CHILDREN'S INFORMATION SEEKING DURING SIGNED AND SPOKEN LANGUAGE PROCESSING

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

How do children comprehend and learn language despite noisy input and limits on their information processing capabilities? Social learning theories argue for the importance of acquiring language from more knowledgeable adults who can constrain the learning task. Statistical learning accounts emphasize the role of children's pattern detection abilities that can take advantage of the structure available in the input. Finally, active learning explanations focus on children's capacity to gather information to support their learning. This thesis presents an integrative explanation that brings together ideas from these theoretical accounts to investigate how children's information seeking adapts to learning within social contexts. I use the formalization of Optimal Experiment Design (OED) as a conceptual tool to bring together ideas from the social and active learning theories. Then, I present a series of empirical studies motivated by the integrative account that ask how children's information seeking adapts to support their language processing across a diverse contexts: signed vs. spoken language (Chapters 2 and 3), (2) speech in clear vs. noisy environments (Chapter 3), and (3) novel words with or without an accompanied social cue to reference (Chapters 4 and 5). The upshot of the empirical work is that children's early information seeking is quite flexible, providing a way to overcome ambiguity in the input by gathering useful information from communicative partners.

Dedication

In loving memory of my Grammy, Sheila Paget, who always encouraged me to ask questions.

Acknowledgments

This dissertation would not have been possible without the support of the Deaf community. I am especially grateful to the California School for the Deaf in Fremont, CA and to the families who participated in this research. A special thank you to Karina Pedersen, Lisalee Egbert, Laura Petersen, Michele Berke, Sean Virnig, Pearlene Utley, and Rosa Lee Timm.

I am also thankful for the developmental community at Stanford. This research was improved thanks to the feedback from my cohort, my labmates in the Language and Cognition Lab, and the members of the Language Learning and Social Cognition labs.

I have been fortunate to work with incredible mentors throughout graduate school. Thank you to Anne Fernald, Virginia Marchman, and Hyo Gweon for your support and for spending many hours providing feedback on the ideas in this dissertation. I am especially grateful to my advisor, Michael Frank, who has been my greatest advocate. Thank you for your patience, kindness, and for showing me how to think clearly.

I am grateful to have made some lifelong friends while completing this research. Kara Weisman, MH Tessler, and Erica Yoon, I'm so glad we went through this experience together, and your support throughout has meant the world to me.

None of what I have accomplished would be possible without the unconditional love and support of my parents, Morag and Richard MacDonald. I also thank my younger sister, Kaitlin, for always being there when I needed to someone to talk to. Finally, thank you to my love, Ellika. My days begin and end with you, and for this, I couldn't be more grateful.

Contents

List of Tables

List of Figures

Introduction

Early language acquisition seems simple. An adult produces a word about something in the surrounding context (e.g., "look at the ball") and the child connects what they hear with the round object that they see in front of them. This characterization of language learning via association belies the complexity of the processing and acquisition challenges that children face. Consider that learning the meaning of concrete nouns, children have to extract the correct units from a continuous stream of linguistic information and map them on another continuous stream of visual information. The word-to-meaning mapping task becomes even more complex when we consider that the co-occurring context does not often disambiguate a speaker's intended meaning. This point was made famous by W.V. Quine's example of a field linguist trying to select the target meaning of a new word ("gavagai") from the set of possible meanings consistent with the event of a rabbit running (e.g., "white," "rabbit," "dinner," etc.) (Quine, 1960). Remarkably, children's word learning proceeds rapidly, with estimates of adult vocabularies ranging from 50,000 to 100,000 distinct lexical concepts (P. Bloom, 2002).

The number of word meanings children will acquire is not the only impressive feature of their lexical development. As children accumulate word knowledge, they also gain the ability to access the conceptual representation linked with a word quite rapidly. Children can understand language even though adults speak at a rate of three words per second. And empirical work shows that both adults and young children shift their visual attention to a familiar object in the scene within hundreds of milliseconds upon hearing its name (Allopenna, Magnuson, & Tanenhaus, 1998; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The speed of lexical access becomes even more impressive when compared with the slower retrieval of other learned, arbitrary facts (e.g., phone numbers) (Pinker, 2003).

Together, these feaatures have made children's language comprehension and word learning fundamental topics of research in cognitive science. How is it that nearly all children growing up under normal circumstances acquire language despite noisy input? What learning mechanisms could account for the robustness and flexibility of language development?

Social learning accounts point out that the child does not have to re-invent language on their own. Instead, children are typically surrounded by parents, other knowledgeable adults, or older peers – all of whom know the target language and want to facilitate their learning (P. Bloom, 2002; E. V. Clark, 2009; Hollich et al., 2000). Statistical learning theories propose that children possess powerful pattern detection skills that can learn the consistent structure in their input and reduce ambiguity in the learning task (Roy & Pentland, 2002; Siskind, 1996; Yu & Smith, 2007). More recent theories highlight the role child as an active learner, controlling aspects of their learning via the selection of behaviors – for example, asking questions or choosing where to allocate visual attention – that change the content, pacing, and sequence of their learning experiences (Gopnik, Meltzoff, & Kuhl, 1999; L. Schulz, 2012). A common thread across these three accounts – social, statistical, and active – is that children have access to information that *constrains* their inferences about new word meanings, thus reducing the problem of indeterminacy.

Empirical work in each of these traditions has often proceeded in parallel, but, in real-world learning, these mechanisms do not operate in isolation. Thus, it becomes important to develop integrative accounts that try to explain how different sources of information might mutually influence one another during language development. But how do we integrate ideas from these proposals that often lack clear definitions and formal theory that generate testable predictions? In this thesis, I argue that answering this question is important because children learn language from interactions with other people, which provides the opportunity to integrate social information with their prior knowledge to select actions that support their language learning.

Figure 1 shows a schematic overview of the contents of this dissertation. Chapter 1 presents the details of an integrative account of how social contexts shape children's active learning. I use the computational 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 children's information seeking. The key insight is that learning in the presence of other people plays a direct role in determining the *usefulness* of different actions. Using this

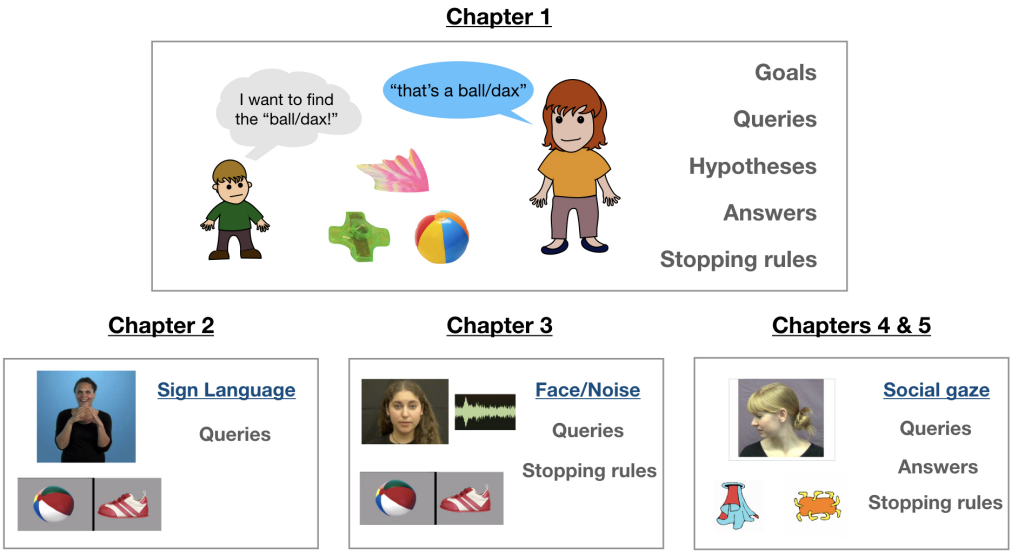


Figure 1: The upper panel shows a schematic overview of the components of an integrative framework of information seeking during language processing within a social context. The lower panels show the different case studies and pieces of the general model that correspond to each chapter of the dissertation.

framework allows us to ask how social and statistical information might selectively affect the different underlying components of children's information selection during lexical comprehension and word learning. Chapters 2-5 describe a diverse set of case studies of children and adult's eye movements during real-time familiar language comprehension and novel word learning. The empirical work focuses on eye movements as one instantiation of an active learning behavior that is particularly relevant for early lexical development. That is, we can characterize decisions about visual fixation as a type of question-asking where perceivers deploy their gaze to reduce uncertainty about the world and to maximize their expected future rewards concerning some goal (M. Hayhoe & Ballard, 2005). Moreover, eye movements are behaviors that children can control relatively early in development and map well on to the ecological task of interest: linking linguistic and visual information about concrete objects perceived via the visual channel. Overall, the goal of the empirical work is to ask how children's real-time visual information selection adapts to the information available across a wide range of processing contexts.

Chapter 1

Children's information seeking within social learning contexts

1.1 Introduction

Early cognitive development is remarkable. Consider that children, despite limitations on their processing capabilities and ambiguity in the input, rapidly acquire new lexical concepts, eventually reaching an adult vocabulary ranging from 50,000 to 100,000 words (P. Bloom, 2002). And they accomplish this all while also developing motor skills, learning social norms, and building causal knowledge. How can we explain the children's prodigious learning abilities?

Social learning accounts point out that children do not solve these problems on their own. Instead, they are typically surrounded by parents, other knowledgeable adults, or older peers – all of whom are likely to know more than they do and want to facilitate their learning. Social learning accounts also emphasize how social contexts can bootstrap children's learning via several, distinct 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 (Yu & Ballard, 2007), and increases levels of sustained attention, which results in better learning outcomes (Kuhl, 2007; Yu & Smith, 2016).

Social contexts can also change the 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 reasoning where the learner thinks about *why* other people performed specific actions. The critical insight comes from knowing that another person has intentionally selected examples, which allows children to make stronger inferences that speed learning (Bonawitz & Shafto, 2016; M. C. 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 actions that generated that evidence were accidental (less informative) or intentionally selected (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 to learn (i.e., teaching) (Shafto, Goodman, & Frank, 2012b).

However, children are not passive recipients of information – from people or the world. Instead, they actively select behaviors – for example, asking questions or choosing where to allocate visual attention – that can change the content, pacing, and sequence of their learning experience. Recent theories of cognitive development have proposed the metaphor of the "child as a scientist" and characterized early learning as a process of exploration and hypothesis testing following principles of the scientific method (Gopnik et al., 1999; L. Schulz, 2012). Moreover, 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 self-directed contexts (active learning) as compared to settings where the learner has less control (passive learning). The upshot of this work is that active contexts often lead to faster learning because of enhanced attention and arousal or because learners select information that is closely linked to their current goals, uncertainty, and skill (see Gureckis & Markant (2012) for a review).

Thus, children are capable of selecting actions to support their learning, and a large body of work shows that social partners can create particularly rich contexts for learning. Empirical work within each framework, however, has often focused on the separate effects of active or social processes on children's learning. An approach that does not reflect the ecological context in which language acquisition unfolds. That is, children acquire their first language from interactions with other people where they have the opportunity to select actions – where to look, where to point, what question to ask – that influence the content and pacing of the experience. This mismatch between the existing

empirical work and the ecological context in which children learn highlights the need for integration across the active and social theoretical accounts.

But how can we integrate these two proposals, which often include definitional issues and a lack of formal foundations? In this chapter, I present an integrated account of active learning within social contexts, arguing that it represents an important step for theories of early cognitive development. To connect the active and social learning proposals, I use the computational frameworks of Optimal Experiment Design (Emery & Nenarokomov, 1998; Lindley, 1956) and Bayesian models of social reasoning (Goodman & Frank, 2016). The key insight is that the presence of another person can modulate the *availability* and *usefulness* of different actions that the active learner could select. Specifically, social interactions can shape learners' goals, hypotheses, actions, answers, and decisions about when to stop gathering information.

In addition to the theoretical account, this chapter provides concrete definitions of active and social learning to clarify the behaviors and phenomena that the account aims to address (Part I). Then, we briefly review the framework of Optimal Experiment Design (OED) (Part II), highlighting how it can be can be used to formalize human information seeking behaviors. We also discuss the empirical evidence for OED-like reasoning in both adults' and children's learning. Finally, we conclude by presenting the integrative account of how social contexts can shape information seeking (Part III). In the final section, we also highlight a series of new links between the social and active accounts, which present a way forward for empirical work that sheds light on how children's active learning operates over fundamentally social input.

1.2 Part I: Learning from others and learning on your own

A diverse set of scholars, theories, and empirical work has considered the contributions of active and social processes to children's learning. One consequence of this range of approaches is that the terms "active" and "social" have become associated with a variety of meanings. Thus, before integrating the two accounts, it is worth defining what we mean by active learning and social contexts. By providing clear definitions, this section aims to clarify the behaviors and phenomena that the integrative account attempts to explain.

1.2.1 Social learning

Learning can 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 the performance of simple tasks and impair the performance of complex tasks (N. B. Cottrell, Wack, Sekerak, & Rittle, 1968; Uziel, 2007). Second, children could learn by looking to others as a guide and observing or imitating their behavior. Children's capacity for faithful imitation has been argued to be 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 psychological reasoning processes that alter the course of learning (M. C. Frank & Goodman, 2014; Goodman & Frank, 2016; Shafto et al., 2012b).

For the integrative account discussed in this chapter, we 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. This broad definition highlights the diverse pathways through which social information could interact with children's active learning. Moreover, this definition captures the variety of contexts in which children develop, i.e., in some cultures children experience high amounts of child-centered interactions while in others they are often expected to learn through observation of older peers and adults (Rogoff et al., 1993).

Social learning theories emphasize 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 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 provide input that is most useful for children (Kline, 2015; Shafto et al., 2012b) and communicate information that is likely to generalize to other contexts (Csibra & Gergely, 2009).

There is a large body of empirical work showing the effects of social input in a variety of domains, including language acquisition, causal learning, and concept learning. For example, newborn infants prefer to look at face-like patterns compared to other abstract configurations (Farroni, Csibra,

Simion, & Johnson, 2002; Johnson, Dziurawiec, Ellis, & Morton, 1991), to listen to speech over non-speech (Vouloumanos & Werker, 2007), their mother's voice over a stranger's (DeCasper, Fifer, Oates, & Sheldon, 1987), and infant-directed speech over adult-directed speech (Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Pegg, Werker, & McLeod, 1992). Moreover, these attentional biases lead to differential learning in the presence of another person. 4-month-olds show better memory for faces that gazed directly at them (Farroni, Massaccesi, Menon, & Johnson, 2007), for an object if an adult gazed at that object during learning (Cleveland, Schug, & Striano, 2007; Reid & Striano, 2005), and perform better at tasks such as segmenting words from infant-directed speech compared to adult-directed speech (Thiessen, Hill, & Saffran, 2005). Kuhl (2007) refer to these effects as "social gating" phenomena since the presence of another person activates or enhances children's underlying learning mechanisms.

In addition to enhancing attention and memory, social contexts can facilitate learning by generating useful information tailored to the child's current developmental stage (Vygotsky, 1987). Empirical work shows that caregivers alter their communication style when speaking to children (e.g.,exaggerating prosody, reducing speed), which in turn can help children learn vowel sounds (Adriaans & Swingley, 2017; De Boer & Kuhl, 2003), segment the speech stream (Fernald & Mazzie, 1991; Thiessen et al., 2005), recognize familiar words (Singh, Nestor, Parikh, & Yull, 2009) and learn new lexical concepts (Graf Estes & Hurley, 2013). Additional evidence comes from Goldstein & Schwade (2008)'s study where they found that infants modified their babbling to produce more speech-like sounds after interacting with caregivers who provided contingent responses, suggesting that contingent, social input was particularly useful because it was closer in time, making it easier to compare discrepancies between the child's vs. adult's speech sound. Finally, converging support comes from research on children's early word learning, which shows that social partners actively reduce the complexity of the visual scene by selecting actions – gaze, points, holding – that make a single object dominant in the visual field (Yu & Smith, 2013; Yu, Ballard, & Aslin, 2005) and by producing labels for objects that children are already attending to (Tomasello & Farrar, 1986).

Another feature of social interactions is that other people's actions are not random; instead, other people intentionally select actions for some goal (e.g., to communicate information). And if children are sensitive to *why* someone performed a behavior, they can use this information to facilitate learning. Recent empirical and modeling work has formalized this idea as a process of

belief updating in Bayesian models of cognition (M. C. Frank & Goodman, 2014; Goodman & Frank, 2016; Shafto et al., 2012b). For example, adults are more likely to think that pressing both buttons is necessary to activate a toy if that action was demonstrated by someone who knew how the toy worked as compared to someone accidentally pressing both buttons (Goodman, Baker, & Tenenbaum, 2009). Shafto et al. (2012b) interpret these results as a psychological reasoning process: "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 and they should avoid performing unnecessary actions.

Similar effects of psychological reasoning on inference occur in word learning (M. C. Frank & Goodman, 2014; Xu & Tenenbaum, 2007b), selective trust in testimony (Shafto, Eaves, Navarro, & Perfors, 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, J. M. Yoon, Johnson, & Csibra (2008) showed that 8-month-olds tend to encode an object's identity when their attention was directed by a communicative point, but they will encode an object's spatial location if directed by a non-communicative reach. And Senju & Csibra (2008) found that infants are more likely to follow another person's gaze when it was accompanied by relevant, communicative cues (e.g., infant-directed speech or direct eye contact).

Finally, several accounts of cultural learning argue that an assumption of *generalizability* is fundamental to social learning via communication, allowing for the accumulation of cultural knowledge across generations (Boyd et al., 2011; Csibra & Gergely, 2009; Kline, 2015). In these explanations, adults generate ostensive signals such as direct gaze, infant-directed speech, and infant-directed actions, which, in turn, direct infants' attention and bias them to expect generalizable information. Experimental work testing predictions of this "Natural Pedagogy" hypothesis shows that children tend to think that information presented in communicative contexts is generalizable (Butler & Markman, 2012; J. M. Yoon et al., 2008), and corpus analyses show that generic language (e.g., "birds fly") is common in everyday adult-child conversations (Gelman, Goetz, Sarnecka, & Flukes, 2008).

The work on social learning reviewed in this section highlight several points that are important for theories of cognitive development. First, from an early age, children are surrounded by other people who know more than they do. Moreover, these more knowledgeable others are invested in

children's development and can provide useful learning opportunities. Second, children are motivated to interact with other people, and these interactions engage and guide attention to relevant information. Finally, social contexts can trigger a set of psychological reasoning mechanisms that lead to stronger inferences, allowing children to get more information out of the same evidence.

Social learning accounts, however, often reflect a relatively passive construal of the learner. And it is evident that children are not just passive recipients of information from the world or other people. Instead, they actively process information and select behaviors that change their learning experience. A growing body of research on children's active learning has developed alongside social learning theories.

1.2.2 Active learning

Learning can also be active in a variety of ways. First, a child could be physically moving, and these actions could change the way they process information. Research on embodied cognition has explored the effects of action experience on infants' learning (see Kontra, Goldin-Meadow, & Beilock (2012) for a review), showing results such as infants who physically hold and manipulate objects will outperform a control group on measures of object attention and exploration (Needham, Barrett, & Peterman, 2002). Second, active learning could refer to children's active internal monitoring and processing of incoming information. For example, young learners do not just accept other people's claims and will reject answers that conflict with their knowledge (Pea, 1982), and children engage in self-generated explanations that can lead to changes in how they process and retain the same incoming information (Lombrozo, 2006). Finally, active learning could refer to a decision-making process where children take actions that control the sequence and pace of their learning experiences.

For the integrative account discussed here, we focus on active learning effects that arise via children's decisions about what action to take next. The key assumption is that active learners take actions that maximize the amount of information they can get while minimizing the costs of effort and time. By limiting the account to active decisions, I do not aim to ignore the importance of other types of active learning; instead, the goal is to constrain the space of connections between active and social learning theories. Moreover, children's action selection captures a rich set of behaviors, including pointing, eye movements, verbal question asking, and causal interventions. Finally, focusing on active decisions connects work on children's learning to a rich tradition of computational and experimental

research in the fields of machine learning, decision theory, and statistics.

The idea that children's actions are important for cognitive development has also been an influential aspect of both classic (Berlyne, 1960; e.g., Bruner, 1961; Piaget & Cook, 1952) and modern (Gopnik et al., 1999; L. Schulz, 2012) theories of learning. Moreover, the effects of active processes have been studied 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). A common thread across these diverse bodies of work is that active contexts lead to different (and often more rapid) learning outcomes because they (a) enhance learners' attention and memory and (b) allow learners to structure experiences that are tuned to their own goals, beliefs, and capabilities (see D. B. Markant, Ruggeri, Gureckis, & Xu (2016) for a review).

One compelling example of how to explore the effects of self-directed choice comes from D. Markant, DuBrow, Davachi, & Gureckis (2014)'s study on "deconstructed" active learning. Adults were asked to memorize the identities and locations of objects hidden in a grid and given different levels of control. They could select: (1) next location to search, (2) item to reveal, (3) duration of the trial, or (4) time between trials. D. Markant et al. (2014) also included a "yoked" control group who saw the same training data that was generated by the active learning group, thus equating the experience while varying the level of control. Across all conditions, there was a memory advantage for active control, providing evidence for the benefits of being able to coordinate the timing of information with internal attentional processes.

Research on language comprehension has also varied levels of active control to tease apart attentional vs. informational effects. For example, Schober & Clark (1989) asked participants to use language to direct another person on how to arrange a set of geometric objects (that did not have familiar labels) in a 4x4 matrix. Critically, the listener could either: (1) talk to the director, (2) listen to a recorded conversation and pause the recording, or (3) listen to the recording but not stop it. Adults who participated in conversation performed better than participants in both passive listening contexts. And the capacity to control the timing of the recording did not improve accuracy, suggesting that active participation in conversation provides an informational advantage such that people could clarify intended meanings before misunderstandings became an issue for later comprehension.

Developmental experiments have found active learning advantages. For example, Ruggeri, Markant,

Gureckis, & Xu (2016) found that control over the timing of new information led to better spatial memory in 6- to 8-year-olds (Ruggeri et al., 2016); Partridge, McGovern, Yung, & Kidd (2015) showed similar effects in the domain of word learning (see also Kachergis, Yu, & Shiffrin (2013) for evidence in adults); and L. Schulz (2012) found that active learning supports preschoolers' understanding of new causal structures. Finally, 16-month-olds show better memory for objects they pointed to compared to an object that they had not actively engaged with (Begus, Gliga, & Southgate, 2014) (see also Stahl & Feigenson (2015)) and adults tend to produce more labels for objects that their infants point to (Z. Wu & Gros-Louis, 2015), providing additional evidence that even young infants can generate actions that elicit information to support their learning.

Research on infants' selective visual attention provides a final compelling example of children's capacity to select useful information. For example, Kidd, Piantadosi, & Aslin (2012) found that 7- and 8-month-olds' prefer to direct gaze at stimuli with intermediate levels of complexity, looking away when the stimulus was either highly predictable or highly surprising (see also Kidd, Piantadosi, & Aslin (2014) for evidence in the auditory domain). Moreover, Gerken, Balcomb, & Minton (2011) found that 17-month-olds increase attention to a stream of input that was learnable but would disengage if the input did not have any stable structure to extract.

1.2.3 The scope of the integrative account

In sum, the empirical work reviewed above shows that both social and active processes can influence learning by enhancing attention/memory, changing inferential processes, and by generating useful learning input. The rest of this chapter presents an integrative account that we scope to focus on children's active learning *decisions* within social contexts. We argue that this is a useful step forward because it allows researchers to ask how specific pieces of social learning affect separate underlying components of decision making.

One major challenge for integrating these two proposals is the large number of factors that could influence how active learning interacts with social reasoning, which in turn creates an extensive space of possibilities for researchers to test. One way to constrain this problem is to use a formal model of information seeking that take an ideal-observer approach (Geisler, 2003). The ideal-observer model defines the structure of the learning task and asks how well learners can take advantage of the information in the environment. This approach pushes researchers to explicitly specify the processes

of active and social learning as separable, underlying components, which can then become targets for empirical work.

Research in cognitive psychology has used the ideal-observer approach to understand human active learning. One particularly useful model has been a formal account of scientific reasoning named Optimal Experiment Design (OED). The OED model was initially designed to help scientists select the best experiment from the set of possible experiments, where "best" means the experiment leading to the most information gained. Researchers have used this formalization to ask whether people's information-seeking behaviors match predictions from OED models.

Moreover, OED has the benefit of using the same mathematical framework as recently-developed models of social learning: Bayesian models of learning, teaching, and communication (M. C. Frank et al., 2009; Goodman & Frank, 2016; Shafto et al., 2012a). These parallel formalizations suggest a productive way to integrate active and social learning theories. This point is the primary focus of the integrative account that we present in Part III of this chapter. Before discussing the details of the account, I briefly review the OED formalization and evidence for OED-like reasoning in human information seeking.

1.3 Part II: 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

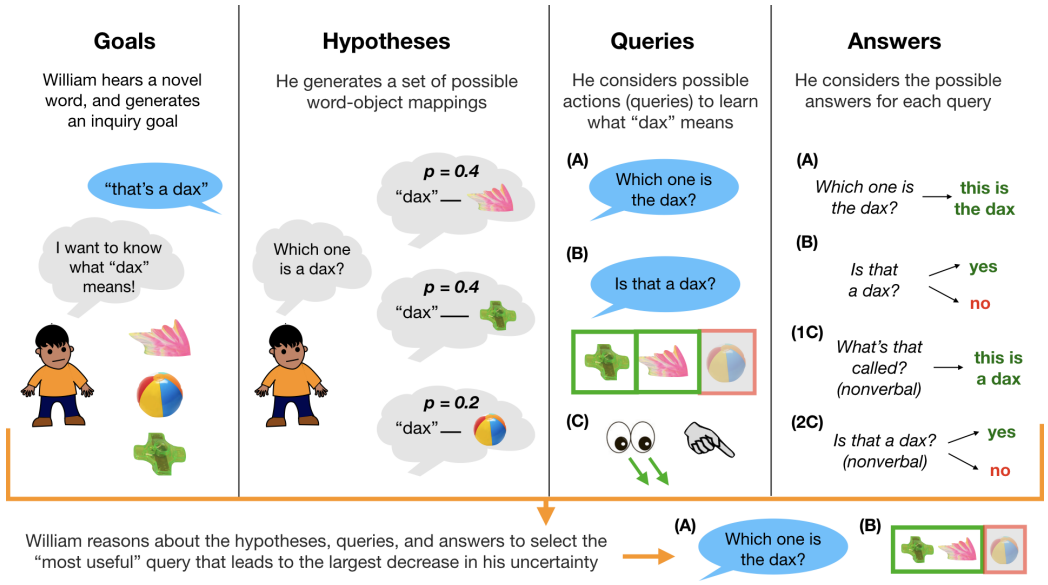


Figure 1.1: Schematic of an active word learning context using the decomposition of Optimal Experiment Design. Social input (hearing a new word) triggers an inquiry goal. Then the learner considers potential hypotheses for the candidate word-object links, weighting each hypothesis by its prior probability. 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. 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). 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.

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. Nelson, McKenzie, Cottrell, & Sejnowski (2010) demonstrate the usefulness of taking an OED approach for differentiating competing theories of information seeking

in 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 different predictions, and thus speeded the rate of learning from their experiments.

Coenen, Nelson, & Gureckis (2017) provide a thorough review of the OED framework and its links to human information seeking. They describe the four parts of an OED model, which are necessary to compute the expected value of an action: (1) hypotheses, (2) queries, (3) a generative model of the types of answers that each query could elicit, and (4) a way to score each answer with respect to the learning goal. There are also two components that are external to the utility computation but important for an account of human inquiry. First, having a learning goal, which instantiates the process of reasoning about hypotheses, questions, and answers. And second, having a stopping rule, which refers to a threshold that causes the learner to stop seeking information and generate an action. Part III defines each component of the OED model, and the mathematical details can be found in the Appendix for this chapter.

1.3.1 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 proposal is that people make decisions, they engage in a similar process of evaluating the usefulness of different actions relative to their learning goals. Using this evaluation process, learners can then select behaviors that maximize the potential for gaining information. One of the successes 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 (C. Cook, Goodman, & Schulz, 2011), and decisions about where to look during scene understanding (Najemnik & Geisler, 2005). Figure 1.1 shows a schematic overview of how OED principles shape the learning process for the task of word learning.

