Self-directed learning in language development: Interactions of linguistic complexity, learner attention, and language socialization

by

Ruthe J Foushee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mahesh Srinivasan, Co-chair
Professor Fei Xu, Co-chair
Professor Susanne Gahl
Professor Michael C. Frank

Fall 2020

Self-directed learning in language development:
Interactions of linguistic complexity, learner attention, and language socialization

Abstract

Self-directed learning in language development:
Interactions of linguistic complexity, learner attention, and language socialization

by

Ruthe J Foushee

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Mahesh Srinivasan, Co-chair

Professor Fei Xu, Co-chair

Children are famously scrappy learners: curious, active, and resourceful. And yet when we consider their development of language — a complex social system that children are highly motivated to master — we tend to study them as passive recipients of adult guidance. This focus on the child's 'receipt' of linguistic input is a loss not only for our science (the one-on-one pedagogical contexts that research and public policy tend to emphasize are distinctly W.E.I.R.D.[1]), it overlooks language development as a fruitful domain in which to study children's self-directed learning, as well as insights that recent active learning frameworks could bring to our understanding of how language is learned. In this dissertation, I discuss language development as a coordinated process between communicative adults and increasingly active learners. In particular, I see children's learning from speech not directed to them, but rather *overheard*, as a uniquely ecologically valid test case of their self-directed learning capabilities.

A combination of experimental and computational studies from this perspective speak to an apparent paradox in the language development literature: while studies testing correlations between sources of language input in toddlers' home environments and later vocabulary growth have been taken to indicate that overheard speech is ineffective for word-learning, numerous experimental studies show learning from simplified indirect speech in laboratory settings during the same period. The idea that children may disattend to stimuli that are too complex for their current level of competence may help explain these conflicting results. That is, young rational learners may initially learn little from overhearing because the speech that surrounds them is too complex to maintain their attention — especially when compared to the speech that they regularly receive from caregivers.

[1] *Western - Educated - Industrialized - Rich - Developed*

A first study compares multiple empirically-motivated metrics of speech complexity in large-scale longitudinal child-directed corpora, and "overheard speech" simulated via corpora of adult-adult conversations. I find that words in simulated overheard speech are likely to be more abstract, unpredictable, later-acquired, and lower frequency than words in speech to children. This is likely to be true through at least the first four years of life, spanning the period when researchers have consistently failed to find a correlation between measurements of the overheard speech in children's environments and their future vocabularies.

Across three behavioral experiments in the second chapter, I test children's ability to learn from dense, naturalistic overheard speech in a context designed to place significant demands on their self-directed learning abilities, including their spontaneous recognition of an "information gap," and their ability to independently gather the information to fill it. In contrast to previous laboratory experiments — but consistent with many overhearing opportunities day-to-day — the speech I tested included *multiple* pieces of novel linguistic information, embedded in diverse sentence structures, and delivered in the register and rate typical of adult conversations. While all children in the sample were able to learn a set of 5–6 novel facts, only older preschoolers ($M_{\text{age}} = 5.1$ years) demonstrated robust learning of novel *words* through overhearing. Analyses of children's play and gaze behavior during the overhearing episode suggest that older children's success may be owed at least in part to their enhanced ability to coordinate attention between the referential context and the nearby speech.

In the third chapter, I develop a novel eyetracking method to test the classic idea that children learn best from information that is of an appropriate level of complexity for them — and in particular the role that children themselves might play in actively selecting and attending to potential sources of (linguistic) information. By measuring children's attention to a story narrated at distinct levels of verbal complexity — operationalized in terms of words' estimated age of acquisition — I find evidence that children attend more to speech that is more appropriate for their level of competence. Furthermore, while previous research has assumed that children's attention and learning are meaningfully related, this new method makes direct evidence possible. I find a strong correlation between children's self-directed attention to the story narration and their ability to recall its plot and to learn new words from it.

Inspired by qualitative studies typically limited to child-directed speech, in the fourth chapter, I develop a coding scheme that enables us to characterize the full range of potential sources of language accessible to a given child, in terms of their relative utility for word-learning. In applying this scheme to longitudinal video data from the home of a single English-learning child, I find that features that contribute to the referential transparency and salience of an utterance are not exclusive to child-directed speech, but rather occur with some lower frequency in overheard speech as well. In light of this, my analyses suggest a functional role for caregivers' exaggerated prosody as a self-reinforcing cue to language that is intended for the child, and therefore to where the child's attention is more likely to be rewarded (*i.e.,* because the speech was designed for them). Through this fine-grained coding

of individual utterances in context, our results uncover dynamics in how adults and children co-structure the early language environment — and how the landscape itself shifts with the child's maturation — that are invisible to entirely quantitative approaches. Ongoing work extends the ideas in the dissertation to new contexts and populations, beginning by employing the same multidimensional qualitative coding scheme to describe crosslinguistic learning environments, thereby facilitating contact with more humanistic fields like anthropology.

The fifth and final chapter tests the learner's ability to adapt their learning to the affordances of their language environment by measuring implicit lexical knowledge in Tseltal Maya infants, whose primary exposure to spoken language is through overhearing. While the preceding four chapters challenge our assumptions of how language is typically learned (*i.e.,* by emphasizing the role of overhearing, in contrast to receiving directed speech), this work aims to expand our (testable) notions of what counts as legitimate language knowledge by testing infants' knowledge of not only common nouns, but of honorific terms embedded in culturally specific sociolinguistic routines.

The studies in this dissertation draw on methods from natural language processing (NLP), computational linguistics, developmental psychology, psycholinguistics, and anthropology. The experimental studies share a focus on using naturalistic speech and ecologically valid learning contexts, and together point to the role of domain-general processes like attention, information processing, and adaptation in the course of language development.

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to my committee for all of the forms that their guidance and mentorship took: Mahesh Srinivasan, who above all has shown me incredible patience, Fei Xu, in whose active learning seminar during my first semester of graduate school I learned the vocabulary to express these ideas, Susanne Gahl, whose direct (anonymized) quotes I presented to faculty two years in a row as models of bringing the same rigor to theoretical discussions as to equity in the seminar room, and Mike Frank, whose thoughtful comments on this dissertation will serve me long after it is submitted.

— and to everyone who supported and helped me develop this research: Terry Regier, Yang Xu, Tom Griffiths, Celeste Kidd, and the López family.

I am grateful for all of the arbitrary reasons that I was lucky enough to get to do this.

I am grateful to mentors and role models and introducers of possibilities: Peggy Li, Linda Abarbanell, Susan Carey, Jesse Snedeker, Lauren Eby-Clemens, Michael Becker, Suzi Lima, Laura Wagner, Margaret Bridges, Luvy Vanegas-Grimaud, Helen Hadani, Lisa Branum, Tucker Hiatt, B & L, Mary Crane — and Margaret Wilch, in whose Tucson High science class I designed my first language experiment.

. . . my parents and alloparents: A & C, G & E, A & J, T — and Aunt Sally, who is still who I want to be when I grow up.

. . . all of the family I lost during graduate school (B, M, G, P), and all of the family I gained during graduate school (Sara Al-Mughairy, Frances Nkara, Rachel Jansen. . .)

. . . my siblings and graduate school siblings and lab siblings: Zi Lin Sim, Azzurra Ruggeri, Ariel Starr, Monica Ellwood-Lowe, Roya Baharloo, Ellie Kaplan, Stephan Meylan, Antonia Langenhoff, Katie Kimura, Shaun O'Grady, Neta Gottlieb, Hannah Bosley, Jon Wehry, Meg Bishop, Catherine Berner, Dorsa Amir, Gabriela Horton, Julene Paul, Anna Murphy, Paul Whang, Xanthia Tucker, Katherine Agard, Tyler Haddow, Pearl Bhatnagar, Ellen Danforth.

. . . my interlocutors: PC, who made it very easy to remember what I really care about, JK, the only other speaker of Shuh Shuh — and SK, the ultimate self-directed learner.

. . . my one phone call, Ben Whitney, and my fellow children's book critic, Mariel Goddu.

. . . the undergraduates I have had the privilege of working with, and especially Allison Fong, Grace Horton, Akil Ismael, Jacqueline Nguyen, Xiaoyuan Zhang, Saya Sivaram, Claudia Valdivia, and Tory Rose Full.

. . . all of my kids:

N, who I had never heard produce a multi-word utterance until he came skidding into the kitchen to stretch naked on the floor and declare that he was "lounging."

Z, who pulled her nightie over her knees the first night we hung out and informed me that we were going to talk about sharks.

N, who at 4 asked whether I was a kid or a grown-up, then cut me off with:
"Those are not grown-up pants."

BW, who inspired me to start a retirement account.

. . . and my always and forever co-learner, Nina.

For Baba.

The Children Are Swearing More in Quarantine.
The New Yorker headline, 2020

At the age of twenty-one months she tipped her hat and said, "how do you do."
A child's vocabulary and its growth
J. R. Grant, 1915

. . . Is coronavirus really popular right now?
A four-year-old, 2020

Introduction

OK, Linda.
Linda, listen, listen, listen.

A viral video[2] shows a kid of three or four looking up at his mother, half defiant, half exasperated. His neck is craned to make wide-eyed eye contact with her, and his hands leave the small of his back only to gesture his urgency (he's being denied cupcakes).

But Linda, honey, honey, lookit.
Linda lookit listen to me, ugh.
Look it Linda.

The video alternately disturbed and delighted online viewers (Navarrette, 2014). To me, that this demonstration of adult verbal behavior was so surprising reveals our tendency to see language learners as passive recipients of adult guidance, rather than highly motivated scavengers, experimentalists, and explorers of their rich linguistic worlds. It's easy to see why this perspective on children as passive language learners is so intuitive. For one, there is the observation that adults often direct simplified, exaggerated speech to children, which can look from the outside like *teaching* children to talk. For another, in the modern era, billboards, bus stops, and radio advertisements dominate the public sphere with propaganda framing language directed to young children as "nutrition." Imagery of caregivers spoonfeeding their prelinguistic infants invite analogies between baby food and new vocabulary. Of course, the idea that children have little to do with their own development of language long predates bus stop public service announcements: for example, Wilhelm Wundt wrote in 1897 that the "larger part of the process [of language development] depends on those about him, rather than on the child himself" (p. 342). One naturalistic source of evidence that we might see as challenging this view is precisely the young internet sensation's knowledge of not only his mother's first name, but of how to engage in an argument (the tone of voice, the performative restraint, the gestures). I like this video because both *Linda* and the choreography of an adult argument represent language knowledge that must typically be learned by *overhearing*, rather than by being taught directly.

[2] https://youtu.be/aFYsJYPye94

In this dissertation, I explore how adopting a view of the child as an "active" learner might give us purchase on basic questions about how language development unfolds, and in particular, what makes language input *effective.* Decades of research dedicated to this question have taken one-on-one interactions between caregivers and young children as the default context for learning, employing diverse experimental and observational methods to identify the features most associated with learning new words. These include social dimensions of the caregiver's behavior, like establishing joint attention (Carpenter et al., 1998; Hirsh-Pasek et al., 2015; Morales et al., 2000; Scott et al., 2013; Tomasello & Farrar, 1986) or imitating the child (Tamis-LeMonda et al., 2001), linguistic dimensions, like using exaggerated prosody (Fernald, 1984; Ramírez-Esparza et al., 2014) and diverse syntax (Barnes et al., 1983; Hoff & Naigles, 2002; Huttenlocher et al., 1991; Huttenlocher et al., 2010; Rowe et al., 2017), and referential dimensions, like gesturing to disambiguate an utterance or to elicit the child's attention (Cartmill et al., 2013; E. V. Clark & Estigarribia, 2011).

Yet dyadic interactions represent but one of the myriad potential learning contexts in which children find themselves across the day and around the world. Relative to what we know about the language that children hear *directly* and its consequences for children's development, we know almost nothing about the language that children hear *around* them, and how children's own development might interact with it to shape learning. Similarly, relative to what we know about how infants and toddlers learn when their attention is being directed by an adult, we know almost nothing about how children learn when they must manage their own attention across their complex linguistic landscapes.

These 'knowledge gaps' (Loewenstein, 1994) are surprising when we consider that child-directed speech practices vary considerably across households and societies (P. Brown, 1998; Casillas et al., 2019; Cristia et al., 2019; de León, 1998; Lieven, 1994; Mastin & Vogt, 2016; Ochs & Schieffelin, 1995; Ochs, 1982, 1990; Pye, 1986a, 1986b; Schieffelin, 1990; Shneidman & Goldin-Meadow, 2012a; Sperry et al., 2019; Vogt et al., 2015; Ward, 1971; Weisleder & Fernald, 2013), but also because of foundational ideas in developmental science more broadly of the child as an active participant and driver of their own learning (e.g., Berlyne, 1960; Bruner, 1961). Indeed, growing evidence demonstrates that children are robust self-directed learners in other, non-linguistic domains, where they regularly gain new information through their own strategic exploration of their environments (e.g., Begus et al., 2014; Cook et al., 2011; Saylor & Ganea, 2018; Schulz & Bonawitz, 2007; Sim & Xu, 2017).

By expanding the scope of our attention beyond language spoken to young children directly, I argue that we come closer to recognizing the genuine complexity of the language-learning environment, of the linguistic system, and of the practice that is linguistic communication. In addition, a scientific focus on learning language through overhearing encourages us to expand the developmental contexts we consider — and consider as normative.

The studies in the following chapters are interested in overhearing as a language-learning context that may provide us with a natural illustration of children's active learning skills, including children's rational deployment of attention, and their independent gathering of information to reduce uncertainty in their environments (Gureckis & Markant, 2012; Loewenstein, 1994; Saylor & Ganea, 2018). I take a multidisciplinary approach, drawing on methods and insights from natural language processing (NLP), computational linguistics, developmental psychology, psycholinguistics, and anthropology:

Chapter 1 explores the learnability of child-directed and simulated overheard speech across transcribed language corpora of adults speaking to children versus of adults speaking to other adults. This chapter establishes both that adults calibrate their speech to young children, and that it may be rational for children to ignore overheard language input early in development by virtue of its complexity relative to the speech that they receive directly from caregivers.

Chapter 2 uses the control of the laboratory to ask whether children whose language development is more advanced can learn novel linguistic information from relatively complex, naturalistic overheard speech.

Chapter 3 tests the hypothesis that children's attention to natural language is responsive to its complexity, introducing children's selective attention as a potential determinant of the "effectiveness" of different linguistic inputs, and a potential explanation for results showing no relation between the speech toddlers overhear and their vocabulary development.

Chapter 4 revisits the relative learnability of child-directed and overheard speech, this time by analyzing fine-grained variation in the daily language experiences of a single child, using a system designed to capture diversity in the structure of the language environment for children across the world.

Chapter 5 discusses ongoing work dedicated in part to bringing the literature on language *socialization* (e.g., P. Brown, 2011; P. Brown & Gaskins, 2014; Ochs & Schieffelin, 1984; Schieffelin & Ochs, 1987; Solomon, 2011) into contact with quantitative and experimental methods in developmental psycholinguistics. This effort includes field experiments testing infants' knowledge of language that could *only* have been acquired through overhearing, in an indigenous Mayan context where caregivers are in constant contact with their prelinguistic infants, but rarely engage them verbally.

Across dissertation chapters, the experimental studies share a focus on using naturalistic speech and ecologically valid contexts, and point to the role of domain-general processes like attention and information processing in the course of language development.

What Counts as Effective Input?

Of the topics at the intersection of active learning and language development, I see learning from overhearing as particularly compelling, as it relates to one of the most basic questions in the cognitive science of language: that is, *how much does language* ***knowledge*** *rely on language* ***input***? This question has attracted scholars from all human-interested fields, from psychology and linguistics to sociology and anthropology (though how each might be inclined to phrase the theoretical question might differ). In a relevant passage of a 2007 review of the literature on infant-directed speech, Soderstrom writes:

> Ask a formal linguist or a developmental psychologist about the characteristics of "the input," and you will get widely divergent answers. Are we talking about a formal characterization of the structural properties of the language being learned? Or are we talking about "speech" in all its ambiguous, degenerate and disfluent glory? What about the speaker — should we only consider maternal input? What about the father, other caregivers, the nanny, the older sibling, the local shopkeeper or the ubiquitous television? Should we only consider speech directed at the infant, or all of the speech bouts produced in hearing range? For speech directed at the infant, what age should we consider? In order to answer how the input is relevant to the process of language development, we must have a clear understanding of what constitutes "the input."

Tied to the ambiguity of what counts as the input is the question of what counts as language knowledge: that is to say, "...effective input for *what*?" Answers to *this* question have varied by discipline, and arguably dictate the degree to which researchers are interested in input effects, along with the empirical methods that they employ to uncover them.

In one telling of the story, the modern subfield of language acquisition owes its origin to Chomsky's (1959) articulation of both the abstract complexity of his and others' linguistic intuitions, and the impossibility of learning them from the environment (Baars, 1986; Chomsky, 1957, 1959; Gardner, 1987). Much of Chomsky's focus was on the system of grammatical rules that, despite predicting how well-formed entirely novel sentences would be judged by a native speaker, no such speaker could spontaneously produce. Language knowledge, therefore, meant *abstract syntax*, of the kind possessed by all the world's speakers.

Chomsky's formulation of the learning problem suggested that productive grammatical knowledge could not be inferred solely on the basis of the linguistic evidence that children received (evidence frequently described as "degenerate" and "ill-formed"), and that therefore acquisition must owe to biologically endowed mechanisms constraining all the world's languages. Chomsky made a related distinction between individuals' competence and their performance, such that individuals' linguistic behavior should not be taken as evidence of their language knowledge. This extended even to child language: "It is commonly assumed that there is a two-word stage, a three-word stage, and so on, with an ultimate Great Leap Forward to unbounded generation. That is observed in performance, but it is also observed

that at the early stage the child understands much more complex expressions" (Chomsky, 2004). Together, these commitments meant that there was little to learn from studying learners' language environments, or from measuring language 'outcomes' across individuals.

> We know that the grammars that are in fact constructed vary only slightly among speakers of the same language, despite wide variations not only in intelligence but also in the conditions under which language is acquired. As participants in a certain culture, we are naturally aware of the great differences in ability to use language, in knowledge of vocabulary, and so on that result from differences in native ability and from differences in conditions of acquisition; we naturally pay much less attention to the similarities and to common knowledge, which we take for granted. But if we manage to establish the requisite psychic distance, if we actually compare the generative grammars that must be postulated for different speakers of the same language, we find that the similarities that we take for granted are quite marked and that the divergences are few and marginal.

Arguably as a consequence of the field's focus on aspects of language knowledge that speakers tend to *share*, linguists have typically paid less attention to a conception of child-directed speech as non-trivially related to a child's acquisition of language. That is, on one definition of language — the Chomskyan one — there is indeed very little variability in outcomes, as all children eventually become adults, endowed with the linguistic intuitions that describe the grammar of their mother tongue. In any conversation about 'effective input', this is an important place to start.

In contrast, the relation between the child's environment and their mastery of language has been a major interest in more prescriptive or quantitatively-minded fields, like education and psychology — where what is deemed important about language knowledge and where we see variability in so-called language outcomes are more likely to coincide. In education, for example, language is most visible as a means of imparting, acquiring, and demonstrating knowledge, suggesting *vocabulary* as a theoretically meaningful and practically measurable target of language learning. In psychology, where statistics are integral to overcoming individual differences and arriving at often-normative generalizations about populations, variability in a measure like vocabulary begs explanation from other, equally quantifiable variables. Through controlled experiments in the lab, psychologists have identified contexts associated with more or less successful word-learning in samples of children at different ages, while a growing body of research seeks correlations between coarse measures of the child's typical language environment and their later outcomes.

This background is relevant because it informs an important subgoal of this dissertation, which is to begin to make contact between the literatures on linguistic input that are currently siloed in linguistics, in psychology, and in anthropology. So as to maintain a connection to existing bodies of quantitative evidence, the bulk of the dissertation focuses on the implications of linguistic input for learning new *words*, though I embrace more expansive notions of language knowledge in Chapter 5 and in ongoing work.

In the following section, I review the psychological literature on child-directed speech ('in all its ambiguous, degenerate and disfluent glory;' Soderstrom, 2007), as context for the research reported in the chapters that follow.

Talking with Children

1.	(a)	You should try this new food. It tastes like chicken except it's a bit spicy. I recommend you give it a try.	(b) Try this new food buddy. Try at least 3 bites to see if you like it.
2.	(a)	Just load the clothes in the front, put the soap in the drawer and then pick the settings you want — I usually set it to "normal wash" and "cold" but you can press the buttons and turn the knob to change it to whatever you want.	(b) The clothes go in first. Ok... good job! Now close the door. Open this little door. Yup, put the soap right in there.
3.	(a)	Get the first aid kit under the bathroom sink, then call an ambulance. Our address is 111 Pine street in Hopetown.	(b) Hey booboo, go look under the bathroom sink for the white box with a red plus sign on it like this and get it. Then get Mom's phone and dial the 9 and the 1 and the 1. Tell them you need help.

Responses in (a) and (b) are from the same speakers. In (a) responses, adults wrote scripts for talking to a close friend. In (b) responses, adults imagined they were talking to their child,[3] whose age they later provided (Appendix A). Even in this contrived context, comparing the (a) and (b) samples illustrates generalizable trends in adult speech to children. The first observation is that they are different. Across languages, places, and cultures, adults alter their speech to children, relative to when they are speaking with other adults (R. Brown, 1973; Cameron-Faulkner et al., 2003; Ringler, 1981; Snow, 1972, 1977).

[3] "Please take a moment to imagine your child in front of you. // Now, imagine having to explain [...]"

Child-directed modification happens across modalities (Holzrichter & Meier, 2000) and at all levels of linguistic structure, such that for every level of structure, there exists an adult speaker population that regularly adjusts it when talking to children (Bateson, 1975; P. Brown, 1998; Choi, 2000; C. A. Ferguson, 1977; Phillips, 1973; Pye, 1986b; Schieffelin, 1990; Snow, 1972). Even in contexts where caregivers are typically thought of as *not* modifying their speech to young children (i.e., because they do not typically exaggerate their pronunciation), adults adjust the pragmatic demands of their utterances, supplying children with phrases to be repeated, engaging them in ritualized language games (de León, 1998), or speaking 'for' them until they can produce speech on their own (Schieffelin, 1990).

Adjustment at some levels of linguistic structure are more salient than others. For example, what early researchers called "baby talk" (later "motherese," then "parentese;" see Saxton, 2009; Soderstrom, 2007; Solomon, 2011 for a review) often *sounds* very different from speech directed to other adults (Cristia, 2013; C. A. Ferguson, 1977; Fernald, 1984; Fernald et al., 1989). In diverse languages, the pitch of speech to children tends to be higher, more variable, and to occur in a common set of shapes, or 'contours' (Broesch & Bryant, 2015, 2018; Farran et al., 2016; Fernald & Morikawa, 1993; Grieser & Kuhl, 1988; Niwano & Sugai, 2002). The specific prosodic features modulated in speech to children vary across languages. Nevertheless, prosodic contours apparently have enough (a) relation to meaning, and (b) commonality, that English-speaking toddlers in a 1993 study could interpret an admonishment or approval from an Italian caregiver whose speech had a similar pitch contour to their mothers', even if the children had never before heard Italian (Fernald & Morikawa, 1993).

