

Kuhl, P. K. (2007). Is speech learning 'gated'by the social brain? *Developmental Science*, *10*(1), 110–120.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, *26*(4), 512–559.

Legare, C. H., Mills, C. M., Souza, A. L., Plummer, L. E., & Yasskin, R. (2013). The use of questions as problem-solving strategies during early childhood. *Journal of Experimental Child Psychology*, *114*(1), 63–76.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. `http://doi.org/10.18637/jss.v069.i01`

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 986–1005.

Lockhart, K. L., Goddu, M. K., Smith, E. D., & Keil, F. C. (2016). What could you really learn on your own?: Understanding the epistemic limitations of knowledge acquisition. *Child Development*, *87*(2), 477–493.

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

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

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). Springer.

MacDonald, K., Blonder, A., Marchman, V., Fernald, A., & Frank, M. C. (2017a). An information-seeking account of eye movements during spoken and signed language comprehension. In *Proceedings of the 39th annual conference of the cognitive science society.*

MacDonald, K., LaMarr, T., Corina, D., Marchman, V. A., & Fernald, A. (2018a). Real-time lexical comprehension in young children learning american sign language. *Developmental Science*, e12672.

MacDonald, K., Marchman, V., Fernald, A., & Frank, M. C. (2018b). Adults and preschoolers seek visual information to support language comprehension in noisy environments. In *Proceedings of the 40th annual conference of the cognitive science society.*

MacDonald, K., Schug, M., Chase, E., & Barth, H. (2013). My people, right or wrong? Minimal group membership disrupts preschoolers' selective trust. *Cognitive Development*, *28*(3), 247–259.

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017b). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67–84.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

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*(3), 142–152.

Markant, D., DuBrow, S., Davachi, L., & Gureckis, T. M. (2014). Deconstructing the effect of self-directed study on episodic memory. *Memory & Cognition*, *42*(8), 1211–1224.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 643–655.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178.

McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22.

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, *46*(3), 229–259.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831.

Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making*, *7*(2), 119.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014–9019.

Mills, C. M., Danovitch, J. H., Grant, M. G., & Elashi, F. B. (2012). Little pitchers use their big ears: Preschoolers solve problems by listening to others ask questions. *Child Development*, *83*(2), 568–580.

Mills, C. M., Legare, C. H., Bills, M., & Mejias, C. (2010). Preschoolers use questions as a tool to acquire knowledge from different sources. *Journal of Cognition and Development*, *11*(4), 533–560.

Mills, C. M., Legare, C. H., Grant, M. G., & Landrum, A. R. (2011). Determining who to question, what to ask, and how much information to ask for: The development of inquiry in young children. *Journal of Experimental Child Psychology*, *110*(4), 539–560.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387.

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using "sticky mittens" enhances young infants' object exploration skills. *Infant Behavior and Development*, *25*(3), 279–295.

Nelson, J. D. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4).

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science*, *21*(7), 960–969.

Oakes, L. M. (2011). *Infant perception and cognition: Recent advances, emerging theories, and future directions*. Oxford University Press, USA.

Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2008). Mere belief in social action improves complex learning. In *Proceedings of the 8th international conference on international conference for the learning sciences-volume 2* (pp. 132–139). International Society of the Learning Sciences.

Olson, J., & Masur, E. F. (2011). Infants' gestures influence mothers' provision of object, action and internal state labels. *Journal of Child Language*, *38*(5), 1028–1054.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*(4), 301–332.

Ouyang, L., Tessler, M. H., Ly, D., & Goodman, N. (2016). Practical optimal experiment design with probabilistic programs. *arXiv Preprint arXiv:1608.05046*.

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2015). Young children's self-directed information gathering on touchscreens. In *Proceedings of the 37th annual conference of the cognitive science society, austin, tx. Cognitive science society.*

Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two-and three-year olds. *Journal of Child Language*, *9*(3), 597–626.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, *15*(3), 325–345.

Peterson, C. C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language*, *15*(1), 123–145.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231.

Quine, W. V. (1960). 0. Word and object. *111e MIT Press*.

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*(4), 785–798.

Ramirez-Loaiza, M. E., Sharma, M., Kumar, G., & Bilgic, M. (2017). Active learning: An empirical study of common baselines. *Data Mining and Knowledge Discovery*, *31*(2), 287–313.

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *European Journal of Neuroscience*, *21*(6), 1763–1766.

Rogoff, B., Mistry, J., GÃűncÃij, A., Mosier, C., Chavajay, P., & Heath, S. B. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*, i–179.

Rohde, H., & Frank, M. C. (2014). Markers of topical discourse in child-directed speech. *Cognitive Science*, *38*(8), 1634–1661.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956–970.

Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development*, *74*(6), 1807–1822.

Rothe, A., Lake, B. M., & Gureckis, T. M. (2015). Asking useful questions: Active learning with rich queries. In *CogSci*.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning.

Developmental Science, *12*(1), 182–187.

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science*, *41*(S1), 162–179.

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.

Ruggeri, A., Markant, D. B., Gureckis, T. M., & Xu, F. (2016). Active control of study leads to improved recognition memory in children. In *Proceedings of the 38th annual conference of the cognitive science society. Austin, tx: Cognitive science society.*

Sage, K. D., & Baldwin, D. (2011). Disentangling the social and the pedagogical in infants' learning about tool-use. *Social Development*, *20*(4), 825–844.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389.

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

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *6*(1), 1–114.

Settles, B., Craven, M., & Friedland, L. (2008). Active learning with real annotation costs. In *Proceedings of the nips workshop on cost-sensitive learning* (pp. 1–10).

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012a). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012b). Learning from others the consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89.

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654–666.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1), 39–91.

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25–49.

Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition*, *34*(2), 411–419.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454–1463.

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R.

(2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117–135.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Uziel, L. (2007). Individual differences in the social facilitation effect: A review and meta-analysis. *Journal of Research in Personality*, *41*(3), 579–601.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Vredenburgh, C., & Kushnir, T. (2016). Young children's help-seeking as active information gathering. *Cognitive Science*, *40*(3), 697–722.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Vygotsky, L. (1987). Zone of proximal development. *Mind in Society: The Development of Higher Psychological Processes*, *5291*, 157.

Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided play: Where curricular goals meet a playful pedagogy. *Mind, Brain, and Education*, *7*(2), 104–112.

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development*, *12*(3), 252–261.

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, *47*(5), 1220.

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136.

Wu, Z., & Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language*, *42*(3), 538–561.

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in bayesian word learning. *Developmental Science*, *10*(3), 288–297.

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as bayesian inference. *Psychological Review*, *114*(2), 245.

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, *105*(36), 13690–13695.

Yoon, T., Erica J, & Frank, M. C. (2017). "I won't lie, it wasn't amazing": Modeling polite indirect speech. In *CogSci*.

Yu, B., Yue, & Shafto, P. (2017). Pedagogical questions in parent-child conversations. *Child Development*.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, *29*(6), 961–1005.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Yu, C., & Smith, L. B. (2012a). Embodied attention and word learning by toddlers. *Cognition*.

Yu, C., & Smith, L. B. (2012b). Modeling cross-situational word–referent learning: Prior questions.

Psychological Review, *119*(1), 21.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One*, *8*(11), e79659.

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966.