One compelling use case of OED metaphor as a model of human behavior comes from Nelson (2005)'s study of eye movements during 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. They

modeled eye movements as question-asking to gather 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.

Developmental work has used OED models to ask whether children are capable of selecting useful behaviors that maximize learning goals. For example, Legare, Mills, Souza, Plummer, & Yasskin (2013) found that 4- to 6-year-old children consistently asked yes/no questions to figure out the identity of an object hidden in a box. Children produced a high proportion of questions that generated useful, constraining information (e.g., "Is it red?") as opposed to questions that provided redundant information (see also Mills, Legare, Grant, & Landrum (2011); Mills, Legare, Bills, & Mejias (2010)).

Children are also capable of generating efficient actions to learn about the causal structure of objects. For example, C. Cook et al. (2011) found that preschoolers were capable of designing good tests for successfully distinguishing between different hypotheses for how to activate a music box. Moreover, children's behaviors suggested that they were reasoning about a decision that would influence their future opportunity to generate useful information, providing a compelling example of OED-like reasoning in early learning.

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 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 the limitations of the OED approach can be productively reconstrued as opportunities for understanding how learning from other people could scaffold children's active learning. We focus on integrating research and theory on social learning with five critical components of the OED model: goals, hypotheses, questions, answers, and stopping rules (see Figure 1.2 for a schematic overview of the account).

1.4 Part III: Information seeking within social contexts

Why connect active learning processes with social learning effects? First, children do not re-invent knowledge of the world, and while they can learn a tremendous amount from their actions, much of their generalization and abstraction is shaped by input from other people. Moreover, social learning can be the only way to learn something (e.g., first language acquisition). 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) found 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, Bramley, Frosch, Patrick, & Lagnado (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 the learner's understanding of the task determines the quality of active exploration: if the representation is weak, then self-directed learning will be biased and ineffective. Coenen et al. (2017) pursue this line of argument even further, proposing a set of specific challenges for research on active learning. Here is a sample of those open questions that are most relevant to the ideas in this chapter:

- 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 an answer good?
- How do people generate and weight possible answers to their queries?
- How does learning from answers affect query selection and belief change?

In this section, I argue that ideas from research on social learning can address these challenges. We start from the OED model and use it as a conceptual tool to integrate social contexts and active learning. The contribution of this formalization is that it makes the different components of active learning explicit while highlighting the aspects that might be particularly challenging for young learners. Moreover, the "challenges" to the OED account can be reconstrued as opportunities for understanding the benefits of learning from other people. In each of the following sub-sections, I define the challenge for active learning, discuss how social contexts could address it, and highlight prior or future empirical work that connects active and social learning accounts (see Figure 1.2 for an overview).

1.4.1 Goals

In the OED framework, an inquiry goal refers to the underlying motivation for information seeking. Researchers often operationalize these goals as a search for the target hypothesis amongst a set of candidate hypotheses. Some examples of inquiry goals that children might hold are:

- What is this speaker referring to? (word learning)
- What types of objects are called "daxes?" (category learning)
- How does this toy work? (causal learning)
- Is this person reliable? (selective learning)

Specifying learning goals is essential since without them the learner will struggle to evaluate whether an action leads to progress. Coenen et al. (2017) illustrate this point by saying,

> 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 that goal. Similarly, in developmental studies, children are often explicitly asked to answer specific questions, solve a particular problem, or choose between a set of actions. (p. 32-33)

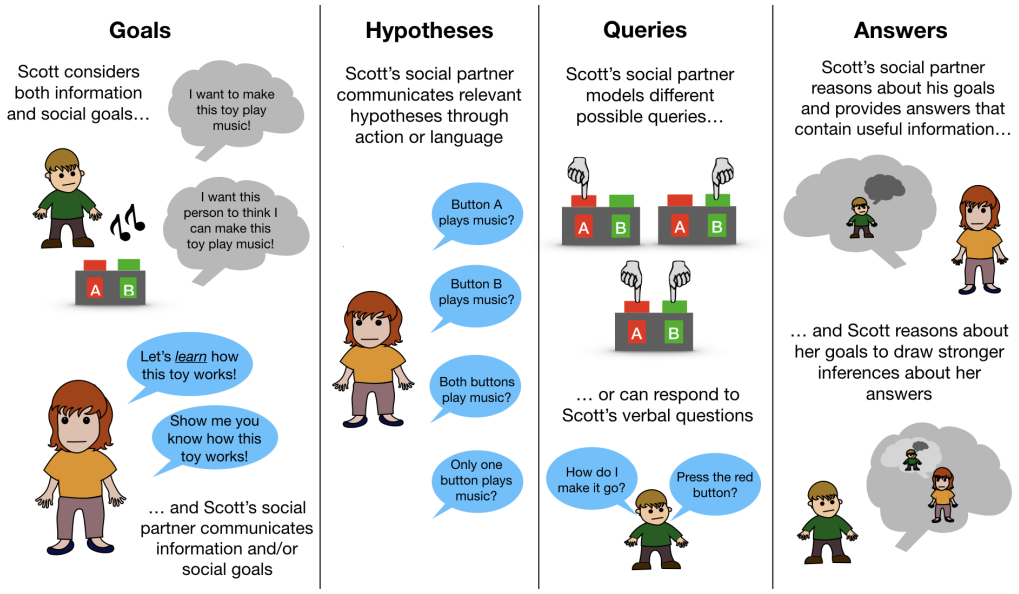


Figure 1.2: Schematic of active learning within a social context. Each panel shows how social information could influence a different component of the active learning process. These social effects occur in-the-moment of learning or over developmental time. Also, the cause of the social effect varies 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.

Characterizing children's goals, however, is not trivial since they could consider a range of goals at any moment with no guarantee that learning progress is one of them. One line of theorizing about the OED hypothesis argues that useful information seeking should only occur in the presence of precise tasks and goals. Consider the example of a parent handing their child a new toy with several buttons on it and saying, "Let's figure out how this toy works!" In this case, it becomes possible to ask whether the child selects actions that are more or less useful for increasing their knowledge of the toy's causal structure.

This example illustrates how other people can trigger learning goals. Empirical work shows how the actions of a social partner can increase children's tendency to detect their uncertainty (see Lyons & Ghetti (2010) for a review). The upshot of this research is that children's ability to realize when they do not know something is relatively slow to develop. For example, Markman (1979) showed that elementary school-aged children were unable to detect apparent inconsistencies in written paragraphs

(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."). Interestingly, if the experimenter gave children an explicit warning or challenge to find a problem with the essay, then they were more likely to monitor their uncertainty and asked more clarifying questions. This finding can be reconstrued as an effect of the social context, shifting children's expectations, highlighting the learning goal, and facilitating their information seeking actions. Converging evidence comes S. Kim, Paulus, Sodian, & Proust (2016), who showed that 3- to 4-year-olds are more likely to monitor their uncertainty about the contents of an opaque box when they were told that they would have to teach another person about the contents of the box (see also Lombrozo (2006)). In this case, the anticipation of social interaction may have increased children's tendency to focus on their goal to learn the identity of the objects.

In addition to triggering inquiry goals, social partners can directly communicate the *value* of learning goals relative to the other goals that children may pursue. This connection draws on theoretical and empirical work exploring how children's environments can shape their intuitive frameworks for processing information and generating goals (Dweck & Leggett, 1988). Under these accounts, children's beliefs about the malleability of traits influence whether they focus on taking actions that increase their competence (prioritize learning) or that demonstrate their fixed abilities (prioritize performing). Empirical work shows that when adults use language to emphasize a learning goal (e.g., "If you pick the task in this box, you'll probably learn a lot of new things." vs. "If you pick this box, although you won't learn new things, it will really show me what kids can do."), then children tend to select the more difficult task [elliott1988goals]. Moreover, both lab-based experiments and observational work provide evidence that the language adults use when praising children influences their tendency to adopt inquiry goals (Cimpian, Arce, Markman, & Dweck, 2007; Gunderson et al., 2013).

While social partners can emphasize learning, the presence of another person can also engage learners' psychological reasoning about their mental states. This reasoning process can then elicit social goals that take priority over learning goals. Consider the "Goals" panel of the schematic active-social learning context shown in **??**. If the learner is worried about what his social partner thinks of him, then he might prioritize actions that reduce the chance of appearing unknowledgeable and choose not to ask about the novel objects. Moreover, he could seek out easy tasks that demonstrate

his competence at the expense of selecting actions that help him learn. The OED framework typically focuses on informational goals with progress measured as a reduction in uncertainty over candidate hypotheses. However, recent empirical and modeling work has begun to move beyond information-specific utility functions to include other goals that learners may hold (e.g., saving time, money, or cognitive resources) [meder2012information].

Recent work in modeling pragmatic communication[1] provides a way to integrate social goals within the OED framework. For example, E. J. Yoon, Tessler, Goodman, & Frank (2017) modeled speakers' decisions to use polite speech as opposed to direct speech (e.g., indirect language such as "we don't think that dress looks phenomenal on you" as opposed to "It looks terrible") as a tradeoff between maximizing informational and social goals, showing that speakers are balancing these two when deciding what to say.

In both OED and RSA, people are assumed to select actions that maximize utility, but the politeness model allows social information to play a role in the utility computation. Future research could adapt this utility-theoretic approach and model the effects of social contexts as changes to the weight children place on social goals, which, in turn, leads to selecting easier tasks where they can appear competent (see E. J. Yoon, MacDonald, Asaba, Gweon, & Frank (2018) for an example of this type of approach). This integration builds on the goal-orienting accounts reviewed above (Dweck & Leggett, 1988) since behavior is as an output of a mixture of goals as opposed to children holding either a performance or a learning goal.

Another open question about learning goals is how often children experience contexts where these goals are apparent or emphasized in their everyday experience. We do not yet have a reliable estimate of the prevalence of situations that would lead children to generate learning goals. One exception, however, is research on *guided participation* by Rogoff et al. (1993) provides evidence that parents from a diverse set of cultural backgrounds produce high rates of structuring and goal-orienting behaviors during activities such as cooking, shopping, working, etc. It is interesting to consider whether variability in children's exposure to goal-orienting behaviors could influence children's tendency to generate learning goals spontaneously.

[1] 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)

1.4.2 Hypotheses

Once children have a learning goal, the next step in the OED model is to figure out what hypotheses to consider. Intuitively, a hypothesis is a candidate explanation for how the world works. For example, consider the schematic learning context shown in **??**. Here, the child might generate the following hypotheses about the meaning of a new word "Dax": (1) Dax = object A, (2) Dax = object B, or (3) Dax = object C.[2]

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 (see Appendix A for details) works by comparing the learner's uncertainty over hypotheses before and after they see an answer. Without a defined hypothesis space, however, it is challenging to select the best action to reduce uncertainty. Put another way; the OED framework does not readily deal with situations where learners might have to consider a large space of hypotheses or might perform actions without considering any hypotheses at all. These challenges are especially relevant for developmental accounts that draw on OED principles since these scenarios are plausible for young learners. Social partners, however, can address this challenge by providing children with a constrained set of hypotheses that facilitate the comparison of different information-seeking actions.

Consider the case of children's early word learning where even the simplest of words, concrete nouns, are often used in complex contexts with multiple possible referents, which creates the potential for an (in principle) unlimited number of hypotheses that children could entertain when trying to figure out word meanings (Quine, 1960). Social-pragmatic theories of lexical development often start from the idea that adults constrain the learning task by providing social cues (eye gaze, pointing, etc.) that connect words to their referents during labeling (P. Bloom, 2002; E. V. Clark, 2009; Hollich et al., 2000). Some of our empirical work shows that the number and fidelity word-object hypotheses that learners store tracks with the strength of social cues available during the labeling moment (MacDonald, Yurovsky, & Frank, 2017b). Other empirical work has found that 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 (D. A. 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 (Yu & Smith, 2012).

[2]This hypothesis space is simplified since it only considers the possibility of one-to-one word-object mappings and that the new word refers to the co-occurring visual context.

Social input can also shape the hypotheses that children generate by revising their intuitive theories about the world. Gerstenberg & Tenenbaum (2017) describe an intuitive theory as, "an ontology of concepts, and a system of (causal) laws that govern how the different concepts interrelate..." (p. 3). Importantly, these theories shape how children interpret incoming information and, in turn, the set of candidate explanations they could explore. Empirical work on conceptual change across development has studied the ways that children integrate input from other people with their current theories (Gelman, 2009). For example, elementary school-aged children tend to hold a mixture of beliefs about the shape of the earth, ranging from a flat earth theory to the adult-like, sphere model (Vosniadou & Brewer, 1992). Interestingly, some children hold intermediate beliefs such as a theory where there are two earths, suggesting that they actively integrate aspects of their initial theory with the information they get from other people (see Opfer & Siegler (2004) for another example from intuitive theories of biological concepts).

In sum, children's candidate hypotheses can be shaped by social information. This effect can occur during the moment of learning (e.g., referential gaze during word learning) and over a developmental timescale (e.g., intuitive theory revision). Critically, social partners can facilitate active learning by constraining the set of hypotheses under consideration, which makes comparing the utility of information-seeking actions more tractable. An open question is how children's information seeking changes as a function of the size of the hypothesis space. Moreover, similar to the research on children's goals, it would be useful to move beyond lab-based studies and towards observational research that measures how children's everyday social interactions constrain the hypotheses they decide to target with their exploration behaviors.

1.4.3 Queries

Queries refer to the set of experiments that a scientist could conduct to gather information about their hypotheses. In children's active learning, queries are the set of actions they could take to collect information from the world. Empirical work on active learning has explored a variety of behaviors, 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 that it provides general principles that could explain such a wide range of behaviors.

The challenge for young learners, however, is to discover what behaviors are available and figure

out which of those actions are useful for learning. One way to address this challenge is for children to look to older peers and adults to figure out what actions can help them learn. Moreover, social learning contexts can change the space of possible queries by adding the option of seeking information from a social partner, which expands the set of possible learning-related actions children could take.

Research on children's verbal questions provides insight into how social partners could model useful queries. First, it is a truism that 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. Finally, 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 had children who asked better questions during book reading sessions at school (Birbili & Karagiorgou, 2009). One explanation for these associations is that wh-questions led children to produce more complex responses that build verbal abilities. Another intriguing causal pathway, however, is that the frequency and type of parents' questions provided children with a template for how to ask useful questions.

Even if children can generate different queries, they still have to evaluate the relative usefulness of each action concerning their learning goal. Empirical work with adults shows a substantial gap between people's question-generating and question-evaluation skills. For example, Rothe, Lake, & Gureckis (2015) used an OED model to measure the expected information gain of adults' natural language questions in a modified "Battleship" game and found that people rarely produced high information gain questions. In a follow-up experiment, however, Rothe et al. (2015) showed that when a different group of adults had access to the list of questions generated by the participants in the unconstrained version, they were quite good at recognizing questions that would generate good information.

Developmental work provides additional evidence of this production-comprehension asymmetry in question asking. For example, children younger than the age of three have difficulty generating useful verbal questions compared to their older peers in lab-based experiments (Mills et al., 2010, 2011), but even the youngest children in those studies are capable of comprehending information

elicited by others' questions to succeed at parallel tasks (Mills, Danovitch, Grant, & Elashi, 2012). Moreover, Mills et al. (2011) manipulated whether children were pre-exposed to adults who modeled useful queries and found that this led to even the youngest children producing useful questions at a much higher rate. Finally, converging evidence comes from research showing that direct instruction (i.e., understanding the objectives of inquiry and identifying useful questions) is necessary for elementary-school-aged children to develop skill in designing informative experiments for learning about physics concepts (Kuhn & Pease, 2008).

In addition to providing a model of queries, social contexts also expand the set of possible questions by adding a social target for information seeking actions. That is, the child has an additional choice: to gather information from the non-social world or other people. The "Questions" panel in Figure 3 shows a child playing with a set of concrete objects. If there is no social partner present, they could still interact with the objects to learn about their shape, texture, or functional properties. But if another person is present, the child now has the option to ask verbal questions or to seek information from the other speaker by gathering their nonverbal cues (e.g., pointing, eye gaze, or 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) showed that children know when to query other people to gain information that they could not get on their own (e.g., invisible properties of a novel social category). Additional evidence comes from work by Lockhart, Goddu, Smith, & Keil (2016) where they showed that 5- to 11-year-olds could reliably determine the kinds of things a person "growing up on their own" could learn from personal experience (that the sky is blue) vs. required interactions with other people (that the earth is round). Finally, work on children's help-seeking behaviors shows that both preschoolers and even infants seek help systematically when completing a task, 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 (Gweon & Schulz, 2011; Vredenburgh & Kushnir, 2016).

Some of our work has investigated how the presence of a social partner changes the set of information seeking behaviors available to the learner. By measuring changes in children's eye movements during familiar language processing, we asked how real-time information seeking adapts to support comprehension. We found that children who were learning a visual-manual language

gathered more information before generating a language-driven gaze shift away from a signer as compared to children processing spoken language while fixating on a speaker (MacDonald, Blonder, Marchman, Fernald, & Frank, 2017a). This result suggests that sign language learners were sensitive to the higher value of fixations in a visual-manual language where disengaging from the signer reduces access to any following linguistic information. We found a similar adaptation of information seeking in both children and adults' processing of spoken language within noisy acoustic environments where looking to a speaker's face provided visual information that could support the comprehension of the less reliable acoustic signal. Critically, listeners would not have had access to this information seeking action outside of a face-to-face, social interaction (e.g., listening to a noisy audio recording).

1.4.4 Answers

Once the learner generates a set of possible queries, they can simulate the set of possible answers they could get in return. In the OED framework, an answer is useful if it decreases the learner's uncertainty about their hypotheses. The challenge for the child as an active learner is two-fold: (1) figure out what which answers are likely for each query and (2) decide how much to update their beliefs after seeing an answer. In this section, we discuss how social contexts can affect each component of the answer-evaluation process.

Defining a "good" answer is challenging. Intuitively, a good answer provides the learner with information that they did not already know, that they were interested in learning, and that is likely to be useful beyond the current context. Even within the formal OED framework, there have been a variety of ways to instantiate this utility function (e.g., information gain, probability gain, and Kullback-Leibler divergence) (Nelson, 2005). These information-theoretic utility functions take into account the learner's prior beliefs, which are represented as probability distributions over hypotheses and compute the effect of each answer on the learner's beliefs represented as a conditional probability distribution.

Social learning theories often start from the idea that interactions with other people increase the probability of getting useful information. For example, evolutionary models argue that knowledge transfer between generations allows for the gradual accumulation of small improvements that eventually led to sophisticated 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 example

of a hunter stumbling across the link between the color/texture of ice and its stability, saying that this type of rare information would be much harder for individuals to re-discover on their own. But social transmission allows humans to pass these discoveries down to subsequent generations, allowing humans to move beyond the information that the world is likely to provide.

Csibra & Gergely (2009)'s theory of "Natural Pedagogy" argues that an assumption of *generalizability* is a fundamental component of the answers that children get when they interact with adults. Under their account, when adults provide ostensive cues (eye gaze or child-directed speech) to signal generalizable knowledge children update their beliefs differently. Empirical work shows that infants are more likely to generalize the valence of an object to a new person when learning in the presence of an ostensive, pedagogical cue (Gergely, Egyed, & Király, 2007). Moreover, 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 to that object (J. M. Yoon et al., 2008).

In addition to an increased chance of seeing generalizable information, features of the social context can also modulate the way active learners evaluate possible answers. This link is one of the more developed connections between the OED and social learning accounts. That is, researchers have made progress in modeling the influence of different assumptions that a learner could make about the process through which other people generate answers. Shafto et al. (2012b) describe these different sampling assumptions, placing them along a continuum that varies regarding how much a learner should update their beliefs:

- *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)

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 intentions.[3]

Empirical support for the pedagogical sampling account comes from a range of domains/tasks, including word learning (M. C. Frank et al., 2009), pragmatic inference (M. C. Frank & Goodman, 2012), and causal reasoning (Bonawitz et al., 2011). These empirical findings connect directly to the OED model. Consider Xu & Tenenbaum (2007a)'s finding that when a knowledgeable teacher generates object labels, learners assumed the examples indicated the true word meaning, making it more likely that the teacher would draw three samples from the broader, basic category (as opposed to the smaller subordinate category). They modeled this effect by modifying the likelihood function in a Bayesian cognitive model. In this case, the likelihood function for the teacher-driven condition was designed to capture the idea that learners should prefer more restrictive hypotheses if they are confident that the teacher generated labels based on the actual word meaning. This formalization makes direct contact with the OED model of human inquiry since learners consider how much answers will update their beliefs, capturing the idea that responses generated with the goal to teach carry more information.

Another line of empirical work has focused on children's decisions about whom to learn from. Even 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 useful information in the past (Koenig, Clement, & Harris, 2004). For example, Chow, Poulin-Dubois, & Lewis (2008) found that 14-month-olds are less likely to follow the gaze of a person who had consistently directed gaze towards an empty location. Moreover, children prefer to learn from familiar rather than unfamiliar adults (Corriveau & Harris, 2009), adults rather than peers (Rakoczy, Hamann, Warneken, & Tomasello, 2010), and ingroup rather than outgroup members (MacDonald, Schug, Chase, & Barth, 2013). Moreover, Gweon, Pelton, Konopka, & Schulz (2014) found that 14-month-olds explore objects more after interacting with a teacher who had been under-informative in their previous demonstrations, providing evidence of an early ability to link the quality of answers to active learning decisions.

[3]See the "Answers" panel of Figure 1.2 for an illustration of this recursive reasoning process within an active learning context.

Interestingly, the outcome measures in these studies of selective learning – whom to direct questions towards or how long to explore a novel toy – are information-gathering decisions. While this is an implicit link to the OED account, researchers have modeled selective learning phenomena as modifications to the likelihood function in Bayesian cognitive models. For example, Shafto et al. (2012a) developed a model in which children reason about both the helpfulness and knowledgeability of speakers when deciding whom to learn from and were able to capture several qualitative findings from the selective trust literature. This approach parallels the word learning and pedagogical sampling models reviewed above, highlighting how social reasoning could change the utility of the answers learners get from other people.

One candidate for future research is to ask how social contexts change children's decisions about whether to initiate information seeking behaviors in the first place. That is, if their social partners are unlikely to provide useful answers, then active learning (even if children are capable of selecting informative actions) becomes less valuable. The dual consideration of costs and benefits in active learning has been the focus of much research in machine learning, which tries to select the set of questions to ask taking into account how long it will take a human to respond (Haertel, Seppi, Ringger, & Carroll, 2008). Moreover, recent lab-based studies and theorizing in social cognition suggest that children are capable of "intuitive utility maximization" reasoning that others take actions to increase rewards while minimizing costs regarding time, effort, etc. (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). It would be interesting to bring these cost-based approaches to bear on questions about how social contexts modify children's active learning.

1.4.5 Stopping rules

A stopping rule is a threshold that when exceeded causes the learner to stop gathering information and generate an action. These rules fall into two categories: information or time-based. For example, when searching for information on the internet, a learner might create the time-based stopping rule – spend one hour searching – or the information-based rule – explore until we find the definitions that we need. The function of a stopping rule is to balance information gained with reducing unnecessary time/effort put into the search task.

The challenge for both the child as active learner and researchers interested in explaining stopping rules is both operationalizing and estimating the critical pieces needed for developing a stopping rule:

the rate of information gain and the cost of information search. To address this challenge, researchers have used ideas from theories of animal foraging to model adults' decisions of when to stop gathering information (Pirolli & Card, 1999). For example, Manohar & Husain (2013) used a "patch" model to explain the timing of adults' decisions about when to fixate and re-fixate on one of two symbolic gambles displayed on a computer monitor. They proposed that fixating on an item results in a leaky gain in precision and that people should shift their gaze once their information gain rate falls below a pre-defined threshold (see also Hills, Jones, & Todd (2012) for evidence from semantic memory). Moreover, much progress has been made by modeling perceptual decisions as a noisy evidence accumulation process with responses generated when information crosses a pre-defined threshold (Ratcliff & McKoon, 2008).

Social partners play a key role in children's decisions to stop information seeking since they are a valuable resource of information. Empirical work has primarily focused on how children persist in seeking information from social partners if their initial request is not satisfied. For example, Frazier, Gelman, & Wellman (2009) used a corpus analysis of parent responses to children's *how* and *why* questions and found that preschoolers were more than twice as likely to re-ask a question after getting a non-explanatory response (e.g., CHILD: "Why you put yogurt in there?" ADULT: "Yogurt's part of the ingredients") compared to an explanatory answer (e.g., CHILD: "How do you get sick?" ADULT: "we don't know.") (see also Deborah, Louisa Chan, & Holt (2004)). These studies suggest that even toddlers are sensitive to when they have gathered sufficient information to address their information seeking goals within social contexts.

In addition to providing information, social partners can take actions that shape 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). Children's stopping behavior is reasonable since their social partner communicated that there was no other information to be gained by taking actions on the object. Moreover, Butler & Markman (2012) showed that an adults' pedagogical demonstration led to an increase in children's object exploration when testing whether a hidden property (magnetism) generalized to a new but similar looking object. These results highlight the complex ways social interactions can modulate children's thresholds for ending their information search.

Our work on where children choose to look during real-time language comprehension could be construed as an effect of social contexts on children's stopping decisions (MacDonald, Marchman, Fernald, & Frank, 2018b). We found that both adults and children fixate more on a speaker's face when processing familiar words in a noisy auditory environment, an adaptation that resulted in a higher proportion of gaze shifts to the named referent. We used a cognitive model of decision making (Drift-Diffusion Model Ratcliff & McKoon (2008)) to provide evidence that an increase to listeners' decision thresholds, and not processing efficiency, best explained the behavioral results. This approach provides an example of bringing cognitive models of decision making into contact with empirical questions about children's language comprehension within face-to-face social interactions.

Another promising approach for future research is to ask how children's developing capacity to reason about the cost of actions changes their active learning behaviors. In the OED framework, cost (e.g., monetary value of running an additional experiment) is critical to deciding when to stop gathering additional information, but this is complex to operationalize in human information seeking. However, 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.

An important next step to understanding children as efficient active learners is to explore who they balance the desire want to ask questions that maximize information gain while taking 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 mental states. This line of research falls directly out of an integrated active-social learning framework.

1.5 Conclusions: Eye movements as a case study

In this chapter, we presented concrete definitions of active and social learning and described a unifying framework of information seeking within social learning contexts. We used Optimal Experiment Design (OED) as a theoretical tool to propose that the presence of another person can affect

children's information seeking by:

1. instantiating learning goals, communicating the value of learning goals, or triggering social goals
2. constraining the hypothesis space
3. modeling useful queries
4. providing useful answers
5. changing thresholds for stopping information seeking

The account described in this chapter argues that the social and active learning theories have much to be gained from considering the other. Active learning accounts will benefit by understanding how learners incorporate social information into their decision making, allowing researchers to develop more sophisticated utility functions that might better capture what people care about when learning something new. This approach seems especially crucial 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 phenomena can benefit from advances in the study of active learning by connecting their ideas to the rich traditions of machine learning, decision theory, and statistics. Moreover, research on social learning phenomena often uses information-seeking decisions as dependent variables in experiments. Thus, a secondary benefit would be a better understanding of the factors that influence the measurement of social learning effects.

Finally, the evidence reviewed in this chapter shows that research has mostly focused on lab-based experiments where social and active learning effects are measured in isolation during highly constrained task settings. An important next step is to estimate the presence of learning goals, tractable hypothesis spaces, and the quality of answers in children's everyday experiences. This line of inquiry will require a shift from relying on highly-controlled lab experiments to leveraging large-scale, observational datasets. And while this approach adds complexity, I think that the benefit will be a far greater understanding of how children's active learning operates over fundamentally social input.

The rest of this thesis presents a series of empirical studies of familiar language comprehension and novel word learning, which we situate within the integrative framework presented in this chapter.

The majority of the experiments use eye movements as a case study of children's early information seeking. Eye movements are well-suited for studying the connection between active processes and social contexts because (1) visual attention is important for early lexical learning, (2) control over gaze is earlier to develop as compared to other forms of information seeking (e.g., verbal question asking), and (3) there is already a body of work that had characterized visual fixations as a form of question-asking in research on goal-based vision (M. Hayhoe & Ballard, 2005).

At the beginning of each chapter, I show a schematic overview of the link between the integrative framework and each case study, and I highlight the relevant piece of the broader framework that the empirical work addresses. Chapter 2 investigates how children decide where to look while comprehending a visual-manual language (American Sign Language: ASL). In the grounded, sign processing context, children must use visual fixations to gather information about the referents and the linguistic signal, creating competition for visual attention (i.e., queries) that is not present in spoken language comprehension.