When talking to child audiences, adults also tend to speak more slowly, lengthening their syllables, and producing longer pauses and fewer disfluencies (Broen, 1972; Broesch & Bryant, 2018; Cross & Morris, 1980; Pine, 1994; Ratner & Pye, 1984). In many cases, sentence structure is also leaner, with fewer subordinate and relative clauses, and less negation (Newport et al., 1977; Phillips, 1973; Sherrod et al., 1977): what is one sentence for an adult audience in (2a) becomes three sentences for a child listener in (2b). Sentences are often partially repeated, such that common constructions reoccur across adjacent utterances. In one study of sentence-level patterns in child-directed speech, fully half of all caregivers' utterances belonged to the same set of 52 sentence frames (Cameron-Faulkner et al., 2003). Analogues for many of these modifications appear in child-directed sign language, where a sample of ASL-signing caregivers, for example, tended to produce signs of greater duration and repetitiousness to their 8–12-month-old children (Holzrichter & Meier, 2000).

The examples we elicited from online participants illustrate another frequent way in which adults modulate their speech to children, which is through the *words* they use. In a 1964 review, Ferguson highlights cross-linguistic commonalities in prototypical "baby words" like *dada* and *choo choo* from languages across language families: for example, *baba* in Marathi and Arabic, *tata* in Spanish, *papi* in Comache, and even examples from Latin, e.g., *pappa* for food. Similar reduplicated forms often characterize terms of affection like *boo boo*, and diminutives like *buddy* (C. A. Ferguson, 1977). Like the respondents in (2) and (3), above, adults commonly intersperse these terms throughout the literal content of their speech, often

trading off with repetitions of the child's own name (Broen, 1972; Ervin-Tripp, 1978; C. A. Ferguson, 1964, 1997; Pye, 1986b). Even outside of baby words, when speaking to children, adults tend to use words that are conceptually simpler than they use when conversing with other adults (e.g., "dog" versus "collie;" Blewitt, 1983).

Child-Directed Modification: What Is It Good For?

Evidence that adults' prosodic, syntactic, lexical, and phonetic modifications in their speech to children *promote* language development is highly mixed. One conclusion is that it depends in part on how the learning problem is formulated (e.g., A. Clark & Lappin, 2013; Guevara-Rukoz et al., 2018; A. Martin et al., 2015; Rafferty & Griffiths, 2010). For example, the expanded vowel space of many child-directed registers might help infants identify the relevant sound categories of their language (Eaves et al., 2016; Eimas, 1971; Werker & Tees, 1984), but not the boundaries for words, or syntactic phrases (Cristia, 2013). On another prominent hypothesis, the exaggerated pitch contours of child-directed speech make "prosodic bootstrapping" possible, enabling the child to parse incoming speech into meaningful syntactic units marked by prosodic boundaries (Fisher & Tokura, 1996; Nelson et al., 1989; Seidl, 2007). However, adult utterances to infants and young children are also typically shorter, which limits the syntactic phrase boundaries and intra-utterance structure that prosody could help identify. Meta-analyses suggest that if there are language benefits to caregivers' exaggerated prosody, they are largely to pre-linguistic outcomes like vocal imitation, and gone by the end of the first year (Spinelli et al., 2017).

There is also some evidence that adults' tendency to simplify the syntax of their utterances might benefit learners (Bohannon et al., 1982; Cross & Morris, 1980; Furrow et al., 1979; Murray et al., 1990; Raneri et al., 2020). For example, the greater repetition often observed in child-directed speech might aid children's segmentation and processing of embedded words (Newman et al., 2016). When Fernald and Hurtado (2006) tested infants' knowledge of familiar words in a looking-while-listening experiment, infants looked to the correct image faster when its label was presented in one of the frames used most frequently by caregivers than when it was presented in isolation. In another study, parents' use of "baby talk" words like *choo choo* and *baba* was correlated with their nine-month-old's vocabulary growth (Ota et al., 2018).

Where Does Child-Directed Speech *Come From?*

Notably, adults adjust their speech less and less as the child gets older (Henning et al., 2005; Liu et al., 2009; Smith & Trainor, 2008). Caregivers' pitch settles. They repeat themselves less (Sherrod et al., 1977). Their utterances get longer and more complex, filled with more, and more different, words (Adi-Bensaid et al., 2015; Ervin-Tripp, 1978; Fraser & Roberts, 1975). And they talk faster (Ko, 2012; Raneri et al., 2020). It can look like caregivers are altering their speech *for the sake of* the child learning language (Bateson, 1975; Bohannon et al., 1982; Cross, 1977; Cross & Morris, 1980). This interpretation of adults' child-directed speech as *pedagogical* is especially intuitive from the view of children as passive, rather than active, learners. However, an alternative is to explain adults' accommodation of young children with reference to the same processes that we know are in operation elsewhere, while recognizing that adults tend to be egocentric in speech production (Horton & Keysar, 1996).

The contrast between the above two styles of explanation rhymes with a long-standing contrast between parallel traditions in psycholinguistics: what Brennan and Hanna (2009) describe as the "language as product" and "language as action" traditions. Roughly, one perspective on language development suggests language as something given (or fed!) to young children by adults. The other frames language development as occurring in the course of normal sociocultural practice. That is, adults' modified speech may arise out of the very same communicative practices that they engage in with other adults — only, when talking to children, adults' conversational partners know very little about the world, can communicate even less, and exhibit highly inconsistent attention.

If we also choose to adopt the view that language is an inherently communal enterprise (*à la* H. Clark & Brennan, 1991; Pickering & Garrod, 2004), then we might say that in adult-child exchanges, adults take on more of the conversational labor early on, out of a genuine need to get their message across. And there is some evidence for caregivers' genuine need to communicate with their children from studies of caregivers' productions. In one study of the discourse functions of caregiver speech, for example, over a quarter of caregivers' utterances to their two-year-old children were coded as 'directives,' a category that included prohibitions like "don't touch that," as well as commands like "give me the toy" (Rowe et al., 2004).

If nothing else, the frequency of directive utterances suggests that caregivers often use speech in contexts where their children's comprehension is important (and likely to influence the speech itself). In the case of directive speech acts, parents receive immediate feedback from the child's behavior on whether their message has been received. However, even comparatively subtle cues to the success versus futility of our communicative goals can affect how we produce speech online. For example, psycholinguistic experiments show that evidence our interlocutor is distracted or confused results in not only verbal repair behaviors to regain common ground (Clark, 2014), but apparently subconscious effects on our own speech production that lead speakers to do things like selectively emphasize critical words when giving task instructions (Rosa et al., 2013). This work suggests that the forms that our messages take depend on feedback from our interlocutors that they are paying attention and comprehending. Critically, research also suggests that the attentiveness to our inter-

locutor that gives rise to well-calibrated input is itself *costly* (Abel & Babel, 2016; Branigan et al., 2007; Branigan et al., 2011; Fischer, 2016; Yurovsky et al., 2016a), further predicting that the distribution of communicative labor between adults and children should initially skew toward adults, but become gradually more and more equal as children's domain-general cognitive capacities develop.

Another thing we would expect from this demystified origin of child-directed speech is for caregivers' speech to children to be conditioned on the same factors that condition both the caregivers' need to communicate and their attention (see also Ellwood-Lowe et al., *under review*). This appears to be borne out across human communities (Gutiérrez & Rogoff, 2003; Rogoff et al., 2003). For example, infant-directed speech is reported to be rare in Mayan households, where mothers and infants are in almost constant physical contact and non-verbal communication. Infants begin to receive directed speech once they have matured out of this developmental period and begin walking (P. Brown, 2008; Shneidman & Goldin-Meadow, 2012a) — putting them at a distance from the mother, and opening the possibility of mishaps, mischief, and small requests from one party to the other. That is, communication becomes *relevant*. Even within a given developmental window, caregivers' propensity to engage their children arguably varies with their attentional resources: for example, caregivers may speak less to their children when experiencing greater financial scarcity or money worries, relative to their own patterns when finances are less of a concern (Ellwood-Lowe et al., 2020).

Different family organizations offer additional illustrations of the opportunistic nature of speech to young children. For example, infant-directed speech is more common in nuclear family structures, where a caregiver is more likely to find themselves at home with only an infant to talk to, long before that infant can meaningfully reply. In contrast, in extended-family dwellings — the norm across much of the world — caregivers are more likely to find a more suitable conversational partner nearby. As a result, the child will likely hear less caregiver speech (but will be spoken to by a greater number of household members, and overhear more conversations; P. Brown, 2011; P. Brown and Gaskins, 2014; Ochs and Schieffelin, 1984, 1995; Shneidman et al., 2013).

A Puzzle in the Existing Literature

The experimental and observational studies in the following chapters shed light on a puzzle in the literature on effective language input. That is, in spite of children's self-directed learning skill in other domains, correlational studies repeatedly suggest that overheard speech is *not* a source of word-learning (Ramírez-Esparza et al., 2014, 2017; Shneidman et al., 2013; Shneidman and Goldin-Meadow, 2012a; Weisleder and Fernald, 2013; see Table 1). These studies record samples of children's language environments when they are between 18 and 30 months of age, coding all recorded utterances as 'child-directed' or 'overheard.' When counts of the number of tokens (total words) or types (unique words) are used to predict the same children's vocabulary sizes at 36–42 months, measurements of the speech categorized

as 'child-directed' reliably correlate with later vocabulary size, while the same measurements of 'overheard,' or 'adult-directed,' speech do not. The inference that differences in children's later vocabulary sizes reflect words learned *from* child-directed speech is partially supported by additional correlational evidence relating words' frequency in child-directed speech to their typical age of acquisition (e.g., Braginsky, Yurovsky, et al., 2019; Swingley & Humphrey, 2018). Surprisingly, the independence of overheard speech and vocabulary size persists even in contexts where the majority of children's input is overheard (Shneidman & Goldin-Meadow, 2012a), which characterizes many cultural communities (Correa-Chávez & Rogoff, 2009; Heath, 1983; Ochs, 1982; Schieffelin, 1990; Ward, 1971).

Across studies, then, number of tokens of child-directed speech accounts for variance in children's vocabulary outcomes... Why shouldn't overheard speech, which represents valuable data about the system the child is acquiring, and undoubtedly also affects vocabulary outcomes? Adding to this puzzle is the fact that children can learn a novel word from overheard speech in experimental contexts by as early as 18 months (Akhtar, 2005; Akhtar et al., 2001; Floor and Akhtar, 2006; Gampe et al., 2012; Martínez-Sussmann et al., 2011; Shneidman et al., 2009; see Appendix E). What makes lab-presented overheard speech so different from its naturalistic counterpart?

Drawing on the idea that children preferentially attend to stimuli that are at a manageable level of complexity (Gerken et al., 2011; Kidd et al., 2012, 2014), I propose that the apparent language-learning disadvantage of overheard speech early in development is owed in part to its complexity relative to other sources of input, leading children to disattend to it until it is of equivalent complexity to the child-directed speech they regularly receive. A combination of results derived from novel experimental paradigms, computational methods, and observational analyses suggest this proposal holds promise.

Chapter 1

Children May Filter for Complexity, Ignoring Overheard Speech Until It Is Subjectively Learnable

Abstract

In this chapter, we explore the idea that infants may pay less attention to and learn less from overheard speech, due in part to its complexity relative to child-directed speech. We employ multiple empirically-motivated operationalizations of 'complexity' or, conversely, *learnability*, that can be computed on decontextualized speech corpora. These include lexical variables like concreteness, frequency, and age of acquisition, associated with faster processing in production and comprehension in psycholinguistic research with adults. Additional information-theoretic measures capture the predictability versus *entropy* of randomly sampled child- and adult-directed speech. If caregivers' child-directed speech offers an accurate estimate of the speech complexity appropriate for a given child, and if children's attention to different sources of spoken language input is at least partly responsive to their efficiency processing them, then we might expect overheard speech at or below the level of the child's typical child-directed speech to be eligible as a target of the child's attention and a reliable source of new vocabulary. Our analyses suggest that for most children this may not be until after language development is well underway, and after the point at which previous studies have unsuccessfully used measurements of overheard language in children's daily lives to predict vocabulary growth.

1.1 Introduction

> Talking with children isn't easy... Each child's ability to speak, understand, and converse is a moving target: it changes not just yearly, monthly, weekly, and daily, but moment to moment, and we cannot know where he or she is at any moment. How, then, do we manage to talk with children?
>
> How to Talk with Children
> *H. H. Clark, 2014*

One way of thinking about language development is as a coordination problem (Yurovsky, 2018). Adults alter their speech to children in the course of *grounding*, or establishing common knowledge, with them (H. Clark & Brennan, 1991; Clark, 2014). 'Active' child learners, for their part, play a role in shaping the language they are exposed to, and may monitor the different sources of language around them, selectively attending to the language that will best support their learning. Early in development, the speech that they receive directly — over, for example, the speech that they might overhear. This is in part because when speaking with toddlers and young children, adults seek to engage them, and receive feedback in the course of their interaction on, e.g., which words are already in their vocabularies (C. A. Ferguson, 1977). Researchers at least since Snow (1977) have argued that much of children's remarkable success at language-learning should be attributed to adults' coordination with them (e.g., Yurovsky, 2018). Compared to adult-directed speech, the resulting child-directed register is simplified, attention-getting, slower, and contains fewer unique words and more repetition (Broen, 1972; Ervin-Tripp, 1978; C. A. Ferguson, 1964; Soderstrom, 2007).

Research aimed at characterizing what counts as effective input repeatedly finds that such *child-directed speech* appears to advantage learning, eliciting and maintaining children's attention in infancy (Graf Estes & Hurley, 2013; Ma et al., 2011; Singh et al., 2009), and augmenting their language processing efficiency and vocabularies as toddlers (Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a; Weisleder & Fernald, 2013). One body of work relates snapshots of children's early language environments to their language outcomes six months or a year later, consistently finding that the amount of child-directed speech that toddlers receive is predictive of their vocabularies, but not the total amount of language that they hear from other sources, including speech classified as "overheard" (Shneidman et al., 2013; Shneidman and Goldin-Meadow, 2012a; Weisleder and Fernald, 2013; see Table 1. Experimental overhearing studies have typically presented novel words in highly simplified speech, where the new words are utterance-final, stressed, and embedded in explicit labeling or directive sentence frames (Akhtar, 2005; Akhtar et al., 2001; Floor & Akhtar, 2006; Gampe et al., 2012; Martínez-Sussmann et al., 2011; Shneidman et al., 2009). Thus, elegant as they are, previous experimental studies may not have represented the naturalistic phenomenon.

Table 1: *Correlational Studies Using Child-Directed and Overheard Speech Quantity to Predict Early Language Outcomes*

		Input		Language Outcome	
Study	Population	Age*	Measure	Age*	Measure
Shneidman and Goldin-Meadow, 2012a	Yucatec Mayan	24	adult† word types	35	Peabody Picture Vocabulary Test (PPVT)‡¶ Expressive One-Word Picture Vocabulary Test (EOWPVT)§‖
Shneidman et al., 2013	English-learning; single and multiple primary caregivers	30	adult word types and tokens	42	Peabody Picture Vocabulary Test (PPVT)‡¶
Weisleder and Fernald, 2013	Spanish-learning	19	adult word tokens	19	looking-while-listening task#
				24	MacArthur-Bates Communicative Development Inventories (MCDI)** ‖

Note. In all studies, INPUT measures are correlated with LANGUAGE OUTCOME measures, but only for speech classified as *child-directed*, and not for speech classified at *overheard*.

* in months
† $\geq$ 11 years of age
‡ Dunn and Dunn, 1981
¶ receptive vocabulary measure
§ Brownell, 2000
‖ productive vocabulary measure
\# speech processing efficiency; Fernald et al., 2008
** Fenson et al., 2007; Spanish-language version: Jackson-Maldonado et al., 2003

That children as young as 18 months (Floor & Akhtar, 2006; Gampe et al., 2012) attend to and learn from overheard utterances that are at least as simple as typical child-directed ones supports the idea that they might otherwise be filtering for complexity.

In this chapter, we extend findings that infants preferentially attend to stimuli of intermediate complexity (Gerken et al., 2011; Kidd et al., 2012, 2014) to explain gaps in our understanding of the relative value of different input types for language development (see also Räsänen et al., 2018). Specifically, we quantify the difference in complexity or learnability between child- and adult-directed speech across developmental time. We focus our analyses on the lexico-semantic and distributional properties of the *words* adults use, reasoning that higher proportions of words that are easier to process for adults should translate to facilitated comprehension in children (Swingley et al., 1999), along with facilitated online processes implicated in word-learning, including speech segmentation and prediction (de Carvalho et al., 2019; Hills & Adelman, 2015; Reuter et al., 2019; Reuter et al., 2018; Weisleder & Fernald, 2013). We capitalize on existing corpora of transcribed naturalistic speech to test whether and for how long the words adults use in their speech to children are significantly simpler than the words they use with other adults. The analyses that we report can therefore be seen as quantifying (1) the degree to which adults simplify their speech when talking to children, and (2) if adult spoken language corpora approximate the speech that children might overhear, the differential in child-directed and overheard speech in children's daily lives.

Outside of child-directed speech, experimental research suggests that infants selectively attend to stimuli at a manageable level of complexity: neither too simple nor too complex (termed the 'Goldilocks effect': Kidd et al., 2012, 2014). In looking time studies with 7–8-month-olds, for example, children's probability of looking away from a visual display is lowest when the visual or auditory events are within an intermediate range of predictability. In another study, Gerken et al. (2011) tested infants' attention in an artificial language-learning paradigm where the linguistic stimuli differed in their *learnability*. There, 17-month-old infants exhibited greater attention for language samples whose morphological patterns were rule-governed and learnable, compared to language samples whose morphological system was either perfectly uncompressable or was empirically unlearnable by previous samples. Such studies indicate that children may be implicitly tracking the complexity of different stimuli and budgeting their attentional resources, attending only to those stimuli that fall in a learnable range, according to their knowledge state. Authors of the aforementioned studies describe infant's selective attention as a rational general mechanism for learning, which prevents them from wasting precious cognitive resources attempting to compress patterns that are too unpredictable, and therefore potentially unreliable. Even for highly complex stimuli that is patterned and could theoretically support learning, infants' progress will be inefficient compared to the learning progress they could be making elsewhere (or at a later date).

We propose that part of children's own role in language development may be to selectively direct their attention to the most effective learning data, a hallmark of the 'active' learner (Gerken et al., 2011; Gottlieb et al., 2013; Kidd & Hayden, 2015; Kidd et al., 2012, 2014).

Attention is arguably a prerequisite for word-learning (Graf Estes & Hurley, 2013; Ma et al., 2011; Singh et al., 2009). Thus, if children are monitoring the subjective complexity of different inputs "in the wild," and if overheard, adult-directed speech is consistently more complex, children may not learn new words from it early on in life because they do not direct their limited attentional resources to it. Over years, though, speech directed to a child becomes speech directed to an adult. When speech around a given child is of comparable complexity to speech directed to her, it may become eligible as a target of the child's sustained attention and a source of new words. Focusing exclusively — as previous studies have done — on overheard speech in the first two years of life, and without a unifying explanation for the advantage of child-directed speech versus impotence of overheard input, may have led to the premature conclusion that overheard speech is permanently ineffective for learning. But language development is a protracted process, and vocabulary development in particular continues throughout the lifespan. Given the significant variability in the composition of children's early language environments (e.g., Casillas et al., 2019; Weisleder & Fernald, 2013), it is critical to explain why different sources are more or less conducive to learning, and how their contributions might morph with the child's own development.

Our focus on input *complexity* places greater emphasis on the child as a rational learner and filterer of the sense data around them. In this way, it is consistent with growing interest in children as agents of their own learning, actively exploring mysteries in their environments and selecting information relevant for the specific questions whose answers they crave (e.g., Begus et al., 2014; Cook et al., 2011; Gottlieb et al., 2013; Partridge et al., 2012; Saylor & Ganea, 2018; Schulz & Bonawitz, 2007; Sim & Xu, 2014). These studies frame the child as molding and curating sources of information, rather than merely analyzing the data given to them directly.

In addition to the active, rational learner, focusing on input complexity highlights the role of the child's mature interlocutor. Implicit in the idea that child- and adult-directed speech are differentially complex in a way that is relevant for learning (and therefore filtering), is the idea that competent speakers accommodate their listeners in the service of communication. When we speak, we speak to transmit our message *to* someone, which means how we formulate and deliver it is necessarily going to depend on what we think they will be able to receive (Lam and Kitamura, 2009; Yurovsky, 2017; though see Keysar et al., 1998). Maintaining the child's sustained attention is one of adults' goals in trying to communicate, leading them to subconsciously adapt their speech complexity in a way that is sensitive to it. That children learn from child-directed speech suggests that we might think of speech directed to the child at a given stage as a benchmark or lower bound for the complexity range from which they can learn. We suggest that any source of language input may be at a disadvantage with respect to a child's learning and attention for at least as long as it remains more complex than the speech that the child receives directly. This might be the case for at least two reasons. First, if caregivers' linguistic modification is responsive to children's online attention, then speech directed to children may provide an accurate (if indirect) reflection of their levels of language development. Second, *relative* complexity may be the relevant determiner of children's attention, such that the presence of simplified speech

leads children to ignore more complex language, even if it might also have been a source of word-learning. There is some evidence for this idea: in one study of dual-language immersion kindergarten classrooms, for example, teachers' tendency to translate instructions into students' dominant language had the effect of reducing children's attention to speech in their non-dominant language (e.g., Wong Fillmore, 1982).

The study we report here is interested in characterizing the complexity or learnability within children's language environments, with the hypothesis that variation along this dimension is determinant of children's attention. We compute multiple text-based metrics of complexity across longitudinal child-directed and conversational adult corpora to understand how input complexity changes with the age of the child, as well as when overheard speech might resemble the speech that children receive directly. By comparing the complexity or learnability of these two potential sources of data for early language learners, we aim to inform predictions about when overhearing may become a consistently viable source of word-learning.

1.2 Method

To explore the viability of the idea that infants may pay less attention to overheard speech, due in part to its complexity relative to child-directed speech, we employ multiple empirically-motivated operationalizations of 'complexity' or, conversely, ease of processing, that can be computed on decontextualized speech corpora. These include lexical variables associated with faster processing in production and comprehension in psycholinguistic research, as well as information-theoretic measures capturing the predictability versus *entropy* of the speech.

We begin by aggregating model child- and adult-directed speech from a variety of sources.

Speech Corpora

All child- and adult-directed utterances were tokenized and stemmed for analysis using the nltk package in Python. Preprocessing and analysis scripts can be found in our online repository (https://osf.io/hy5z2/).