Chapter 3 builds on the sign language work and directly compares the gaze dynamics of children learning ASL to children learning spoken English during familiar language processing. Specifically, we measure when children decide to stop fixating on a social partner to seek named referents in the visual world (i.e., stopping rules). Chapter 3 also describes a study of English-learning children and adult's eye movements in noisy auditory contexts, which could make the visual information gathered from looks to a social partner (i.e., answers) more useful for language understanding.

Chapters 4 and 5 explore how the information seeking framework generalizes to processing words in the context of social cues to reference. In Chapter 4, we present a series of large-scale word learning experiments, showing that the presence of a social cue modulates how much information (i.e., answers) adults remember from labeling events. Finally, Chapter 5 describes a set of eye-tracking studies that measure how children's decisions to stop gather information from a social partner (i.e., stopping rules) changes as a function of (a) whether the social partner provides a cue to reference (i.e., answers) and (b) their knowledge of the word-object mappings (i.e., hypotheses).

The goal of the empirical work is to ask how children's gaze patterns adapt to a diverse set of contexts that change the value of seeking information from social partners. The common thread across this research is that children's real-time visual information seeking is quite flexible and can adapt to the informational structure of the environment to support efficient language processing.

Chapter 2

Children's distribution of visual attention during real-time American Sign Language comprehension[1]

This chapter presents a study of eye movements in a visual-manual language (American Sign Language: ASL). ASL is an interesting case because visual information about the referents and the linguistic signal are gathered through the same mechanism: visual fixations. In contrast, children learning spoken language can look away from their social partners while gathering more linguistic information. Within our broader active-social framework, this study asks whether ASL-learners' visual information seeking (i.e., queries) would look dramatically different given the constraints of processing a visual language in real time or by having differential access to auditory information in day-to-day life.

To answer this question, we measured eye movements during real-time ASL comprehension of

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

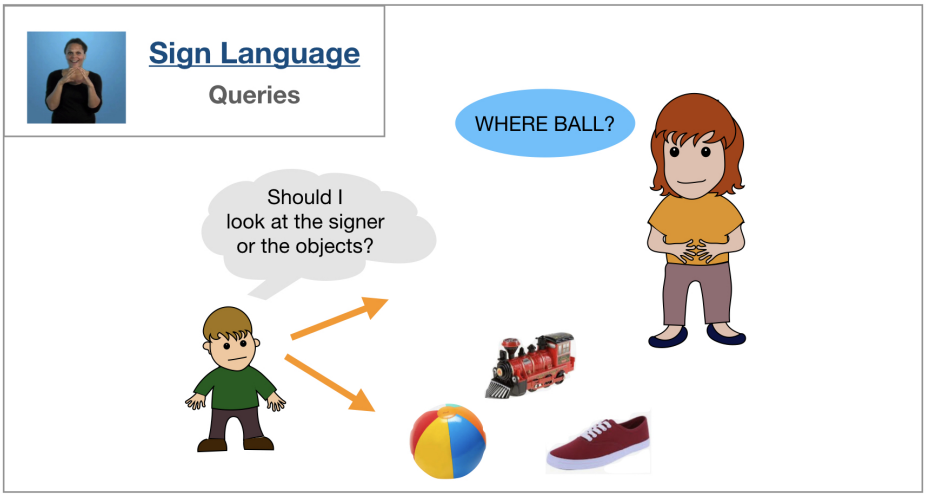


Figure 2.1: A schematic showing the components of the OED model captured by the case studies in Chapter 2.

29 native ASL-learning children (16-53 mos, 16 deaf, 13 hearing) and 16 fluent deaf adult signers. All signers showed evidence of incremental language comprehension, tending to initiate an eye movement before sign offset. Moreover, Deaf and hearing ASL-learners showed remarkably similar gaze patterns. Finally, variation in children's ASL processing was positively correlated with age and vocabulary size. These results suggest that, despite competition for attention within a single modality, signers will shift visual attention to seek named objects in ways that parallel spoken language processing and that the timing and accuracy of these fixations reflect language-relevant information processing skills.

2.1 Introduction

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, 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, Gutiérrez-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 Study

2.2.1 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

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 1.5 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.

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

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

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.

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

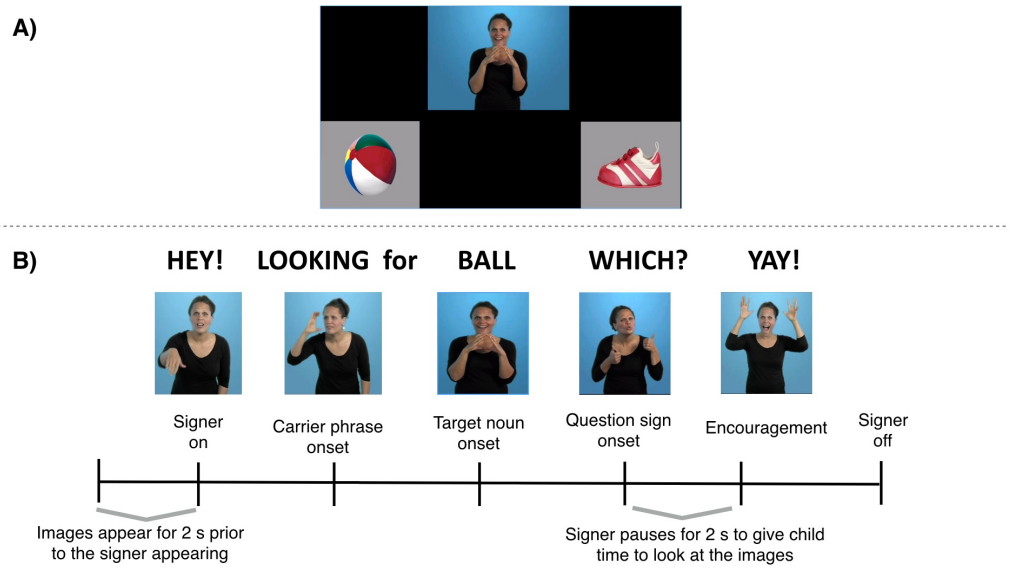


Figure 2.2: Configuration of visual stimuli (1A) and trial structure (1B) for one question type (sentence final wh-phrase) shown in the central video on the VLP task.

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 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.

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]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. This analysis is presented in the Appendix for this chapter.

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.

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.

2.2.2 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).

[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.

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* $\sim$ *age* and *processing measure* $\sim$ *vocabulary*.

2.2.3 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?

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,

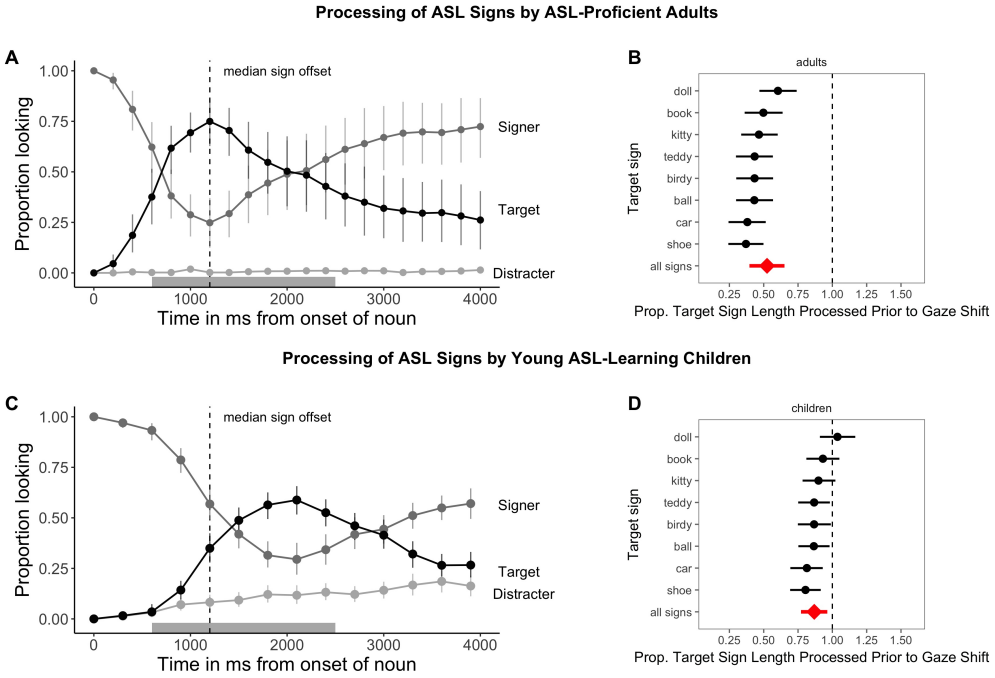


Figure 2.3: The time course of looking behavior for ASL-proficient adults (2A) and young ASL-learners (2C). The curves show mean proportion looking to the signer (dark grey), the target image (black), and the distracter image (light grey). The grey shaded region marks the analysis window (600-2500ms); 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.

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. 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.

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.

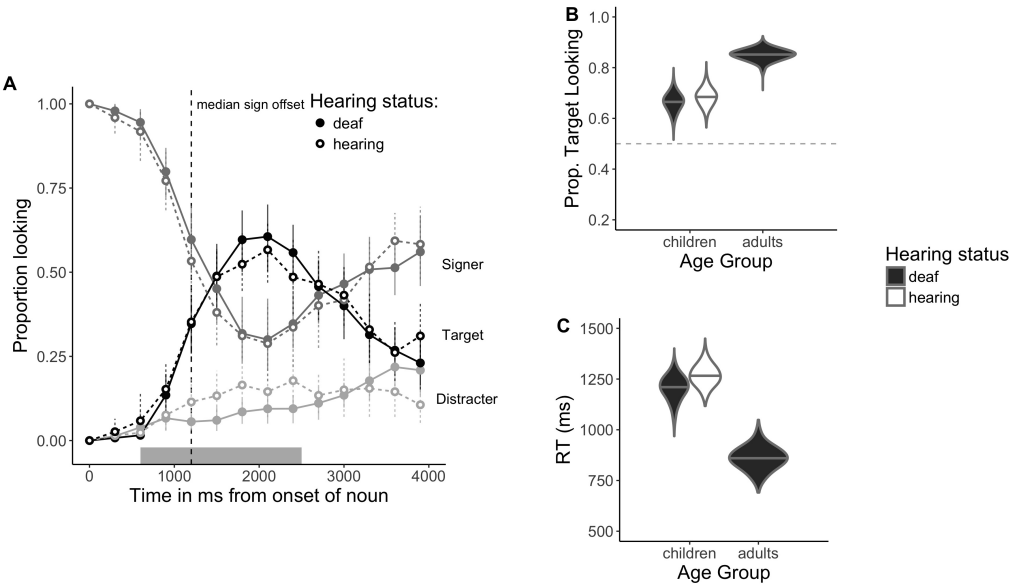


Figure 2.4: The time course of looking behavior for young deaf and hearing ASL-learners (3A). Filled circles represent deaf signers, while open circles represent hearing signers; All other plotting conventions are the same as in Figure 2. Panels B and C show full posterior distributions over model estimates 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).

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 ($M_{hearing} = 0.68$, $M_{deaf} = 0.65$; $\beta_{diff} = 0.03$, 95% HDI $[-0.07, 0.13]$) or RT ($M_{hearing} = 1265.62$ ms, $M_{deaf} = 1185.05$ ms; $\beta_{diff} = 78.32$ ms, 95% HDI $[-86.01ms, 247.04ms]$), 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. 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 Accuracy (Figure 3B) and RT (Figure 3C). Overall, adults were more accurate ($M_{adults} = 0.85$, $M_{children} = 0.68$, $\beta_{diff} = 0.17$, 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; $\beta_{diff} = -367.76$ 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.

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

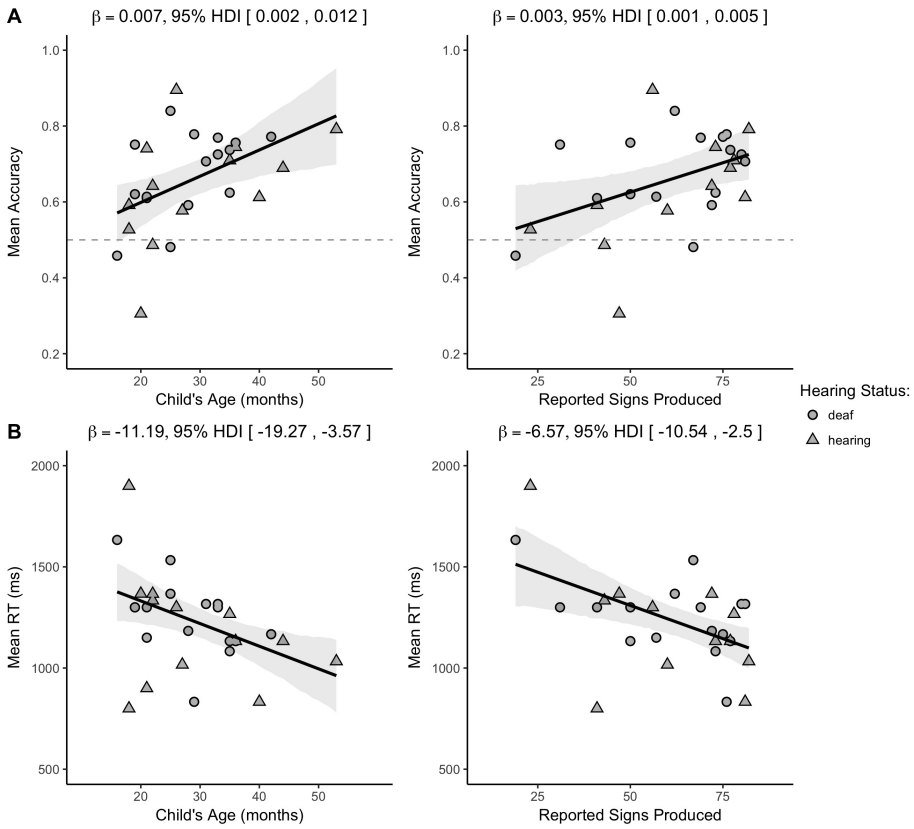


Figure 2.5: Scatterplots of relations between children's age and vocabulary and measures of their mean accuracy (4A) and mean RT (4B) in the VLP procedure. Shape represents children's hearing status. The solid black line is the maximum a posteriori model estimate for the mean accuracy at each age point. The shaded gray regions represent the 95% Highest Density Interval (range of plausible values) around the regression line.

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 β 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]

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).

2.3 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).

2.3.1 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.

2.4 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

Children flexibly seek visual information during signed and spoken language comprehension[1]

In this chapter, we present two studies of eye movements during real-time familiar language processing. Within our broader active-social framework, these studies explore whether children adapt their eye movements to query locations in response to the utility of that location for their goal of rapid language comprehension. Moreover, these studies investigate children's decisions to stop fixating on a social partner and seek a named referent, synthesizing threshold models of decision making, stopping rules, and language-driven visual attention.

Real-time language comprehension involves linking the incoming linguistic signal to the visual world. Information that is gathered through visual fixations can facilitate the comprehension process. But do listeners flexibly select what visual information to gather? Here, we propose that children flexibly adapt their gaze to seek visual information from social partners to support language understanding. We present evidence for our explanation using two case studies of eye movements during

[1]Parts of this chapter are published as MacDonald et al. (2017a) An information-seeking account of eye movements during spoken and signed language comprehension and as 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.

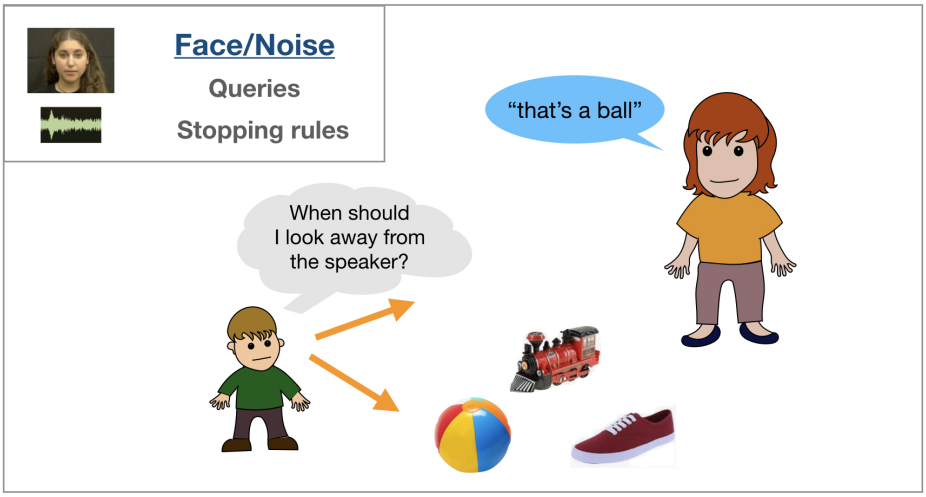


Figure 3.1: A schematic showing the components of the OED model captured by the case studies in Chapter 3.

real-time language processing: children learning spoken English vs. young ASL-learners and spoken English in noisy vs. clear auditory environments. Across both studies, we found that listeners adapted their gaze to fixate longer on a social partner when it was useful for language comprehension. Fixating longer on their social partner led to a higher proportion of gaze shifts landing on the named objects, and more language-driven, as opposed to random, shifts. These results suggest that children can adapt their information gathering thresholds to seek additional visual information from their social partners that support their real-time language comprehension.

3.1 Introduction

Extracting meaning from language represents a formidable challenge for young learners. Consider that even in the simple case of understanding grounded, familiar language (e.g., "look at the ball"), listeners must quickly integrate linguistic and non-linguistic information from two continuous streams of input. Moreover, words occur within dynamic interactions where there is often insufficient information to figure out what is being said, and yet listeners must decide how best to respond. Even young children are capable of linking language to the world quite efficiently, shifting visual attention

to a named object in a scene within hundreds of milliseconds upon hearing its name in controlled laboratory conditions (Allopenna et al., 1998; Fernald, Perfors, & Marchman, 2006; Spivey et al., 2002; Tanenhaus et al., 1995). But how do children comprehend language when processing demands are higher and the input is noisy?

In interactive theories of language comprehension, listeners integrate noisy input from multiple sources of information to constrain the set of possible interpretations of an utterance (M. C. MacDonald & Seidenberg, 2006; McClelland & Elman, 1986). Under this interactive account, listeners comprehend words by partially activating several candidates that are consistent with incoming perceptual information. As more information arrives, words that do not match the perceptual signal are no longer considered, and words that do match become more strongly activated until a single interpretation is reached (see McClelland, Mirman, & Holt (2006) for a review). Critically, multiple sources of information – the linguistic signal, visual world, and conceptual knowledge – mutually influence one another to constrain the listener's interpretation of an utterance (e.g., the "McGurk effect" (J. MacDonald & McGurk, 1978) and parsing syntactically ambiguous utterances (Tanenhaus et al., 1995)).

Thus, information gathered from the visual world can facilitate language comprehension. The incoming linguistic signal is transient, however, and rather than randomly fixate a scene, listeners might strategically deploy their fixations to informative locations that maximize successful comprehension. Consider a speaker who asks you to "Pass the salt" in a noisy restaurant where it is difficult to perceive what she is saying. Recent theoretical and empirical work suggests that children and adults handle this sort of noise in the signal by integrating what they perceive with their prior beliefs about the plausibility of a speaker's intended meaning (Fourtassi & Frank, 2017; Gibson, Bergen, & Piantadosi, 2013; Yurovsky, Case, & Frank, 2017). Here, we pursue the idea that the strategic gathering of visual information via looks to specific locations in the visual scene could support comprehension. Returning to the noisy restaurant example, the listener could facilitate understanding by looking at the objects on the table (e.g., the type of food the speaker is eating) or by looking at the speaker directly (e.g., reading her lips or the direction of her gaze), but not by allocating attention to other, perhaps quite salient, areas of the visual scene (e.g., a flashing light above the dinner table).

Another compelling case where strategic use of visual fixations becomes necessary is the comprehension of visual-manual languages such as American Sign Language (ASL). In ASL, fixations to the language source are highly informative because signers must process all linguistic information via the visual channel. Further, the decision to look away from a signer to the rest of the visual world could be risky because this behavior might reduce visual access to subsequent linguistic information, thus complicating the listener's decision to gather visual information that might constrain the meaning of a noisy utterance. Finally, ASL represents a compelling case where there is direct competition between allocating visual attention to the non-linguistic information – gesture, eye gaze, and facial expressions – that has been suggested to support spoken language comprehension.

These examples highlight how we can characterize eye movements as an active decision-making process where listeners select fixations to gather language-relevant information. In the current work, we pursue this idea and propose that listeners are sensitive to the value of different fixation behaviors for the goal of grounded language understanding. We hypothesize that even young children can flexibly adapt the dynamics of their gaze to seek higher value visual information that supports comprehension. Several research programs inspired this hypothesis, including work on language-driven shifts in visual attention (Allopenna et al., 1998; Tanenhaus et al., 1995), goal-based accounts of eye movements in everyday tasks (M. Hayhoe & Ballard, 2005), and language perception as a process of multisensory cue integration (Vigliocco, Perniss, & Vinson, 2014). In the following sections, we briefly review each literature to motivate our explanation of information-seeking eye movements in grounded signed and spoken language comprehension.

3.1.1 Vision-language interactions during language comprehension

The study of eye movements during spoken language comprehension has provided insight into the interaction between concepts, language, and visual attention. The majority of this work has used the Visual World Paradigm (VWP) where listeners' eye movements are recorded at the millisecond timescale while processing language and looking at a set of objects (see Salverda, Brown, and Tanenhaus (2011) for a review). Crucially, these analyses rely on the fact that listeners will initiate gaze shifts to named referents with only partial information, in contrast to waiting until the end of a cognitive process (Gold & Shadlen, 2000). Thus, the time course of eye movements can provide a window onto how and when people integrate information to reach an interpretation of the incoming

linguistic signal.

A classic finding using the VWP shows that listeners will rapidly shift visual attention upon hearing the name of an object ("Pick up a beaker.") in the visual scene with a high proportion of shifts occurring soon after the target word begins (Allopenna et al., 1998). Moreover, adults will look at a phonological onset-competitor ("beetle") early upon hearing the word "beaker," suggesting that they activate multiple interpretations and resolve ambiguity as the stimulus unfolds. Finally, empirical work shows that information from the visual world can constrain interpretation by activating listeners' conceptual representations before the arrival of the linguistic signal (Dahan & Tanenhaus, 2005; Yee & Sedivy, 2006). These results fall out of predictions made by interactive models of speech perception where information from multiple sources is integrated rapidly to constrain language understanding (M. C. MacDonald & Seidenberg, 2006; McClelland et al., 2006).

In addition to work in adult psycholinguistics, the VWP has been useful for studying developmental change in language comprehension skill in children. Researchers have adapted the task to measure the timing and accuracy of children's gaze shifts as they look at two familiar objects and listen to simple sentences naming one of the objects (Fernald, Zangl, Portillo, & Marchman, 2008; Venker, Eernisse, Saffran, & Weismer, 2013). Such research finds that children, like adults, shift gaze to named objects soon after the acoustic information is sufficient to enable referent identification. Further, individual differences in the speed and accuracy of eye movements predict vocabulary growth and later language and cognitive outcomes (Fernald et al., 2006; Marchman & Fernald, 2008; Rigler et al., 2015).

In the current studies, we use the VWP to ask whether children strategically adapt their information seeking in response to the processing demands in their environments. This approach reflects a shift from construing eye movements in the VWP task as an index of the interaction between language and visual attention to an index of behaviors that gather information to support real-time language comprehension. This construal dovetails with a body of work on goal-based accounts of vision that start from the idea that eye movements are an active information-gathering process driven by perceivers' internal task goals (M. Hayhoe & Ballard, 2005).

3.1.2 Goal-based accounts of eye movements in everyday tasks

Under goal-based accounts of vision, people deploy gaze to reduce their uncertainty about the world and to maximize their expected future rewards concerning some goal. For example, M. Hayhoe & Ballard (2005) review evidence that adults fixate on locations that are most helpful for their current task (e.g., looks to an upcoming obstacle when walking) as opposed to other aspects of a visual scene that might be more salient (e.g., a flashing light). Moreover, empirical work shows that adults gather task-specific information via different visual routines as they become useful for their goals. For example, Triesch, Ballard, Hayhoe, & Sullivan (2003) found that adults were less likely to collect and remember visual information about the size of an object when it was not relevant to the task of sorting and stacking the objects.

M. Hayhoe & Ballard (2005)'s review also highlights how perceivers learn to deploy efficient gaze patterns as they become more familiar with a task. They point out that visual routines develop over time, and it is only when a task becomes highly-practiced that people allocate fewer looks to less-relevant aspects of the scene. For example, Shinoda, Hayhoe, & Shrivastava (2001) show that skilled drivers learn to spread visual attention more broadly at intersections to better detect stop signs. Other empirical work shows that the visual system rapidly learns to use temporal regularities in the environment to control the timing of eye movements to identify goal-relevant events (Hoppe & Rothkopf, 2016). Finally, the timing of eye movements in these tasks often occurs before an expected event, suggesting that gaze patterns reflect an interaction between people's expectations, the information available in the visual scene, and their task goals.

Recent theoretical work has argued for a stronger link between goal-based perspectives and the work on eye movements during language comprehension reviewed above. Salverda, Brown, & Tanenhaus (2011) highlight the immediate relevance of visual information for language understanding, suggesting that listeners' goals should be a key predictor of fixation behaviors. Further, they point out that factors such as the difficulty of executing a real-world task should change decisions about where to look during comprehension. One example of applying a goal-based approach is Nelson & Cottrell (2007)'s study of gaze patterns during category learning. Nelson & Cottrell (2007) modeled eye movements as a type of question-asking behavior and found that when participants became more familiar with novel concepts, their gaze patterns shifted from exploratory to efficient, suggesting that fixations changed as a function of goals during the task.

Pursuing this connection further, in our current studies, goal-based models of eye movements predict that gaze during language comprehension should adapt to the processing context. That is, listeners should change the timing and location of eye movements when a fixation area becomes more useful for understanding. This proposal, which we test, also connects with a growing body of research that explores the effects of multisensory (gesture, prosody, facial expression, and body movement) integration on language perception and comprehension.

3.1.3 Language perception as multisensory integration

The final line of research that informs our studies is work exploring the process of language comprehension as multisensory integration. This research starts from the idea that language understanding does not just involve a single stream of linguistic information. Instead, face-to-face communication provides the listener with access to a set of multimodal cues that can shape language understanding (for a review, see Vigliocco, Perniss, and Vinson, 2014). For example, empirical work shows that when gesture and speech provide redundant cues to meaning, adults are faster to process the information and make fewer comprehension errors (Kelly, Özyürek, & Maris, 2010). Moreover, developmental work shows that parents use visual cues such as gesture and eye gaze to structure language interactions with their children (Estigarribia & Clark, 2007). And, from a young age, children also produce gestures such as reaches and points to share attention with others to achieve communicative goals (Liszkowski, Brown, Callaghan, Takada, & De Vos, 2012).

Most developmental accounts of early language acquisition begin from the ecological context of children grounding language within multimodal, social interactions (E. V. Clark, 2009; Tomasello & Farrar, 1986). This literature has often focused on how children integrate social cues processed in a modality different from the linguistic signal (i.e., spoken words are auditory while a speaker's eye gaze or points are visual). The case of ASL, which we explore in this work, highlights how the process of integrating social cues with the linguistic signal is not necessarily cross-modal. When children comprehend ASL, both signs and social cues are visual and could compete for fixations. And yet we know little about how young ASL-learners deploy fixations to gather information about signs, social signals, or the contents of the visual world.

Additional support for the role of multisensory processing in language comes from work on audiovisual speech perception. These studies show that visual information from a speaker's mouth

can shape spoken language perception. In a review, Peelle & Sommers (2015) point out that mouth movements provide a clear indication of when someone has started to talk, which cues the listener to allocate additional attention to the speech signal. Further, a speaker's mouth movements convey information about the phonemes in the acoustic signal. For example, visual speech information distinguishes between consonants such as **/b/** vs. **/d/** and place of articulation can help a listener differentiate between words such as "cat" or "cap." Finally, classic empirical work shows benefits for audiovisual speech perception compared to auditory- or visual-only speech perception, especially in noisy listening contexts (Erber, 1969).