Child-Directed Speech (CDS)

Child-directed speech from CHILDES. We compiled speech directed to children from childes-db (Sanchez et al., 2018), an open online database of the Child Language Data Exchange System (CHILDES) Database (Brent & Siskind, 2001; R. Brown, 1973; MacWhinney, 2000; Peters, 1987; Rollins, 2003; Rollins & Trautman, 2006, 2011; Rose & MacWhinney, 2014; Wilson & Peters, 1988), using the childes-r package (Braginsky, Sanchez, et al., 2019). We filtered the utterances in the English-language transcripts to exclude all child productions, as well as all adult utterances that we inferred to be directed to another adult (see *Adult-directed speech from CHILDES*, below). When making general comparisons between

child- and adult-directed speech, we include *all* children in our analyses. This decision has a dual motivation, as it both maximizes the number of tokens for analysis, and provides a more stringent test of our hypotheses. This speech comprises almost six million (5,637,187) tokens, pulled from 84 corpora and 1,187 children ($M_{\text{age}} = 38.18$ months, $SD_{\text{age}} = 17.36$ months), with an average of a little over one thousand words per target child. When analyzing differences across age, we limit ourselves to speech in the two-year age range reliably associated with later outcomes; 12–36 months. This subset consisted of 3,053,440 tokens spoken to 643 children ($M_{\text{age}} = 27.70$ months, $SD_{\text{age}} = 5.28$ months).

Child-directed speech from the VanDam Corpus. Next, we pulled transcripts from the VanDam Corpus (VanDam, 2016), for an additional high-quality source of child-directed speech. The VanDam corpus was collected in the homes of 53 toddlers ($M = 29.8$ months, $SD = 2.8$ months), across the spectrum of hearing status. While the bulk of the corpus is diarized data output by a widely used language processing software (LENA; Cristia et al., 2020), five minutes of dense parent-child interaction from each child's home recording session was transcribed by a human coder, and has been made publicly available on Homebank (VanDam et al., 2016). Families in the original study were visited three times apiece, for a total 159 five-minute segments and 35,045 tokens for us to use in our analyses. The documentation for the corpus reports upper-middle class maternal socio-economic status, averaging 7.81 ($SD = 2.36$), on a 12-point scale. All but two families identified as "white/Caucasian," with only one family identifying as "Black/African-descent," and another identifying as "mixed race."

Child-directed speech from the Manchester Corpus. Last, we considered the Manchester Corpus (Theakston et al., 2001), also hosted on CHILDES, for use as case study. The Manchester corpus records a study of 12 monolingual English-speaking children from middle-class households in Manchester, UK, from ages 20 to 36 months. This corpus was selected because of its dense, longitudinal, and naturalistic nature. Mothers and children were audio-recorded playing freely in their homes two times every three weeks for a year, for a maximum of 34 sessions per child. Recordings of these sessions were transcribed in CHAT format, and tagged with the age of the child and the individuals present. We analyzed only the speech of the primary caregivers in this case study (81,302 tokens from 12 mothers).

Adult-directed Speech (ADS)

The ideal corpus for this study would include large amounts of child and adult-directed speech, from the same speakers, in the same context. Unsurprisingly, however, adult-directed utterances are rare in corpora designed to illuminate *child* language development. In order to triangulate on an accurate characterization of overheard adult speech, we compute our measures on spoken corpora from three sources, each representing distinct trade-offs in terms of their size versus representativeness of speech children might typically have opportunity to overhear.

Adult-directed speech from CHILDES. We first compiled a corpus of the sparse inter-adult speech in CHILDES. While small, this corpus had the advantage of containing utterances we know were spoken in front of a child (whose age we also knew). However, our method for ensuring that they were truly adult-directed was inferential: following previous work (Yurovsky et al., 2016b), we first combined all adjacent utterances from the same adult speaker, then classified as "adult-directed" any adult utterance that was immediately followed by an utterance by a different adult. In this way, we hoped to be selecting utterances that represented adult conversational turns. There were a total of 6,150 tokens, which occurred in the presence of 133 different children, from 22 to 176 months of age.

Adult-directed speech from the VanDam Corpus. We took a similar approach in the VanDam Corpus, resulting in a modest set of 2,229 tokens of speech presumed to be spoken by one adult in the home to another.

Adult-directed speech from the Santa Barbara Corpus. A slightly larger but still highly controlled corpus of inter-adult speech meant to reflect the speech that children might overhear came from the Santa Barbara Corpus (Du Bois et al., 2000), a database of transcribed audio recordings of American English conversations from diverse contexts and regions. We included all files representing speech that was (1) informal, (2) English-only, and (3) adult-directed. This meant excluding recorded academic lectures, negotiations at car dealerships, and job trainings, where we estimated a child's presence would be less likely. Importantly, these excluded transcripts are also characterized by clusters of dense technical words, which might skew our estimate of typical adult-directed speech complexity. Thus, in addition to providing a more representative picture of speech around children, the subset of the corpus we analyzed provides a more conservative test of our hypotheses. The final set of 19 transcripts included speech during long-distance phone calls, birthday parties, and conversations while preparing dinner, from a total of 67 speakers (87,496 tokens).

Adult-directed speech from the British National Corpus. Finally, we considered a significantly denser corpus. The full British National Corpus totals over 100 million words of written and spoken English, and is used as a standard in many studies. We limit ourselves to the spoken subset of the corpus (Love et al., 2017) — approximately 11 million words. A random sample of tags in the corpus includes such quotidian topics as "bills," "moving homes," "portion distribution," "Andy's grandmother," and "Joey's broken arm."

Capturing Complexity

Our first analysis takes advantage of previously collected ratings of words along several subjective dimensions, as well as of estimates of words' frequency, a dimension we would expect to systematically differ between speech directed to children and to adults. Normative ratings for individual words like those we employ here are often used to standardize experimental stimuli and test hypotheses in psycholinguistic research (e.g., Balota et al., 2007). *Concreteness* ratings served as a proxy for semantic accessibility; we argue that *age of acquisition* ratings do something similar. *Valence* ratings were intended to capture affective adjustment on the part of caregivers at the level of the word. Importantly, norms for frequency, concreteness and age of acquisition are reliably predictive of adult performance on a variety of psycholinguistic tasks (Bonin et al., 2001; Brysbaert & Biemiller, 2017; Ghyselinck, Lewis, et al., 2004; Izura & Ellis, 2002; Morrison & Ellis, 2000). We included valence ratings specifically as a constraint on our hypotheses. While our other measures are transparently related to learnability, we speculated a weaker relation between valence and learning during the early stages of language development, in part because caregivers often use prosody to convey the valence of their utterances (Saint-Georges et al., 2013), which we reasoned might imply lesser reliance on the semantic valence of the words they choose. In addition, effects of valence are inconsistent in lexical processing (e.g., Delaney-Busch et al., 2016; Imbir et al., 2016; Kuchinke et al., 2005; Kuchinke et al., 2007; Moors et al., 2013; Nasrallah et al., 2009; Yao et al., 2016) and memory tasks (e.g., Aquino & Arnell, 2007).

For all lexical variables, our data are limited to the lemmas for which we have ratings. We note the percentage of tokens available for each dataset in the subsections below. For complete type and token counts by corpus (all datasets) or speaker (Manchester Corpus) at each stage of transcript cleaning and analysis, see Appendix C.

Concreteness

We first considered *concreteness* as a means to compare differences in semantic complexity between child- and adult-directed speech. Concreteness ratings came from the Brysbaert et al. (2014) set of concreteness ratings for 40,000 English lemmas. The authors collected ratings on a scale from 1 ('language-based') to 5 ('experience-based') from native English-speaking, current U.S. residents using Amazon's Mechanical Turk. In the Brysbaert and colleagues' (2014) norming study, raters were explained the scale as follows:

> A concrete word [...] refers to something that exists in reality; you can have immediate experience of it through your senses (smelling, tasting, touching, hearing, seeing) and the actions you do. The easiest way to explain a word is by pointing to it or by demonstrating it (e.g. [...] To explain "jump" you could simply jump up and down [...] To explain "couch" you could point to a couch or show a picture of a couch). An abstract word comes with a lower rating and refers to something you cannot experience directly through your senses or actions. Its meaning depends on language. The easiest way to explain it is by using other words (e.g. There is no simple way to demonstrate "justice;" but we can explain the meaning of the word by using other words that capture parts of its meaning).

We see concreteness as relating to the child's attention and learning in at least two ways. First, speech about the child's immediate context is likely to be more engaging due to its relevance for the child. Such speech is also likely to be more concrete, because it is about objects, etc. in the child's own environment, rather than abstract concepts and mental states, which are necessarily decontextualized. Second, if children's attention is in part maintained by learnability, concrete language may maintain their attention precisely because its meaning is more learnable. Indeed, the above description illustrates the potential learning of advantage of highly concrete words: they can be demonstrated by a caregiver, and their meanings are more likely to be inferable from a scene without the need for pre-existing linguistic knowledge or scaffolding.

It is also important to note that concreteness is a dimension that cuts across lexical categories: nouns vary in their concreteness, but so do verbs (e.g., "jump," in the example above, with a rating of 4.52, versus "hope," rated 1.25) and adjectives ("hairless," rated 4.52, versus "fake," 1.97). The concreteness rating we use for each lemma represents the mean of ratings by at least 20 raters familiar with the word (37,058 total ratings; accounting for 94.9% of CHILDES CDS, 82.2% of VanDam CDS, 95.8% of Manchester CDS, 94.5% of CHILDES ADS, 80.4% of VanDam ADS, 77.7% of the Santa Barbara Corpus, and 75.9% British National Corpus).

Valence

To compare any observed trends in word type concreteness to another lexical semantic variable whose relation to learning is more opaque, we analyze *valence* ratings collected by the same group (Warriner et al., 2013). A word's valence is a reflection of how pleasant it is. While it is possible that child- versus adult-directed speech differs along this dimension, it is not clear which values (more positive or more negative words) we would expect to better capture children's attention, and facilitate their language-learning. Valence ratings were elicited on a scale from 1 ('unhappy') to 9 ('happy'). For illustration, the top-rated word is "vacation," with a mean rating of 8.53, and the lowest rated word is "pedophile" (1.26). The values we use represent the mean of at least 15 adult raters. There was notably less coverage for the valence-rated words than for other measures (13,915 total words; CHILDES CDS: 41.6%, VanDam: 33.2%, Manchester: 41.4%, CHILDES ADS: 41.5%, VanDam ADS: 32.7%, Santa Barbara Corpus: 29.2%, British National Corpus: 24.7%), so we interpret these results with caution.

Age of Acquisition

Like scores of concreteness and valence, age of acquisition (AoA) values came from a large-scale study of English-speaking adults (Kuperman et al., 2012). Following a series of calibration items, online participants were presented with 300 words, and provided subjective ratings for all of the words that they also reported knowing. For each word, adults were asked to enter the age, in years, at which they thought they had learned the word — in the words of the study, "the age at which you would have understood that word if somebody had used it in front of you, *even if you did not* use, read or write it at the time." The earliest rated word is, unsurprisingly, "momma" (1.49), followed by "potty' (2.28), "yes" (2.31) and "water," (2.37). Median (10.5) words include "ricochet," "blondish," and "suede," while loanwords ("eisteddfod;" 25), regional items ("saguaro;" 18.2), and technological terms ("app;" 18.3) dominate the upper ranks. Twenty-five adults rated each word; we use the mean of their responses (5,175 total words; CHILDES CDS: 93.4%, VanDam: 75.5%, Manchester: 94.6%, CHILDES ADS: 93.0%, VanDam ADS: 73.7%, Santa Barbara Corpus: 70.4%, British National Corpus: 67.0%).

Age of acquisition norms are reliably correlated with objective measures of lexical development (Gilhooly & Gilhooly, 1980; Gilhooly & Logie, 1980). For our purposes, however, it may not matter whether adults' age of acquisition estimates accurately reflect their own linguistic histories. That is, age of acquisition norms might be especially well-poised to answer questions about the degree to which adults alter their speech to children, as the values may better reflect the order in which adults expect certain words and concepts to be appropriate for the children in their charge. Consensus across many adults may reflect an intuitive and multidimensional theory of complexity (see Appendix B for items where adults' age of acquisition estimates are especially inaccurate, yet revealing).

Frequency

Our measure of word frequency is derived by Brysbaert and New 2009 using a 51-million-word corpus of American movie and television subtitles (SUBTLEX). The SUBTLEX frequency norms are commonly used to accurately model individual words' frequency in everyday language use, and have been shown to reliably predict adult processing latencies in psycholinguistic tasks (Brysbaert & New, 2009). Frequency effects are pervasive in adult language processing more broadly, where higher frequency words are associated with faster and more accurate word recognition and production (e.g., Balota & Chumbley, 1984) as well as throughout first language acquisition (see Ambridge et al., 2015, for a review), where higher-frequency words are consistently acquired earlier (though see Morrison & Ellis, 1995, for alternative interpretations of the explanatory power of frequency versus age of acquisition). For our analyses, we normalize the counts of individual words' occurrences within the SUBTLEX corpus (i.e., by dividing by the total number of words), then take the logarithm of the resulting value. Coverage of the words used in our child- and adult-directed speech corpora was especially high for this variable (48,411,930 total words, covering 96.2% of CHILDES CDS, 94.1% of VanDam CDS, 96.2% of Manchester CDS, 96.1% of CHILDES ADS, 93.4% of VanDam ADS, 85.4% of the Santa Barbara Corpus, and 89.1% of the British National Corpus).

Lexical Complexity

Building off of our age of acquisition metric, our next measure defines complexity in terms of the ratio of words children likely already know or are on the verge of acquiring, to those they do not. We use the MacArthur Bates Communicative Development Inventory (M-CDI; Fenson et al., 2007) to obtain a set of child-friendly words typically acquired by the end of our age range. So as to not make assumptions about the age-related vocabularies of specific children, we include *all* words on the M-CDI "Words and Sentences" instrument, intended for children 16–30 months.

Of course, appearance on the M-CDI does not guarantee that even a child of three knows it. Standardized assessments include words to which the answer for most assessees will be "no," by design. But they are also designed to be normative. An examination of the set of American English administrations archived online (5,846 at the time of writing; Frank et al., 2015) suggests that the vast majority of M-CDI items are produced by 30 months. That is, over half of the parents of 30-month-olds report that their child produces 603 out of the 680 total items. This set of words also does not fully account for the child's vocabulary, but it does provide us with a standard to apply to all speech sources.

To derive our metric of *lexical complexity* for each speech source, we pulled 100 random sample of 1,000 tokens each. For each 1,000–token sample, we calculated the proportion of tokens that appeared on the M-CDI. To obtain a measure of complexity that increased with the density of words *outside* the set, we took the negative log of the probability of belonging to the M-CDI set. For an example, say 600 of the 1,000 words in a sample of child-directed

speech were among the "Words and Sentences" items. The proportion of known words would be (600/1000), or 0.60. The complexity measure for that sample would be $(-log_2(0.60))$, or 0.51. To assess this measure's trajectory with age, we additionally sample each month in child age in CHILDES child-directed speech from 12 to 30 months.

Unfamiliar Entropy

Our final measure is designed to reflect the diversity or unpredictability within the set of words that we have less reason to believe might be familiar to the child. Currently, two utterances with the same proportion of tokens on the M-CDI receive the same lexical complexity score, regardless of the composition of the words *outside* the M-CDI set. However, we might think that an utterance where all of the non–M-CDI tokens are new, unique types should be considered 'more complex' than an utterance where all of the words in the same-sized set are repetitions of the same unique word. We use what we term *unfamiliar entropy* as our final measure to reflect this intuition. Since its introduction by Shannon (Shannon, 1948), variants on entropy have been used as indices of structural diversity in natural language processing, machine learning, and ecology (Hale, 2016; Masisi et al., 2008). Sampling from child- and adult-directed speech in the manner described above, we calculated the entropy in the set of tokens *not* on the M-CDI.

Analysis & Predictions

For each of our metrics, we ask:

(1) Is the speech adults direct to children reliably less complex than the speech they direct to one another?

(2) Does child-directed speech change with the child's age?

(3) Does child-directed speech *converge* with adult-directed speech?

We restrict our results to the most conservative and interpretable tests of our hypotheses that can be supported by all of our speech sources. To evaluate the difference in speech complexity, we compare the mean values for each of our measures in child- and adult-directed corpora. We predict higher frequency and mean concreteness in child-directed speech, and lower age of acquisition, lexical complexity, and entropy. We do not anticipate a significant difference in valence between words in the two speech types. To evaluate whether child-directed speech complexity increases with child age, we compute each of our metrics for each month in child age from 12 to 48 months, and report the correlation across all child-directed speech. To evaluate the robustness of any effect and capture variability across speakers, we use linear mixed effects models, fit to the data for the twelve caregivers in the Manchester corpus. Finally, we divide our child-directed corpus into three twelve-month age bins, from 12–24, 24–36, and 36–48 months. We again compute the difference in means for the same

measures in all adult-directed speech. We predict a greater difference between child- and adult-directed speech in the early age group, and ask whether the difference persists when considering only the second and third years of this range, or whether child- and adult-directed speech 'converge.'

For all questions, we use permutation tests to assess significance and to obtain the p-values we report. All resampling tests were performed with 10,000 iterations. For additional analyses and visualizations — including of words' relative *information content* — please see the *Supplemental Online Materials*.

1.3 Results & Discussion

Is adult-directed speech reliably more complex?

To assess whether speech between adults is reliably more complex than speech to children, we compare the mean value for each of our measures across corpora of child- and adult-directed speech (see Table 2). Compared to the set of adult-directed speech tokens ('All ADS'), words in child-directed speech ('All CDS') were reliably higher frequency (*difference in means*: −0.370), and rated more concrete (*difference in means*: −0.109), earlier acquired (*difference in means*: 0.124), and happier (*difference in means*: 0.029). Samples across child-directed speech also received lower scores on our measures of lexical complexity (*difference in means*: 0.157) and unfamiliar entropy (*difference in means*: 0.050; all *ps* here and elsewhere $< .001$, unless otherwise noted). With the exception of *valence*, which showed no significant difference, these results were echoed when comparing means for individual children in the Manchester corpus to the All ADS mean values, as well as in pair-wise comparisons between child- and adult-directed speech within CHILDES and the VanDam corpus.

That our results obtained within the VanDam corpus is especially notable. While sparse in terms of adult-directed speech, the VanDam corpus represented especially high-quality data for our research questions, as all speech came from the same set of adult speakers, eliminating the possibility that qualitative differences between child- and adult-directed speech corpora might be an artifact of differences in speaker context or population. Even more favorably, the speech samples in the VanDam corpus were originally captured by an audio recorder worn by the target child in the transcript, thereby establishing any adult-directed speech as legitimately *overheard* (or overhearable) by that child. This is in contrast to other sources that we use to simulate speech that might be available to the child to overhear. For example, while the Santa Barbara and British National corpora enable us to accurately characterize speech between adults, we can't know whether adults would use the same language if a child were present (e.g., Kempe, 2009).

Having confirmed that our metrics capture real dimensions of variability in child- and adult-directed speech, we examined their relation to children's age.

Table 2: *Mean Complexity across Corpora*

	Concreteness	AoA	Valence	Frequency	Lexical Comp.	Entropy
Child-Directed						
CHILDES	2.81	4.35	6.01	−5.90	0.34	6.88
	(2.81, 2.81)	(4.35, 4.35)	(6.01, 6.01)	(−5.90, −5.90)	(0.30, 0.39)	(6.63, 7.11)
VanDam	2.83	4.29	6.02	−6.13	0.31	7.09
	(2.82, 2.85)	(4.27, 4.30)	(5.99, 6.04)	(−6.92, −6.86)	(0.45, 0.58)	(6.87, 7.30)
Manchester	2.78	4.41	5.97	−5.82	0.33	6.66
	(2.78, 2.78)	(4.41, 4.41)	(5.96, 5.97)	(−5.97, −5.96)	(0.29, 0.38)	(6.40, 6.90)
All CDS*	2.81	4.35	6.01	−5.90	0.34	6.88
	(2.81, 2.81)	(4.35, 4.35)	(6.01, 6.01)	(−6.06, −6.05)	(0.30, 0.39)	(6.63, 7.12)
Adult-Directed						
CHILDES	2.73	4.45	6.05	−6.09	0.41	7.10
	(2.73, 2.73)	(4.45, 4.45)	(6.04, 6.05)	(−6.10, −6.09)	(0.36, 0.46)	(6.87, 7.31)
VanDam	2.81	4.34	5.99	−6.95	0.58	7.26
	(2.79, 2.83)	(4.32, 4.36)	(5.95, 6.02)	(−7.00, −6.89)	(0.52, 0.64)	(7.06, 7.44)
SBC	2.62	4.76	5.88	−6.70	0.75	7.62
	(2.61, 2.62)	(4.75, 4.77)	(5.87, 5.89)	(−6.72, −6.69)	(0.68, 0.83)	(7.43, 7.80)
BNC	2.52	4.67	5.96	−6.30	0.75	6.80
	(2.52, 2.52)	(4.67, 4.67)	(5.95, 5.96)	(−6.30, −6.29)	(0.68, 0.82)	(6.55, 7.03)
All ADS*	2.56	4.63	5.98	−6.27	0.69	6.93
	(2.56, 2.57)	(4.63, 4.63)	(5.98, 5.98)	(−6.27, −6.27)	(0.62, 0.76)	(6.70, 7.15)

* Weighted mean across corpora.

Does child-directed speech complexity track with child age?

We next sought evidence that adults calibrated their speech to their developing interlocutors by relating each of our metrics with the age of the target child, across child-directed corpora. Months in target child age showed a significant positive correlation with age of acquisition (Pearson's $r = .02$) and lexical complexity (Pearson's $r = .51$), a significant negative correlation with concreteness (Pearson's $r = -.026$) and unfamiliar entropy (Pearson's $r = -.75$), and no significant correlation with frequency (Pearson's $r = -.002$, $p = .39$) or valence (Pearson's $r = -.001$, $p = .06$).