In sum, work on multisensory processing shows that auditory and visual information interact to shape language perception. These results parallel the predictions of interactive models of language processing reviewed earlier (M. C. MacDonald & Seidenberg, 2006; McClelland et al., 2006), and they suggest that visual information from a social partner is an essential input to children's language comprehension. Finally, this work highlights the importance of studying language understanding within face-to-face communication, where listeners can choose to look at their social partners to gather language-relevant information.

3.1.4 The present studies

The current studies explore an information-seeking explanation of eye movements during grounded signed and spoken language comprehension. We propose that the timing of gaze shifts is related to the goal of gathering language-relevant visual information from a speaker balanced with fixating on the surrounding visual scene. We draw on models of eye movements as active decisions that collect information to achieve reliable interpretations of incoming language and test predictions of our account using two case studies: processing of signed vs. spoken language and processing spoken language in noisy vs. clear auditory environments. These cases, while superficially different, share a key feature: The interaction between the listener and their environment changes the value of fixating on the source of language to support comprehension. For example, in comparing ASL to spoken language, the value of looking to an interlocutor is higher since all of the language-relevant information is located at that point in the visual world; whereas a young child processing spoken language can fixate on other locations in the visual scene while still processing linguistic information via the auditory channel.

A secondary goal of this work was to test whether children and adults would show similar patterns of gaze adaptation in response to changes in the value of looking to a social partner for language understanding. Recent developmental work shows that, like adults, preschoolers will flexibly adjust how they interpret ambiguous sentences (e.g., "I had carrots and *bees* for dinner.") by integrating information about the reliability of the incoming perceptual information with their expectations about the speaker (Gibson et al., 2013; Yurovsky et al., 2017). While children's behavior paralleled adults, they relied more on top-down expectations about the speaker, perhaps because their perceptual representations were noisier. These developmental differences provide insight into how children succeed in understanding language despite having partial knowledge of word-object mappings.

The critical behavioral prediction is that children and adults will adapt the timing of their eye movements to facilitate word recognition. We hypothesized that as fixations to the source of language – either a signer or a speaker – provide higher value visual information, listeners should prioritize looking to their social partner. Concretely, in a noisy auditory environment, looks to a speaker's face should be more useful compared to the same behavior without background noise where it is easier to perceive the acoustic signal. In this case, we predict that listeners would be (a) slower to shift gaze away from the speaker's face, which in turn would lead to (b) more consistent shifts to named objects and (c) fewer early, nonlanguage-driven eye movements to the *rest* of the visual world.

3.2 Analytic approach

Before describing the empirical work, it is worth motivating our analytic approach. To quantify evidence for our predictions, for each experiment we present four analyses: (1) the time course of listeners' looking to each area of interest (AOI), (2) the Reaction Time (RT) and Accuracy of listeners' first shifts away from the signer/speaker, (3) an Exponentially Weighted Moving Average (EWMA) of first shifts, and (4) a Drift Diffusion Model (DDM) of first shifts.[2]

First, we analyzed the time course of participants' looking to each AOI in the visual scene as the target sentence unfolded. Proportion looking reflects the mean proportion of trials on which participants fixated on the signer/speaker, the target image, or the distracter image at every 33-ms interval of the stimulus sentence. We tested condition differences in the proportion looking to the

[2]All analysis code can be found in the online repository for this project: `https://github.com/kemacdonald/speed-acc`.

language source – signer or speaker – using a nonparametric cluster-based permutation analysis, which accounts for the issue of taking multiple comparisons across many time bins in the timecourse (see Maris & Oostenveld (2007) for details about the technique). This analysis tests the binary hypothesis of a difference between two time series and provides a high-level overview of how changes in the processing context modulated listeners' looking behavior. A higher proportion of looking to the language source across the trial would indicate listeners' prioritization of seeking information from the signer/speaker.

Next, we analyzed the RT and Accuracy of participants' initial gaze shifts away from the signer/speaker. RT corresponds to the latency of shifting gaze away from the central stimulus to either object measured from the onset of the target noun. We trimmed all reaction time distributions to between zero and two seconds and modeled RTs in log space. Accuracy corresponds to whether participants' first gaze shift landed on the target or the distracter object. This analysis of accuracy does not focus on the amount of time spent looking at the target vs. the distracter images – a measure typically used in analyses of the Visual World Paradigm. Moreover, while we use the term accuracy as a label for this dependent measure in our task, we do not want to claim that fixations to the distracter object are incorrect in a general sense since these behaviors could help listeners to encode the identity of objects in the scene better.

We chose to analyze first shifts because we think that they tend to reflect rapid decisions driven by accumulating information about the identity of the named object. This measure provides a window on to changes in the underlying dynamics of how listeners decide what kind of visual information to gather. If listeners generate slower but more accurate gaze shifts, this provides evidence that gathering more visual information from the signer/speaker led to more robust language comprehension.

We used the `rstanarm` (Gabry & Goodrich, 2016) package to fit Bayesian mixed-effects regression models. The mixed-effects approach allowed us to model the nested structure of our data – multiple trials for each participant and item, and a within-participants manipulation – by including random intercepts for each participant and item, and a random slope for each item and noise condition. We used Bayesian estimation to quantify uncertainty in our point estimates, which we communicate using a 95% Highest Density Interval (HDI). The HDI provides a range of credible values given the data and model. Finally, to estimate age-related differences, we fit two types of models: (1) age group (adults vs. children) as a categorical predictor and (2) age as a continuous predictor (measured

in days) within the child sample. In the main text, we report specific effects and contrasts of interest for our hypotheses, but, in the Appendix, we report the full model output for each analytic model in the paper.

Following the behavioral results, we present two model-based analyses. The goal of each model is to move beyond a description of the data and to map behavioral differences to underlying psychological processes. The EWMA models changes in the tendency to generate random gaze shifts as a function of when they occurred in the RT distribution (Vandekerckhove & Tuerlinckx, 2007). For each RT, the model generates two values: a "control statistic" (CS, which captures the running average accuracy of first shifts) and an "upper control limit" (UCL, which captures the pre-defined limit of when we consider accuracy to be better than guessing). Here, the CS is an expectation of random shifting to either the target or the distracter image (nonlanguage-driven shifts), or a Bernoulli process with probability of success 0.5. As RTs get slower, we assume that participants have gathered more information and should become more accurate (i.e., language-driven), or a Bernoulli process with probability success > 0.5. Using this model, we can quantify the proportion of gaze shifts that were classified as language-driven as opposed to guessing. If listeners seek more visual information from the language source, then they should generate more language-driven shifts and fewer random responses.

Finally, following Vandekerckhove & Tuerlinckx (2007), we selected the gaze shifts categorized as language-driven by the EWMA and fit a hierarchical Bayesian Drift-Diffusion Model (HDDM). The DDM is a cognitive model of decision making developed over the past forty years (Ratcliff & McKoon, 2008) that can help to quantify differences in the underlying decision process that lead to different patterns of observable behavior. The model assumes that people accumulate noisy evidence in favor of one alternative with a response generated when the evidence crosses a predefined decision threshold. We chose to implement a hierarchical Bayesian version of the DDM using the HDDM Python package (Wiecki, Sofer, & Frank, 2013) since we had relatively few trials from child participants and recent simulation studies have shown that the HDDM approach was better than other fitting methods for small data sets (Ratcliff & Childers, 2015). Here, we focus on two parameters of interest for our hypotheses: *boundary separation*, which indexes the amount of evidence gathered before generating a response (higher values suggest more cautious responding) and *drift rate*, which indexes the amount of evidence accumulated per unit time (higher values

Table 3.1: Age distributions of children in Experiment 1. All ages are reported in months.

Center Stimulus	Mean	Min	Max	n
ASL	27.90	16.00	53.00	30.00
Face	26.00	25.00	26.00	24.00
Object	31.90	26.00	39.00	40.00
Bullseye	26.10	26.00	27.00	16.00

suggest more efficient processing). If listeners have a higher boundary separation estimate, this provides evidence that changes in information accumulation, as opposed to processing efficiency, led to higher accuracy rates.

3.3 Experiment 1

In Experiment 1, we compared eye movements of children learning American Sign Language to children learning a spoken language using parallel real-time language comprehension tasks. In the task, children processed familiar sentences (e.g., "Where's the ball?") while looking at a simplified visual world with three fixation targets (a center stimulus that varied by condition, a target picture, and a distracter picture; see Figure 3.2). The spoken language data are a reanalysis of three unpublished data sets, and the ASL data are reported in MacDonald et al. (2018a). Our primary question of interest is whether processing a sign language like ASL would increase the value of fixating on the language source and decrease the value of generating exploratory, nonlanguage-driven shifts even after the disambiguation point in the linguistic signal. If ASL learners are sensitive to the cost of shifting gaze away from a signer, then they would show evidence of prioritizing accuracy over and above the speed of shifting gaze to the named object.

3.3.1 Methods

Participants

Table 1 contains details about the age distributions of children in all four samples.

Spoken English samples. Participants were 80 native, monolingual English-learning children divided across three samples with no reported history of developmental or language delay.

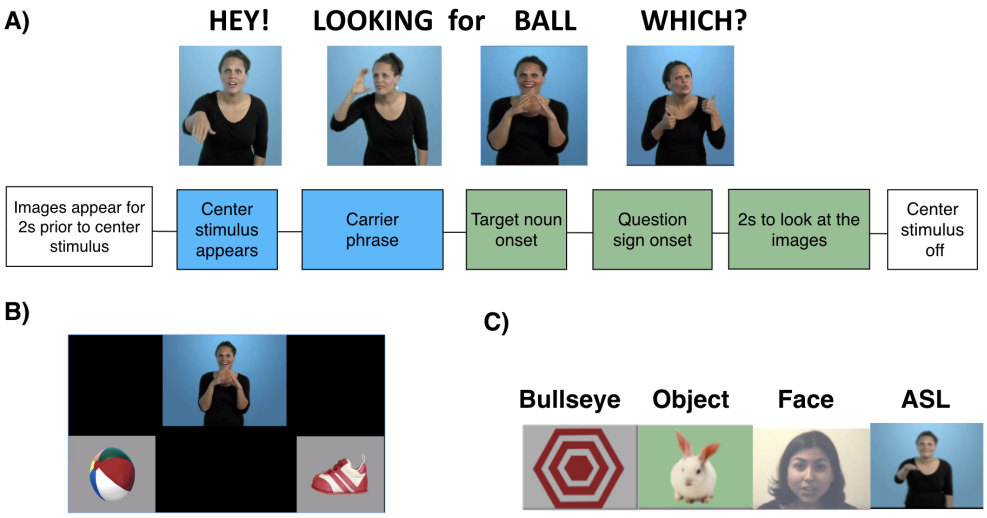


Figure 3.2: Stimuli for Experiments 1 and 2. Panel A shows the timecourse of the linguistic stimuli for a single trial for children learning American Sign Language. Panel B shows the layout of the fixation locations for all tasks: the center stimulus, the target, and the distracter. Panel C shows the four center stimulus items: a static geometric shape (Bullseye), a static image of a familiar object (Object), a person speaking (Face), a person signing (ASL).

ASL sample. Participants were 30 native, monolingual ASL-learning children (18 deaf, 12 hearing). 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. The ASL sample included a wider age range compared to the spoken English samples because this is a rare population.

Stimuli

There are differences between ASL and English question structures. However, all linguistic stimuli shared the same trial structure: language to attract participants' attention followed by a sentence containing a target noun.

ASL linguistic stimuli. We recorded two sets of ASL stimuli, using two valid ASL sentence structures for questions: 1) Sentence-initial wh-phrase: "HEY! WHERE [target noun]?" and 2) Sentence-final wh-phrase: "HEY! [target noun] WHERE?" Two female native ASL users recorded several tokens of each sentence in a child-directed register. Before each sentence, the signer produced

a common attention-getting gesture. Mean sign length was 1254 ms, ranging from 693 ms to 1980 ms.

English linguistic stimuli. All three tasks (Object, Bullseye, and Face) featured the same female speaker who used natural child-directed speech and said: "Look! Where's the (target word)?" The target words were: ball, banana, book, cookie, juice, and shoe. For the Face task, a female native English speaker was video-recorded as she looked straight ahead and said, "Look! Where's the (target word)?" Mean word length was 786.7 ms, ranging from 600 ms to 940 ms.

ASL and English visual stimuli. The image set consisted of colorful digitized pictures of objects presented in fixed pairs with no phonological overlap (ASL task: cat—bird, car—book, bear—doll, ball—shoe; English tasks: book-shoe, juice-banana, cookie-ball). Side of target picture was counterbalanced across trials.

Design and procedure

Trial structure. Children sat on their caregiver's lap and viewed the task on a screen while their gaze was recorded using a digital camcorder. On each trial, the child saw two images of familiar objects on the screen for two seconds before the center stimulus appeared. This time allowed the child to visually explore both images. Next, the target sentence – which consisted of a carrier phrase, target noun, and question sign – was presented, followed by two seconds without language to allow the child to respond to the signer's sentence. The trial structure of the Face, Object, and Bullseye tasks were highly similar: children were given two seconds to visually explore the objects prior to the appearance of the center stimulus, then processed a target sentence, and finally were given two seconds of silence to generate a response to the target noun. Participants saw 32 test trials with several filler trials interspersed to maintain interest.

Coding. Participants' gaze patterns were videotaped and later coded frame-by-frame at 33-ms resolution by 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

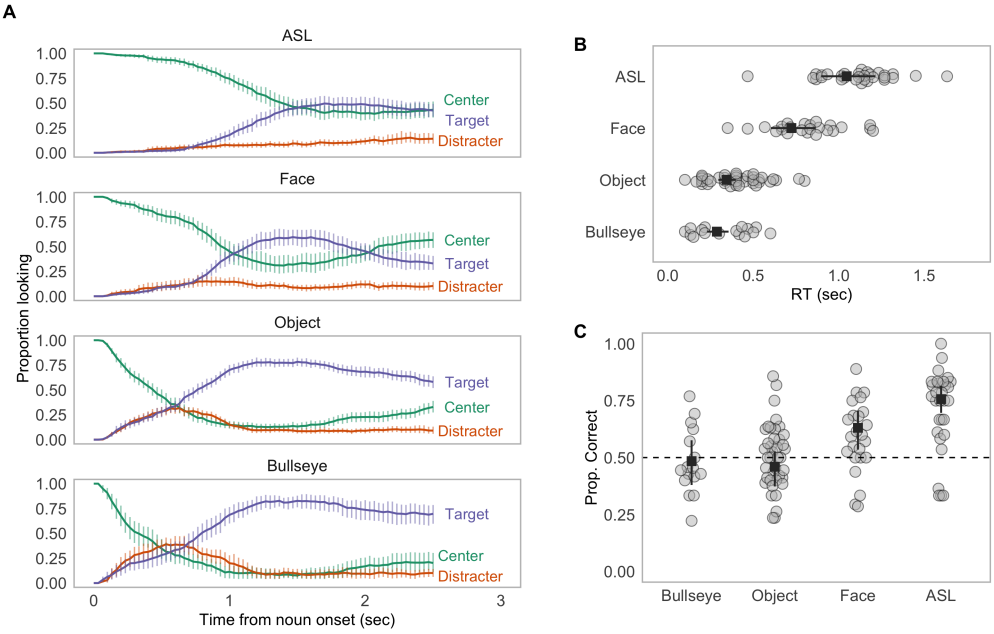


Figure 3.3: Timecourse looking, first shift Reaction Time (RT), and Accuracy results for children in Experiment 1. Panel A shows the overall looking to the center, target, and distracter stimulus for each context. Panel B shows the distribution of RTs for each participant. Each point represents a participant's average RT. The black squares represent the group means. And the error bars represent 95% Highest Density Intervals around the group means. Panel C shows the same information but for participants' first shift accuracy.

assessments.

3.3.2 Results

Behavioral analyses

Timecourse looking. The first question of interest was how do young ASL and English learners distribute attention across the three fixation locations while processing language in real-time? Figure 3.3A presents an overview of children's looking to each AOI for each processing context. This plot shows changes in the mean proportion of trials on which participants fixated the center stimulus, the target image, or the distracter image at every 33-ms interval of the stimulus sentence. At target-noun onset, children tended to look at the center stimulus. As the target noun unfolded, the mean proportion looking to the center 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 all four contexts.

After looking to the target image, participants tended to shift their gaze back to the center, shown by the increase in proportion looking to the center around two seconds after target-noun onset. There were several qualitative differences in looking behavior across the different center stimulus types. First, both ASL- and English-learners who processed sentences from a video of speaker spent more time looking to the center as indicated by the shallower slope on their center-looking curves. Second, when the center stimulus was a static geometric object (Bullseye) or a static familiar object (Object), spoken language learners were more likely to look at the distracter image, especially early in the timecourse of the target noun as indicated by the parallel increase in target and distracter-looking curves in Figure 3.3A. In contrast, spoken language learners in the Face context spent less time looking at the disracter, and ASL-learners rarely looked to the distracter image at any point in the trial. This pattern of behavior provides qualitative evidence that children adapated their gaze depending on language-relevant information available in the center stimulus loaction.

Based on a nonparametric cluster-based permutation analysis, the center-looking curve for the ASL learners was significantly different from all other conditions (all $p < .001$). Within the spoken language groups, children's looking to a speaker's face was different from looking to the Bullseye and the Familar object ($p < .001$). Finally, the Object and Bullseye center-looking curves were not different from one another, with no significant differences at any point in the timecourse. Next, we ask how these different processing contexts changed the timing and accuracy of children's initial decisions to shift away from the center stimulus.

RT. Figure 3.3B shows the full RT data distribution. To quantify differences across the groups, we fit a Bayesian linear mixed-effects regression predicting first shift RT as a function of center stimulus type controlling for age: `Log(RT)` $\sim$ `center stimulus type + age + (1 | subject) + (1 | item)`. ASL learners generated slower RTs compared to all of the spoken English samples ($\beta =$ 595.2 ms, 95% HDI [444.6, 760.8]). Moreover, ASL learners' shifts were slower compared directly to children processing spoken language in the Face condition ($\beta = 323.1$ ms, 95% HDI [132.3, 522.6]). Finally, children in the Face context shifted gaze slower compared to participants in the Object and Bullseye contexts ($\beta = 408.2$ ms, 95% HDI [286.6, 546.2]).

Accuracy. Next, we compared the accuracy of first shifts across the different tasks (Figure 3.3C)

by fitting a mixed-effects logistic regression with the same specifications and contrasts as the RT model. We found that (a) ASL learners were more accurate compared to all of the spoken English samples (β = 0.23, 95% HDI [0.17, 0.29]), (b) ASL learners were more accurate when directly compared to participants in the Face task (β = 0.13, 95% HDI [0.04, 0.23]), (c) children learning spoken language were more accurate when processing language from dynamic video of a person speaking compared to the Object and Bullseye tasks (β = 0.16, 95% HDI [0.07, 0.24]), and (d) English-learners' first shifts were no different from random responding in the Object (β = -0.04, 95% HDI [-0.13, 0.03]) and Bullseye (β = -0.02, 95% HDI [-0.12, 0.08]) contexts.

Model-based analyses

EWMA. Our third question of interest was how the tendency to generate random vs. language-driven (i.e, accurate) gaze shifts evolved as a function of reaction time across the different processing contexts. Figure 3.4A shows changes in the control statistic (CS) and the upper control limit (UCL) as a function of RT. Each CS starts at chance performance and below the UCL. In the ASL and Face tasks, the CS value begins to increase with RTs around 0.7 seconds after noun onset and eventually crosses the UCL, indicating that responses $>$ 0.7 sec were on average above chance levels. In contrast, the CS in the Object and Bullseye tasks never crossed the UCL, indicating that children's shifts were equally likely to land on the target or the distracter, regardless of when they were initiated. This result suggests that first shifts measured in the Bullseye/Object tasks were qualitatively different behaviors than those in the ASL and Face contexts. That is, these shifts are likely the result of a different generative process such as gathering more information about the referents in the visual world.

Next, we compared the EWMA model fits for participants in the ASL and Face processing contexts since these groups showed evidence of language-driven responding. We found that ASL learners generated fewer shifts when the CS was below the UCL compared to children learning spoken language (β = 0.14, 95% HDI [0.08, 0.23]). This result indicates that ASL-learners were more likely to have gathered sufficient information about the linguistic signal prior to shifting gaze away from the language source. We found some evidence that ASL learners started producing language-driven shifts earlier in the RT distribution as indicated by the point at which the CS crossed the UCL (β = 0.22 sec, 95% HDI [0.05, 0.39]), indicating that these children were less likely

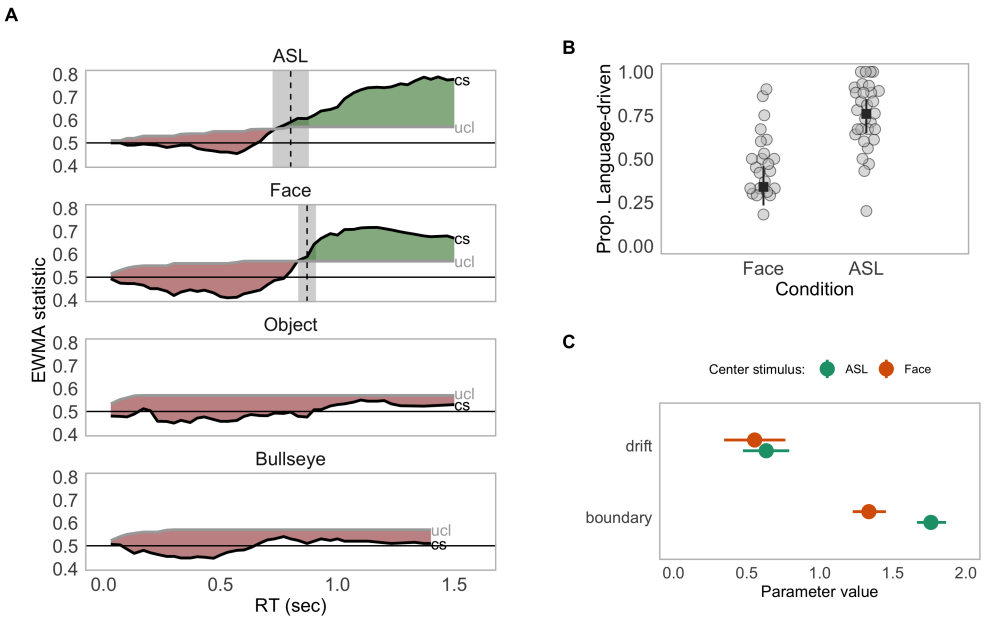


Figure 3.4: Results for the model-based analyses in Experiment 1. Panel A shows a control chart representing the timecourse of the EWMA statistic. The black curve represents the evolution of the control statistic (CS) as a function of reaction time. The grey curve represents the upper control limit (UCL) or the pre-defined upper limit on random responding. The vertical dashed line is the median cutoff value (point in the RT distribution when children's gaze shifts were no longer random). The grey shaded area represents the 95% Highest Density Interval around the estimate of the median cutoff point, and the shaded ribbons represent the classification of responses as guesses (red, below the UCL) and language-driven (green, above the UCL). Panel B shows a summary of the proportion of shifts that were categorized as language-driven for the Face and ASL processing contexts. Panel C shows the point estimate and 95% Highest Density Intervals for the boundary and drift rate parameters for the Face and ASL contexts.

to generate early, random gaze shifts away from the signer.

HDDM. We fit a hierarchical Drift Diffusion Model using only the gaze shifts categorized as language-driven by the EWMA. This allowed us to ask what underlying decision processes are likely to account for the measured differences in First Shift Accuracy and RT.[3] ASL learners had a higher estimate for the boundary separation parameter compared to children processing spoken English from a speaker (ASL boundary = 1.76, 95% HDI [1.65, 1.88]; Face boundary = 1.34, 95% HDI [1.21, 1.47]), with no overlap in the HDIs (see Figure 3.4C). This suggests that ASL learners' higher

[3]We chose not to interpret the HDDM fits for the Bullseye or Face tasks since there was no suggestion of any non-guessing signal from the EWMA analysis.

accuracy was driven by accumulating more evidence about the linguistic signal before generating an eye movement. We found high overlap for estimates of the drift rate parameter, indicating that both groups processed the linguistic information with similar efficiency (ASL drift = 0.63, 95% HDI [0.44, 0.82]; Face drift = 0.55, 95% HDI [0.3, 0.8]).

3.3.3 Discussion

The behavioral and model-based analyses provide converging support that ASL learners were sensitive to the value of delaying eye movements away from the language source. Compared to spoken language learners, ASL learners prioritized accuracy over speed (HDDM), produced fewer nonlanguage-driven shifts away from the center stimulus (EWMA), and were more accurate with these gaze shifts (behavioral). Importantly, we did not see evidence in the HDDM model fits that these accuracy differences could be explained by differential efficiency in processing the linguistic information. Instead, the pattern of results suggests that ASL learners increased their decision threshold to gather more information before shifting gaze away from the signer and to a named object.

We hypothesized that prioritizing accuracy of gaze shifts when processing a visual-manual language is an adaptive response. That is, to map referential language to the visual world in ASL involves competition for visual attention. When ASL learners choose to shift their gaze away from a signer, they are leaving an area that provides a great deal of useful information. Further, unlike children learning spoken languages, ASL learners cannot gather more of the linguistic signal if their gaze is directed away from a signer. Thus, it seems reasonable that ASL learners would adapt the timing of their gaze shifts to gather additional information that increases certainty in their comprehension before seeking a named object.

These findings, however, were based on exploratory analyses, and our information-seeking explanation was developed to explain this pattern of results. There are several, potentially significant differences between the stimuli, apparatus, and populations that limit the strength of our interpretation and the generality of our account. We also cannot make causal conclusions because of the observational nature of research comparing children who are learning different languages. Thus, we designed Experiment 2 to address these concerns by constructing a well-controlled situation that created information-seeking demands that are analogous on some dimensions to the modality-based differences in Experiment 1.

3.4 Experiment 2

In Experiment 2, we created a well-controlled experimental context where we could manipulate the information-seeking demands in ways that parallel some aspects of the differences between young ASL- and English-learners.[4] We measured adults and children's eye movements during a real-time language comprehension task where participants processed familiar sentences (e.g., "Where's the ball?") while looking at a simplified visual world with three fixation targets. Using a within-participants design, we manipulated the signal-to-noise ratio of the auditory signal by convolving the acoustic input with brown noise (random noise with higher energy at lower frequencies). We chose to manipulate background noise because it allowed us to increase the value of looking to a speaker for language comprehension, and it could be used with both adults and children.

We predicted that processing speech in a noisy context would make participants less likely to shift before collecting sufficient information indexed by a lower proportion of shifts flagged as random/exploratory in the EWMA analysis. We also predicted a developmental difference: that children would produce a higher proportion of random shifts and accumulate information less efficiently compared to adults (indexed by lower estimates of drift and boundary separation in the DDM), and a developmental parallel: that children would also adapt gaze to gather additional visual information from the speaker in the noisier auditory environment.

3.4.1 Methods

Participants

Participants were native, monolingual English-learning children (n = 39; 22 F) and adults (n = 31; 22 F). All participants had no reported history of developmental or language delay and normal vision. 14 participants (11 children, 3 adults) were run but not included in the analysis because either the eye tracker failed to calibrate (2 children, 3 adults) or the participant did not complete the task (9 children).

Stimuli

Linguistic stimuli. The video/audio stimuli were recorded in a sound-proof room and featured two female speakers who used natural child-directed speech and said one of two phrases: "Hey! Can you

[4] See `https://osf.io/g8h9r/` for a pre-registration of the analysis plan.

find the (target word)" or "Look! Where's the (target word). The target words were: ball, bunny, boat, bottle, cookie, juice, chicken, and shoe. The target words varied in length (shortest = 411.68 ms, longest = 779.62 ms) with an average length of 586.71 ms.

Noise manipulation. To create the stimuli in the noise condition, we convolved each recording with Brown noise using the Audacity audio editor. The average signal-to-noise ratio (values greater than 0 dB indicate more signal than noise) in the noise condition was 2.87 dB compared to the clear condition, which was 35.05 dB.

Visual stimuli. The image set consisted of colorful digitized pictures of objects presented in fixed pairs with no phonological overlap between the target and the distracter image (cookie-bottle, boat-juice, bunny-chicken, shoe-ball). The side of the target picture was counterbalanced across trials.