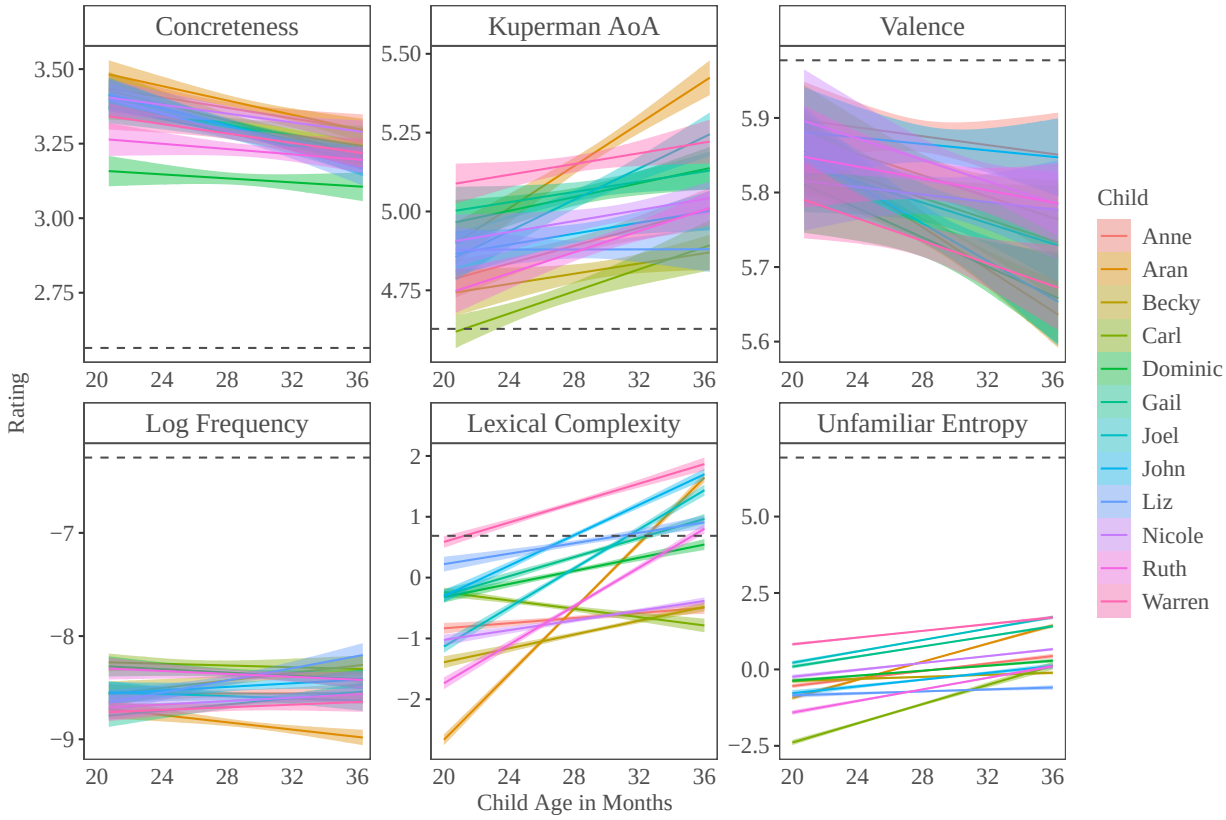


Figure 1: *Linear Regressions for Individual Children in the Manchester Corpus.*

Note. Shading indicates standard error; dashed line marks mean value for All ADS. Average correlations with child age: Concreteness: −.05 [−.05, −.04]; AoA: .04 [.03, .04]; Valence: −.02 [−.02, −.01]; Frequency: −.002 [−.007, .003]; Lexical Complexity: .27 [.26, .28]; Unfamiliar Entropy: .23 [.21, .24]. Outcome variables have been centered and scaled to facilitate comparison across panels.

The Manchester corpus enabled us to evaluate the robustness of these effects at the level of individual caregivers, in speech to their children from 20–36 months in age. Figure 1 plots linear regression lines fit to each child's longitudinal data. Concreteness in speech to all 12 children showed a negative slope (*Range*: −0.009, −0.004). Slopes for the estimated age of acquisition of child-directed words were typically positive (*Range*: 0.0005, 0.024), but this was reliably the case for only 9/12 mothers (95% confidence intervals for the age coefficients of 3/12 caregivers spanned 0). Results for valence, frequency, and unfamiliar entropy were even more variable. Valence reliably decreased with age for only 5/12 children (*Range*: −0.009, −0.010), while frequency decreased as expected for 1/12 children, increased for 3/12, and neither increased nor decreased for the remaining 8/12 children (*Range*: −0.017, 0.029) Similarly, entropy increased as expected for 7/12 children (*Range*: 0.002, 0.006), decreased for 2/12 (*Range*:−0.0007, −0.0002), and showed no reliable trend for 3/12 children. Finally, lexical complexity increased reliably in all but one household (*Range*: 0.020, 0.280), where it decreased with age ($\beta = -0.04$ [−0.05, −0.03]).

To evaluate trends *across* caregivers, we used the lme4 package in R (Bates et al., 2015) to fit a single mixed effects linear model to the data for each measure, with age as the sole fixed effect in each model, and random intercepts for maternal speaker. Model results appear in Table 3, and confirm the relation between age and words' increasing processing demands, controlling for variability between speakers.

In the next section, we take a different look at the relation between speech complexity and age by pooling tokens across each year between children's first and fourth birthdays.

When will child- and adult-directed speech *converge*?

In our last analysis, we binned child-directed speech in the years between children's first and fourth birthdays. Our choice of age bins is motivated by the evidence, reviewed in the Introduction, that measurements of overheard speech quantity during this period are unrelated to child vocabulary (Ramírez-Esparza et al., 2014; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a).

Table 4 shows the All CDS means, and difference from All ADS at 12–24, 24–36, and 36–48 months in age. The first thing to notice about these data is that, with the exception of frequency and entropy, the difference from adult-directed speech is greater (in terms of absolute value) in the first age bin, compared to the second and third. The next thing to notice is that the difference at 36–48 months remains significant for all variables but valence (see Appendix D for results by corpus). A surprising result of this analysis is that unfamiliar entropy was significantly *lower* in all ADS than all CDS for each year-long age bin (that is, the parenthetical differences representing All ADS − All CDS in Table 4 are negative). While unexpected, we hypothesize that this trend might be a further reflection of ways in which adults adjust their speech to children: for example, if caregivers frequently use (a) the child's name (E. V. Clark & Estigarribia, 2011), and/or (b) idiosyncratic terms, like *binkie* or *baba* (Mayor & Plunkett, 2011) — both of which will vary across households — then entropy across the entire set of tokens might be higher in child-directed speech than in

Table 3: *Mixed Effects Linear Regressions on Complexity in the Manchester Corpus*

	Dependent variable:					
	Concreteness	AoA	Valence	Frequency	Lexical Comp.	Entropy
Constant	3.58^{***}	4.56^{***}	6.00^{***}	-8.59^{***}	0.23^{***}	5.30^{***}
	(3.52, 3.64)	(4.46, 4.66)	(5.94, 6.07)	(−8.73, −8.46)	(0.21, 0.24)	(5.06, 5.55)
Age	-0.01^{***}	0.01^{***}	-0.01^{***}	0.002	0.004^{***}	0.03^{***}
	(−0.01, −0.01)	(0.01, 0.02)	(−0.01, −0.01)	(−0.001, 0.01)	(0.003, 0.004)	(0.03, 0.03)
Observations	156,635	153,689	108,220	158,526	15,000	15,000
Sd(Caregiver)	0.06	0.13	0.04	0.16	0.03	0.48
Log Likelihood	−247,536.00	−286,234.30	−167,227.10	−359,462.60	30,566.85	2,973.90
AIC	495,080.00	572,476.60	334,462.10	718,933.20	−61,125.69	−5,939.79
BIC	495,119.90	572,516.30	334,500.50	718,973.10	−61,095.23	−5,909.33

Note. Model includes random intercepts for caregiver. *p<0.05; **p<0.01; ***p<0.001

adult-directed speech, where conventional terms might be more dominant. If this is the case, we might not expect the same trend when computing this metric household by household, where idiosyncratic terms are likely to reoccur. In support of this idea, entropy displayed the expected relationship when comparing child- and adult-directed speech within the VanDam corpus: that is, adult-directed speech was higher entropy, but the difference decreased with age (*differences in means* between 0.210 and 0.507, ps < 0.001).

In summary, these data suggest that adult-directed speech may remain significantly more complex into the fourth year of life, with potential implications for overheard speech as an eligible source of language learning.

1.4 General Discussion

Our exploration of adult speech to children versus to other adults was motivated by a puzzle in the literature on language development: though speech *around* children but not directed *to* them would appear to offer valuable language data for young learners, studies that have explicitly measured overheard speech in children's environments from 12–30 months consistently find no correlation between it and vocabulary outcomes at 18–36 months (Ramírez-Esparza et al., 2014, 2017; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a; Weisleder & Fernald, 2013). The hypothesis that we advanced drew on work suggesting that infants' attention is responsive to variation in formalizeable notions of complexity or learnability (e.g., Gerken et al., 2011; Kidd et al., 2012, 2014; see also Räsänen et al., 2018). More generally, our study represents a novel approach to the question of whether and how adults simplify their speech to children — particularly in such a way as would plausibly advance

Table 4: All CDS *Mean Values and* All ADS–All CDS *by Age Bin*

	Concr.	AoA	Valence	Freq.	Lexical	Entropy
Age	M[†](Diff)[‡]	M[†](Diff)[‡]	M[†](Diff)[‡]	M[†](Diff)[‡]	M[†](Diff)[‡]	M[†](Diff)[‡]
12–24	2.885	4.253	6.029	−6.148	0.297	6.862
	(−0.321)	(0.374)	(−0.052)	(−0.122)	(0.390)	(−0.068)
24–36	2.810	4.361	5.997	−6.038	0.329	6.565
	(−0.245)	(0.266)	(−0.020)[§]	(−0.233)	(0.358)	(−0.365)
36–48	2.786	4.371	6.002	−6.006	0.343	6.606
	(−0.221)	(0.256)	(−0.025)[§]	(−0.264)	(0.344)	(−0.324)

[†] All CDS mean.
[‡] Observed difference in means, All ADS − All CDS.
[§] Value does **not** significantly differ from zero (all ps > 0.05).

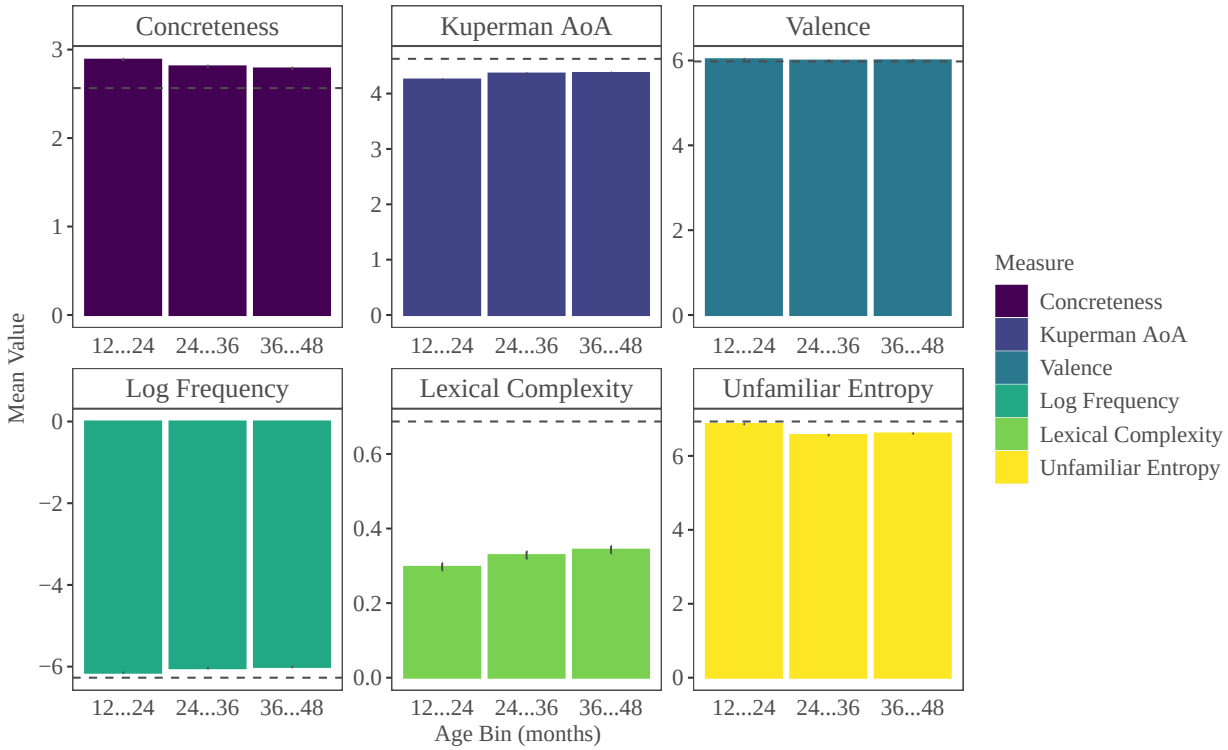


Figure 2: All CDS *Mean Complexity at 12–24, 24–36, & 36–48 Months.*

Note. Error bars indicate 95% bootstrapped confidence intervals; dashed line marks All ADS mean value.

learning. Much of the previous longitudinal literature has focused on characterizing caregivers' syntactic and prosodic adjustment in small cohorts of children (e.g., Bornstein et al., 1992; Huttenlocher et al., 1991; Sherrod et al., 1977), while previous work comparing adult- and child-directed speech using corpora has typically focused on acoustic or distributional properties of the signal (e.g., Adi-Bensaid et al., 2015; Genovese et al., 2020; Guevara-Rukoz et al., 2018; Huttenlocher et al., 2007; Ko, 2012; Liu et al., 2009; though see Yurovsky et al., 2016a). Here, we examined adults' simplification at the level of lexical semantics, in line with methodologically related studies predicting the order of acquisition of individual words from their semantic and distributional features (Braginsky, Yurovsky, et al., 2019; Goodman et al., 2008; Hills et al., 2010; Swingley & Humphrey, 2018). As proxies for speech complexity, we relied on pre-existing ratings of individual words' concreteness (Brysbaert et al., 2014), estimated frequency (Brysbaert & New, 2009), and age of acquisition (Kuperman et al., 2012), which show facilitative processing effects in adult production and comprehension (A. W. Ellis & Morrison, 1998; Gilhooly & Gilhooly, 1980; Jaeger & Tily, 2011). We and others (Dominey & Dodane, 2004; Hills & Adelman, 2015) hypothesized that higher values

along these dimensions should enhance the learnability of samples of speech. In addition to these indices of relative semantic complexity, we employed minimally adapted metrics of information density to capture the degree to which caregivers favor words more likely to be familiar to the child, and "cushion" later-acquired words with otherwise repetitious speech. These measures are founded on the idea that the presence of familiar words makes unfamiliar words more learnable (e.g., Sullivan & Barner, 2015).

Comparing our complexity metrics across corpora, our results suggest that speech between adults is more abstract, less predictable, and includes more words (believed to be) learned earlier in life. The majority of these measures remain significantly different at 24–36 and 36–48 months, overlapping (and exceeding) the period in which previous studies have measured the quantity of overheard speech in children's environments, and found it unrelated to children's vocabulary growth.

Of the lexical variables, effects for *valence* were the weakest and most inconsistent. It is possible that this merely reflects the relative data sparsity for this measure, as valence norms existed for fewer than half of the tokens in each of our corpora. However, it is also consistent with recent models of cross-linguistic acquisition order, where *concreteness* is a robust predictor of order of acquisition across languages, while valence is not (of the words on the M-CDI; Braginsky, Yurovsky, et al., 2019; Fenson et al., 2007). In fact, valence may be an especially dispensable variable in accounting for the child-directed speech advantage at the word level, given that caregivers' frequently employ prosody to convey similar affective information (e.g., Saint-Georges et al., 2013; Trainor et al., 2000; Trainor & Desjardins, 2002).

Unfamiliar entropy was another outlier variable in our data. In combination with the lexical complexity results, we suggest that there might be a common learning story. Lexical complexity showed a reliable decrease with child age, suggesting that caregivers are using fewer and fewer of the set of hyper-familiar words contained on the M-CDI as the child matures. Nonetheless, comparing values for child- and adult-directed lexical complexity suggests that caregivers continue to favor using M-CDI words in speech to their children, even at the end of our child age range (Table 4). That unfamiliar entropy values during the same period were sometimes equivalent to, or even higher than, adult-directed values suggests that adults may have introduced a variety of new words into their child-directed speech in the meantime, thereby providing some signal of calibration to their maturing addressees.

Remarkably, each of our metrics of complexity showed significant correlations with the age of the child addressee. Trends with age were more variable when we considered longitudinal data for a single child at a time, using the Manchester corpus. Previous work suggests that how individual differences in complexity trajectories manifest depends on the particular measure and its learning trajectory. For example, in one longitudinal study of caregiver speech, 7/8 measures of syntactic complexity showed linear trajectories, with only different intercepts needed to capture inter-caregiver variability, while the remaining measure exhibited quadratic growth (Huttenlocher et al., 2007).

Interestingly, in the midst of findings that all caregivers simplify their speech to children (Shneidman & Goldin-Meadow, 2012b), studies of individual differences in caregiver

complexity imply that more complex speech is better: that is, greater syntactic complexity and lexical diversity in child-directed speech correlate with more advanced child vocabularies and later verbal productions (Hoff, 2003; Huttenlocher et al., 2010). In a 2006 review of the literature, Hoff addresses the apparent contradiction between our assumption that language learning should be advantaged by simplified, 'just-right' speech complexity, and findings that *more* complex language input predicts greater learning by invoking precisely children's strategic deployment of attention:

> Despite the findings that simpler maternal speech is not associated with more rapid language development than more complex maternal speech, it still may be the case that the average degree of simplification in child-directed speech benefits language acquisition. All of the observed benefits of complexity were obtained within the range of complexity in child-directed speech. Furthermore, children may filter out, by not processing, input that is too complex — with no negative consequences to language development — so long as sufficient processable input is available. In contrast, children have no way to make up for input that is too simple (p. 75).

Thus, the greatest variability between caregivers may be in the responsiveness of their speech to children's growth and language potential. That metrics for all maternal caregivers in the Manchester corpus were lower than corresponding values in adult-directed corpora — yet showed distinct relations to child age — is consistent with this conclusion. Moving forward, understanding the causes of individual differences in complexity trajectories is an important research goal, as it promises to shed light on the mechanisms that determine effective language input (e.g., Belsky et al., 1980; Kaplan et al., 2015; Spinelli et al., 2015). These are unlikely to be products of either parent or child alone: as discussed in the Introduction, well-calibrated child-directed speech likely arises out of the caregivers' genuine need to communicate with the child, and requires a combination of attentional investment on the part of the caregiver, and signaling of their degree of comprehension on the part of the child. These drivers will in turn be conditioned by a variety of cultural and contextual factors, like the spatial relation between caregiver and child, the presence of competition for either party's focus of attention, and even the time of day (Casillas et al., 2019). While our results raise questions about individual differences in child-directed speech, while suggesting that caregiver speech complexity tends to track with children's age, the approach we took in this study does not enable us to evaluate caregivers' *degree* of calibration — in the terminology of the past literature, how 'finely tuned' their speech is (Cross, 1977). In examining the frequency of the words on a commonly used parental report of early vocabulary, we aimed to capture the adults' preferential use of words they believed the child already knew, or would soon. The ideal analysis of this nature would employ a more sensitive model of the child's lexicon. Given that adult estimates of age of acquisition are reliably correlated with objective measures of lexical acquisition, one way of achieving this might be to use evidence of the child's productions to impute knowledge of all lower-ranked words. An alternative method might estimate the probability of word knowledge based on models of the lexical

network. Previous work suggests that high phonological connectivity (sounding like many other words) is predictive of acquisition order (Stella et al., 2017), as is high semantic connectivity (sharing meaning associations with many other words; Beckage & Colunga, 2016; Sizemore et al., 2018). Network connectivity in the input can be modeled (e.g., Beckage et al., 2011; Fourtassi et al., 2020) to inform more accurate representations of children's own lexical networks — though child-specific lexical structure (like dense, interconnected knowledge of dinosaur names; Chi & Koeske, 1983), representing precisely the topics they would be most likely to "tune in" to, would remain difficult to capture.

By analyzing child- and adult-directed speech at this macro level, we evade some of the issues with generalizability present in previous work (e.g., Cross, 1977; R. Ellis & Wells, 1977), and share new limitations with other studies of this nature (e.g., Braginsky, Yurovsky, et al., 2019; Goodman et al., 2008). For one, our metrics of complexity are coarse, as are their inputs. All of our metrics approach individual words as unitary constructs, when — to adapt Walt Whitman (1855) — individual lemmas 'contain multitudes.' That is, the same string of sounds can be used to express a great number of different meanings. To illustrate this oversight, we can look to the case of *polysemy*, where a single word can be used to express multiple related senses (Y. Xu et al., 2020). Individual senses of the same polysemous word may vary in how intuitive they are: for example, we know that the sense of *milk* in "milking a cow" is typically acquired later than the sense used in "drinking milk" (Srinivasan & Barner, 2013). Nevertheless, all occurrences of the lemma "milk" receive the same values on our indices of semantic complexity. Similarly, ratings of a variable like concreteness will collapse literal and non-literal uses of the same word, despite the intuition that instances of figurative language like metaphor might be considered more complex. When the sick child says "there's a fire-engine in my stomach" (Roeper, 2013), *fire-engine* would receive the same rating in our study as it would when the child reports "a fire-engine on the floor."

1.5 Conclusion

Our study lays the groundwork for future investigations of how linguistic complexity trades off with other variables to promote learning. When children are very young, all speech may be too complex and unpredictable to maintain their attention on learnability alone. While overheard speech may have little means, beyond falling within the child's "Goldilock's zone," of eliciting their attention, speech directed *to* children may continually regain their attention via other means. For example, the exaggerated prosody of many children's earliest directed speech may compensate for what would otherwise be irredeemable complexity (Soderstrom, 2007). As we have discussed, as much as the issue is that language is too complex and needs to be simplified, we might expect overheard speech to be at less of a disadvantage for learning when it approximates child-directed speech in simplicity, which could occur in a single learning episode (as in the case of experimental overhearing studies: Akhtar, 2005; Akhtar et al., 2001; Baldwin, 1991; Fitch et al., 2020; Floor & Akhtar, 2006; Gampe et al., 2012; Martínez-Sussmann et al., 2011; Shneidman et al., 2009), when the adult is

speaking simply to another child (Forrester, 2002), or simply as the child gets older. And if attention is the mediator of learning, other features might make a speech stream more or less supportive of learning. A preschooler may learn more from a complex overheard dialogue about their next birthday party — a topic that interests them — before they learn from an equally complex discussion of a neighbor's marriage. Finally, mindful of the loss of contextual determinants of complexity in our analyses here, in Chapter 4 and ongoing work, we go beyond text-based metrics to capture qualitative dimensions of how supportive individual instances of language are for learning. These dimensions, and the ongoing reward they promise for children's attention, undoubtedly evolve as children mature.

Chapter 2

Self-Directed Learning by Preschoolers in a Naturalistic Overhearing Context

Abstract

Three studies investigated preschoolers' self-directed learning ability in a naturalistic context: learning from overheard speech. In Experiment 1, 4.5– to 6-year-olds were exposed to 4 novel words and 6 arbitrary facts corresponding to a set of co-present toys; in Experiment 2, 3– to 4.5–year-olds heard 5 nouns and 3 facts. In the Pedagogical conditions, children were taught the information with the aid of multiple pedagogical cues, but in the Overhearing conditions, children had to listen in to one side of a phone call to learn the information. Older preschoolers (Experiment 1) learned all items above chance in both conditions. Younger preschoolers (Experiment 2) learned words and facts above chance in the Pedagogical condition but were at chance at learning words in the Overhearing condition, despite reliably learning facts from overhearing. Experiment 3 demonstrated that younger children's difficulty at learning new words from overhearing could not be explained by only being able to hear one side of the phone conversation, as they similarly struggled when the phone call took place over speakerphone. Measures of children's touch behavior suggest that older children were better able to coordinate their attention between the overheard speech and objects, though even younger children showed evidence of attention to the overheard speech. Together, our results demonstrate that by age 5, children can learn multiple new words and facts via overhearing. This self-directed learning ability depends on being able to coordinate attention between speech and the surrounding environment, a capacity that develops throughout preschool.

2.1 Introduction

Since Jerome Bruner's (1961) description of "discovery learning," the idea of self-directed learning has been influential in the educational and psychological communities, and, more recently, the machine learning community. The self-directed learner, in contrast to the passive learner, selects the information they want to receive (Gureckis & Markant, 2012). Studies with children in this vein support the idea that they are curious and exploratory learners. For example, infants and young children selectively attend to some auditory or visual inputs over others, and selectively explore objects, suggesting that children choose the information they want to receive from early in life (e.g., Gerken et al., 2011; Golinkoff et al., 1987; Kidd et al., 2012, 2014; Piantadosi et al., 2014; Sim & Xu, 2017; Stahl & Feigenson, 2015). As they mature and expand the scope of their attention, children amass information about the world around them by observing, asking questions, and performing physical interventions on their environments (Gopnik & Wellman, 2012; Piaget, 1954; Schulz, 2012; F. Xu, 2019).

To date, the majority of research on children's ability to direct their own information-gathering has focused on their independent investigation of causal systems, rather than social or linguistic systems (but see Partridge et al., 2015; Ruggeri et al., 2019). Although these studies have provided insight into children's developing self-directed learning abilities, causal systems arguably require less social learning to master, and may therefore be particularly amenable to self-directed learning. For example, a child alone in the crib can discover that a twitch of their leg causes an object suspended overhead to move (Rovee & Rovee, 1969), but will have to learn from another person that the object is called a "mobile." The present studies ask whether the self-directed learning abilities demonstrated in previous studies of causal learning — children's recognition of and attraction to unknown information, and their capacity to acquire relevant information through their own selective attention and action (e.g., Cook et al., 2011; Schulz & Bonawitz, 2007; Sim & Xu, 2017) — extend to a more social domain, like language development. Previous work suggests that in teaching contexts, children are selective in who they trust as credible sources of new linguistic information (Koenig et al., 2004; Koenig & Harris, 2005; Luchkina et al., 2018). But what about in the real world, when children have to not only evaluate new information from potential sources of learning, but also recognize learning opportunities in the first place, and selectively 'tune in' to them?

Like causal learning, language development is a domain in which children are surrounded by relevant information for learning, namely, the language spoken by speakers around them. This naturally occurring speech provides potential opportunities for self-directed learning, as there will be many utterances that are available to but not yet understood by the child, and that speakers around them do not explicitly help them comprehend. Speech that is not directed to a child — but that the child can overhear — can take many different forms, including an adult directing speech to a sibling, conversations among other children, television monologues, and speech among adults. Our experiments focus on what children can

learn from overheard speech between adults because this presents an especially challenging information source to learn from. Compared to when they are speaking to another adult, an adult directing speech to a child will take more responsibility for maintaining their addressee's attention and monitoring their understanding (Schober & Clark, 1989; Tomasello et al., 2005). Thus, child-directed speech can be thought of as guiding a child's attention, similar to the way experimenters in previous studies explicitly demonstrated how a novel toy worked for a child's benefit (e.g., Bonawitz et al., 2011; Schulz & Bonawitz, 2007; Sim & Xu, 2017). Learning language from adult-directed overheard speech, on the other hand, can be thought of as analogous to leaving a toy for a child to explore and learn from on their own. Learning in this context would seem to require many self-directed learning skills, as it requires children to (a) preferentially allocate their attention to the overheard speech without support from the speaker (e.g., because the speech is typically not marked as relevant for the child), (b) recognize how information in the overheard speech could fill children's own knowledge gaps (e.g., words for novel objects), and (c) learn from that information (e.g., mapping a novel word to its referent).

Although learning from overhearing can be seen as a paradigm example of self-directed learning, it is not typically studied as such. The under-emphasis on self-directed learning in the language domain likely stems in part from thinking about language acquisition as the product of the child receiving speech. Indeed, while a great deal of research suggests that children readily learn from speech that is directed to them, it is less clear what they are able to learn from speech that they overhear in their daily environments (Golinkoff et al., 2019; Shneidman et al., 2013). This question is of central importance because overheard speech constitutes a significant portion of the linguistic input for children across the world, and a larger proportion of the input than child-directed speech in many communities (e.g., Casillas et al., 2019; Shneidman and Goldin-Meadow, 2012a; see also P. Brown, 1998; Cristia et al., 2019; de León, 1998; Mastin and Vogt, 2016; Ochs and Schieffelin, 1995; Pye, 1986a, 1986b; Sperry et al., 2019; Vogt et al., 2015; Weisleder and Fernald, 2013). For example, in one Yucatec Maya community, up to 80% of words that 12-month-olds heard were overheard (Shneidman and Goldin-Meadow, 2012a; see also Casillas et al., 2019). And in a diverse group of families from across the United States, overheard speech represented between a 54% and 210% increase over the average number of words that were directed to children by their primary caregivers (Sperry et al., 2019). Children's ability to learn from overheard speech is also important because it may provide a valuable source of information about the target language, since overheard speech is likely to contain different words and grammatical constructions from child-directed speech (Soderstrom, 2007), and is arguably a more accurate model of the language used by the target community (Sperry et al., 2019).

Although prior studies have failed to show a correlation between the quantity of overheard speech in children's home environments and their later vocabularies (Ramírez-Esparza et al., 2017; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a; Weisleder & Fernald, 2013), a number of experimental studies have shown that from at least 18 months of age, children are able to learn a new word equally well regardless of whether they have been taught the word directly, or have learned it via overhearing (Akhtar, 2005; Akhtar et al., 2001;

Baldwin, 1991; Fitch et al., 2020; Floor and Akhtar, 2006; Gampe et al., 2012; Martínez-Sussmann et al., 2011; Shneidman et al., 2009; for a review see Shneidman et al., 2016).[1] Together, these experimental studies provide important evidence that young children do not have to be engaged in joint attention toward a new word's referent in order to learn that word. Moreover, these studies show that children can track the referent of a novel word heard around them, even when the speaker is labeling the object for someone else, and when there is little indication that the utterance will be directly relevant to the child. Young children are even able to learn a new word from overheard speech when they have been given a distracting toy to play with (Akhtar, 2005).

While these prior experimental studies of learning from overhearing laid the groundwork for our experiments, they were not designed to test the degree to which children can learn new words from the complex, adult-directed speech that is likely to be present in children's daily environments, where demands on self-directed learning abilities are likely to be higher. For example, in prior studies (see Table E in the Appendix), children often only needed to learn a single novel word (Akhtar et al., 2001; Floor & Akhtar, 2006; Shneidman et al., 2009). This word was repeated as many as nine times and embedded in a small number of explicit labeling or directive sentence frames (Akhtar, 2005; Akhtar et al., 2001; Floor & Akhtar, 2006; Martínez-Sussmann et al., 2011; O'Doherty et al., 2011; Shneidman et al., 2009), and was sometimes presented using the cadence characteristic of child-directed speech (e.g., Shneidman et al., 2009), even though the speaker was talking to another adult. Further, experimenters often engaged with the child before beginning the conversation that the child was going to observe (Floor & Akhtar, 2006; Martínez-Sussmann et al., 2011), and interacted with the referents of the novel words and/or facts directly during the overheard conversation. Thus, while the ambient interactions in these previous studies were between third parties, they often resembled pedagogical child-directed interactions, and the early experimenter–child familiarization periods may have suggested that the context was one that children would be able to learn from (Gampe et al., 2012; see Appendix E for examples of how the experimental procedures of previous studies may have reduced demands on self-directed learning).

Building on this prior work, we aimed to design a conservative and more naturalistic test of children's self-directed learning from overhearing, to compare to learning in pedagogical, adult-guided contexts. Our experiments compared learning of multiple words and facts from conditions representing two extremes in terms of the demands they impose on self-directed learning: (1) an adult-guided interaction in which children were explicitly taught words and facts about a set of objects, and (2) a situation in which children could overhear an adult's phone conversation about the objects (which employed the same words and facts), but in which the adult did not look at the objects or the child. Given the intentionally challenging nature of our overhearing task (and informed by piloting with younger children), we tested preschoolers aged three to six. This was in contrast to previous experimental

[1] In the General Discussion, we return to the question of why children may show evidence of learning new words via overhearing in experimental lab studies, but not in their home environments.

studies of learning from overhearing, which have focused on children 18 to 30 months in age (see Table E). Our goal was in part to determine the lower bound with respect to age at which children can learn from overhearing when demands on self-directed learning are high.

2.2 The Present Studies

Across three experiments, we asked how learning from an explicitly pedagogical adult-guided interaction compares to self-directed learning from complex, naturalistic overheard speech during the preschool years. Following previous overhearing experiments, in Experiments 1 and 2, we employed a between-subjects design to compare learning in a highly pedagogical interaction (Pedagogical condition) to self-directed learning (Overhearing condition) by 4.5– to 6.0–year-olds (Experiment 1) and 3.0– to 4.5–year-olds (Experiment 2). In both conditions, children were first familiarized with a set of familiar and novel objects. In the Overhearing conditions, an experimenter received a phone call while the child played with the objects. The experimenter's half of the dialogue — which was directed to an unseen adult interlocutor in the Overhearing condition — was directly addressed to the child in the Pedagogical condition. In the Overhearing condition, the experimenter described the objects without looking at or manipulating them: she indirectly provided a novel label for each of the unfamiliar objects (e.g., "I brought a purple pimwit today"), and an idiosyncratic fact corresponding to each of the unfamiliar and familiar objects (e.g., "The purple pimwit is my sister's favorite"). In contrast, in the Pedagogical condition, the experimenter used child-directed speech, engaged in joint attention with the child and the objects, and pedagogically demonstrated each toy as she introduced its associated label and fact. Children in both conditions were then tested on whether they had linked the new labels and facts to the target objects via an explicit object request task. Finally, Experiment 3 followed up on the results of Experiments 1 and 2 to explore whether 3.0– to 4.5–year-old children would be better able to learn from overhearing if they had access to *both* ends of the phone call, and thus overheard a dialogue as opposed to a 'halfalogue' (Emberson et al., 2010).

Our overhearing conditions were designed to simulate what it might be like to learn from speech directed from one adult to another (indeed, multiple parents received phone calls during their child's participation in the lab). First, since conversations between adults are likely to contain multiple pieces of information that are unknown to children, children in our studies overheard multiple novel words and facts (Experiment 1: four words and six facts; Experiments 2 and 3: three words and five facts). Second, these novel words and facts were embedded in a variety of sentential contexts and were spoken in a conversational, adult-directed speech style, rather than with the pace and prosody of child-directed speech. Third, although the novel words and facts referred to objects that were present in the scene, these objects were displaced from the experimenter, who did not look at or manipulate them. There is evidence that this is a common feature of real-world word occurrences, at least for verbs: in one naturalistic study of toddlers' verb-learning, over 60% of the verbs caregivers produced were in reference to absent events (Tomasello & Kruger, 1992). Following criticism

by prior researchers that the early familiarizing interactions with the experimenter in previous studies might open a pedagogical frame, in our experiment, the child did not engage with the experimenter until after the phone call was over, and instead interacted only with an adult confederate.

We see our overhearing context as analogous to a variety of naturalistic ones. For example, when driving, an adult's conversation (in person or on the phone) with another adult or an older sibling will often be audible from the backseat. Likewise, when preparing food or orchestrating bedtime, adults may discuss objects present in the scene (ingredients, dishware, bath supplies . . .) without interacting with those objects directly, and while their attention is half-focused on another task. Anecdotally, when caregivers answered a phone call when we tested in lab or at museums, their speech often included some explanation or description of their immediate whereabouts ("We came in to do a study at Berkeley" / "There's a broken car toy here that she's obsessed with" / "Somehow we got here with only three shoes between them" / "I'm regretting having brought such sticky snacks"). In order to learn the new words and facts, children in our Overhearing condition had to recognize that the overheard speech was relevant to their situation, coordinate their attention between the overheard speech and the objects, and use the linguistic context to establish correspondences between the words, facts, and objects. Our three-year age range enabled us to examine how children's developing attention might influence their efficacy at recognizing and seizing this learning opportunity.

Inspired by previous research, we included different kinds of learning targets — i.e., new words for novel objects, and new facts for novel and familiar objects — to understand the factors that might affect learning from overheard speech (Markson & Bloom, 1997). We hypothesized that it would be easier for children to learn facts for the novel objects (e.g., that a novel object was "found in the garden") than words for those objects (e.g., that a novel object is "a zav") because only the latter require children to encode and retain a novel phonological form in memory (e.g., Deák & Toney, 2013). Extending this logic, we predicted that children might also be more successful at learning facts corresponding to familiar objects — comprised entirely of known words — compared to facts corresponding to novel objects, which might be more difficult to both map and remember. Our overhearing context requires that children attend to both the overheard speech and the objects in front of them, suggesting that the task of mapping overheard facts might be especially difficult when the objects themselves are unfamiliar and have to be identified. To understand how attention might affect learning, we also monitored what children looked at and touched as they overheard the experimenter's phone conversation while playing with the objects, and explored both how this changed with age and whether it was related to children's performance at test.

2.3 Experiment 1

Method

Participants

Participants were 68 children learning English as their primary language between 4.5 and 6.0 years of age (31 female; 4.5–5.9 years, $M = 5.1$ years, $SD = 0.5$ years). Our target sample size was 64 children; however, once an additional child had participated in the study, we recruited an additional three participants to maintain our equal sample sizes between conditions and counterbalanced orders. Our target sample size was determined because it provided us with at least 85% power to detect the most conservative of the effect sizes reported by Gampe and colleagues (2012; Cohen's $d = 0.55$, at $\alpha = 0.05$), using a one-sample t-test comparing children's learning from overhearing to chance. Power was calculated using the pwr package (Champely, 2014) in R (R Development Core Team, 2020).

Participants were assigned to one of two conditions, Overhearing ($n = 34$, 14 female; 4.5–5.9 years, $M = 5.1$, $SD = 0.5$) or Pedagogical ($n = 34$, 17 female; 4.5–5.9 years, $M = 5.2$, $SD = 0.4$). There was no difference in age between the two conditions ($t(100) = 0.1$, $p = .9$; Cohen's $d = -0.019$). Families were recruited and tested in lab or at a local preschool or museum. When parents gave permission, study sessions were filmed, so that videos of the Overhearing condition could be coded (30 videos in the Overhearing condition total). Eight additional children participated, but were excluded due to failing a familiar label control trial (4; see *Procedure* section, below), having already witnessed another child participate (2), failing to complete the study (1), or experimenter error (1).

Stimuli

Our stimuli consisted of six toys: four novel objects, and two familiar objects, shown in Figure 3. Children were exposed to new words for each of the four novel toys, and idiosyncratic facts for each of the entire set of six toys (see Table 5). Within each condition, children were assigned to one of two mappings between the words, facts and objects (see Table S1 in the *Supplementary Online Materials*). This made it less likely that overall learning of any specific novel word or fact could be due to its natural fit with any particular object. Also to guard against this possibility, we created facts that were not transparently related to any perceptual features of the objects.

The novel objects were purchased from a hardware store and subsequently altered to appear more novel. Each object had a distinct dominant color. The *pimwit/zav* was a French whisk with a circular metal face and purple pom-pom hair, which could stand on its own or be bounced on the table. The *toma/fep* was a large button light decoupaged lime green and rimmed with pipe cleaner spirals. Children could make the light turn on by pressing the green felt star on the object's domed surface. The *fep/pimwit* was a blue microfiber duster with the handle removed, leaving two sleeves children could slip their

fingers inside. The *zav/toma* was a wooden finial painted yellow and covered in multicolored Velcro diamonds that could be removed to reveal felt of a different color, and then replaced elsewhere. Finally, the two familiar objects were a small plush dog and a plastic toy cup of milk. Initial piloting with this set of toys confirmed that children of this age did not recognize or know category labels for any of the novel objects, but were consistently able to recognize and name the two familiar objects.

Figure 3: *Stimuli Used in Experiment 1.*

Note. Novel objects appear in the top row, familiar objects in the bottom row.

Table 5: *Words and Facts Used in Experiment 1*

Word	**Fact**
fep	*. . . I got from Disneyland*
pimwit	*. . . my sister's favorite*
toma	*. . . my uncle gave me*
zav	*. . . I found in the garden*
dog	*. . . I bring to school*
cup	*. . . I've had for two years*

Procedure

The procedure for Experiment 1 consisted of three phases: *familiarization*, *learning*, and *test* (Figure 4). In the familiarization phase, children were seated at a table and introduced to the set of objects, without labels, either by the experimenter (Pedagogical condition) or a confederate (Overhearing condition). In the learning phase, children were exposed to a set of mappings between labels, facts, and the array of objects, either through direct instruction (Pedagogical) or a phone conversation that they could overhear (Overhearing). In the test phase, the experimenter tested children's learning of the mappings in a series of requests for objects. Four control trials interspersed throughout the test phase probed children's ability to give the correct object in response to familiar nouns (two requests each for "dog" and "cup"). Children who failed one or more of these trials were excluded from analysis ($n = 4$).

Parents were asked to complete a brief questionnaire regarding their child's typical language environment, modeled after interviews used to assess children' overhearing experience by Shneidman and colleagues (2009). We obtained completed questionnaires from 54 of our participants. The questionnaire, along with summary statistics regarding this subset of our sample — including caregivers' estimates of their child's exposure to overheard phone calls — can be found in Table S2 in the *Supplementary Online Materials*.

Overhearing condition. Each participant in the Overhearing condition entered the testing room with the confederate, who sat across from them at a low table. Caregivers and siblings, when present, were asked to sit quietly out of the child's direct line of sight.

Familiarization phase. Once the child and confederate were seated, the experimenter entered the room, placing a box containing the six toys in the center of the table and announcing, "These are my toys!" To diminish any potential for the interaction to be interpreted as pedagogical, the experimenter did not make eye contact with either the child or confederate. She then walked to a chair placed against the wall 3–4 ft from the table, where she began "working" on a laptop that had been resting there, surreptitiously starting a timer on her phone for 1 min. The confederate meanwhile pulled the box of toys toward her and commented on their unfamiliarity: "I've never seen these toys before, these are [Experimenter name]'s toys!" The confederate then removed each toy from the box individually, drawing the child's attention to it as she placed it on the table between them. If the child asked the confederate a question about the objects, she replied, "I don't know! These are [Experimenter name]'s toys." When all the toys had been removed from the box, the confederate set the box on the floor and excused herself, but encouraged the child to continue playing: "I have to go do some work now, but it was nice playing with you. You can keep playing with [Experimenter name]'s toys." The confederate sat behind the child, where she filled out paperwork associated with the visit.

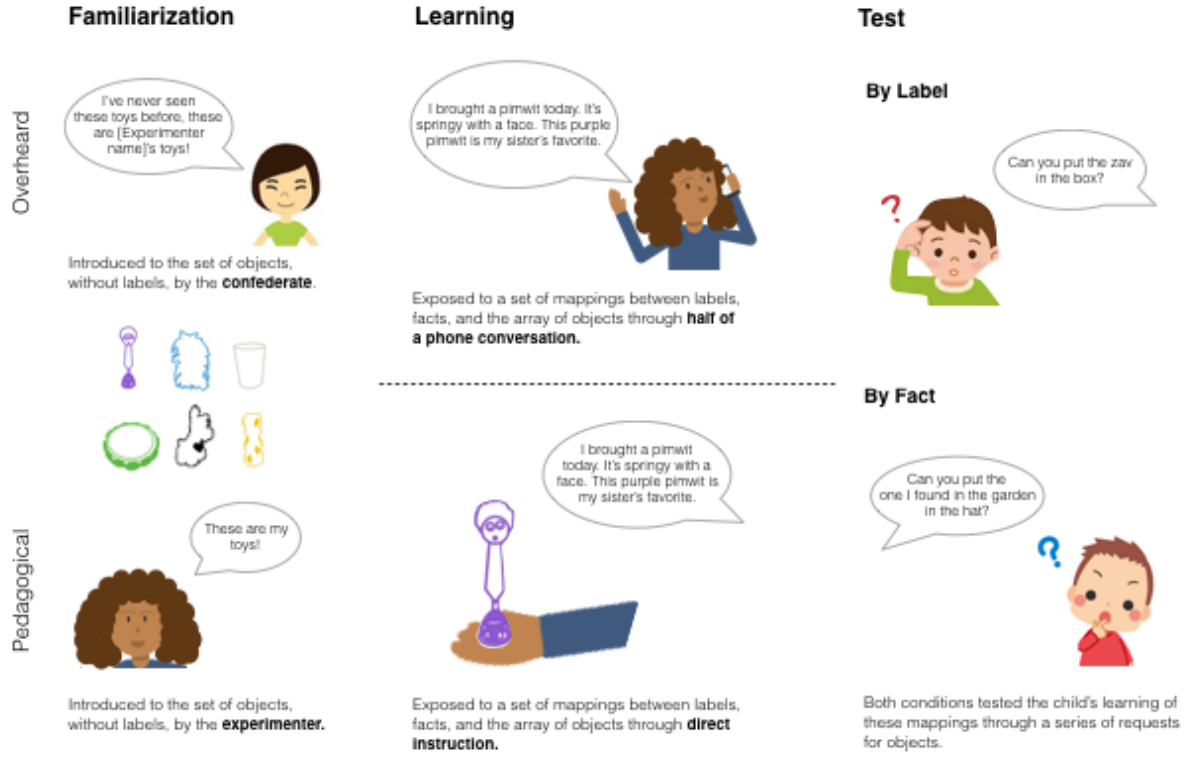


Figure 4: *Experimental Procedure for Experiments 1 and 2.*

Learning phase. While the child was playing with the toys, the experimenter's phone rang. The experimenter answered the phone, and casually described each of the toys, as if in conversation with a friend (see Appendix). The other side of the conversation could not be heard; children were thus exposed to a halfalogue. Following an exchange of pleasantries, the experimenter listed the objects, then spent approximately 15s discussing each in turn, never looking toward them. Within each 15-second segment, the experimenter referred to physical properties of the object (e.g., its color and shape), and uttered its novel label three times, and its fact once. The target fact was always mentioned toward the end of the segment of speech for that object. At the end of the phone call, the experimenter briefly mentioned the novel labels and their associated facts again. In total, each novel word was used five times, embedded in a variety of sentential frames, while each fact was uttered twice (further repetitions of the facts made the script substantially less naturalistic). The experimenter avoided making eye contact with the child through this entire phase, but following the phone call, turned to them and apologized for having taken the call, asking if the child was ready to play a game. When the child answered yes, the experimenter moved to the chair formerly occupied by the confederate, and proceeded to the testing phase.

Pedagogical condition. Children in the Pedagogical condition entered the testing room with the experimenter, and sat across from her at the table. Caregivers and siblings sat behind the child.

Familiarization phase. The experimenter placed the box of toys on the table between her and the child, and said, "These are my toys!" She removed each toy from the box, sharing attention with the child toward it, and then set the empty box on the floor.