Design and procedure

Participants viewed the task on a screen while their gaze was tracked using an SMI RED corneal-reflection eye-tracker mounted on an LCD monitor, sampling at 30 Hz. The eye-tracker was first calibrated for each participant using a 6-point calibration. On each trial, participants saw two images of familiar objects on the screen for two seconds before the center stimulus appeared. Next, they processed the target sentence – which consisted of a carrier phrase, a target noun, and a question – followed by two seconds without language to allow for a response. Child participants saw 32 trials (16 noise trials; 16 clear trials) with several filler trials interspersed to maintain interest. Adult participants saw 64 trials (32 noise; 32 clear). The noise manipulation was presented in a blocked design with the order of block counterbalanced across participants.

3.4.2 Results and Discussion

Behavioral analyses

Timecourse looking. Figure 3.5A presents an overview of looking to the speaker, target, and distracter images for the noisy and clear processing contexts from the start of the target noun. Similar to the results in Experiment 1, participants tended to fixate on the speaker at target-noun onset. As the target noun unfolded, the mean proportion looking to the center decreased rapidly as participants shifted their gaze to the objects. Proportion looking to the target increased sooner and

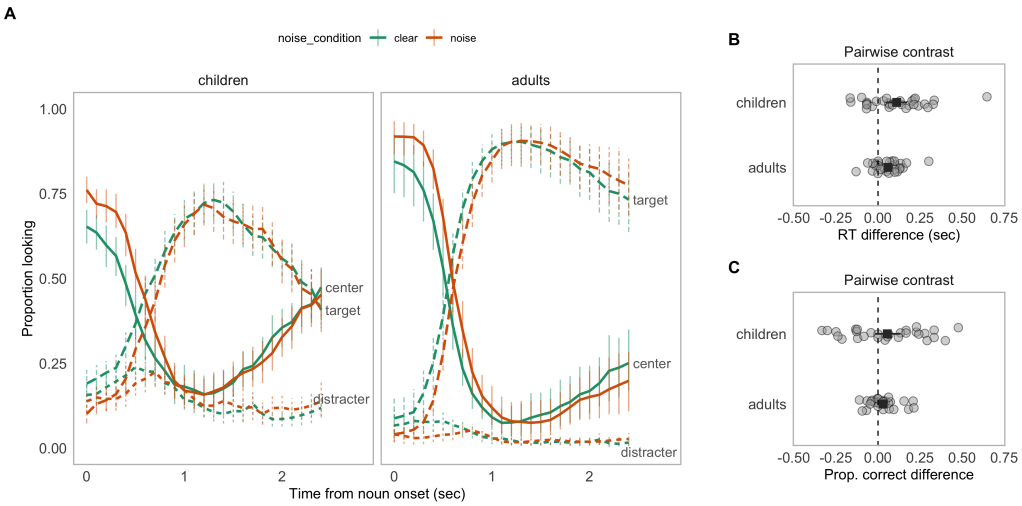


Figure 3.5: Behavioral results for children and adults in Experiment 2. Panel A shows the overall looking to the center, target, and distracter stimulus for each processing condition and age group. Panel B shows the distribution of RTs for each participant and the pairwise contrast between the noise and clear conditions. The square point represents the mean value for each mesure. The vertical dashed line represents the null model of zero condition difference. The width each point represents the 95% HDI. Panel C shows the same information but for first shift accuracy.

reached a higher asymptote compared to proportion looking to the distracter for both processing contexts and age groups. After looking to the target image, participants tended to shift their gaze back to the speaker as shown by the increase in center looking curve around 1 second.

There are several developmental differences to highlight. First, children tended to look more to the objects at noun onset, as indicated by the lower intercept of children's center-looking curves. Second, children's target looking curves reached a lower asymptote as compared to adults and they spent relatively more time fixating on the distracter image, whereas adults rarely looked at the unnamed object after 0.5 seconds in the timecourse of the trial. And third, children showed a stronger tendency to shift back to the speaker after looking to the named object.

Visual inspection of the center looking curves suggests a difference in looking behavior in the noisy processing context. Both children and adult's spent more time fixating on the speaker when the auditory signal was less reliable as indicated by the rightward shift of the center-looking curves in the noisy condition. A cluster-based permutation test confirmed that there was evidence of a significant difference in looking to the speaker between the Noisy and Clear conditions ($p < .05$).

This pattern of behavior provides evidence that reducing the quality of the auditory signal increased looking to the speaker early in the timecourse of the target noun.

RT. Figure 3.5B shows the full distribution of the estimated RT differences between each participants' performance in the noisy and clear contexts. Both children and adults were slower to identify the target in the noise condition (Children M_{noise} = 500.2 ms; Adult M_{noise} = 595.2 ms), as compared to the clear condition (Children M_{clear} = 455.7 ms Adult M_{clear} = 542.4 ms). RTs in the noise condition were 48.8 ms slower on average, with a 95% HDI ranging from 3.7 ms to 96.3 ms, and not including the null value of zero condition difference. Older children responded faster than younger children (β_{age} = -0.44, [-0.74, -0.16]), with little evidence for an interaction between age and condition within the child sample.

Accuracy. Next, we modeled adults and children's first shift accuracy using a mixed-effects logistic regression with the same specifications (Figure 3.5C). Both groups were more accurate than a model of random responding with the null value of 0.5 falling well outside the lower bound of the 95% HDI for each group mean. Adults were more accurate (M_{adults} = 90%) than children ($M_{children}$ = 61%). Interestingly, both groups showed evidence of higher accuracy in the noise condition: children (M_{noise} = 67%; M_{clear} = 61%) and adults (M_{noise} = 92%; M_{clear} = 90%). Accuracy in the noise condition was on average 4% higher, with a 95% HDI from -1% to 12%. Note that the null value of zero difference falls at the very edge of the HDI. But 95% of the credible values are greater than zero, providing evidence for comparable, if not higher, accuracy in the noise condition. Within the child sample, there was no evidence of a main effect of age or an interaction between age and noise condition.

Model-based analyses

EWMA. Figure 3.6A shows the proportion of shifts that the model classified as random vs. language-driven for each age group and processing context. On average, 41% (95% HDI: 32%, 50%) of children's shifts were categorized as language-driven, which was significantly fewer than adults, 87% (95% HDI: 78%, 96%). Critically, processing speech in a noisy context caused both adults and children to generate a higher proportion of language-driven shifts (i.e., fewer random, exploratory shifts away from the speaker), with the 95% HDI excluding the null value of zero condition difference (β_{noise} = 11%, [7%, 16%]). Within the child sample, older children generated fewer random,

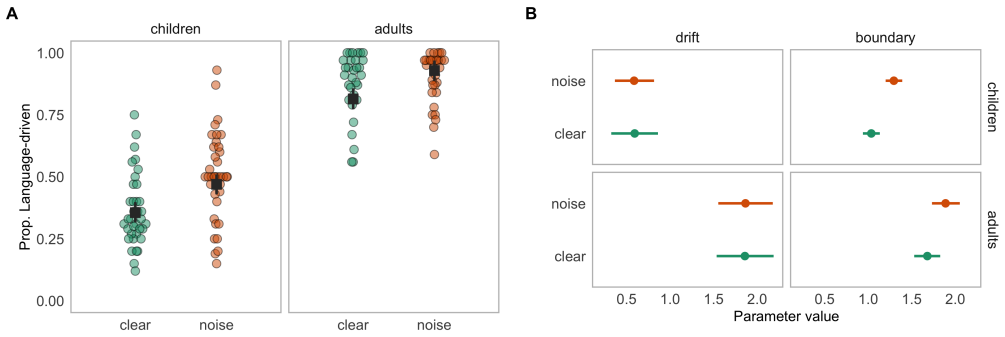


Figure 3.6: Results for the model-based analyses for Experiment 2. The plotting conventions are the same as Figure 3.

early shifts (M_{age} = -0.21, [-0.35, -0.08]). There was no evidence of an interaction between age and condition. This pattern of results suggests that the noise condition caused participants to increase visual fixations to the language source, leading them to generate fewer exploratory, random shifts before they had accumulated sufficient information to respond accurately.

HDDM. Figure 3.6B shows the full posterior distributions for the HDDM output. Children had lower estimates of drift rate (children $M_{drift} = 0.59$; adults $M_{drift} = 1.9$) and boundary separation (children $M_{boundary} = 1.16$; adults $M_{boundary} = 1.67$) as compared to adults, suggesting that children were less efficient and gathered less information. The noise manipulation selectively affected the boundary separation parameter, with higher estimates in the noise condition for both age groups ($\beta_{noise} = 0.26$, [0.1, 0.42]). This result suggests that participants in the noise condition prioritized information accumulation over speed when generating an eye movement in response to the incoming language, and this increased decision threshold led to higher accuracy. Moreover, the high overlap in estimates of drift rate suggests that participants were able to integrate the visual and auditory signals such that they could achieve a level of processing efficiency comparable to the clear processing context.

The behavioral and model-based results provide evidence for our information-seeking explanation of eye movements during grounded language comprehension. Processing speech in noise caused both children and adults to look longer at their social partner, which in turn, resulted in a higher proportion of accurate gaze shifts to a named object. Moreover, we observed a similar pattern of behavior in children and adults, with both groups producing more language-driven shifts (EWMA) and prioritized accuracy over speed (HDDM) in the more challenging noisy environment.

Our analysis plan was preregistered, but there were some cases where we deviated or did not predict a particular result. We successfully predicted that the noise manipulation would cause listeners to gather more information by looking longer at the speaker's face (slower RTs) and that this behavior would lead listeners to produce a higher proportion of language-driven shifts as indexed by the EWMA analysis. We did not, however, predict that first shifts would be *more* accurate in the noisier context and that the noise manipulation would selectively affect the boundary separation parameter in the HDDM. Finally, based on time and resource constraints, we did not collect our planned sample size (42) for each age group (3-, 4-, and 5-year-olds) in the target age range; instead, we decided to use age as a continuous covariate and collected a single sample that included children across the entire age range.

3.5 General Discussion

Language comprehension in grounded, social contexts provides children access to a rich set of multimodal cues that could support the linking of linguistic information to the world. But do children flexibly select what information to gather? In this work, we proposed that listeners adapt their gaze to seek visual information from their social partners when it was especially useful for language comprehension. We presented evidence for this explanation by measuring changes in how children chose to allocate visual attention across two diverse language processing contexts. In Experiment 1, we found that, compared to children learning spoken English, young ASL-learners delayed their gaze shifts away from a language source, were more accurate, and produced a higher proportion of language-driven eye movements. In Experiment 2, we showed that 3-5 year-olds and adults delayed the timing of gaze shifts away from a speaker's face while processing speech in a noisy auditory environment. This slower response resulted in fewer nonlanguage-driven eye movements and more accurate gaze shifts.

These results synthesize ideas from several research programs, including work on language-driven visual attention (Tanenhaus et al., 1995), goal-based accounts of vision during everyday tasks (M. Hayhoe & Ballard, 2005), and work on language perception as multisensory integration (Vigliocco et al., 2014). Our findings also parallel the results of several recent studies that measure the adaption of visual processes in response to different auditory experiences. First, Heimler et al. (2015) compared Deaf and hearing adults' performance on an oculomotor singleton detection paradigm where

participants made speeded eye-movements to a unique target embedded among distracters that varied in saliency. Deaf adults were slower to generate a gaze shift away from the center fixation and, as a result, they were less affected by high saliency distracters. Second, McMurray, Farris-Trimble, & Rigler (2017) found that individuals with cochlear implants, who are consistently processing degraded auditory input, are more likely to delay the process of lexical access as measured by slower gaze shifts to named referents and fewer incorrect gaze shifts to phonological onset competitors. McMurray et al. (2017) also found that they could replicate these changes in adults with typical hearing by using noise-vocoded speech stimuli that shared features with the output of a cochlear implant.

Our findings also connect to the literature investigating how experience with a visual-manual language may change basic cognitive processes (see Bavelier, Dye, & Hauser (2006) for a review). The upshot of this work is that the effects of Deafness are dissociable from the effects of learning a signed language. Specifically, Deaf individuals show selective enhancement in peripheral visual attention as evidenced by higher sensitivity to peripheral distracters on spatial orienting tasks. In contrast, learning to sign results in several specific changes such as enhanced mental imagery (Emmorey, Kosslyn, & Bellugi, 1993), mental rotation (Emmorey, Klima, & Hickok, 1998), and face processing (Bettger, Emmorey, McCullough, & Bellugi, 1997). The results of Experiment 1 suggest that ASL learners adapt the timing of when they disengage from a language source to increase their certainty before seeking named object. It is an open question as to whether ASL-learners' differential responding is best explained by a lack of access to auditory information or learning a visual-manual language.

Finally, our results dovetail with recent developmental work by Yurovsky et al. (2017). In their study, preschoolers, like adults, were able to integrate top-down expectations about the kinds of things speakers are likely to talk about with bottom-up cues from auditory perception. Yurovsky et al. (2017) situated this finding within the framework of modeling language as a *noisy channel* where listeners combine expectations with perceptual data and weight each based on its reliability. In Experiment 2, we found a similar developmental parallel in language processing: that 3-5 year-olds, like adults, adapted their gaze patterns to seek additional visual information when the auditory signal became less reliable. This adaptation allowed listeners to generate comparable, if not more, accurate responses in the noisy context.

In sum, the work reported here shows that young listeners can seek visual information to support language comprehension. These results fit well with the interactive models of language perception reviewed in the Introduction (M. C. MacDonald & Seidenberg, 2006; McClelland et al., 2006). These studies also highlight the value of integrating observational and experimental approaches. In Experiment 1, we compared language comprehension across populations of children who had very different language experiences (signed vs. spoken) to generate a new explanation of observed differences in children's gaze dynamics. We were able to better understand this observational result by designing a well-controlled, follow-up experiment that allowed us to make stronger claims about the generality of our hypothesis.

3.5.1 Limitations and future work

Our results provide evidence that young listeners can adapt their gaze patterns to the demands of different processing environments to seek visual information from social partners that support language comprehension. We cannot, however, make claims about how children's behavior in our task (the Visual World Paradigm: VWP) would generalize to their decisions about how to distribute attention within real-world learning environments. There is a growing body of research showing meaningful links between children's gaze behavior in the VWP and relevant outcome measures such as vocabulary development (Fernald et al., 2006; Marchman & Fernald, 2008; Rigler et al., 2015). Nonetheless, a valuable next step for our work would be to leverage tasks that move closer to the ecological context in which children process and learn language such as using head-mounted cameras and eye trackers that would allow measurement of where children choose to look during everyday interactions (Fausey, Jayaraman, & Smith, 2016; Franchak, Kretch, Soska, & Adolph, 2011).

This work has several other limitations. First, we chose to focus on a single decision about visual fixation to provide a window onto the dynamics of decision-making across different language processing contexts. But our analysis does not consider the rich information present in the gaze patterns that occur leading up to this decision. In our future work, we aim to measure how changes in the language environment might lead to shifts in the dynamics of gaze across a longer timescale. For example, perhaps listeners gather more information about the objects in the scene before the sentence in anticipation of allocating more attention to the speaker once they start to speak.

Second, we chose one instantiation of a noisy processing context – random background noise.

But we think our findings should generalize to settings where other kinds of noise – e.g., uncertainty over a speaker's reliability or when processing accented speech – make gathering visual information from the speaker more useful for language understanding. Moreover, we used a simple visual world, with only three places to look, and simple linguistic stimuli. Thus it remains an open question how these results might scale up to more complex language interactions and visual environments. It could be that looks to a speaker become even more useful for disambiguating reference in complex visual environments.

Third, we do not yet know what might be driving the population differences between children learning ASL and children learning spoken English found in Experiment 1. It could be that ASL-learners' massive experience dealing with competition for visual attention leads to changes in the deployment of eye movements during language comprehension. Or, it could be that the in-the-moment constraints of processing a visual language cause different fixation behaviors. This question could be addressed by studies that measure how quickly listeners adapt the dynamics of gaze when visual information becomes more useful. Another interesting approach would be to measure eye movements in hearing children learning both a signed and a spoken language (bimodal bilinguals). Our prior work found that hearing and Deaf native ASL learners show remarkably similar looking patterns over the time course of processing familiar signs (MacDonald et al., 2018a). It would be interesting if hearing signers also prioritize accuracy over speed when comprehending their spoken language. This result would suggest that experience with a visual-manual language is changing a general response strategy, but if gaze dynamics looked different across language modalities within an individual, then this would favor an explanation based on the in-the-moment constraints of processing a visual-manual language in real-time.

Finally, our eye tracking paradigm removes an important component of successful communication: dynamic interaction between the speaker and listener. It is interesting to consider how speakers might adapt their behavior present the listener with useful visual information in challenging comprehension contexts. For example, in noisy environments, speakers will exaggerate mouth movements (Fitzpatrick, Kim, & Davis, 2011) and increase the frequency of gestural cues such as head nodding (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004), and parents exaggerate mouth movements during infant-directed speech (Green, Nip, Wilson, Mefferd, & Yunusova, 2010). Moreover, observational studies of parent-child interactions in signed languages show variability in how

sensitive adult signers are to the competing demands on children's visual attention (M. Harris & Mohay, 1997). That is, some interactions contain many utterances that young signers miss because they are fixating on objects; whereas other interactions are marked by adaptations that accommodate the demands on visual attention by parents displacing signs onto the objects that are currently the focus of children's attention (similar to follow-in labeling effects Tomasello and Farrar, 1986). Thus it is an open question how interacting with a speaker that adapts to increase the availability and utility of visual information might change children's decisions about visual fixation.

3.6 Conclusion

In this paper, we presented an information-seeking explanation for the differences in the dynamics of eye movements during grounded signed vs. spoken language comprehension. We started from an interesting, observational result: that ASL learners, compared to English-learning children, generate slower but more accurate gaze shifts away from a language source and to a named referent. We then tested the generality and causal claims of this explanation by experimentally manipulating the value of seeking visual information for language comprehension. We found that young listeners can adapt the dynamics of their gaze to gather visual information when it is useful for language understanding.

While we chose to start with the domain of familiar language processing, this approach could generalize to the acquisition context. Consider that early in language learning, children are acquiring novel word-object links while also learning about visual object categories. Both of these tasks produce different goals that should, in turn, modulate children's decisions about where to allocate visual attention – e.g., seeking nonlinguistic cues to reference such as eye gaze and pointing become critical when you are unfamiliar with the information in the linguistic signal. More generally, this approach presents a way forward for explaining decisions about visual fixation during language comprehension and acquisition across a broader set of processing contexts and at different stages of development.

Chapter 4

Social cues modulate attention and memory during cross-situational learning[1]

In this chapter, we present a series of studies exploring adults' word learning in the presence of social cues to word meanings. Within our broader active-social framework, these experiments investigate how social information can constrain the strength and the number of word-object hypotheses that learners store, which, in turn, could shape subsequent information seeking behaviors. Overall, this line of work brings together social and statistical accounts of word learning to ask how statistical learning mechanisms operate over social input.

Because learners hear language in environments that contain many things to talk about, figuring out the meaning of even the simplest word requires making inferences under uncertainty. A cross-situational statistical learner can aggregate across naming events to form stable word-referent mappings, but this approach neglects an important source of information that can reduce referential uncertainty: social cues from speakers (e.g., eye gaze). In four large-scale experiments with adults, we tested the effects of varying referential uncertainty in cross-situational word learning using social cues. Social cues shifted learners away from tracking multiple hypotheses and towards storing only

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

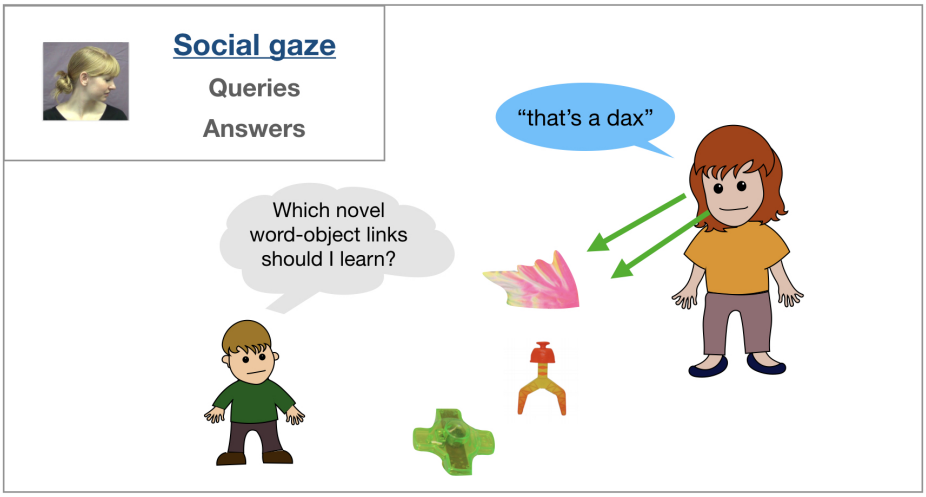


Figure 4.1: A schematic showing the components of the OED model captured by the case studies in Chapter 4.

a single hypothesis (Experiments 1 and 2). Also, learners were sensitive to graded changes in the strength of a social cue, and when it became less reliable, they were more likely to store multiple hypotheses (Experiment 3). Finally, learners stored fewer word-referent mappings in the presence of a social cue even when given the opportunity to visually inspect the objects for the same amount of time (Experiment 4). These results suggest that the representations underlying cross-situational word learning of concrete object labels flexibly respond to uncertainty in the input. And when ambiguity is high, learners tend to store a broader range of information.

4.1 Introduction

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

[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.

learning proceeds despite this uncertainty, with estimates of adult vocabularies ranging between 50,000 to 100,000 distinct words (P. 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; 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, 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 (K. 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, Medina, Hafri, & Gleitman, 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 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 (P. Bloom, 2002; E. V. 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 (D. A. 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, Nagell, Tomasello, Butterworth, & Moore, 1998). Moreover, studies outside the domain of language acquisition have shown that the presence of social cues: (a) produce better spatial learning of audiovisual events (R. 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 (J. M. 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

[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.

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.

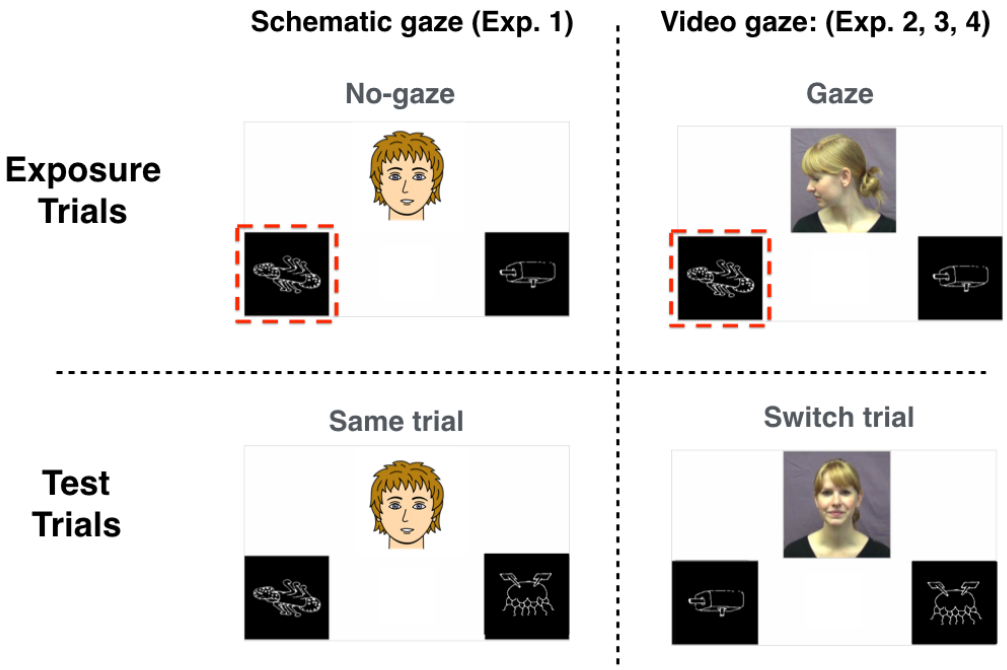


Figure 4.2: Screenshots of exposure and test trials from Experiments 1-4. The top left panel shows an exposure trial in the No-gaze condition using the schematic gaze cue (Experiment 4.1). The top right panel shows an exposure trial in the Gaze condition using the video gaze cue (Experiments 4.2-4.4). 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.

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 ^{TM}(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 $\alpha = .05$

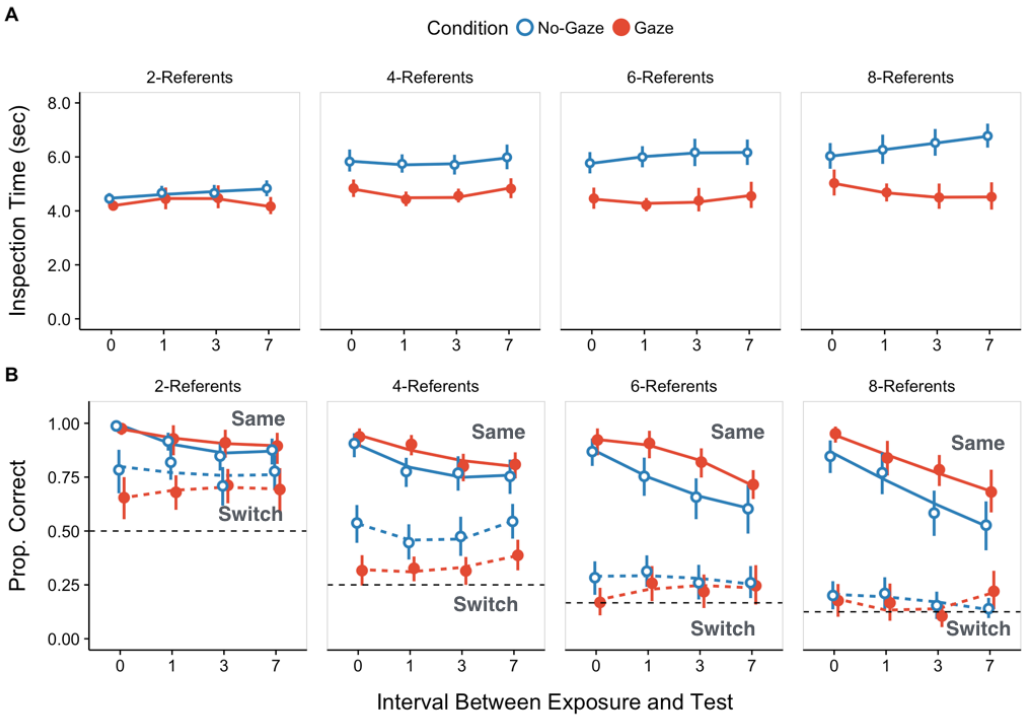


Figure 4.3: Experiment 4.1 results. The top row shows average inspection times on exposure trials for all experimental conditions as a function of the number of trials that occurred between exposure and test. Each panel represents a different number of referents, and line color represents the Gaze and 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.

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 ~ 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) ~ (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 (β = -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 (β = -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 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

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.

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 ~ (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

[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, $p = 0.006$).

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 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 ($M_{Exp1} = 0.8$, $M_{Exp2} = 0.91$), 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 β = 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 ($\beta = -2.99$, $p < .001$).

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

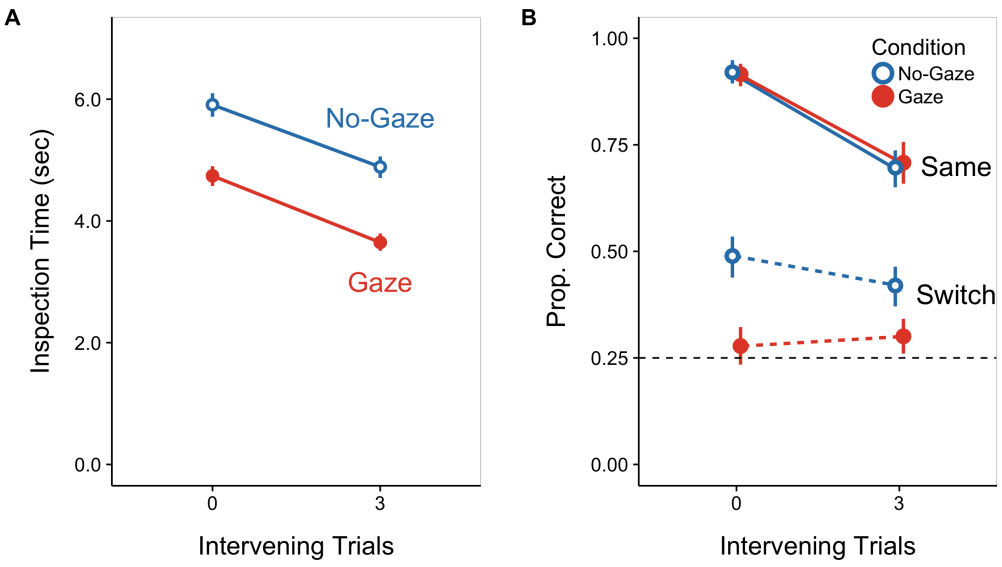


Figure 4.4: Experiment 2 results. Panel A shows inspection times on exposure trials with and without gaze. Panel B shows accuracy on Same and Switch test trials. All plotting conventions are the same as in Figure 2. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

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,

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.

$p < .001$). We also found an interaction between inspection time and trial type ($\beta = 0.21$, $p = 0.05$), 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."

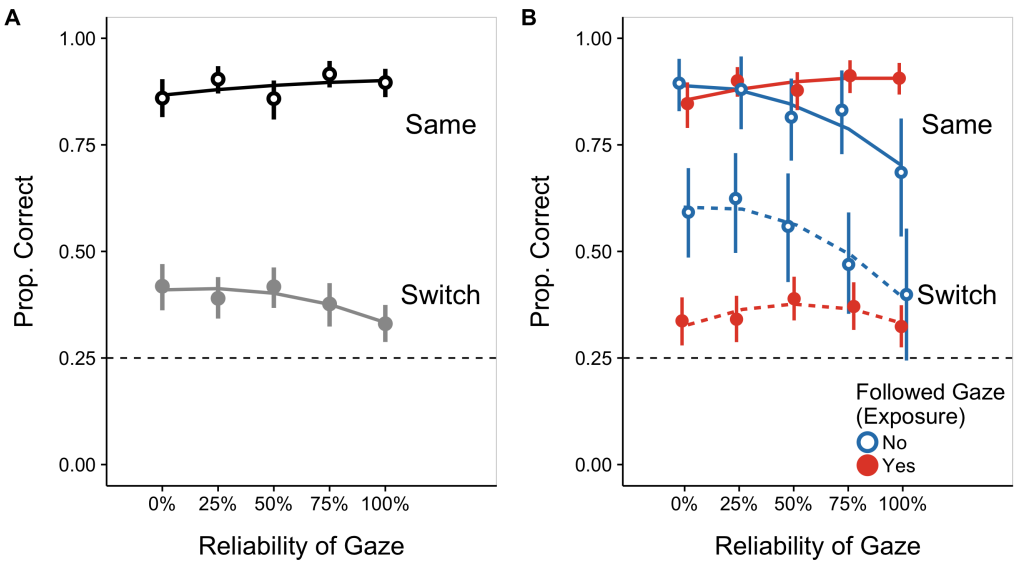


Figure 4.5: Primary analyses of test trial performance in Experiment 3. Panel A shows performance as a function of reliability condition. Panel B shows performance as a function of reliability condition and whether participants chose to follow gaze on exposure trials. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

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.

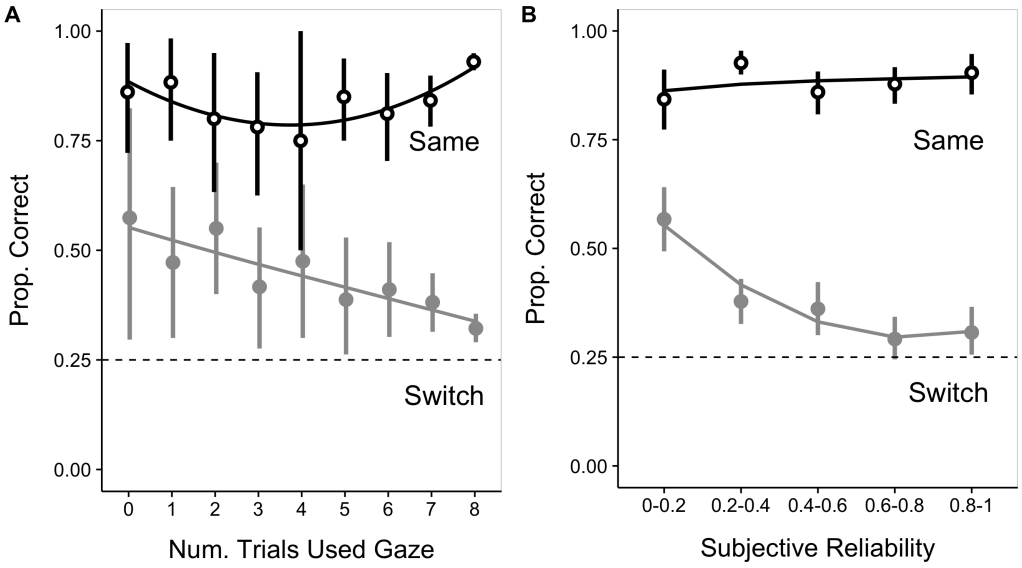


Figure 4.6: Secondary analyses of test trial performance in Experiment 3. Panel A shows accuracy as a function of the number of exposure trials on which participants chose to use the gaze cue. Panel B shows accuracy as a function of participants' subjective reliability judgments. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

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 (β = -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

[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.

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.

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

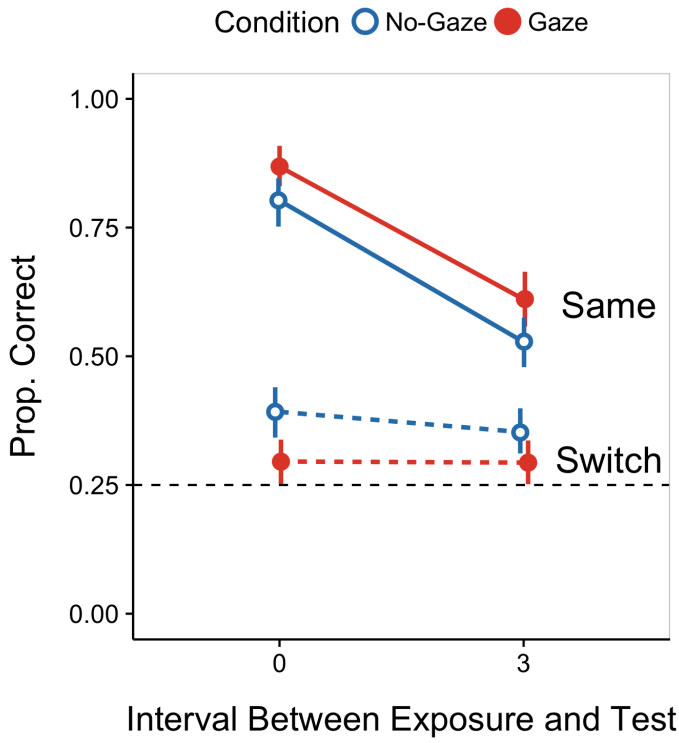


Figure 4.7: Experiment 4.4 results. Accuracy on test trials in Experiment 4 collapsed across the Long and Short inspection time conditions. The dashed line represents chance performance. Color and line type indicate whether there was gaze present on exposure trials. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

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 (P. Bloom, 2002; E. V. Clark, 2009) and to recent work exploring the interaction between statistical learning mechanisms and other types of information (M. C. Frank et al., 2009; Koehne & Crocker, 2014; 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 (R. Wu & Kirkham, 2010), and that even toddlers readily use gaze to infer novel word meanings (D. A. 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 (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 (M. C. 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, Snedeker, Trueswell, & Gleitman, 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 (M. C. 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.

Chapter 5

Integrating statistical and social information during language comprehension and word learning

In this chapter, we present three studies that ask how the presence of a social cue to reference (a speaker's gaze) changes listeners' eye movements during language comprehension and word learning. Within our broader active-social framework, these studies ask how the value of information gained from querying a social partner interacts with learners' developing knowledge of word meanings. The integrative account of active information seeking within social contexts presented in Chapter 1 motivates the empirical approach in this chapter.

Across three studies, we measured how children's decisions about whether to look at a social target (a speaker's face) or objects changed as a function of their word knowledge. First, both children and adults showed parallel gaze dynamics when comprehending familiar words in the presence of a social cue (eye gaze). Second, in a cross-situational word learning task, adults showed stronger memory for word-object mappings learned via a social cue. Finally, in contrast to the familiar words in Experiment 1, both children and adults gathered more visual information from a speaker's face when she provided a useful gaze cue in the context of novel object labeling. This differential looking to a social partner increased throughout the experiment, as learners were exposed to more

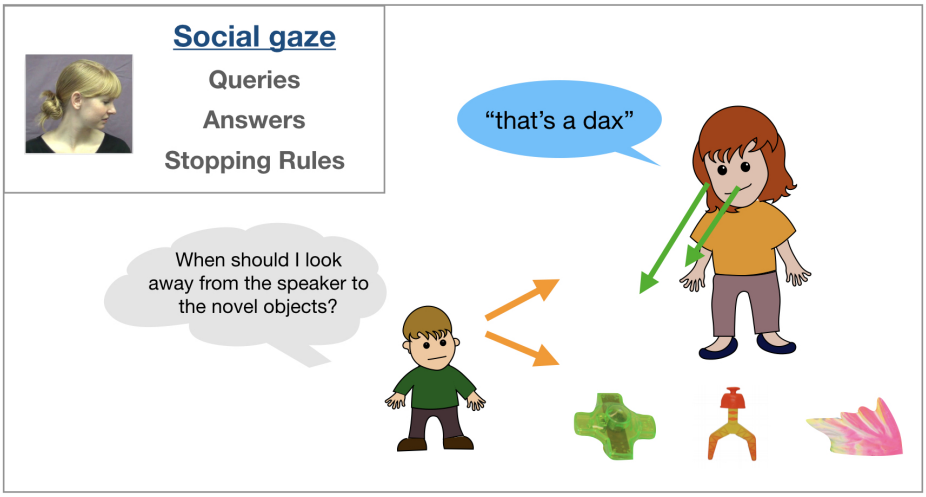


Figure 5.1: A schematic showing the components of the OED model captured by the case studies in Chapter 5.

word-object co-occurrence information. These results provide evidence that learners look more to a communicative partner when it supports their learning goal, and estimating this value involves an integration of prior knowledge of word-object mappings with the availability of social information in the learning context.

5.1 Introduction

How is it that children, who are just learning how to walk, can segment units from a continuous stream of linguistic information and map them to their corresponding conceptual representations. Children's word-to-meaning mapping skill becomes even more striking when we consider that a speaker's intended meaning is mostly unconstrained by the co-occurring context; a point made famous by W.V. Quine's example of a field linguist trying to select the target meaning of a new word ("gavagai") from the set of possible meanings consistent with the event of a rabbit running (e.g., "white," "rabbit," "dinner," etc.) (Quine, 1960).

Research on early lexical development has pursued several solutions to the problem of referential

uncertainty. First, lab-based studies and computational models have explored how children's statistical learning mechanisms can reduce ambiguity during word learning. Under these *cross-situational* learning accounts, learners can overcome referential uncertainty within a specific labeling event by tracking the elements of a context that remain consistent across multiple exposures to a new word (Roy & Pentland, 2002; Siskind, 1996; Yu & Smith, 2007). Experiments with 12-month-old infants find that they are capable of learning novel words via repeated exposures to consistent word-object pairings (Smith & Yu, 2008). Moreover, simulation studies show that models of a simple cross-situational learner can acquire an adult-sized vocabulary from exposures that fall well within the bounds of children's language experience (Blythe, Smith, & Smith, 2010) and even when referential uncertainty is high (Blythe, Smith, & Smith, 2016).

Social-pragmatic theories argue that the complexity of word learning is reduced via children's ecological context of learning words within grounded, social interactions (P. Bloom, 2002; E. V. Clark, 2009; Hollich et al., 2000). Observational studies show that adults are skilled at using gesture and eye gaze to structure language interactions with children (Estigarribia & Clark, 2007). Moreover, from a young age, children can use social cues to infer word meanings (D. A. Baldwin, 1993) and produce gestures such as reaches and points to share attention and elicit labels from other people (Liszkowski et al., 2012). Finally, correlational studies have demonstrated links between early gaze following and later vocabulary growth (Brooks & Meltzoff, 2005; Carpenter et al., 1998).

Thus, both social and statistical information can reduce children's uncertainty about new word meanings. These learning mechanisms, however, are unlikely to operate in isolation, and a sophisticated learning system could integrate the two sources of information to facilitate language acquisition. Several computational models of word learning have pursued integrative accounts of social and statistical word learning. For example, work by Yu & Ballard (2007) found better word-object mapping performance if their model used social cues (eye gaze) to increase the strength of specific word-object associations stored from a given labeling event. Moreover, M. C. Frank et al. (2009) showed that adding social inferences about a speaker's intended meaning to a cross-situational word learning model allowed the model to reproduce a variety of key behavioral findings in early language development (e.g., mutual exclusivity and the use of gaze to disambiguate reference).

The accounts of word learning reviewed above reflect a somewhat passive construal of the learner

where children absorb social information and word-object statistics from their environments. Children, however, are far from passive learners and can exert control over their input via actions such as choosing where to look, pointing, asking verbal questions. A body of research outside the domain of language acquisition shows the benefits of *active learning* or giving learners control to structure their learning experiences (Castro et al., 2009; Gureckis & Markant, 2012; Settles, 2012). The upshot of this work is that active learning can be superior because it allows people to use their prior knowledge and current uncertainty to select the most helpful examples (e.g., asking a question about something that is particularly confusing). Recent empirical and modeling work has begun to explore the role of active control in word learning (Hidaka, Torii, & Kachergis, 2017; Partridge et al., 2015). For example, Kachergis et al. (2013) showed that adults who were able to select the set of novel objects that would receive labels displayed stronger learning compared to adults who passively experienced the word-object pairings generated by the experiment.

In the current paper, we pursue the idea that children seek information from social partners to support language processing. Selecting an entire set of objects to be labeled, however, is a complex form of information seeking; one that might not yet be available for younger word learners. Children, however, are well-practiced at allocating visual attention to their environment. Moreover, grounded language processing involves linking the incoming linguistic signal to the visual world using information gathered through decisions about visual fixation. And recent work has shown that infants' ability to sustain visual attention on objects is a strong predictor of their novel word learning in experimental tasks (Smith & Yu, 2013). Taken together, these findings suggest that children's real-time seeking of visual information is a good case study for exploring how they integrate social and statistical to support early language processing.

5.1.1 Current studies

Here, we present a set of studies that synthesize ideas from social, statistical, and active learning. We ask how children's real-time information selection via eye movements is shaped by social information present in the labeling moment and by statistical information that is accumulated over time. We draw on ideas from theories of goal-based vision that characterize eye movements as information seeking decisions that aim to minimize uncertainty about the world (M. Hayhoe & Ballard, 2005). Under this account, learners should integrate statistical and social information in their choices of

where to fixate by computing the usefulness of an eye movement for their current goal.

The studies are designed to answer several open questions for research on early language processing. First, how do statistical learning mechanisms operate over fundamentally social input? The majority of prior work on statistical word learning has used linguistic stimuli that come from a disembodied voice, removing a rich set of multimodal cues (e.g., gestures, facial expressions, mouth movements) that occur during face-to-face communication. By including a social partner as a fixation target, this work will add to our understanding of how social contexts shape the input to statistical word learning mechanisms.

Second, how do children use visual information to support their language learning? In this work, we frame the learner's task as decision making under time constraints. Using this theoretical framework allows us to bring top-down, goal-based models of vision (M. Hayhoe & Ballard, 2005) into contact with work on language-driven eye movements (Allopenna et al., 1998) that often characterize gaze shifts as the output of the language comprehension process.

Finally, this study will increase our understanding of how children's in-the-moment behaviors, such as decisions about visual fixation, connect to learning that unfolds over longer timescales. Following McMurray et al. (2012), we separate situation-time behaviors (figuring out the referent of a word) from developmental-time processes (slowly forming mappings between words and concepts). Moreover, by studying changes in patterns of eye movements throughout learning, we will add to a recent body of empirical work that emphasizes the importance of linking real-time information selection to longer-term statistical learning (Yu & Smith, 2012).

5.2 Analytic approach

To quantify evidence for our predictions, we present analyses of (1) the time course of listeners' looking to each area of interest (AOI) and (2) the Reaction Time (RT) and Accuracy of listeners' first shifts away from the speaker's face and to the objects.[1]

First, we analyzed the time course of participants' looking to each AOI in the visual scene as the target sentence unfolded. Proportion looking reflects the mean proportion of trials on which participants fixated on the speaker, the target image, or the distracter image at every 33-ms interval

[1] All analysis code can be found in the online repository for this project: `https://github.com/kemacdonald/speed-acc-novel`.

of the stimulus sentence. We tested condition differences in the proportion looking to the language source – signer or speaker – using a nonparametric cluster-based permutation analysis, which accounts for the issue of taking multiple comparisons across many time bins in the timecourse (Maris & Oostenveld, 2007). A higher proportion of looking to the language source in the gaze condition would indicate listeners' prioritization of seeking visual information from the speaker.

Next, we analyzed the RT and Accuracy of participants' initial gaze shifts away from the speaker to objects. RT corresponds to the latency of shifting gaze away from the central stimulus to either object measured from the onset of the target noun. All reaction time distributions were trimmed to between zero and two seconds, and RTs were modeled in log space. Accuracy corresponds to whether participants' first gaze shift landed on the target or the distracter object. If listeners generate slower but more accurate gaze shifts, this provides evidence that gathering more visual information from the speaker led to more robust language processing in the social gaze context.

In Experiments 2 and 3, which measure novel word learning as a function of multiple word-object exposures, we compute proportion looking to the speaker for each trial, which corresponds to the amount of time looking to the speaker over the total amount of time looking at the three AOIs. We interpret a higher looking to the speaker as increased information seeking to gather the social cue. We also compute the proportion looking to the target object, which corresponds to the time spent looking to the target over the total amount of time fixating on both the target and the distracter objects. Higher target looking on Exposure trials with gaze cues indicate that learners followed the gaze cue. We interpret higher target looking on test trials indicates stronger retention for the newly learned word-object links. In all analyses of learning, we treat trial number as continuous and age group – children vs. adults – as categorical.

We used the `brms` (Bürkner, 2017) package to fit Bayesian mixed-effects regression models. The mixed-effects approach allowed us to model the nested structure of our data – multiple trials for each participant and item, and a within-participants manipulation in Experiments 1 and 3. We used Bayesian estimation to quantify uncertainty in our point estimates, which we communicate using a 95% Highest Density Interval (HDI), providing a range of credible values given the data and model.

Table 5.1: Age distributions of children in Experiments 1 and 3. All ages are reported in months.

Experiment	n	Mean	Min	Max
Experiment 1 (familiar words)	38	55.50	35.60	71.04

5.3 Experiment 1

In Experiment 1, we measured the time course of children and adults' decisions about visual fixation as they processed sentences with familiar words (e.g., "Where's the ball?").[2] We manipulated whether the speaker produced a post-nominal gaze cue to the named object. The visual world consisted of three fixation targets (a center video of a person speaking, a target picture, and a distracter picture; see Figure 1). The primary question of interest is whether listeners would delay shifting away from the speaker's face when she was likely to generate a gaze cue. We predicted that choosing to fixate longer on the speaker would allow listeners to gather more language-relevant visual information and facilitate comprehension. In contrast, if listeners show parallel gaze dynamics across the gaze and no-gaze conditions, this pattern suggests that hearing the familiar word was the primary factor driving shifts in visual attention.

5.3.1 Methods

Participants

Participants were native, monolingual English-learning children (n = 38; 19 F) and adults (n = 33; 23 F). All participants had no reported history of developmental or language delay and normal vision. 12 participants (9 children, 3 adults) were run but not included in the analysis because either the eye tracker falied to calibrate (8 children, 2 adults) or the participant did not complete the task (1 children, 1 adults).

Materials

Linguistic stimuli. The video/audio stimuli were recorded in a sound-proof room and featured two female speakers who used natural child-directed speech and said one of two phrases: "Hey! Can you

[2]See `https://osf.io/2q4gw/for` a pre-registration of the analysis plan.

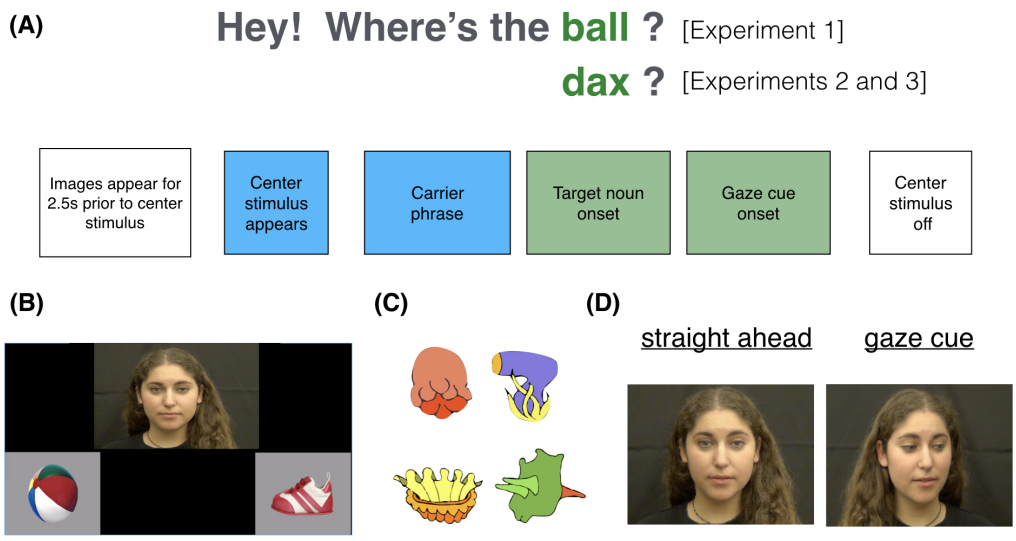


Figure 5.2: Stimuli for Experiments 1, 2, and 3. Panel A shows the structure of the linguistic stimuli for a single trial. Panel B shows the layout of the fixation locations for all tasks: the center stimulus, the target, and the distracter. Panel C shows a sample of the images used as novel objects in Experiment 3. Panel D shows an example of the social gaze manipulation.

find the (target word)" or "Look! Where's the (target word). The target words were: ball, bunny, boat, bottle, cookie, juice, chicken, and shoe. The target words varied in length (shortest = 411.68 ms, longest = 779.62 ms) with an average length of 586.71 ms.

Gaze manipulation. To create the stimuli in the gaze condition, the speaker waited until she finished producing the target sentence and then turned her head to gaze at the bottom right corner of the camera frame. After looking at the named object, she then returned her gaze to the center of the frame. We chose to allow the length of the gaze cue to vary to keep the stimuli naturalistic. The average length of gaze was 2.12 seconds with a range from 1.78 to 3.07 seconds.

Visual stimuli. The image set consisted of colorful digitized pictures of objects presented in fixed pairs with no phonological overlap between the target and the distracter image (cookie-bottle, boat-juice, bunny-chicken, shoe-ball). The side of the target picture was counterbalanced across trials.

Procedure

Participants viewed the task on a screen while their gaze was tracked using an SMI RED corneal-reflection eye-tracker mounted on an LCD monitor, sampling at 30 Hz. The eye-tracker was first calibrated for each participant using a 6-point calibration. On each trial, participants saw two images of familiar objects on the screen for two seconds before the center stimulus appeared. Next, they processed the target sentence – which consisted of a carrier phrase, a target noun, and a question – followed by two seconds without language to allow for a response. Both children and adults saw 32 trials (16 gaze trials; 16 no-gaze trials) with several filler trials interspersed to maintain interest. The gaze manipulation was presented in a blocked design with the order of block counterbalanced across participants.

5.3.2 Results and Discussion

Timecourse looking. We first analyzed how the presence of gaze influenced listeners' distribution of attention across the three fixation locations while processing familiar words. At target-noun onset, listeners tended to look more at the speaker than the objects. As the target noun unfolded, the mean proportion looking to the center decreased as participants shifted their gaze to the images. Proportion looking to the target increased sooner and reached a higher asymptote compared to proportion looking to the distracter for both gaze conditions with adults spending more time looking at the target compared to children. After looking to the named referent, listeners tended to shift their gaze back to the speaker's face. We did not see evidence that the presence of a post-nominal gaze cue changed how children or adults allocated attention early in the target word. Children in the gaze condition, however, tended to shift their focus back to the speaker earlier after shifting gaze to the named object and spent more time fixating on the speaker's face throughout the rest of the trial ($p < .001$; nonparametric cluster-based permutation analysis). Next, we ask how these different processing contexts changed the timing and accuracy of children's initial decisions to shift away from the center stimulus.

First shift RT and Accuracy. To quantify differences across the groups, we fit a Bayesian linear mixed-effects regression predicting first shift RT as a function of gaze condition and age group: *Log(RT)* $\sim$ *gaze condition + age group + (gaze_condition + item | subject).* Both children and adults generated similar RTs in the gaze (children $M_{rt} = 563.159$ ms, adults $M_{rt} = 652.405$ ms) and

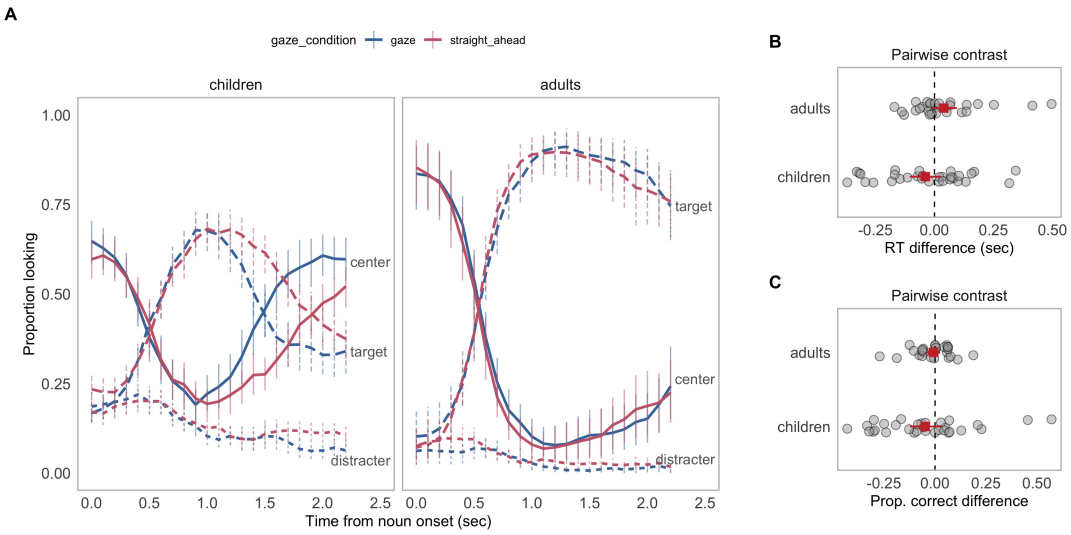


Figure 5.3: Timecourse looking, first shift Reaction Time (RT), and Accuracy results for children and adults in Experiment 1. Panel A shows the overall looking to the center, target, and distracter stimulus for each gaze condition and age group. Panel B shows the distribution of pairwise contrasts between each participant's RT in the gaze and no-gaze conditions. The square point represents the group means. The vertical dashed line represents the null model of zero condition difference. Error bars represent the 95% HDI. Panel C shows the same information but for first shift accuracy.

no-gaze (children M_{rt} = 575.762 ms, adults M_{rt} = 608.314 ms) conditions, with the null value of zero condition differences falling within the 95% credible interval (β = -0.36, 95% HDI [-0.89, 0.06]). Next, we fit the same model to estimate first shift accuracy. Adults generated more accurate gaze shifts (M = 0.9) compared to children (M = 0.64) with the null value falling outside the 95% HDI (β_{age} = -1.76, 95% HDI [-2.19, -1.34]). Similar to the RT analysis, we did not find strong evidence of a difference in performance across the gaze conditions (β = 0.10, 95% HDI [-0.18, 0.41]).