Learning phase. In the Pedagogical condition, the experimenter delivered a nearly identical script to that used in the Overhearing condition, spoken at the same rate, but directed to the child. The experimenter spoke enthusiastically, made eye contact with the child, and held each object in the air between the two of them as she labeled it. The experimenter also demonstrated properties of the objects that appeared in the script (see Appendix). For example, when introducing the *zav/fep*, which has "stickers you can take... on and off," the experimenter peeled and replaced a couple of the Velcro "stickers" as she spoke. When talking about the *toma/zav*, she pointed to the subtle "green star" on its surface and showed how the object "only lights up" when pressed there. These demonstrations amplified the contrast between the Pedagogical condition and the Overhearing condition, where children's attention was self-initiated, rather than elicited and maintained by the experimenter. Following the labeling of the individual objects, the experimenter asked the child if they wanted to play a game with the objects, tapping each one as she provided its associated label and fact a final time.

Test phase. The test phase was identical in both conditions, and consisted of three blocks of six trials each. To initiate each block, the experimenter brought out a single container (a box, bowl, or hat), and asked the child if they were ready to play a game. The toys were arranged on the table immediately in front of the child. On each test trial, the experimenter asked the child to place the toy associated with a particular word or fact into the container: e.g., "Can you put the [*zav/one I found in the garden*] in the [*bowl/box/hat*]?" The experimenter avoided cueing the child toward the target object by maintaining eye contact and refraining from glancing at the objects when asking the test question. After the child placed an object in the container, the experimenter removed it and replaced it on the table with the rest of the toys before moving onto the next trial. The first two blocks always tested children's knowledge of the word-object mappings, providing two data points for each novel word per participant. The third and final block tested children's knowledge of the fact-object mappings. The trials within each block were presented in one of two pseudorandom orders, counterbalanced across conditions and mappings. Finally, to test for the possible influence of children's preferences, the experimenter asked the child to identify their "favorite toy" at the end of the test phase. The experimenter (Pedagogical condition) or confederate (Overhearing condition) noted the object the child provided on each trial.

Coding and Analysis

Results include analyses of children's trial-by-trial test performance, along with analyses of behavioral signatures of attention to the phone call for children in the Overhearing condition. Full documentation of our experimental and data processing procedures can be found at https://osf.io/avyg5/?view_only=33cbb9ab189343a7b6e8f6c7c517026d, along with the raw data and scripts for all analyses outlined below. Study session videos and coding spreadsheets are stored on Databrary.org (linked in the above online repository), and are available to registered users at the access level permitted by each caregiver.

Test performance. When available, children's object choices at test were double-coded from video by a research assistant who had not been present for the study session. Agreement between this second coding and the in-session coding was 100%. For each condition and learning target type, we report means and bootstrapped 95% confidence intervals over all participants' test accuracy, calculated in terms of their proportion of correct critical trials. Independent samples t-tests compare sample means between conditions, for both words and facts.

Comparisons to chance. One-sample t-tests compare sample means to predetermined values for chance. Our selection of chance assumes that children are considering all novel objects (and only novel objects) on every word-learning test-trial, and all possible objects on every fact-learning test trial. We test the validity of this assumption by conducting the same comparison to a learning-target–specific value for chance, but restrict our analysis to the only the first critical trials of each test block (see Independent trials section below).

Mixed effects models. We use mixed effect logit models constructed using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2020) to analyze children's performance at test. These models are fit to the data for children's trial-by-trial accuracy (coded as 0 = incorrect, 1 = correct), with random intercepts per participant. We additionally include fixed effects for our predictors of interest, including condition (Pedagogical, Overhearing), type of learning target (word, fact), and age (in years above our minimum age, to increase the interpretability of our model coefficients). When models with the predictors of interest fail to converge, we refit our model, excluding random effects. We report model coefficients or odds ratios and bootstrapped 95% confidence intervals to assess the magnitude and reliability of the parameters of the winning model. Finally, we use the Anova function in the R car package (Fox & Weisberg, 2019) to report traditional significance levels for our estimated model parameters.

Object familiarity. To test whether fact-learning is affected by whether the relevant object is familiar (i.e., the dog or cup) or novel (i.e., the purple, blue, yellow, or green object), we analyze the trial-by-trial data for facts separately. We follow the same procedure described above for evaluating nested mixed effects logit models, this time including fixed effects for (1) age, (2) condition (Pedagogical, Overhearing), (3) object familiarity (coded as 0 = unfamiliar, 1 = familiar), (4) the interaction of condition and object familiarity.

Behavioral proxies of attention. Pairs of trained research assistants coded videos from the Overhearing condition in Datavyu (Datavyu Team, 2014), focusing especially on the period corresponding to the phone call. We distinguished the initial and final social portions of the call from the segments relating to each object. Each segment for an object began at the onset of the mention of its label, and ended at the onset of the next toy's label. Subsequent passes were coded without audio or transcripts, so that coders of children's behavior were unaware of which toy the experimenter was discussing. After computing interrater reliability for each coded variable, disagreements between coders were resolved by the first author, and these final values were used in all analyses.

Child gaze. Across testing locations, the child was always seated so as to make looks toward the experimenter easy to code (following Martínez-Sussmann et al., 2011). We defined a period of gazing toward the experimenter as beginning when the child turned their head toward the experimenter, and ending when the child turned their head back to the toys. From these periods, we calculated the overall proportion of the phone call — beginning and ending when the experimenter touched her thumb to the phone screen to answer or hang up the phone — that the child spent looking toward the experimenter.

Inspired by previous studies (Martínez-Sussmann et al., 2011; Shneidman et al., 2009), we next asked whether the children who spent more of the overhearing exposure oriented toward the experimenter: (1) performed better at test, and (2) were older. To do so, we calculated the correlation between the percentage of the phone call that the child spent looking toward the experimenter, and their test trial accuracy, using the cor.test function from the R stats package (R Development Core Team, 2020). To test whether children directed more visual attention to the phone call as they got older, we did the same for the child's age in years. Previous results suggest that children's gaze behavior should positively correlate to their test performance. However, because our study involved many objects, we reasoned that gaze to the experimenter might sometimes *impede* children's ability to link the target novel words or facts to their object referents. Thus, as described below, we also coded children's touch behavior as the experimenter was discussing the objects, for evidence of whether children were accurately tracking the referents of the experimenter's speech.

Relation to call. Periods of touching each object were coded as beginning when the child touched an object with either hand, and ending when their hand left it again. To test whether children's touch behavior was likely related to the content of the experimenter's speech, we computed a repeated measures correlation between the order that each object was mentioned (1–6) and the cumulative duration of children's touching of each object, in terms of the number of video frames. We reasoned that if children's attention was drawn to each object following the experimenter's mention of it, the amount of time they spent touching each object should be negatively correlated with its order of mention. That is to say, children should have more time over the course of the call to play with objects that their attention was drawn to early, compared to objects that their attention was drawn to later. We use the eponymous function of the rmcorr package (Bakdash & Marusich, 2017) to report the correlation coefficient, bootstrapped 95% confidence interval, and p-value for the correlation between number of frames and order of mention, across participants.

Matching-object touch. To obtain a single measure reflecting the correspondence between the child's haptic behavior and the content of the experimenter's overheard speech, we first calculated the proportion of each segment of the call during which the experimenter was discussing a particular novel object (e.g., "the purple pimwit"), and the child was touching that object (e.g., the purple whisk). From this, we subtracted the mean proportion that the child was playing with the same object (e.g., the purple whisk) during the remaining three novel-object segments of the call in which it was *not* the object the experimenter was discussing (e.g., concerning the "blue fep," "green toma," and "yellow zav"). Thus, if the child tended to play with objects more when the experimenter was talking about them, compared to when she was not, they would receive a positive score, with the magnitude of the score reflecting the degree to which this was true across novel objects. If, however, the child tended to touch objects more during the times when they were *not* the current topic of the experimenter's speech, their score would be negative.

For illustration, Figure 5 shows the time-course of four children's touch behavior as it aligned with the topic of the experimenter's speech (for an analogous plot of our full sample, see Figure S1). The highest-scorer (*Child A* in Figure 5) touched the purple whisk for 100% of the segment in which the experimenter discussed it (and 53%, 0%, and 0% of the segments in which she discussed the other three novel objects: 18% on average); the blue duster for 63% of the matching segment (and 0%, 100%, and 100% of the other novel-object segments: 67% on average); the green button-light for 79% of the matching segment (0% for all other novel-object segments), and the yellow finial for 71% of the matching segment (0%, 0%, and 16% of the other ones: 5.3% on average). Two children received scores of 0, one because they played with a single object indiscriminately (*Child C*), and another because they never touched the objects at all (*Child B*). The lowest score in Experiment 1 (*Child D*) belonged to a child who only touched objects when the experimenter was not talking about them, earning them a negative score. Average agreement on this measure between pairs of trained research assistants was 82%; disagreements were resolved by the first author, whose final

coding was used in all analyses.

We report means and bootstrapped 95% confidence intervals for this measure. To answer the question of whether children reliably received positive, rather than zero or negative, scores, we conduct an exact binomial test using the binom.test function in the R stats package, against the alternative hypothesis that children should be equally likely to receive a positive score as to not receive a positive score. Finally, as we do for children's gaze proportions, we test for a correlation between children's matching touch score, and both their age in years and accuracy at test.

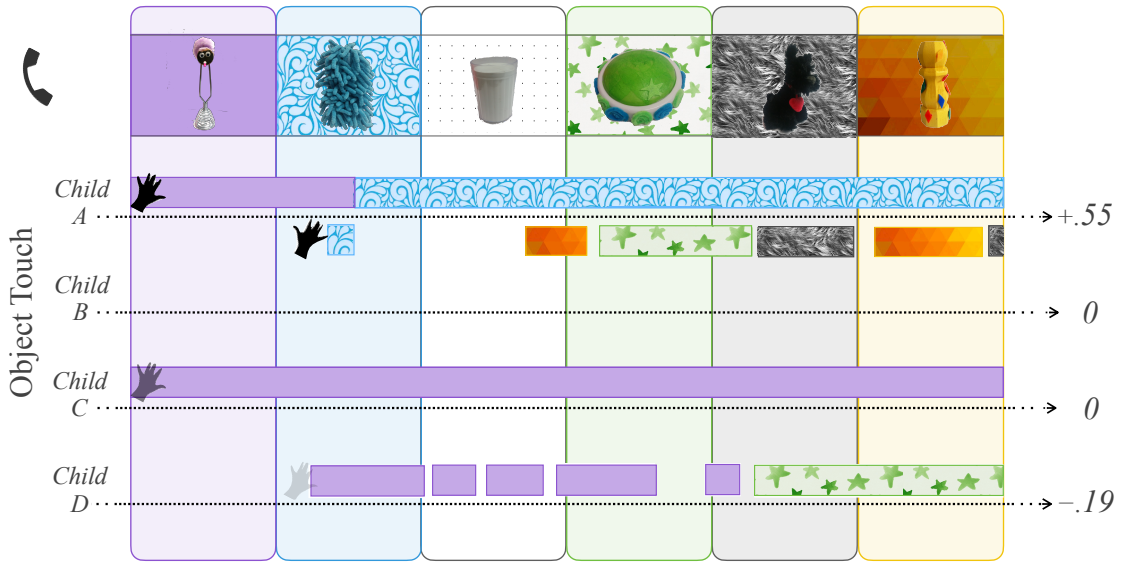


Figure 5: *Experiment 1 Touch Behavior and Matching-Object Touch Scores for Four Participants, Including the Receiver of the Highest (Child A) and Lowest (Child D) Scores.*

Note. Participants' periods of touching each object (horizontal bars for each hand, filled according to which object they were touching) are aligned with the time course of the overheard phone call (speech bubble in top row, divided and filled to reflect the object being discussed). *Child B* never touched any of the objects, and *Child C* touched the same object for the entire duration of the call. Segments of the call during which the experimenter discussed each object are delineated by columns. Matching-object scores corresponding to each participant appear on the right.

Results & Discussion

Test Performance

Preliminary analyses revealed no variation in test accuracy as a function of gender, preferred object (purple, blue, green, yellow, black, white), word form (*pimwit*, *fep*, *toma*, *zav*), test block (1–3), test trial order (1–18), or mapping (1 or 2), so subsequent analyses collapse across these variables.

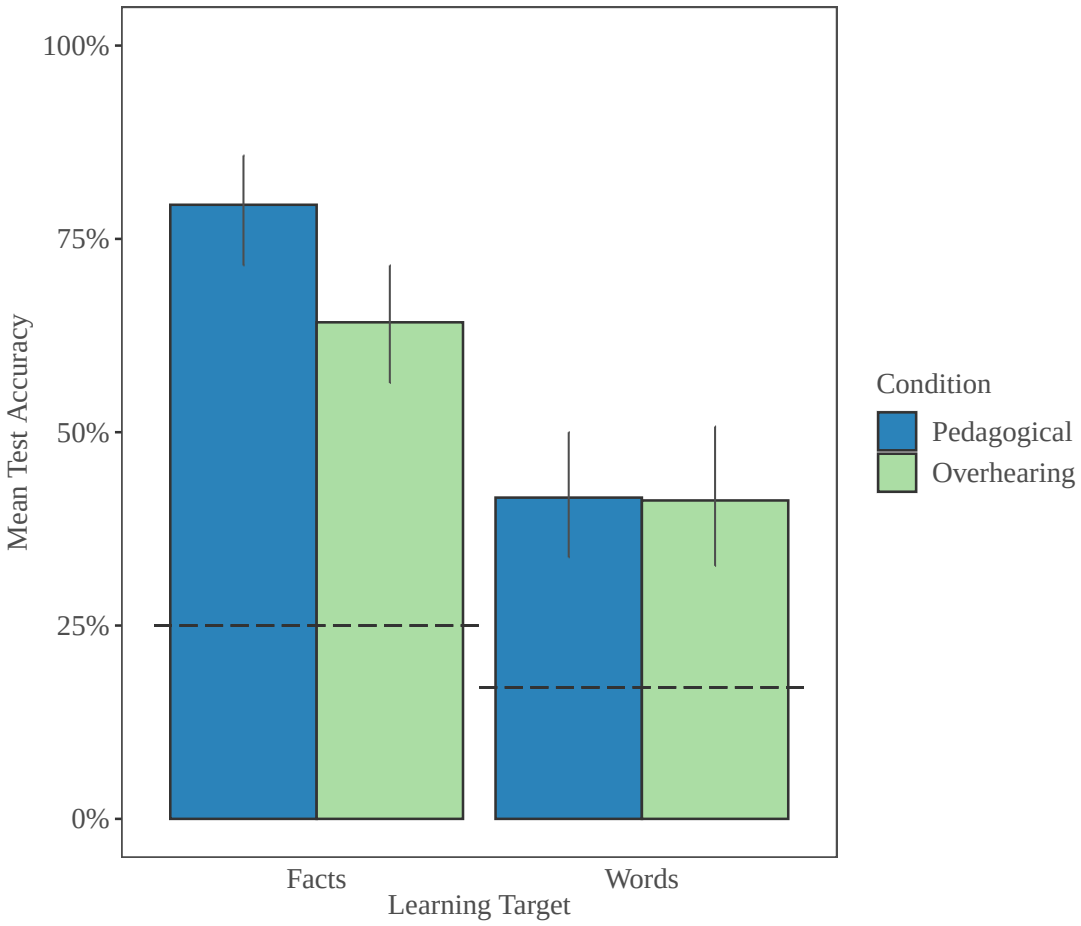


Figure 6: *Experiment 1 Mean Accuracy at Test by Learning Target and Condition.*

Note. Chance for each target type is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

Comparisons to chance. Figure 6 depicts children's accuracy at test, as a function of condition (Pedagogical vs. Overhearing) and learning target (words vs. facts). We considered chance performance for novel words to be 25% (because there were four novel objects to choose from), and chance performance for facts to be 17% (because all six objects were candidate referents). Planned one-sample t-tests revealed that children learned both novel words and facts above chance, in both the Pedagogical condition (Words: 42% [35%, 50%]; $t(33) = 4.1$, $p < .001$, Cohen's $d = 0.71$; Facts: 79% [72%, 86%]; $t(33) = 16$, $p < .001$, Cohen's $d = 2.69$), and the Overhearing condition (Words: 41% [33%, 49%], $t(33) = 3.5, p = .001$, Cohen's $d = 0.61$; Facts: 64% [56%, 72%]; $t(33) = 11$, $p < .001$, Cohen's $d = 1.82$).

Independent trials. To address the concern that choices of objects at test may not have been independent (that is, that children's responses on later trials might be influenced — for better or for worse — by their responses on earlier trials), we looked at performance on the first critical trial of each block. One-sample t-tests confirm that children's first-trial accuracy significantly differed from chance in both conditions (Pedagogical condition, first word-learning trials: 46% [36%, 56%]; $t(33) = 3.6$, $p = .001$, Cohen's $d = 1.36$; first fact-learning trials: 82% [71%, 94%]; $t(33) = -289, p < .001$, Cohen's $d = 1.7$; Overhearing condition, first word-learning trials: 47% [37%, 57%]; $t(33) = 4$, $p < .001$, Cohen's $d = 1.44$; first fact-learning trials: 71% [56%, 85%]; $t(33) = -243$, $p < .001$, Cohen's $d = 1.17$).

Mixed effects models. We next fit a mixed effects logit model predicting trial-by-trial test accuracy (incorrect = 0, correct = 1) from an interaction between condition (Pedagogical or Overhearing) and learning target (word or fact), with random intercepts for subject. Children were more likely to respond accurately for facts overall (Odds Ratio = 5.85 [3.84, 9.08], Wald $\chi^2(1) = 81.88$), suggesting that word learning was more difficult than fact learning in both conditions (Overhearing: $t(400) = -5$, $p < .001$, Cohen's $d = 0.47$; Pedagogical: $t(500) = -9$; $p < .001$, Cohen's $d = 0.83$). Further, children in the Overhearing condition had decreased odds of accuracy for facts at test compared to children in the Pedagogical condition ($OR = 0.46$ [0.26, 0.81]), but the same was not true for words ($OR = 0.99$ [0.65, 1.49]), i.e., the interaction was significant (Wald $\chi^2(1) = 7.14$, $p < .01$). Thus, while children in the Pedagogical condition performed significantly better than those in the Overhearing condition on facts ($t(70) = 3, p < .01$, Cohen's $d = 0.66$), there was no difference in performance between the two conditions for words ($t(70) = 0.1$, $p = .9$, Cohen's $d = -0.015$). This model resulted in a significantly better fit than the null model with no predictors and only random intercepts ($\chi^2(3) = 103.33$, $p < .001$; AIC for model with interaction: 1213.4, AIC for null model: 1310.7), as well as a model which included both learning target and condition, but not their interaction ($\chi^2(1) = 7.23$, $p < .01$; AIC for model without interaction: 1218.6).

The effect of learning target in this analysis is difficult to interpret, given the different baselines for the words versus facts. Nevertheless, one possible reason for why children performed better on facts than words in both conditions is because facts were always tested

after words, when children may have been more familiar with the task and better able to demonstrate their knowledge. According to this logic, children should also have been more accurate when tested on words in the second block of testing than when tested in the first block, but there were no significant block effects for word learning in either condition. Specifically, a mixed effects logit model predicting correct responses on the two word-learning tests did not find a significant effect of block order ($OR = 1.02$ [0.70, 1.47], Wald $\chi^2(1) = 0.01$, $p = .92$). The observed fact advantage also defies an alternative prediction, that learning targets tested further from the learning phase should be recalled with lower accuracy. The fact advantage is also notable because the facts were mentioned fewer times than words (i.e., facts were mentioned only twice, while the new words were mentioned six times).

Instead, children may have exhibited superior learning of the facts because of features of the facts themselves. Unlike the novel words, the facts did not require children to encode and maintain a new phonological form in memory. Further, associations between facts and the relevant objects may have been easier to form because the multiple, familiar content words that comprised the facts (*sister's*, *favorite*) could be mapped directly to the described object (e.g., the *purple*, *springy* toy). As long as the child caught any part of the fact corresponding to that object (e.g., that it related to the experimenter's sister, or was someone's favorite), they could succeed at test. Thus, the length of the facts compared to the words may have afforded the child more opportunities for success, both in listening in, and in remembering what they heard. This explanation accords with previous work comparing fast-mapping of different linguistic items (Deák & Toney, 2013).

As noted in the Introduction, we were also interested in whether children may have performed better on the two facts for the familiar objects with known labels (e.g., "... a cup I've had for two years") than on the four facts for novel objects (which employed novel labels, e.g., "... a zav I found in the garden"). In principle, children could have learned facts for the familiar objects by attending solely to the speech, whereas learning facts for the novel objects additionally required children to determine which object in the scene was being referenced. To test whether it was easier for children to learn facts for familiar objects, we fit a model with age, condition, and object familiarity to the fact learning data, with random intercepts for subjects. Compared to this model, a model which also included an interaction between condition and familiarity resulted in a significantly better fit ($\chi^2(1) = 4.9$, $p < .05$; AIC without interaction: 466, AIC with interaction: 463), and also outperformed a model with condition as the sole fixed effect ($\chi^2(2) = 6.7$, $p < .05$; AIC: 466). Interestingly, facts corresponding to the novel as opposed to familiar objects had decreased odds of accuracy only in the Overhearing condition ($OR = 0.32$ [0.11, 0.88]) but not in the Pedagogical condition. That is, the interaction between object familiarity and condition was significant (Wald $\chi^2(1) = 4.9$, $p < .05$). In the Overhearing condition children were on average 75% [65%, 84%] accurate for familiar object facts, compared to 59% [51%, 67%] for novel object facts; in the Pedagogical condition, accuracy was 76% [66%, 85%] and 81% [74%, 88%] for familiar and novel objects, respectively.

The fact that children performed better on familiar object than novel object facts in the Overhearing condition, but equivalently on familiar and novel object facts in the Pedagogical

condition, suggests that identifying the correct referent as the experimenter spoke was part of the challenge of the overhearing task. To learn facts corresponding specifically to the novel objects, children in the Overhearing condition had to consult the scene to identify the correct object based on the experimenter's description. In the Pedagogical condition, on the other hand, the experimenter drew the child's attention to each object — regardless of familiarity — as she discussed it, reducing the gap in referential ambiguity between the two fact types.

Object familiarity. As noted in the Introduction, we were also interested in whether children may have performed better on the two facts for the familiar objects with known labels (e.g., "...a *cup* I've had for two years") than on the four facts for novel objects (which employed novel labels, e.g., "...a *zav* I found in the garden"). In principle, children could have learned facts for the familiar objects by attending solely to the speech, whereas learning facts for the novel objects additionally required children to determine which object in the scene was being referenced. To test whether it was easier for children to learn facts for familiar objects, we fit a model with age, condition, and object familiarity to the fact learning data, with random intercepts for subjects. Compared to this model, a model which also included an interaction between condition and familiarity resulted in a significantly better fit ($\chi^2(1) = 4.9$, $p < .05$; AIC without interaction: 466, AIC with interaction: 463), and also outperformed a model with condition as the sole fixed effect ($\chi^2(2) = 6.7$, $p < .05$; AIC: 466). Interestingly, facts corresponding to the novel as opposed to familiar objects had decreased odds of accuracy only in the Overhearing condition ($OR = 0.32$ $[0.11, 0.88]$) but not in the Pedagogical condition. That is, the interaction between object familiarity and condition was significant (Wald $\chi^2(1) = 4.9$, $p < .05$). In the Overhearing condition children were on average 75% [65%, 84%] accurate for familiar object facts, compared to 59% [51%, 67%] for novel object facts; in the Pedagogical condition, accuracy was 76% [66%, 85%] and 81% [74%, 88%] for familiar and novel objects, respectively (see Figure 7).