Taken together, the time course and first shift analyses suggest that hearing a familiar noun was sufficient for both adults and children to shift visual attention away from the speaker and seek the named referent. Neither age group showed evidence of delaying their eye movements to fixate on the speaker's face and gather a social cue to reference that could have provided additional disambiguating information. The presence of gaze, however, did change children's looking behavior such that they were more likely to allocate attention to the speaker after processing the familiar noun. While we did not predict these results, it is interesting that listeners did not delay their eye movements to seek

social information when processing familiar words. This behavior seems reasonable if eye movements during familiar language processing are highly-practiced visual routines such that seeking a post-nominal gaze cue becomes less-relevant to the comprehension task. Moreover, if listeners developed an expectation that their goal was to seek out named objects quickly, then fixating on the speaker for longer would become less goal-relevant.

In our previous work, we found that both children and adults fixated longer on a speaker when language in more challenging processing contexts in the presence of background noise (MacDonald, Marchman, Fernald, & Frank, 2018c). We explained this result as listeners adapting to the informational demands of their environment such that they gathered additional visual information to support language comprehension. The results of Experiment 1 can constrain this information seeking explanation by showing that listeners do not always seek social information when it is available; instead, children might take their uncertainty into account and adapt their information seeking when uncertainty is higher. This finding raises an interesting question: Would children gather social information when they do not already know word-object mappings? That is, when the learner is surrounded by novel objects, the value of seeking visual information from a social partner should increase since this action could provide highly-relevant information for decreasing referential uncertainty – a point that has long been emphasized by social-pragmatic theories of language acquisition (P. Bloom, 2002; E. V. Clark, 2009; Hollich et al., 2000). Experiments 2 and 3 explore this case and ask whether learners would adapt their gaze patterns to seek information from social partners in the context of mapping novel words to their referents.

5.4 Experiment 2

Because children hear language in environments with multiple possible referents, learning the meaning of even the simplest word requires reducing this uncertainty. A cross-situational statistical learner can aggregate across ambiguous naming events to learn stable word meanings. But for this aggregation process to work, learners must allocate their limited attention and memory resources to the relevant statistics in the world – how do they select what information to store?

In prior work (discussed in Chapter 4), we found that the presence of a gaze cue shifted adults away from storing multiple word-object links and towards tracking a single hypothesis. Those experiments, however, relied on an offline measurement of word learning (a button press on test

trials) and an indirect measure of attention during learning (self-paced decisions about how long to inspect the visual scene during learning trials). We address these limitations in Experiment 2 by adapting the social cross-situational learning paradigm to use eye-tracking methods. By moving to an eye-tracking procedure, we could ask: (1) how does the presence of gaze alter learners' distribution of visual attention between objects and their social partner? And (2) does the presence of a gaze cue change the strength of the relationship between real-time information selection during learning and long-term retention of word-object links?

5.4.1 Methods

Participants

34 undergraduate students were recruited from the Stanford Psychology One credit pool (17 F). Four participants were excluded during analysis because the eye-tracker did not correctly record their gaze coordinates. The final sample included 30 participants.

Materials

The experiment featured sixteen pseudo-words recorded by an AT&T Natural VoicesTM speech synthesizer using the "Crystal" voice (a woman's voice with an American English accent), as well as 48 novel objects represented by black-and-white drawings of fictional objects from Kanwisher, Woods, Iacoboni, and Mazziotta (1997). Sixteen words were used so that the experiment would be sufficiently long to make within-subject comparisons across trials, and 48 objects were used so that objects would not be repeated across trials. Six familiar objects from the same set of drawings were used for the two practice trials, accompanied by two familiar words using the same speech synthesizer. Finally, the videos of the speaker's face were taken from MacDonald et al. (2017b).

Procedure

We tracked adults' eye movements while they watched a series of ambiguous word-learning events (16 novel words) organized into pairs of exposure and test trials (32 trials total). All trials consisted of a set of two novel objects and one novel word. Participants were randomly assigned to either the Gaze condition in which a speaker looked at one of the objects on exposure trials or the No-Gaze condition in which a speaker looked straight on exposure trials. Every exposure trial was followed

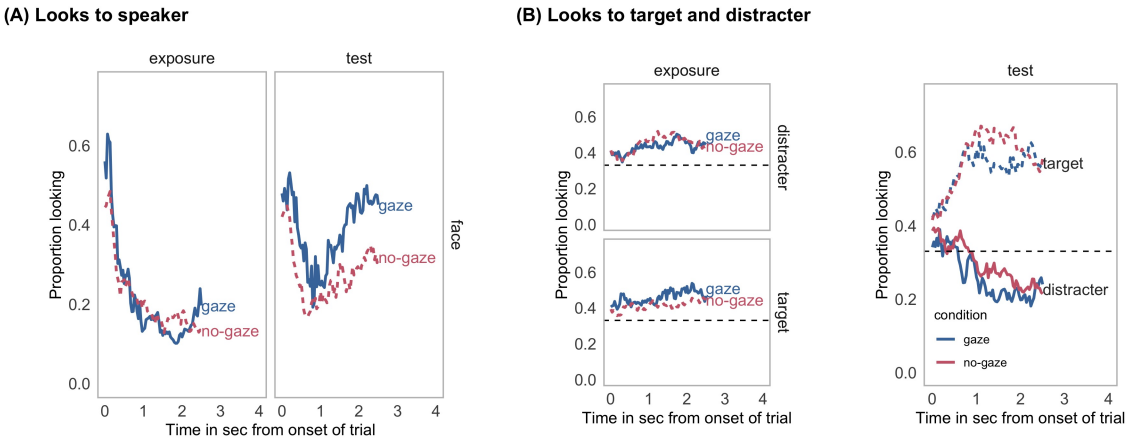


Figure 5.4: Overview of adults' looking to the three fixation targets (Face, Target, Distracter) over the course of the trial. Panel A shows proportion looking to the speaker's face for exposure and test trials. Color and line type represent gaze condition. Panel B shows the same information but for proportion looking to the target and distracter images.

by a test trial, where participants heard the same novel word paired with a new set of two novel objects. One of the objects in the set had appeared in the exposure trial ("target" object), while the other object had not previously appeared in the experiment ("distracter" object). The side of the screen of the target object was counterbalanced throughout the experiment. In the gaze condition, for half of the test trials, the target object was the focus of the speaker's gaze during the exposure trial, while the other half, the target object was the object that had not been the focus of gaze during labeling.

5.4.2 Results and Discussion

Timecourse looking. The first question of interest was how did the presence of a gaze cue change adults' distribution of attention across the three fixation locations while processing language in real-time? Figure 5.4 presents an overview of looking to each AOI for each processing context. At the target-noun onset, adults tended to look more at the speaker's face on both exposure and test trials. As the target noun unfolded, the mean proportion looking to the center decreased as participants shifted their gaze to the target or the distracter images. On exposure, trials tended to distribute their attention relatively evenly across target and distracter images. On test trials, proportion looking to

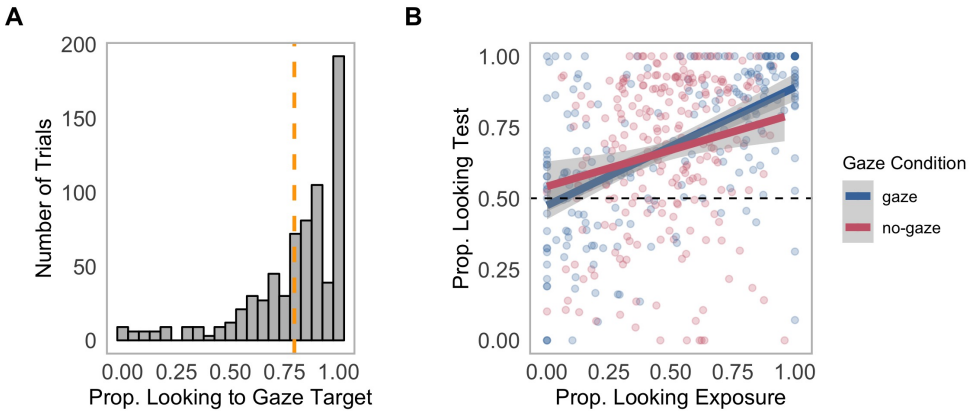


Figure 5.5: Panel A shows participants' tendency to look at the object that was the target of the speaker's gaze on exposure trials. The vertical, dashed line represents the mean proportion of time looking to the gaze target across all trials. Panel B shows the relationship between adults' looking behavior on exposure and test trials for the gaze and no-gaze conditions. The lines represent linear model fits.

the target increased sooner and reached a higher asymptote compared to proportion looking to the distracter for both conditions, suggesting that adults were able to track the consistent word-object links both with and without accompanying social information.

There were several qualitative differences in looking behavior across the different gaze conditions and trial types. First, adults spent more time looking to a speaker's face when she provided a social gaze cue, especially on test trials that were preceded by gaze (Figure 5.4A). Second, adults in the gaze condition looked slightly more to the target image throughout the trial. This behavior is reasonable since half of the trials, the speaker's gaze was focused on the target image that would appear on the subsequent test trial. Third, on test trials, adults looked more to the images in the no-gaze condition, which led to a higher proportion of looking to the target and a higher proportion of looking to the distracter images (Figure 5.4B).

These looking patterns provide evidence that the presence of a gaze cue caused adults to spend more time gathering visual information from the speaker's face, which, in turn, changed how they distributed fixations across the target and distracter objects during subsequent labeling events. We next ask how the presence of gaze modulated learning, which we operationalized as the relationship between proportion looking to the target object on exposure and test trials. *Relationship between performance on exposure and test trials.* When the speaker generated a social cue during labeling,

adults reliably followed that cue and tended to focus their attention on a single object (Figure 5.5A). In contrast, people in the No-gaze condition tended to distribute their attention more broadly across the two objects. For adults in both gaze contexts, more time spent attending to the target object on exposure trials led higher proportion looking to the target, i.e., better recall, at test ($\beta_{exposure}$ = 0.43, 95% HDI [0.36, 0.50]). Critically, there was an interaction between the gaze condition and the effect of exposure looking patterns (Figure 5.5B): When a speaker's gaze guided adults' visual attention, they showed stronger memory for the newly-learned word-object link (β_{int} = -0.19, 95% HDI [-0.33, -0.04]). This result provides evidence that social information does more than change in-the-moment decisions about visual fixation; instead, the presence of gaze modulated the fidelity of information that learners stored during novel object labeling.

Limitations

There were several limitations to this study. First, the linguistic stimulus occurred at the trial onset when the images and the speaker appeared on the screen. This trial structure makes it challenging to interpret learners' initial decisions to stop gathering information from a social target to fixate the objects, a behavior that we have used in our prior work to shed light on how children's information selection adapts to their processing environment (MacDonald et al., 2018c). Second, the linguistic stimuli consisted of pseudowords recorded by a speech synthesizer and presented in isolation, thus removing any sentential context. Presenting isolated words is unlikely to work with the target age range for this research. Finally, we used a minimal cross-situational learning paradigm with only two exposures to each word-object link, which does not allow for measurement of the effect of accumulating statistical information over a longer timescale. Thus, Experiment 3 was designed to address these limitations, allowing us to ask how younger learners' information seeking from social partners changes as a function of increased exposure to consistent word-object mappings.

5.5 Experiment 3

Experiment 3 explores whether learners' real-time information seeking from social partners adapts as they accumulate knowledge of word-object links.[3] We also set out to address the limitations of Experiment 2 discussed above with two key modifications. First, we included more than two

[3]See `https://osf.io/nfz85/` for a pre-registration of the analysis plan and predictions.

exposures to a novel word-object link, allowing us to measure changes in learners' integration of social and statistical information over a longer timescale. Second, we changed the linguistic stimuli to use the trial structure in Experiment 1 such that the novel words occurred within a sentence spoken in a child-friendly register. This change allowed us to analyze children's first gaze shifts away from a social target and ask how the threshold of information gathering changed as a function of statistical learning about word-object mappings.

We aimed to answer the following specific research questions:

1. Does the presence of a social cue to reference (eye gaze) change the dynamics of children's gaze patterns during novel object labeling?

2. Do decisions about where to allocate visual attention (speakers vs. objects) change as a function of repeated exposures to a word-object link?
3. Does social information change the relationship between learners' information selection during labeling and their memory of word-object links?

To answer these questions, we compared the timing and accuracy of eye movements during a real-time cross-situational word learning task where participants processed sentences containing a novel word (e.g., "Where's the *dax*?") while looking at a simplified visual world with three fixation targets (a video of a speaker and two images of unfamiliar objects).

5.5.1 Predictions

We had three key behavioral predictions. First, the presence of a gaze cue will change participants' decisions about visual fixation. We hypothesize that a post-nominal gaze cue increases the value of fixating on a speaker. This manipulation will cause participants to allocate more fixations to the speaker when gaze is present, leading to slower first shift reaction times and higher proportion looking, especially earlier in learning (i.e., lower trial numbers within each block of exposure trials to a novel word-object pairing). This prediction was operationalized as a main effect of Gaze condition on RT, and a trial number by Gaze condition interaction such that the decrease in RT will be greater on exposure trials in the Gaze condition.

Second, for all conditions, participants' distribution of attention to speakers compared to objects will shift throughout learning. Early in the task, participants will allocate more fixations to a speaker

to prioritize gathering visual information that disambiguates reference. After experiencing multiple exposures to a word-object pairing, participants will generate faster saccades, showing signatures of comprehension of the incoming speech. We further predict that later in learning blocks, participants will allocate more fixations to the objects, displaying looking behaviors that support learning long-term associations between words and objects.

Third, the presence of gaze should lead to stronger inferences about the correct word-object mapping, resulting in faster learning that we operationalize as more accurate first shifts, faster RTs, and a higher proportion looking to the target object on test trials as compared to learning words without a gaze cue across both exposure and test trials.

5.5.2 Methods

Participants

Participants were native, monolingual English-learning children (n = NA; NA F) and adults (n = 30; 20 F). All participants had no reported history of developmental or language delay and normal vision. 6 adults were run but not included in the analysis because they were not native speakers of English. 7 children participants were run but not included in the analysis because the participant did not complete more than half of the trials in the task.

Materials

Linguistic stimuli. The video/audio stimuli were recorded in a sound-proof room and featured two female speakers who used natural child-directed speech and said one of two phrases: "Hey! Can you find the (novel word)" or "Look! Where's the (novel word). The target words were four pseudo-words: bosa, modi, toma, and pifo. The novel words varied in length (shortest = 472.00 ms, longest = 736.00 ms) with an average length of 606.31 ms.

Gaze manipulation. To create the stimuli in the gaze condition, the speaker waited until she finished producing the novel word before turning her head to gaze at the bottom right corner of the frame. After looking at the named object, she then returned her gaze to the center of the frame. We chose to allow the length of the gaze cue to vary to keep the stimuli naturalistic. The average length of gaze was 2.06 seconds with a range from 1.74 to 2.67 seconds.

Visual stimuli. The image set consisted of 28 colorful digitized pictures of objects that were

selected such that they would be interesting to and that children would be unlikely to have already a label associated with the objects. The side of the target picture was counterbalanced across trials.

Procedure

Participants viewed the task on a screen while their gaze was tracked using an SMI RED corneal-reflection eye-tracker mounted on an LCD monitor, sampling at 30 Hz. The eye-tracker was first calibrated for each participant using a 6-point calibration. Then, participants watched a series of ambiguous word learning events organized into pairs of one exposure and one test trial. On each trial, participants saw of a set of two unfamiliar objects and heard one novel word.

Each word was learned in a block of four exposure-test pairs for a total of eight trials for each novel word. Critically, on each trial within a word block, one of the objects in the set had appeared on the previous trials (target object), while the other object was a randomly generated novel object not previously shown in the experiment (distracter object). Both children and adults saw 32 trials (16 gaze trials; 16 no-gaze trials) with several filler trials interspersed to maintain interest. The gaze manipulation was presented in a blocked design with the order of block counterbalanced across participants.

5.5.3 Results and Discussion

Timecourse looking

Looking to the speaker. How did the presence of a gaze cue change learners' decisions to fixate on the speaker? Visual inspection of Figure 5.6A) shows that both children and adults tended to start looking at the speaker at noun onset and shifted their gaze away as the noun unfolded, with adults doing so sooner compared to children. Exposure trials when there was a gaze cue, both adults and children tended to look more to the face at noun onset as indicated by the higher intercept of the blue curves. Moreover, around one second after noun onset, listeners tended to shift their attention back to the speaker's face more often and especially so for children. On Test trials that were preceded by an Exposure trial with a gaze cue, children and adults tended to look more to the speaker even though there was no gaze cue present. This pattern of looking suggests that the presence of gaze modulated learners' expectations of being able to gather disambiguating information from the speaker on Test trials.

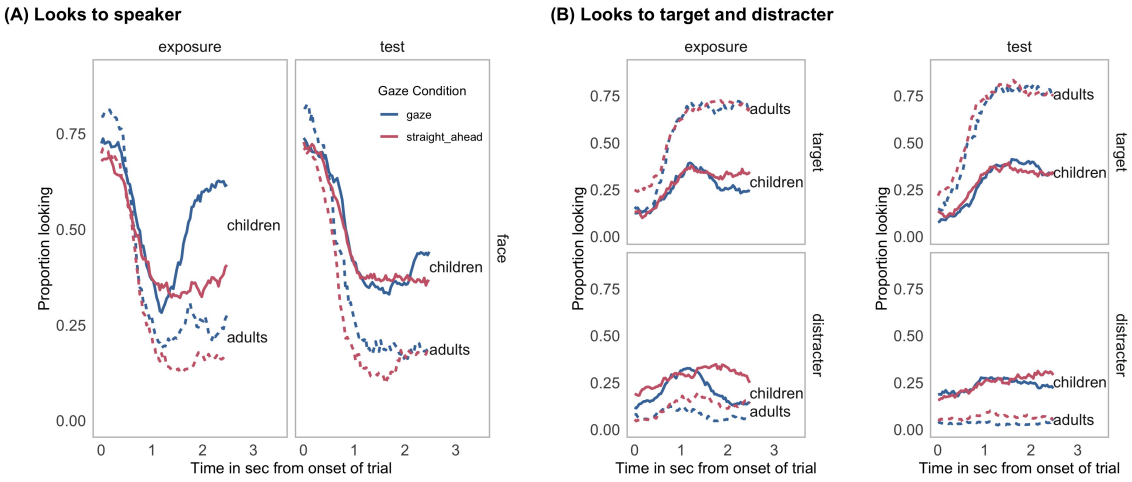


Figure 5.6: Overview of children and adults' looking to the three fixation targets (Speaker, Target, Distracter) over the course of exposure and test trials. Panel A shows proportion looking to the speaker's face with color indicating gaze condition and line type indicating age group. Panel B shows the same information but for proportion looking to the target and distracter images.

Looking to the target and distracter. Next, we asked how learners divided attention between the target and distracter objects. On Exposure trials, looking to both objects increased throughout the trial but more so for looks to the named object as indicated by the higher asymptote of the target looking curves. Adults spent more time looking to the target and less time looking to the distracter as compared to children. The most substantial effect of gaze on the time course of looking was a tendency for learners to allocate fewer fixations to the distracter object when there was a gaze cue present.

Proportion looking

Learning effects. Both children ($M_{gaze} = 0.57$, $M_{no-gaze} = 0.55$) and adults ($M_{gaze} = 0.91$, $M_{no-gaze} = 0.89$) showed evidence of learning the novel word-object links, with the null value of 0.5 falling below the lower bound of the lowest credible interval for children's target looking in the No-gaze context (95% HDI [0.51, 0.60]). Our primary question of interest was how exposure to multiple co-occurrences of word-object pairs would change learners' distribution of attention between the speaker and objects. Figure 5.7) shows proportion looking to the speaker (5.7A) and the target and distracter objects (5.7B) as a function of trial number within a word learning block. Both

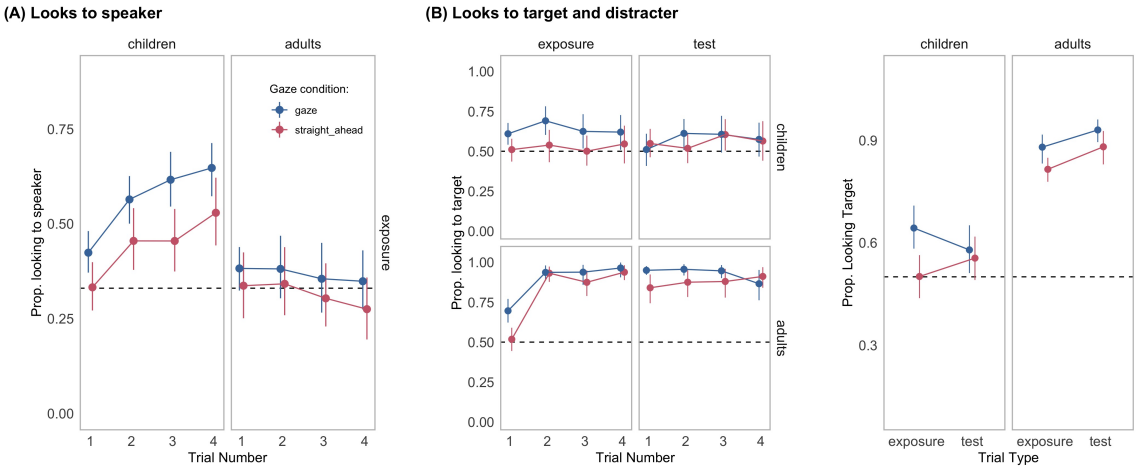


Figure 5.7: Panel A shows participants' tendency to look at the speaker on exposure and test trials as a function of the trial number within a learning block. The horizontal, dashed line represents the tendency to distribute attention equally across the three AOIs. Color indicates gaze condition and error bars represent 95% credible intervals. Panel B shows the same information but for target and distracter looking across the learning block (left) and aggregated over all trials (right).

children and adults were more likely to fixate on the speaker when she provided a gaze cue (β_{gaze} = 0.09, 95% HDI [0.16, 0.01]). Moreover, there was a developmental difference such that children, but not adults, were more likely to increase their fixations to the speaker over the course of the learning block ($\beta_{age:tr.num}$ = -0.07, 95% HDI [-0.11, -0.04]). Overall, looking to the target increased as learners were exposed to more word-object pairings ($\beta_{tr.num}$ = 0.16, 95% HDI [0.09, 0.24]) and was higher when the novel word was accompanied by a gaze cue (β_{gaze} = 0.14, 95% HDI [0.21, 0.06]). Visual inspection of Figure 5.7 shows that on the first Exposure trial, both adults and children used the gaze cue to disambiguate reference, fixating more on the target in the Gaze condition. For children, higher target looking on Exposure trials with gaze remained relatively constant across the learning block. In contrast, adults target looking reached ceiling for both Gaze and No-gaze conditions by trial number two, indicating that they had successfully used the co-occurrence information across trials to map the novel word to its referent. We found an interaction between gaze condition and trial number such that looking to the target increased more quickly in the No-gaze condition ($\beta_{gaze:tr.num}$ = 0.02, 95% HDI [0.00, 0.04]), which reflects (1) the higher intercept of target looking in the presence of gaze and (2) rapid learning of the word-object association via cross-situational information. Finally,

visual inspection of the proportion looking plot suggests that adults tended to look more the target when learning from a gaze cue, only reaching similar levels of accuracy in the no-gaze condition at the end of the learning block. There was not strong evidence for an effect of the gaze manipulation on children's looking behavior on Test trials.

Relationship between looking on exposure and test. For both children and adults, more time attending to the target object on exposure trials led to a higher proportion of looking to the target on test trials, especially for adults ($\beta_{exposure:age} = 0.16$, 95% HDI [0.05, 0.28]) and as the number of word-object exposures increased over the course a learning block ($\beta_{exposure:tr.num} = 0.07$, 95% HDI [0.02, 0.12]). There was evidence that participants in the No-gaze condition showed less learning over the course of each word block ($\beta_{gaze:tr.num}$ = -0.02, 95% HDI [-0.04, 0.00]). This result dovetails with the findings from Experiment 2, providing evidence that the presence of social information did more than change attention on Exposure trials but instead modulated the relationship between attention during learning and later memory for word-object links.

Together, the time course and the proportion looking analyses suggest that the presence of gaze changed how children and adults allocated attention while processing novel words. In the context of unfamiliar objects, children tended to fixate more on a speaker's face when she provided a post-nominal social cue to reference, a difference in looking behavior that increased as they were exposed to more word-object co-occurrences. This result is different from the parallel looking behavior that we found in Experiment 1 where listeners processed highly familiar nouns. Moreover, in the presence of a speaker who provided a gaze cue, children and adults spent less time fixating on the distracter image, which could play a role in the strength of the potential word-object connections that learners could store from labeling event. These changes in gaze patterns, however, did not generalize to performance differences on Test trials for children. Finally, like in Experiment 2, we found that the presence of a social cue increased the strength of the link between attention on exposure and fixations at test.

First shift RT and Accuracy

We next asked how the presence of gaze influenced learners' decision to stop gathering visual information from the speaker and start fixating on the novel objects. To quantify the effect the gaze, we fit a Bayesian linear mixed-effects regression predicting first shift RT as a function of whether there

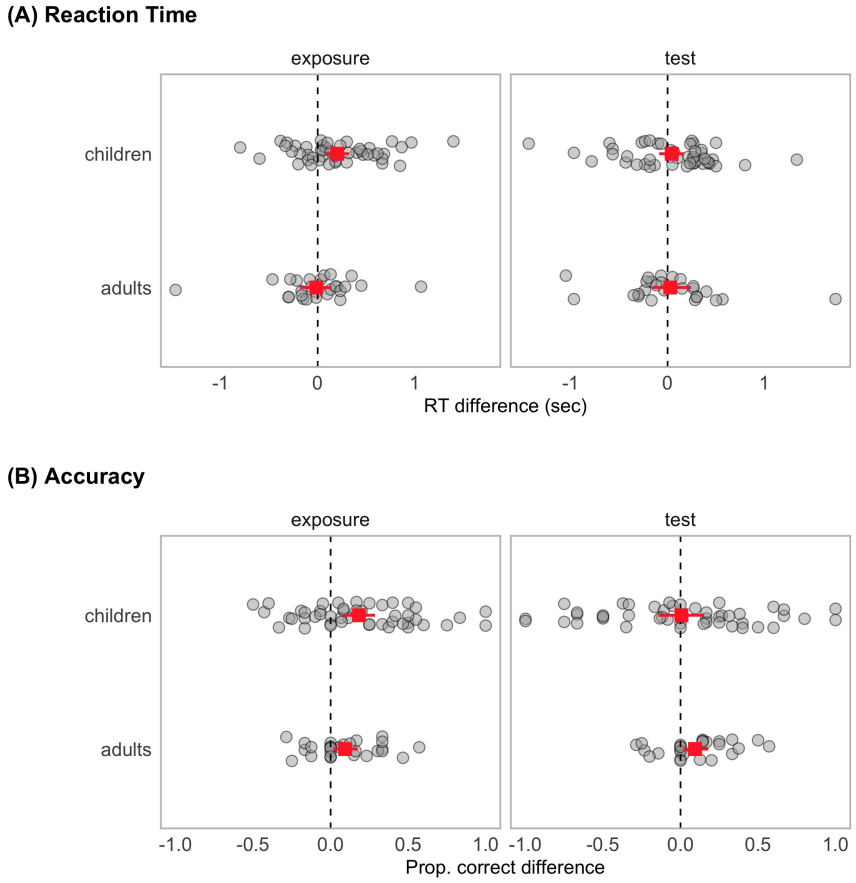


Figure 5.8: First shift Reaction Time (RT), and Accuracy results for children and adults in Experiment 3. Panel A shows the distribution of pairwise contrasts between RTs in the gaze and no-gaze conditions. The square point represents the mean value for each measure. The vertical dashed line represents the null model of zero condition difference. The width each point represents the 95% HDI. Panel B shows the same information but for participants' first shift accuracy.

was a gaze cue present on the trial and age group. Both children (Gaze M_{rt} = 922.4096386 ms, No-gaze M_{rt} = 705.2246604 ms) and adults (Gaze M_{rt} = NA ms, No-gaze M_{rt} = NA ms) fixated longer on the speaker when she provided a gaze cue (β_{gaze} = -0.20, 95% HDI [-0.38, -0.01]). With no evidence of an interaction between gaze condition and age group ($\beta_{age:gaze}$ = 0.27, 95% HDI [0.11, 0.44]). Moreover, both (Gaze M_{acc} = 0.64, No-gaze M_{acc} = 0.49) and adults (Gaze M_{acc} = 0.89, No-gaze M_{acc} = 0.81) generated more accurate first shifts in the gaze condition, indicating they were following the gaze cue on Exposure trials (β = -0.57, 95% HDI [-1.13, 0.00]).

Finally, we asked whether the presence of gaze affected learning by predicting first shift accuracy on Test trials. We found that adults were more accurate than children (β_{age} = 2.24, 95% HDI [1.50, 3.03]), that first shifts became more accurate as learners experienced repeated exposures to word-object pairings ($\beta_{tr.num}$ = 0.21, 95% HDI [-0.02, 0.44]). We did not see evidence for two of our predictions: (1) that children and adults would generate more accurate first shifts when learning from social gaze (β_{gaze} = -0.50, 95% HDI [-1.14, 0.14]) and (2) that learning from gaze would modulate the relationship between accuracy over the course of learning ($\beta_{gaze:tr.num}$ = -0.30, 95% HDI [-0.74, 0.12]), with the null value falling within each credible interval.

Overall, the first shift analyses provide converging, albeit mixed, evidence that learners' modulated their decisions about visual fixation to gather additional a post-nominal gaze cue when it was available. Children but not adults generated slower first shifts away from a speaker's face when there was a Gaze cue to gather. Both children and adults generated a higher proportion of shifts landing on the target image when there was post-nominal gaze cue available. Finally, adults, but not children, generated more accurate first shifts on Test trials that were preceded by Exposure trials with gaze. The absence of condition differences for children's performance on Test trials parallels the timecourse looking analyses and suggests children's learning of the novel word-object links was not strong enough to detect the effect of social information.

5.6 General Discussion

During grounded language processing, fixating on a social partner or ion objects can facilitate comprehension and learning. Do children flexibly seek information to support these goals? And how does children's information seeking adapt as they gain more exposures to consistent word-object pairings? In this work, we pursued the idea that learners flexibly adjust their gaze to seek disambiguating information from social partners when it is useful for their comprehension and learning. We presented evidence for this explanation by tracking children and adults' eye movements as they processed both familiar and novel words accompanied by an ecologically-valid social cue to reference (eye gaze). We also measured how learners' gaze dynamics changed as a function of accumulating statistical information about word-object mappings.

In Experiment 1, we found that children and adults showed parallel gaze dynamics while processing familiar words, shifting attention away from the speaker's face before she produced a post-nominal gaze cue. Experiment 2 showed that the presence of gaze in the context of novel objects focused adults' attention on a single object and modulated the strength of the relationship between visual attention during labeling and later recall for newly learned word-object pairs. Finally, in Experiment 3, we found that both children and adults fixated longer on a speaker to seek a post-nominal gaze cue while processing novel words, which resulted in more attention allocated to the target object and less looking to the distracter. Moreover, both age groups were capable of learning the novel word-object pairings from cross-situational statistics alone, but only adults showed evidence of stronger learning from the less ambiguous social gaze context.

5.6.1 Limitations

This work has several significant limitations. First, we did not see evidence of that the effects of gaze generalized to learning trajectories in children in Experiment 3. Moreover, we did not see evidence of strong uptake of the novel word-object links overall. Our future work will modify this social, cross-situational word learning paradigm to increase children's learning and provide a better opportunity to detect an effect of social information. For example, we plan to make the social cue stronger by increasing the length of time the speaker gazed at the object, which in the current stimulus set was relatively brief social cue (˜2 sec). We also plan to pair the newly-learned novel objects against one another on Test trials, which would reduce any attraction of novelty that pushed children to look at the distracter object that they had not seen before on Test trials in the current design. Finally, we will reduce the number of word-object pairs that children are asked to learn from four to two.

Second, while we did measure the effects of social information on learning over multiple labeling events, it is still a much shorter timescale and a smaller number of exposures relative to children's environmental learning input. Moreover, the visual world paradigm, while well-controlled, is highly constrained relative to the complexity of the information seeking decisions that children make when allocating visual attention in their naturalistic learning environments. Thus, a valuable next step for this work would be to leverage tasks that move closer to the ecological context in which children process and learn language such as using head-mounted cameras and eye trackers that would allow measurement of where children choose to look during everyday interactions. It would be interesting

to measure changes in children's looking to communicative partners when they are first introduced to novel objects in their day-to-day lives.

Third, we used a binary manipulation of the quality of information available in the social context – a fully disambiguating gaze cue or entirely ambiguous label without a gaze cue – which does not reflect the complexity of children's social interactions. That is, children's social partners are more likely to provide intermediate levels of disambiguating information during novel object labeling. Moreover, our prior work suggests that adults are sensitive to the graded changes in the reliability of a gaze cue, storing word-object links with greater fidelity as reliability increased (MacDonald et al., 2017b). It would be useful to know how children's real-time information selection responds to continuous changes in referential ambiguity. This modification would also allow us to measure what children are learning about other people during object labeling. For example, it would be interesting to know if children do more social referencing towards speakers who tend to reduce referential ambiguity during object play.

5.6.2 Conclusions

In this paper, we presented a set of empirical studies that integrated social-pragmatic and statistical accounts of language acquisition with ideas from goal-based accounts of visual. We found that listeners' decisions to seek social information varied depending on their uncertainty over word-object mappings in the visual scene. In the context of processing novel words, learners adapted their gaze dynamics to seek a post-nominal social cue to reference. Moreover, following gaze modulated the relationship between learners' real-time looking behavior during labeling and their retention of word-object labels at a longer timescale. More generally, this work sheds light on how children can use eye movements as an active information gathering process within social contexts, which, in turn, shapes the information that comes into contact with their statistical learning mechanisms.

Conclusion

In this dissertation, I proposed a framework for understanding children's information seeking decisions within social contexts. The core of the argument is that the presence of other people can change the *availability* and *usefulness* of information seeking behaviors by shaping the learner's goals, hypotheses, actions, answers, and decisions about when to stop gathering information. Following the theoretical account, I presented a set of empirical work that explored whether the dynamics of children's real-time information selection via eye movements flexibly adapted to their social context to support language processing.

Chapter 2 investigated how children learning American Sign Language (ASL) allocated attention between language and objects, which both compete for visual attention. Similar to children learning spoken language, ASL learners shifted gaze away from a social partner to seek objects before sign offset, providing evidence that, despite the higher value of fixating on a signer, grounded language drove rapid shifts in attention to named referents during ASL processing. Chapter 3 extended the sign language research by directly comparing ASL learners' gaze dynamics to children learning spoken English on parallel language comprehension tasks. Chapter 3 also described a comparison of English-learning children and adult's eye movements in noisy auditory contexts. Both ASL learners and children processing speech in noise showed parallel adaptations to the higher value of looking to their social partner to gather language-relevant information before shifting gaze seek a named referent. Chapters 4 and 5 explored how the information seeking changed when children and adults were processing words in the context of social cues to reference. These results suggest learners integrate uncertainty over word-object mappings with the availability of social information to select visual fixations that, in turn, modulate the information that comes into contact with statistical learning mechanisms.

Nevertheless, both the integrative account and the empirical work described here are limited in significant ways. First, the majority of this research tested binary hypotheses of behavior change – e.g., sign vs. spoken language; noisy vs. clear speech; word learning with vs. without social gaze – to answer the question of *whether* children would flexibly adapt their information seeking. Chapters 2-5 provide evidence across a diverse set of case studies that they do. However, for the integrative account to be more useful, we would want to develop a fully-specified computational model that can make quantitative predictions for how social contexts would change the utility of information seeking behaviors. This step will require formalizing the value and cost of information seeking actions in a modeling framework that can incorporate the effects of reasoning about other people's mental states. We have taken initial steps towards this goal, showing that a Bayesian model that integrated ideas from Optimal Experiment Design and recursive social reasoning could capture adults' decisions to forego information seeking in favor of more immediately rewarding actions when performance/presentational goals were highlighted by social partners in a causal learning task (E. J. Yoon et al., 2018). I hope that future versions of this active-social learning account will generate graded, testable predictions for behavior across a variety of domains.

Second, we used one formalization of active inquiry that focused on learners' actions given a specific and a set of candidate hypotheses. There are, however, other computational frameworks that have formalized human information seeking in different ways. For example, foraging models pursue the analogy that human information seeking is akin to animals' decisions about where and how long to look for food if their goal was to maximize caloric intake while minimizing effort and time (see Pirolli & Card (1999) for a review). Cognitive scientists have successfully modeled a range of behaviors as foraging, such as searching for concepts in memory (Hills et al., 2012) and decisions about where to direct visual attention (Manohar & Husain, 2013). In addition to search models, recent theories of curiosity-based learning in developmental robotics have developed algorithms that optimize intrinsic estimates of learning progress such that the system will focus on activities and stimuli of intermediate complexity where predictions are steadily improving, and uncertainty is steadily decreasing (Oudeyer & Smith, 2016). One of the challenges for researchers trying to integrate active and social learning is that the space of possible connections is quite large. We hope that by restricting our account to active decision making with social learning contexts will allow for progress in understanding one important sub-component of children's complex information seeking behaviors.

Third, our ultimate goal is to enrich the integrative active-social account to incorporate effects at the developmental timescale. The experiments in this thesis, however, often treated children and adults as two endpoints on a continuum to explore parallels and differences between children's information seeking and our best estimate of the mature state. The final experiment in Chapter 5 represents an exception where we measured children's adaptation of information seeking over the slightly longer timescale of multiple exposures to novel word-object links and compared this to their gaze patterns when processing highly familiar words, which they learned through exposure to many prior labeling events in their day-to-day experience. While this study is a useful first step, future work should measure developmental change over a longer timescale by densely sampling children at different ages and points of cognitive development. For example, it would be useful to know how children's rapidly improving productive language skill provides with a much wider range of information seeking actions in the form of verbal questions. One prediction of our account is that seeking social information via eye movements should become less useful when children can select the action "What is this thing called?" Another example of children's rapid theory of mind development. Our account predicts that young children should focus more on learning goals if they are less skilled at reasoning about others' beliefs. But, as their social reasoning abilities mature and their social environments become more complex, children may start to emphasize performance or presentation goals to update others' beliefs and appear competent.

Finally, the goal of the empirical research described here was to understand how children's information seeking behaviors adapt to support language processing. To accomplish this, we measured changes in children's gaze dynamics as they comprehended and learned words in highly-simplified contexts. This approach has the benefit of providing a high degree of experimenter control and relatively well-understood hypotheses linking external behavior (eye movements) to underlying psychological constructs (e.g., lexical knowledge) (although see XXX for a review of the challenges with interpreting the meaning of eye movements). The risk, however, is that the stimulus responses that we can measure in the lab do not reflect behaviors that support children's learning in their natural environments. That is, children learn their first language from conversation where there is dynamic back-and-forth turn taking and where the child's social partner is also making active decisions that control the flow of experience. This gap suggests two critical next steps for this line of work: (1) measure changes in children's information seeking within free-flowing social interactions

with caregivers (see Franchak et al. (2011) for an example using head-mounted cameras). And (2) develop more realistic lab-based experiments that incorporate more behaviorally-relevant features of children's learning environments such as the capacity to respond contingent on the child's actions (see Benitez & Saffran (2018) for an example of studying word learning with a gaze-contingent eye-tracking paradigm).

In sum, we set out to explore how children's information seeking adapts to a wide range of social contexts during two ecologically-relevant language tasks: familiar word comprehension and novel word learning. We found that children's real-time information seeking is quite flexible: when it was useful for language processing, children adapted their gaze to seek language-relevant information form social partners. Moreover, children and adults showed evidence of differential learning of new words when social gaze directed their visual attention. This work highlights two critical, open challenges for an integrated account of active learning within social contexts: (1) developing a more precise quantitative model of how interactions with other people change the utility of information seeking actions. By formalizing how social contexts can change the cost and value of actions, researchers could make more precise predictions about how social reasoning should interact with children's active learning. (2) Future empirical work should move beyond highly-constrained lab experiments to document information seeking behaviors within social contexts in children's everyday learning environments. Despite these open challenges, the integrative framework presented in this thesis represents a way forward for understanding how children's active learning interacts with social learning environments, which are ubiquitous in children's lives.

Appendix A

Supplementary materials for Chapter 1

A.1 Mathematical details of Optimal Experiment Design

This supplement contains 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 (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[1] (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).[2]

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

[1] 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.

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

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.[3]

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

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

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

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

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 $EU(Q = eatsMeat) > EU(Q = sleepsNight)$.

Appendix B

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

B.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

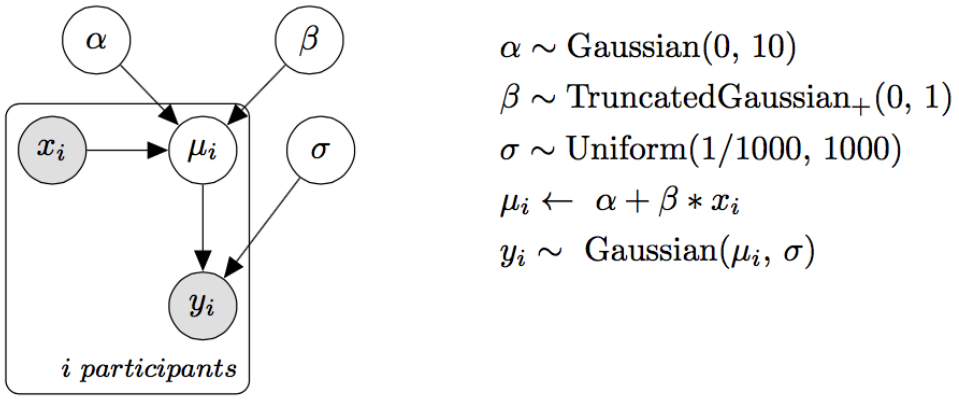


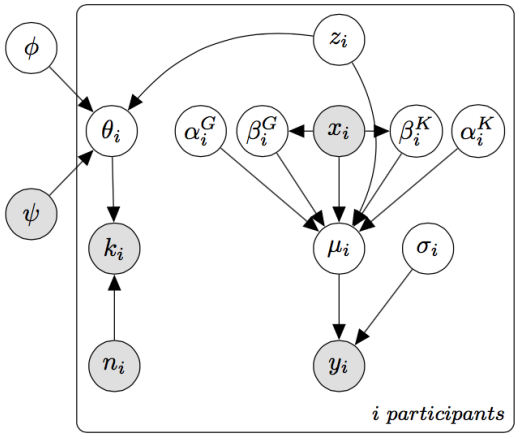
Figure B.1: Graphical model representation of the linear regression used to predict accuracy. The shaded nodes represent observed data (i.e., each participant's age, vocabulary, and mean accuracy). Unshaded nodes represent latent parameters (i.e., the intercept and slope of the linear model).

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

B.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 B.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).

B.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%, ψ, 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).

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

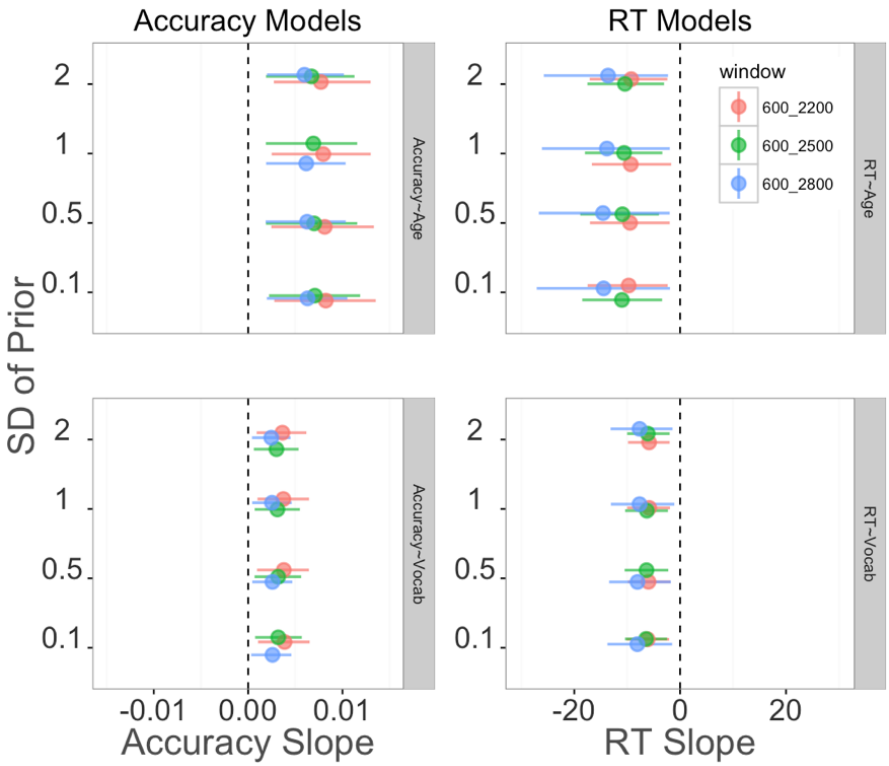


Figure B.3: Coefficient plot for the slope parameter for four different parameterizations of the prior and for three different analysis windows. Each panel shows a different model. Each point represents a coefficient measuring the strength of association between the two variables. Error bars are 95%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

B.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 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])

Table B.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 – 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 – 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 – 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

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.

B.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 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

Table B.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

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.

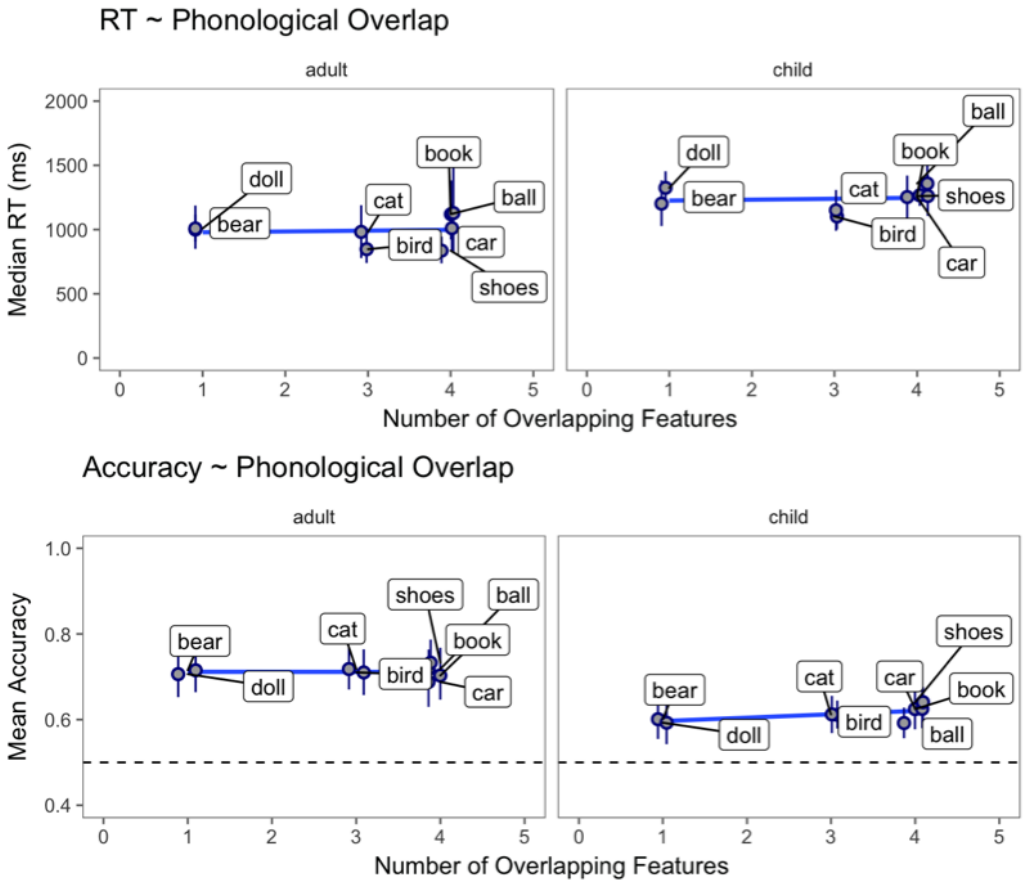


Figure B.4: Scatterplot of the association between degree of phonological overlap and RT (top row) and accuracy (bottom row) for both adults (left column) and children (right column). The blue line represents a linear model fit.

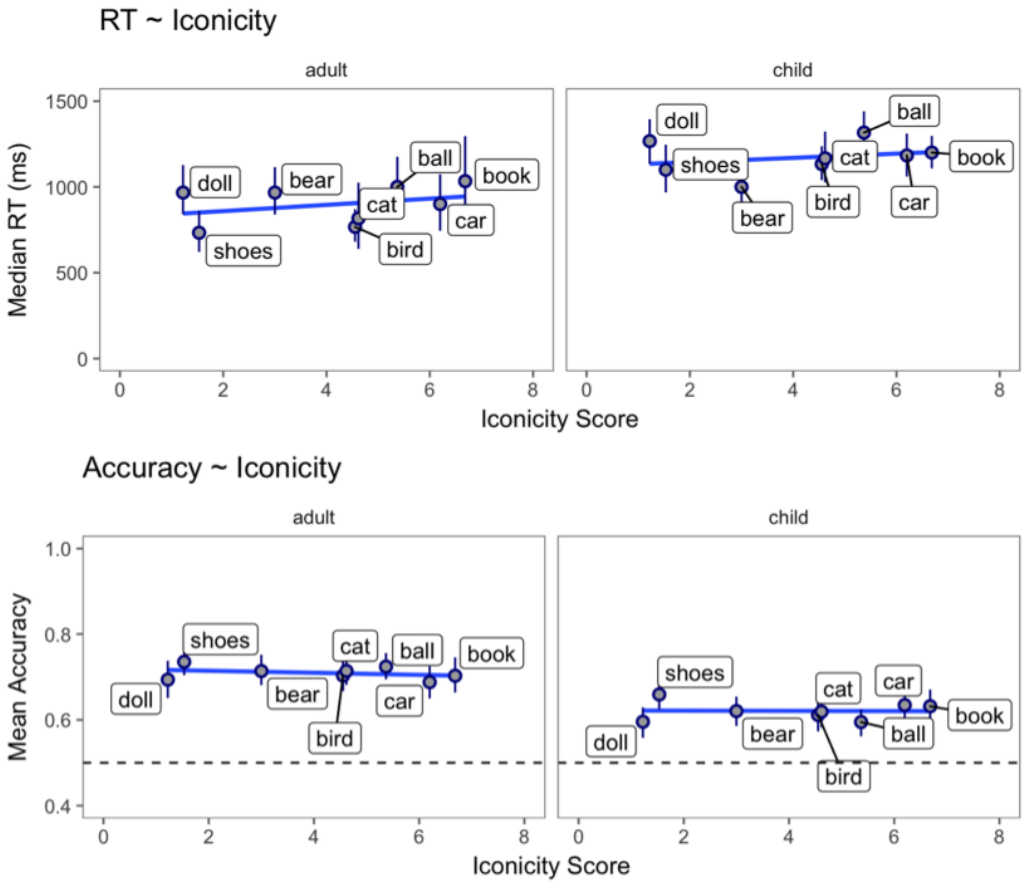


Figure B.5: Scatterplot of the association between degree of iconicity and RT (top row) and accuracy (bottom row) for both adults (left column) and children (right column). The blue line represents a linear model fit.

Appendix C

Supplementary materials for Chapter 4

C.1 Analytic model specifications and output

C.1.1 Experiment 1

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

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

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	

C.1.2 Experiment 2

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	

C.1.3 Experiment 3

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	3.14	0.002	**
Reliability Condition	3.28	1.50	2.19	0.028	*
Subjective Reliability	7.26	1.72	4.22	$< .001$	***
Reliability Condition*Subjective Reliability	-4.58	2.72	-1.69	0.092	.

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	13.09	$< .001$	***
Trial Type	-3.95	0.36	-10.91	$< .001$	***
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	11.58	$< .001$	***
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	0.27	-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	*

C.1.4 Experiment 4

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.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

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).

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, *10*(11), 512–518.

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

Benitez, V. L., & Saffran, J. R. (2018). Predictable events enhance word learning in toddlers. *Current Biology*, *28*(17), 2787–2793.

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

Bettger, J. G., Emmorey, K., McCullough, S. H., & Bellugi, U. (1997). Enhanced facial discrimination: Effects of experience with american sign language. *Journal of Deaf Studies and Deaf Education*, 223–233.

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.

Blythe, R. A., Smith, A. D., & Smith, K. (2016). Word learning under infinite uncertainty. *Cognition*, *151*, 18–27.

Blythe, R. A., Smith, K., & Smith, A. D. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*(4), 620–642.

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.

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*.

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.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

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.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). 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.

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.

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.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459.

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.

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

Emmorey, K., Klima, E., & Hickok, G. (1998). Mental rotation within linguistic and non-linguistic domains in users of american sign language. *Cognition*, *68*(3), 221–246.

Emmorey, K., Kosslyn, S. M., & Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing asl signers. *Cognition*, *46*(2), 139–181.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, *12*(2), 423–425.

Estigarribia, B., & Clark, E. V. (2007). Getting and maintaining attention in talk to young children. *Journal of Child Language*, *34*(4), 799–814.

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.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

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., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, *44*, 97.

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.

Fitzpatrick, M., Kim, J., & Davis, C. (2011). The effect of seeing the interlocutor on auditory and visual speech production in noise. In *Auditory-visual speech processing 2011*.

Fourtassi, A., & Frank, M. C. (2017). Word identification under multidomodal uncertainty. In *Proceedings of the 39th annual conference of the cognitive science society*.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750.

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.

Gabry, J., & Goodrich, B. (2016). Rstanarm: Bayesian applied regression modeling via stan. *R Package Version*, *2*(1).

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., Egyed, K., & Király, 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.

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. *Oxford Handbook of Causal Reasoning*, 515–548.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 201216438.

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.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390.

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

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).

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.

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1529–1542.

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

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.

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).

Harris, M., & Mohay, H. (1997). Learning to look in the right place: A comparison of attentional behavior in deaf children with deaf and hearing mothers. *The Journal of Deaf Studies and Deaf Education*, *2*(2), 95–103.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Heimler, B., Zoest, W. van, Baruffaldi, F., Donk, M., Rinaldi, P., Caselli, M. C., & Pavani, F.

(2015). Finding the balance between capture and control: Oculomotor selection in early deaf adults. *Brain and Cognition*, *96*, 12–27.

Hidaka, S., Torii, T., & Kachergis, G. (2017). Quantifying the impact of active choice in word learning. In. Cognitive Science Society.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431.

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.

Hoppe, D., & Rothkopf, C. A. (2016). Learning rational temporal eye movement strategies. *Proceedings of the National Academy of Sciences*, *113*(29), 8332–8337.

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.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267.

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.

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.

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.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713.

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.

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, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Attention, Perception, & Psychophysics*, *24*(3), 253–257.

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., Marchman, V., Fernald, A., & Frank, M. C. (2018c). Children seek visual information during signed and spoken language comprehension. *Preprint PsyArXiv*.

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.

MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and

sentence comprehension. *Handbook of Psycholinguistics*, *2*, 581–611.

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

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, *7*, 711.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

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.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369.

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.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants

or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147–164.

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.

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.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*(2), 133–137.

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., & Cottrell, G. W. (2007). A probabilistic model of eye movements in concept formation. *Neurocomputing*, *70*(13-15), 2256–2272.

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.

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.

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502.

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.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169–181.

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.

Piaget, J., & Cook, M. T. (1952). The origins of intelligence in children.

Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643.

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.

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

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.

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690.

Rogoff, B., Mistry, J., Göncü, 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.

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., 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.

Roy, D. K., & Pentland, A. P. (2002). Learning words from sights and sounds: A computational

model. *Cognitive Science*, *26*(1), 113–146.

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.

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*(2), 172–180.

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.

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.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, *41*(25-26), 3535–3545.

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., & 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.

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.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632.

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.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 9–9.

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.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*(6), 1011–1026.

Venker, C. E., Eernisse, E. R., Saffran, J. R., & Weismer, S. E. (2013). Individual differences in the real-time comprehension of children with asd. *Autism Research*, *6*(5), 417–432.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. The Royal Society.

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.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*, 14.

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

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.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.

Yoon, E. J., MacDonald, K., Asaba, M., Gweon, H., & Frank, M. C. (2018). Balancing informational

and social goals in active learning. In *Proceedings of the 40th annual conference of the cognitive science society*.

Yoon, E. J., Tessler, Michael Henry, Goodman, N. D., & Frank, M. C. (2017). "I won't lie, it wasn't amazing": Modeling polite indirect speech. In *Proceedings of the 39th annual conference of the cognitive science society*.

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.

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

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

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.

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

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

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*(1), 132–140.

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.