Behavioral Proxies of Attention

Analyses of children's behavior were restricted to the 30 participants in the Overhearing condition for whom we received parental consent to record.

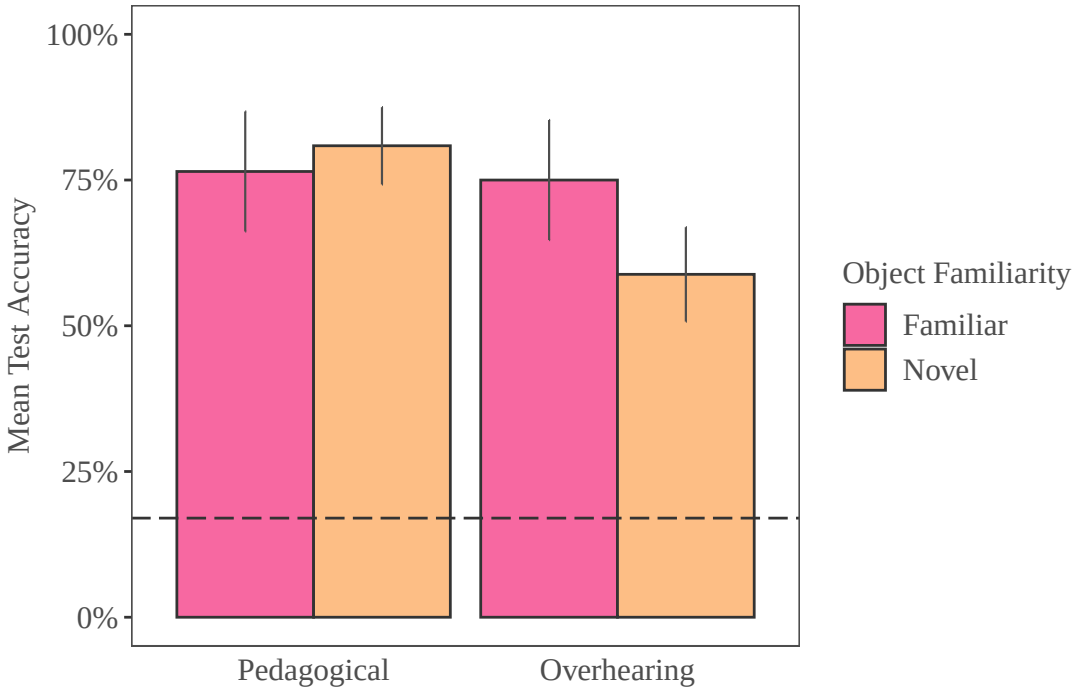


Figure 7: *Experiment 1 Mean Fact Accuracy by Familiarity of Target Object.*

Note. Chance for facts is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

Relation to call. As an initial test of the relation between the content of the overheard phone call and each child's exploratory behavior, we first computed the cumulative sum of frames in which the child was touching each object. As we would expect if children were more likely to attend to objects that they heard described earlier in the phone call, the number of frames in which children touched each object was significantly negatively correlated with its order of mention in the overheard call ($r_{rm}(59) = -0.46$ $[-0.64, -0.23]$; $p < .001$). We also observed that children often perseverated on individual objects in their manual exploration during the phone call, reminiscent of other work on the development of self-directed learning subskills (e.g., question asking, Ruggeri et al., 2016). Children's tendency to focus on single objects makes the significant correlation between touch and phone call more notable, as it means that when children did switch to playing with a new object, their selection was not random, but rather guided by the phone call happening nearby.

Matching-object touch. Twenty-six participants received positive scores on our matching object touch measure (described in *Coding and Analysis*, above), while two did not touch the objects at all (Range: -0.15 – 0.63; $M = 0.26$ $[0.19, 0.33]$; Figure 8). The measure was designed so that children's positive scores suggest they were reliably tracking the referents of the words in the experimenter's speech, as indexed by the objects they were touching, and so that the magnitude of the score might indicate the degree to which they were doing this. An exact binomial test confirmed that children received positive scores significantly more often than zero or negative scores ($p < .001$)[2]. The magnitudes of children's matching-object scores were also significantly correlated with their age (Pearson's $r = .45$ [0.10, 0.69]; $t(30) = 3$, $p = .01$), suggesting children's attention to and processing of the overheard speech improved as children got older. Nonetheless, children's matching-object scores were not significantly correlated with their accuracy at test (Pearson's $r = -0.04$ [−0.40, 0.32]; $t(30) = -0.20$, $p = .8$).

Child gaze. There was substantial variation in the proportion of the phone call that children spent looking toward the experimenter (plotted as points in Figure 8; Range: 0 – 0.47, $M = 0.13$ $[0.11, 0.14]$). However, here, the amount that each child looked toward the experimenter was not significantly correlated with either their age (Pearson's $r = -0.09$ [−0.44, 0.28]; $t(30) = -0.05$, $p = .6$) or their accuracy at test (Pearson's $r = 0.21$ [−0.17, 0.53]; $t(30) = 1$, $p = .3$).

Although this result conflicts with those of previous overhearing studies (Martínez-Sussmann et al., 2011; Shneidman et al., 2009), this is not surprising given the many differences between our study and previous ones. In previous studies, the experimenter manipulated or attended to the novel objects while using the novel labels, such that a child who looked toward the experimenter could attend both to the speech and to the object referents. In our task, on the other hand, children had to choose between looking at the experimenter and looking at the objects, because the experimenter was displaced from the objects she was discussing. Although observation of the experimenter's attention provided referential cues in previous studies, it was not informative in our study, where only the experimenter's speech provided referential cues.

[2]While we see promise in the distribution of positive touch scores, we caution that analyses of the video data in particular should be interpreted as suggestive, given the low sample size.

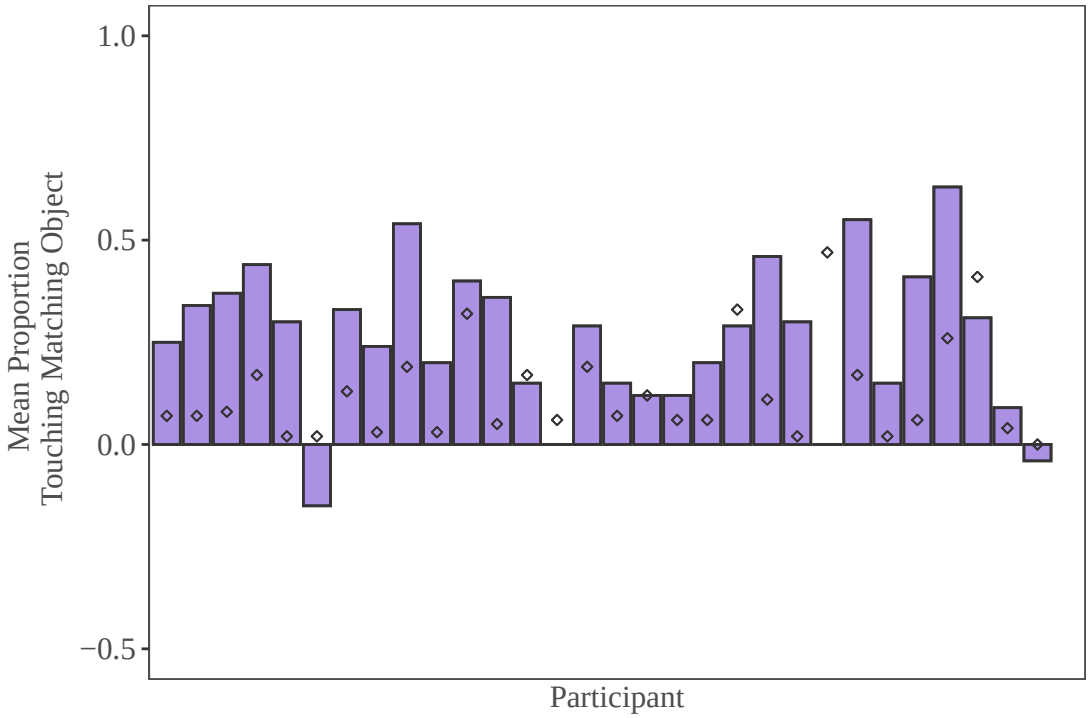


Figure 8: *Experiment 1 Matching-Object Touch and Gaze to the Experimenter.*

Note. Positive values on the matching-touch measure (bars) indicate that the child touched the specific novel object that the experimenter was discussing more often as they was discussing it than when they were not. Overlaid points reflect the proportion of the call each participant spent looking toward the experimenter.

2.4 Experiment 2

Experiment 1 showed that 4.5 to 6–year-olds can learn new words and facts from an entirely self-directed learning context, where they are listening in on complex overheard speech, rather than having their attention directed. Remarkably, children were just as good at learning four new words from overhearing as they were when these words were explicitly taught. They learned six novel facts above chance in both conditions, though they exhibited significantly higher accuracy in the Pedagogical condition. The pattern of matching-object touch results also provides preliminary evidence that children's success at self-directed learning in this context involves their ability to coordinate attention between the speech and the situational context, and that this ability increases with age. Experiment 2 followed up on this developmental trend by extending the task of Experiment 1 to a younger group of children, 3 to 4.5 years of age. Of interest was whether younger children in the Overhearing condition would

be able to meet the attentional demands of having to independently monitor ambient speech and form the appropriate referential mappings online, along with the memory demands imposed by having to learn multiple novel labels and facts. Prior studies suggest that children of this age are impressive information-seekers in other tasks and domains (e.g., Cook et al., 2011; Sim & Xu, 2017); thus, we were interested in whether younger preschoolers could succeed at an analogous task in the language domain.

Method

Participants

64 children aged 3.0 to 4.5 years participated (30 female; 3.0–4.49 years, $M = 3.83$ years, $SD = 0.45$ years). An additional thirteen children participated, but were excluded due to failing at least one familiar object trial (8), not finishing the task (3), or experimenter error (2). As in Experiment 1, participants were randomly assigned to one of two conditions, Overhearing ($n = 32$, 15 female; 3.0–4.46 years, $M = 3.81$, $SD = 0.48$) or Pedagogical ($n = 32$, 15 female; 3.05–4.49 years, $M = 3.85$, $SD = 0.43$). There was no difference in age between conditions.

Procedure

The method for Experiment 2 was identical to Experiment 1, except that the number of novel objects was reduced by one to make it more appropriate for a younger age range. Therefore, in the learning phases of both the Overhearing and Pedagogical conditions, children were exposed to three novel words and five novel facts, which still constitutes a more challenging test of learning from overheard speech than previous experiments have provided (see Table E in the Appendix). Children thus received 15 test trials in three blocks of five trials each. Each of the two word learning blocks included three critical trials and two control trials testing familiar labels (i.e., "dog" and "cup").

Results & Discussion

Comparisons to Chance

Like the older children in Experiment 1, younger children in the Overhearing and Pedagogical conditions of Experiment 2 performed above chance (20%) on fact learning (Overhearing: average 46% [37%, 56%] accuracy, $t(31) = 5.13$, $p < .001$, Cohen's $d = 0.90$; Pedagogical: 74% [66%, 82%], $t(31) = 13.14$, $p < .001$, Cohen's $d = 2.34$). However, while children in the Pedagogical condition performed above chance (33%) on word learning (51% [42%, 61%] accuracy, $t(31) = 3.41$, $p < .01$, $d = 0.61$), children in the Overhearing condition did not (30% [22%, 39%] accuracy, $t(31) = -0.74$, $p = .46$, Cohen's $d = -0.14$; see Figure 9).

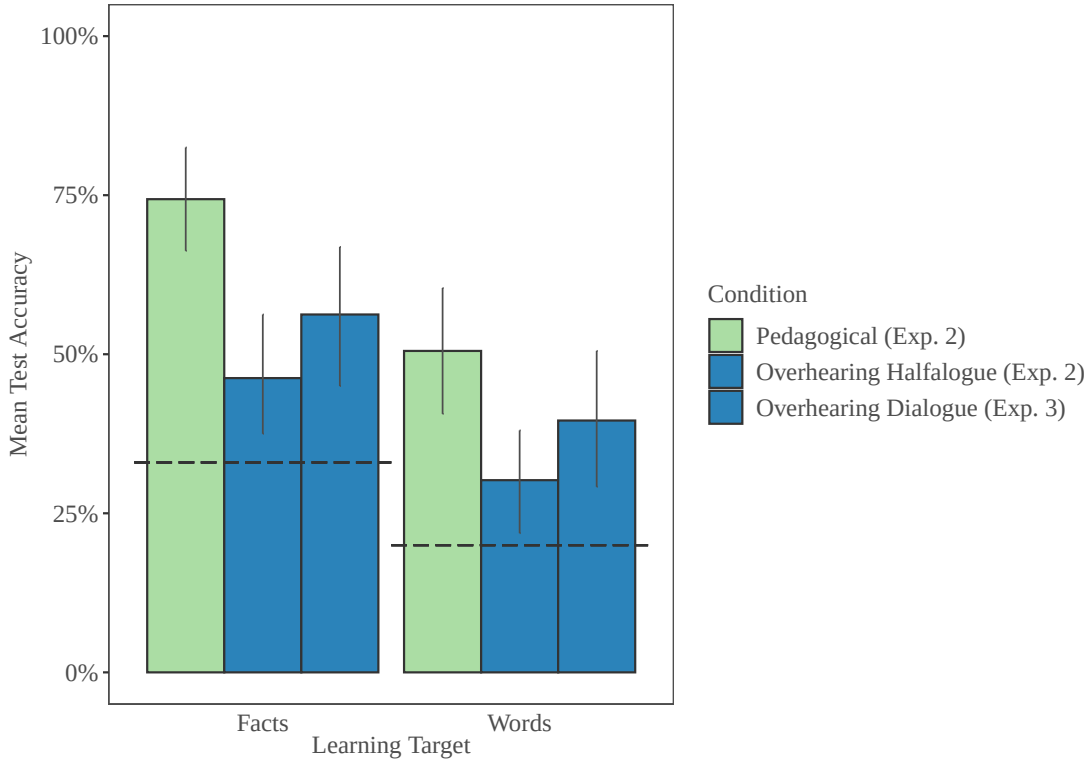


Figure 9: *Experiments 2 and 3 Mean Accuracy at Test by Learning Target and Condition.*

Note. Chance for each target type (20% for facts, and 33% for words) is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

Independent trials. Children's word-learning performance on the first test trials mirrored their performance overall. That is, children in the Pedagogical condition performed significantly above chance, estimated at 33% (47% [34%, 59%]; $t(31) = 2.1$, $p < .05$, Cohen's $d = 1.23$), while children in the Overhearing condition performed no differently from chance (39% [28%, 50%]; $t(31) = 1.1$, $p = .27$, Cohen's $d = 1.27$). As when considering averages across all trials, children's performance in both conditions exceeded chance (20%) on the first fact trials (Pedagogical condition: 72% [56%, 88%]; $t(31) = -301$, $p < .001$, Cohen's $d = 1.14$; Overhearing condition: 47% [31%, 66%]; $t(31) = -274$, $p < .001$, Cohen's $d = 0.69$).

Mixed Effects Models

Models with condition (Pedagogical or Overhearing), learning target (words or facts), an interaction between condition and learning target, and random intercepts for subject were fit to the test data. This model fit the data better than a null model comprised of only random intercepts for subjects ($\chi^2(2) = 51.62$, $p < .0001$; AIC for model with condition and target type: 888.78, AIC for null model: 936.40). In contrast to Experiment 1, children's odds of accuracy were overall lower in the Overhearing condition compared to the Pedagogical condition ($OR = 0.32$ $[0.19, 0.52]$; $\chi^2(1) = 20.21$, $p < .0001$), suggesting younger children experienced a more general advantage of pedagogical instruction. Similar to Experiment 1, children were in general more accurate at learning facts than novel words ($OR = 2.63$ $[1.88, 3.70]$, $\chi^2(1) = 31.84$, $p < .0001$). Finally, condition and learning target did not show a significant interaction ($OR = 0.67$ $[0.34, 1.32]$; $\chi^2(1) = 1.35$, $p = .25$; AIC with interaction: 889.43), suggesting that the impact of condition did not differ substantially by learning target, as it had in Experiment 1.

Object familiarity. As in Experiment 1, we analyzed children's accuracy on the fact-learning test trials to test for the effect of learning facts associated with novel, rather than familiar, objects. Also in parallel to Experiment 1, the best-fitting model included age, condition (Pedagogical versus Overhearing), object familiarity (familiar versus novel), and an interaction between condition and object familiarity (AIC for model without interaction: 397; with interaction: 393; $\chi^2(1) = 6.2$, $p = .01$). Children's odds of accuracy were lower in the Overhearing condition overall ($OR = 0.59$ $[0.28, 1.23]$, Wald $\chi^2(1) = 15.5$, $p < .001$), and children were especially bad at learning a fact associated with an unfamiliar object through overhearing (interaction $OR = 0.27$ $[0.12, 0.83]$; Wald $\chi^2(1) = 6.2$, $p < .05$). In the Overhearing condition children were on average 55% [47%, 64%] accurate for familiar object facts, compared to 49% [42%, 56%] for novel object facts; in the Pedagogical condition, accuracy was 70% [59%, 81%] and 77% [69%, 85%] for familiar and novel object facts, respectively (see Figure 10). Finally, children's odds of accuracy improved significantly with age ($OR = 2.32$ $[1.37, 4.01]$; Wald $\chi^2(1) = 6.3$, $p = .01$).

Behavioral Proxies of Attention

We coded the videos of 26 children from the Overhearing condition whose parents consented to video recording.

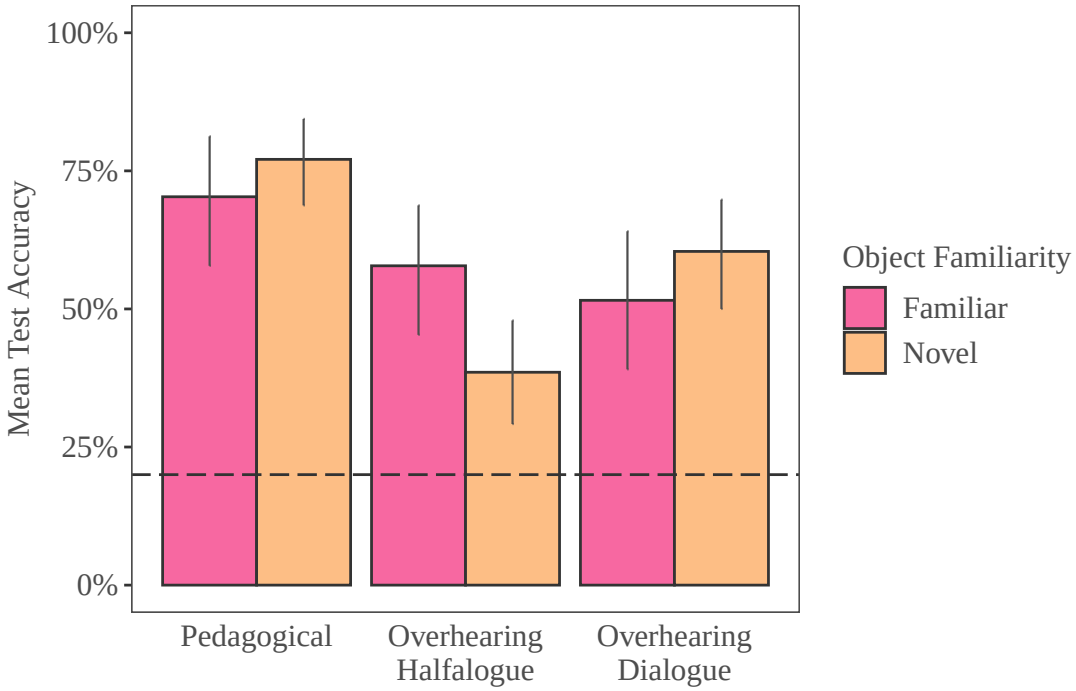


Figure 10: *Experiments 2 and 3 Mean Fact Accuracy by Familiarity of Target Object.*

Note. Chance for facts (20%) is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

Relation to call. We first tested the overall correlation between the number of video frames in which children were touching each object and that object's order of mention. If children were influenced by the experimenter's speech, they would be more likely to spend more time playing with objects that were mentioned earlier, resulting in a negative correlation. There was a significant negative correlation between total frames and order of mention ($r_{rm}(27) = -0.67$ $[-0.84, -0.39]$, $p < .001$), providing evidence that children's exploratory behavior was related to the speech they overheard.

Matching-object touch. Children in Experiment 2 received significantly lower scores on our touch measure (Range: -0.28 – 0.44, $M = 0.15$ $[0.07, 0.22]$) compared to children from the Overhearing condition of Experiment 1 ($t(52.49) = 2.30$, $p < .05$), suggesting that the younger children of Experiment 2 (Figure 12) may not have been coordinating their attention between the overheard speech and referential context as consistently as the older children of Experiment 1. Still, children generally received positive touch scores: 19 children received positive scores, five children received negative scores, and two never touched any of the objects. An exact binomial test confirmed that there was a greater proportion of children that had positive scores compared to negative or zero scores ($p < .05$), suggesting that children were indeed coordinating their attention between the overheard speech and object referents. However, children's matching-object touch scores were not correlated with their test accuracy (Pearson's $r = 0.01$ $[-0.38, 0.40]$; $t(20) = 0.05$, $p = 1$), nor were they correlated with age (Pearson's $r = 0.03$ $[-0.36, 0.42]$; $t(20) = 0.2$, $p = 0.9$). The fact that children in the Overhearing condition were at chance when tested on words despite showing a relation between their touch behavior and the content of the call raises the possibility that they may have formed some word-object mappings during the learning phase, but had difficulty retaining these mappings until the test phase of the experiment.

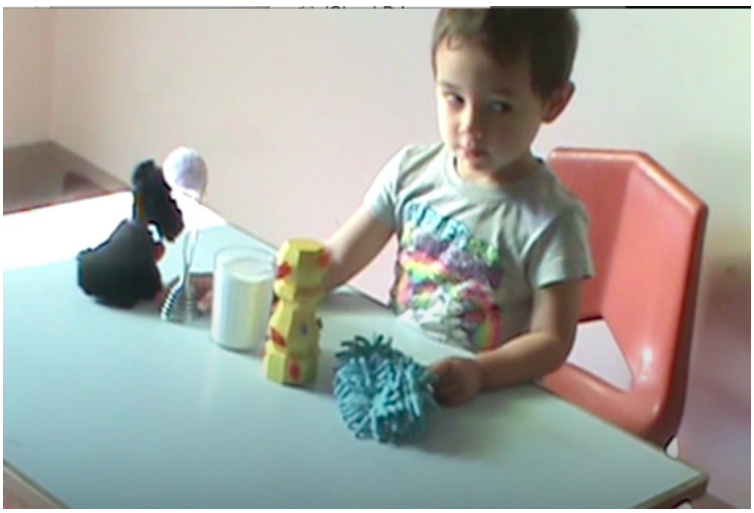


Figure 11: *A Three–Year-Old Eyes the Experimenter during Overheard Call.*

Child gaze. The children in Experiment 2 looked toward the experimenter for up to half of the duration of the phone call (Range: 0.01 – 0.49, $M = 0.16$ $[0.11, 0.21]$). Children's gaze proportions exhibited no significant correlation with their mean test trial accuracy (Pearson's $r = 0.33$ $[-0.07, 0.63]$; $t(20) = 2$, $p = .1$), nor their age (Pearson's $r = -0.08$ $[-0.45, 0.32]$; $t(20) = -0.4$, $p = .7$). These results suggest that in our experiment, merely looking frequently toward the experimenter may not be a good indicator that children have recognized the speech as relevant.

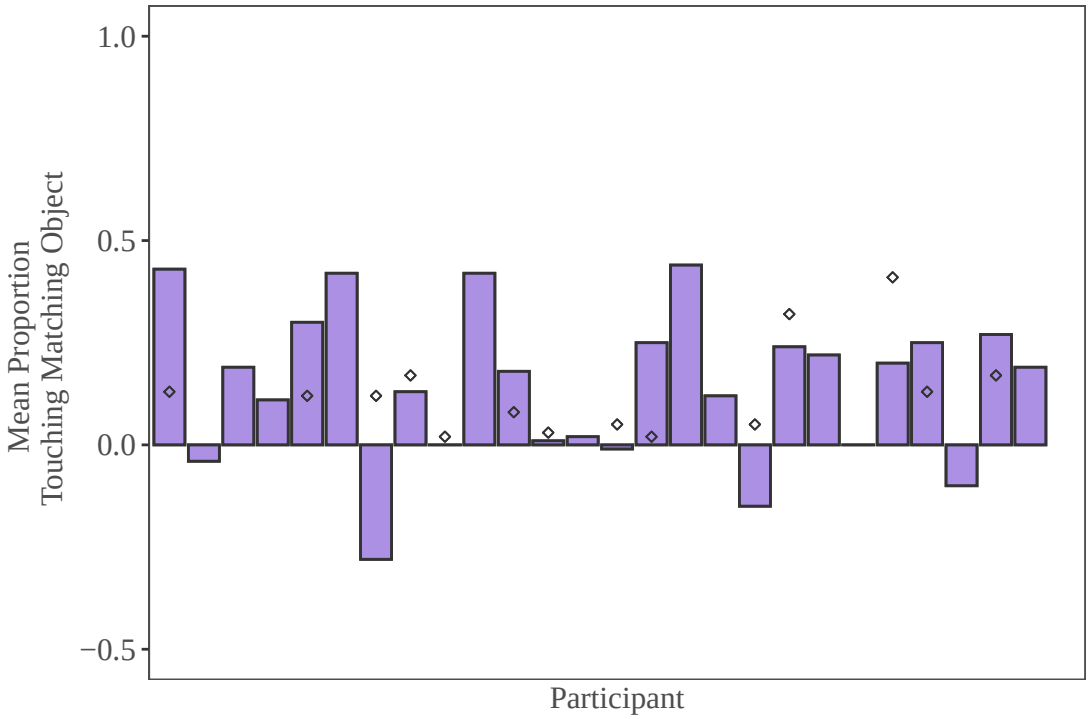


Figure 12: *Experiment 2 Matching-Object Touch and Gaze toward Experimenter.*

Note. Bars represent mean proportion of matching-object touch by participant; points indicate the proportion of the overheard call that children looked toward the experimenter.

2.5 Experiment 3

Experiment 2 found that 3– to 4.5–year-olds struggled to learn from overhearing compared to when learning targets were presented pedagogically. In contrast to the older preschoolers of Experiment 1, younger preschoolers in Experiment 2 were at chance at learning three new words in our overhearing task, though they were able to learn a set of five facts above chance. Across words and facts, younger children's performance was significantly better in the Pedagogical condition compared to the Overhearing condition. One possibility for why children had difficulty learning from overhearing in Experiment 2 is because they could only hear the experimenter's side of the phone conversation (a halfalogue). While the survey we administered to parents suggests that phone calls are frequent in many children's environments, they may be difficult for younger preschoolers to learn from.

Though no study to date speaks directly to the question of whether overheard halfalogues are more difficult to learn novel linguistic information from compared to overheard dialogues, previous research opens the possibility that the phone calls we used in Experiments 1 and 2 might have impeded children's ability to learn new words. Suggestive evidence comes from multiple sources. In one study, toddlers failed to learn a novel word taught to them in person by their mothers when the mother picked up a phone call during instruction (Reed et al., 2017). Other work has shown that adults' performance is impaired in an attention task when they simultaneously overhear a halfalogue, consistent with the idea that overheard halfalogues might be more distracting than dialogues (Emberson et al., 2010). In the context of our study, this latter finding might predict that an overheard halfalogue should be *easier* to learn from than an overheard dialogue, because it is more attention-getting; alternatively, it might predict a learning disadvantage, if a halfalogue is so attention-getting that it limits children's ability to coordinate their attention between the overheard speech and the objects. Still other studies emphasize the importance of contingent interaction in learning episodes (e.g., Roseberry et al., 2014). This perspective predicts decreased learning from an overheard halfalogue, not because children might be distracted, but because they might fail to recognize that there is an opportunity to learn at all, in the absence of a reciprocal social interaction (O'Doherty et al., 2011).

Also motivating the question of whether children are better able to learn from overheard dialogues than from halfalogues are psycholinguistic accounts which emphasize how interlocutors collaborate on meaning in conversation (Fusaroli et al., 2014; Linell, 2009; Pickering & Garrod, 2004) and imply that comprehension given only one side of a conversation should be uniquely difficult. Importantly, in contrast to halfalogues, dialogues may allow children to rely on feedback between interlocutors to establish word mappings (Tolins et al., 2017). This may be especially important for helping young learners assess whether a newly-introduced word is conventional. Backchannels may also attract children's attention, when, for example, addressees react with surprise to novel information from the speaker. For both children and adults, having access to the full process of *grounding*, or the establishment of mutual knowledge between interlocutors (Clark & Brennan, 2004; Fox Tree, 1999), is also known

to aid comprehension — even when the conversation that overhearers are listening in on is one where the addressee plays a limited role, as in listening to a story or receiving instructions (Schober & Clark, 1989; Tolins & Fox Tree, 2016). Indeed, in one study where both overheard interlocutors were visible, two-year-olds learned a novel word when the overheard addressee was visibly attentive and following along, but not when they were visibly distracted (Fitch et al., 2020).

To determine whether using an overheard halfalogue might have suppressed younger preschoolers' learning from overhearing in Experiment 2, we tested learning from a minimally different overheard dialogue in Experiment 3. We conducted the overheard conversation over speakerphone, thereby maintaining control of the speech, referential cues, and number of copresent experimenters, while transforming the halfalogue to a dialogue via a second, audible interlocutor. This context, where both sides of the conversation are audible but only one speaker is visible, happens in the real world not only on speakerphone and video chat, but also when parents are talking between rooms or over the child's head. To increase the social, reciprocal nature of the overheard call and to guard against concerns from previous work, the experimenter and caller were actively engaged with one another, periodically asking each other questions and expressing surprise (see Appendix H). If children are better at learning new information from language when this information is embedded in a reciprocal social interaction that children can access (e.g., O'Doherty et al., 2011; Roseberry et al., 2014), we expect children in Experiment 3 to demonstrate significantly greater learning than their same-age peers in Experiment 2.

Method

Participants

Participants were 32 children learning English as their primary language between 3.0 and 4.5 years of age (16 female; 3.1–4.5 years, $M = 3.8$ years, $SD = 0.4$ years). A total of four children were excluded due to failing at least one familiar label control trial (1), having already witnessed another child participate (2), or experimenter error (1). For clarity in the sections below, we distinguish between the "Overhearing Halfalogue" condition of Experiment 2, and the "Overhearing Dialogue" procedure that all children received in Experiment 3. There was no difference in the age composition of participants in these two groups ($t(70) = -0.2$, $p = .8$).

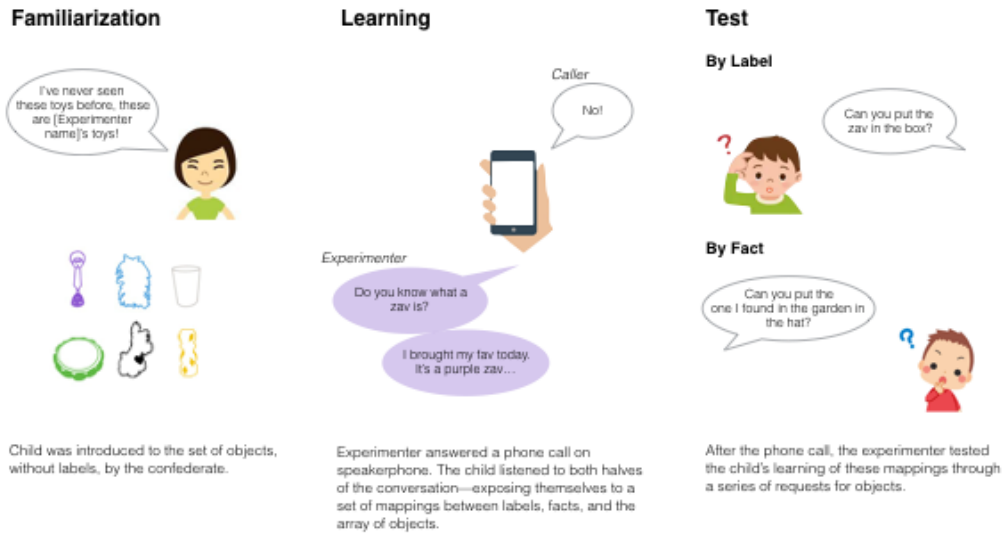


Figure 13: *Overview of Experimental Procedure for Experiment 3.*

Procedure

The Overhearing Dialogue procedure for Experiment 3 differed from the Overhearing Half-alogue procedure of Experiment 2 in that the experimenter picked up a genuine call from a caller, rather than setting a timer and pretending to have a conversation with an invisible other (see Figure 13). The caller called thirty seconds after receiving a warning text from the experimenter, and delivered scripted responses to the experimenter's speech, which was itself identical to the script in Experiment 2 (Appendix F). The experimenter, apparently busy on their laptop, put the caller on speakerphone at maximum volume, making it so that the child could hear the caller at roughly the same volume as the experimenter (see our online repository at https://osf.io/avyg5/?view_only=33cbb9ab189343a7b6e8f6c7c517026d for links to videos of this procedure stored on Databrary.org, along with experimenter scripts for all conditions).

As an additional edit to our procedure, we introduced head-mounted cameras for children to wear, having seen the value of high-quality video data for coding children's attentional behavior in Experiments 1 and 2. These videos were synced after the fact with up to two additional video recordings of the experimental session, one recorded from a tripod, and another recorded from an overhead camera. All video coding was completed using composite videos combining all three angles. The increase in video quality was reflected in

the 93% inter-rater reliability for children's touch behavior. Composite videos and coding spreadsheets can be found archived on Databrary.org (linked also in our OSF repository: https://osf.io/avyg5/?view_only=33cbb9ab189343a7b6e8f6c7c517026d).

Results & Discussion

Comparisons to Chance

Like the same-aged children in the Overhearing Halfalogue condition of Experiment 2, children in Experiment 3 performed above chance (20%) on fact learning (57% [46%, 68%] accuracy, $t(31) = 6.4$, $p < .001$, Cohen's $d = 1.13$), but not on word-learning (chance = 33%; average accuracy 39% [29%, 49%], $t(31) = 1.1$, $p = .27$, Cohen's $d = 0.20$).

Independent trials. We found a similar pattern when we analyzed the first trials as when we analyzed all trials at once: children were at chance (33%) on words (42% [30%, 55%]; $t(31) = 1.4$, $p = 0.18$, Cohen's $d = 0.24$), and above chance (20%) on facts (62% [44%, 78%]; $t(31) = -223$, $p < .001$, Cohen's $d = 0.86$).

Mixed Effects Models

We next fit a mixed effects logit model to the trial-by-trial test data (coded as incorrect = 0, correct = 1), with age and learning target type (word versus fact) as fixed effects, and random intercepts by subject. This model fit the data significantly better than a null model using only participants' own means ($\chi^2(2) = 24$, $p < .001$; AIC for null model: 478, AIC for full model: 458). Including learning target type in our model also significantly improved model fit compared to a model with only age ($\chi^2(1) = 13$, $p < .001$; AIC for model without type: 468). Children's odds of accuracy increased as they got older ($OR = 3.68$ $[1.79, 8.00]$, Wald $\chi^2(1) = 13$, $p < .001$), and, as in Experiments 1 and 2, their odds of accuracy were significantly higher for trials testing facts ($OR = 2.27$ $[1.44, 3.60]$, Wald $\chi^2(1) = 12$, $p < .001$). See Figure 14 for a visualization of age trends across experiments.

Object familiarity. In contrast to both previous experiments, a mixed effects logit model fit to the fact data alone yielded no advantage for facts associated with familiar objects over facts associated with novel objects (52% [38%, 66%] and 60% [47, 73] accuracy, respectively; see Figure 10). That is, while age was a significant predictor of fact accuracy ($OR = 8.93$ $[2.60, 40.30]$, Wald $\chi^2(1) = 11$, $p < .001$), adding object familiarity (familiar versus novel) to the model did not significantly improve fit ($\chi^2(1) = 1.7$, $p = .19$, AIC for model with age as sole fixed effect: 198, AIC for model including object familiarity: 198). This is in contrast to Experiments 1 and 2, where familiar object facts were easier to learn in the Overhearing conditions in particular. In our discussion of our previous results, we suggested that the selective advantage of familiar object facts in the Overhearing condition might reflect their relative ease of being processed in the moment, such that they could

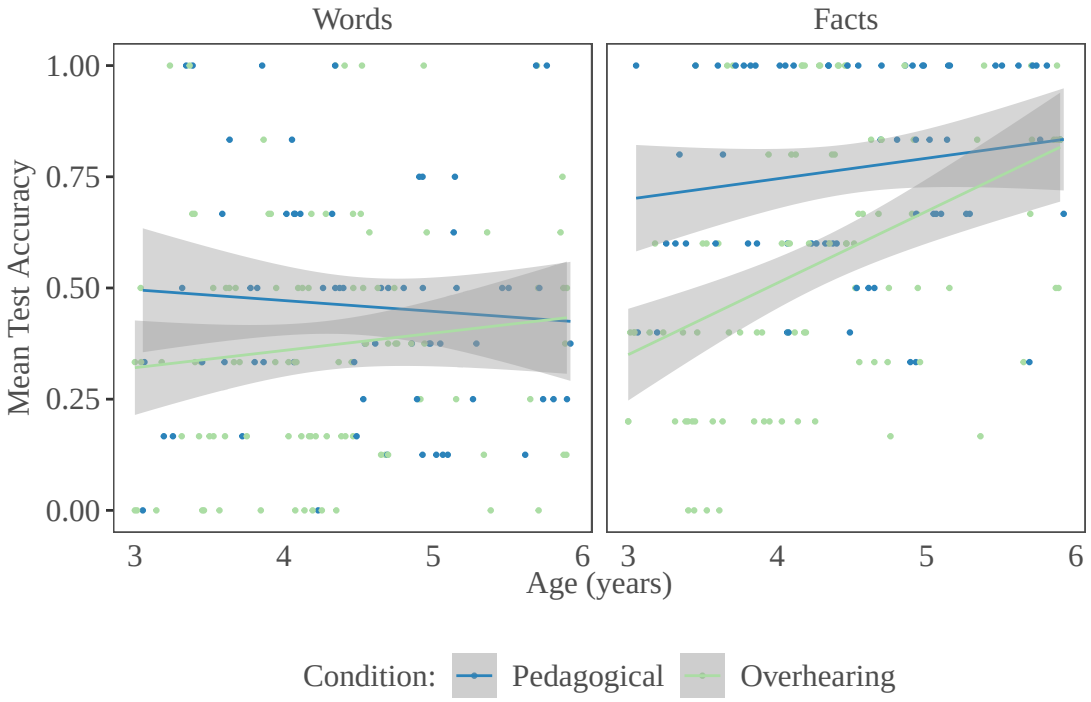


Figure 14: *Mean Accuracy by Child Age, Across Experiments 1–3.*

Note. Overhearing Halfalogue and Overhearing Dialogue conditions are combined. Shaded region indicates standard error.

be mapped to the correct referent — a task children needed to do on their own in the Overhearing, but not Pedagogical, conditions.

Behavioral Proxies of Attention

Videos from 24 participants were coded to capture behavioral proxies of children's online attention to the overheard speech.

Relation to the call. To assess whether children's pattern of object touches suggested influence from the overheard phone call, we computed the correlation between the number of video frames that children ($n = 24$) touched each object, and that object's order in the call. This correlation was significant, and in the predicted direction ($r_{rm}(71) = -0.46$ $[-0.62, -0.25]$, $p < .001$), suggesting that children's exploration of the objects was likely driven by their auditory attention to the overheard call.

Matching object touch. Like their peers in the Overhearing Halfalogue condition (Experiment 2), children in the Overhearing Dialogue condition (Experiment 3) received significantly lower scores on our touch measure (Range: -0.34 – 0.78, $M = 0.13$ $[0.04, 0.23]$) compared to older Overhearing Halfalogue participants (Experiment 1; $t(50) = 2$, $p < .05$), but equivalent scores to same-age Overhearing Halfalogue participants (Experiment 2; $t(50) = 0.3$, $p = .8$). This comparison provides further evidence that the ability to coordinate attention between overheard speech and a scene improves with age. Despite the lack of difference in the magnitude of children's scores compared to their peers in Experiment 2, in this sample, 16 children received positive touch scores, seven received negative scores, and five received scores of 0. An exact binomial test concluded that here, children were no more likely to receive positive scores than negative or zero ones ($p = .6$). That children's sequence of objects touched still correlates with the experimenter's speech suggests they were attending to the call, but the distribution of touch scores we see calls into question either our speculation that the overheard dialogue was easier to process, or our interpretation of our measure. In particular, the greater quantity of zero scores (children who never touched any object) is difficult to interpret, as comprehending the overheard speech does not necessitate touching the objects at all, merely attending to them. Consistent with this, there was no significant correlation between children's touch scores and test accuracy (Pearson's $r = 0.37$ $[-0.01, 0.65]$, $t(30) = 2$, $p = .05$) or age (Pearson's $r = 0.33$ $[-0.05, 0.63]$, $t(30) = 2$, $p = .09$).

Child gaze. Children in Experiment 3 spent variable proportions of the call looking at the experimenter (Range: 0 – 0.49, $M = 0.19$ $[0.14, 0.09]$). This variability did not significantly correlate with children's age (Pearson's $r = 0.29$ $[-0.10, 0.60]$; $t(30) = 2$, $p = .1$), nor their test performance (Pearson's $r = -0.14$ $[-0.48, 0.25]$; $t(30) = -0.7$, $p = .5$).

Comparing Experiments 2 and 3

Planned comparisons yielded no difference in test accuracy between the Overhearing Dialogue condition of Experiment 3 and the Overhearing Halfalogue condition of Experiment 2, for either words ($t(60) = 0.2$, $p = .8$, Cohen's $d = -0.05$) or facts ($t(60) = 0.5$, $p = .6$, Cohen's $d = -0.11$). To model influences on test performance across the two experiments, mixed effect logit models were fit to children's overhearing test data, with fixed effects for learning target (word or fact) and experimental condition (Overhearing Halfalogue or Overhearing Dialogue). Model parameters suggested no difference between the two experimental

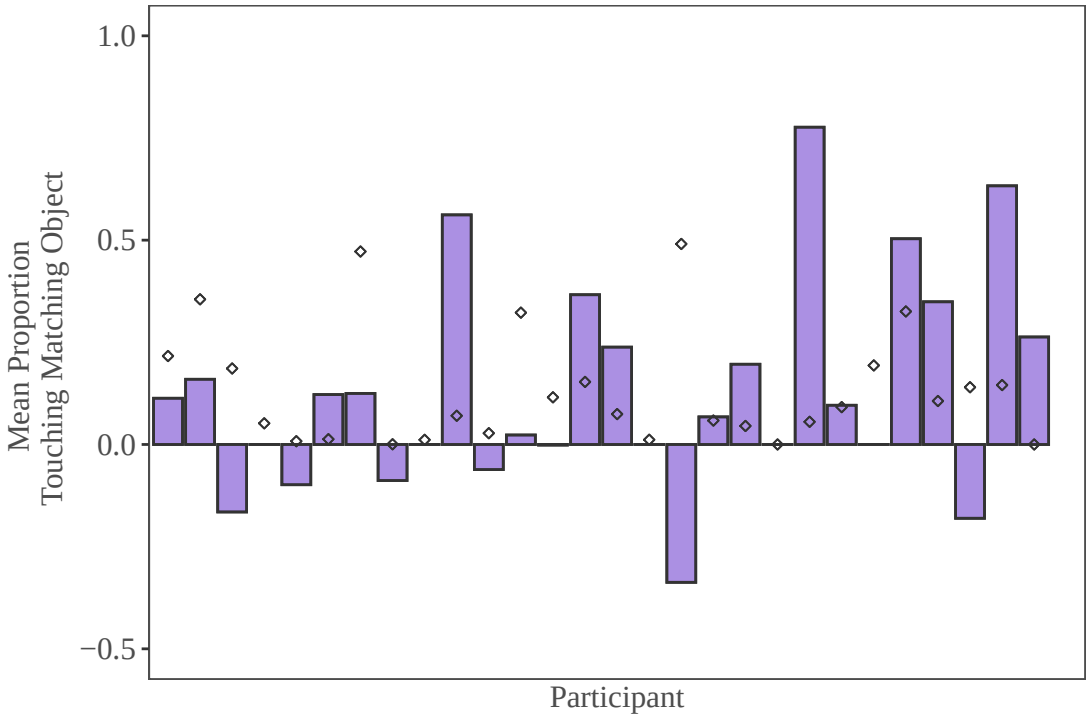


Figure 15: *Matching Object Touch and Gaze Proportion by Participant in Experiment 3.*

Note. Magnitude of computed matching-object touch score is shown in bars, proportion of phone call in which child gazed at experimenter indicated with points.

overhearing conditions ($OR = 1.54$ $[0.92, 2.62]$), but reliably better performance for facts, compared to words ($OR = 2.24$ $[1.61, 3.12]$), across experiments. Nested model comparisons showed that including experimental condition as a predictor did not significantly improve fit compared to a model with learning target as the sole fixed effect ($\chi^2(1) = 2.73$, $p = .10$). In terms of their self-directed learning behavior, children in the Overhearing Halfalogue and Dialogue conditions of Experiments 2 and 3 also did not significantly differ in their matching-touch scores ($t(50) = 0.3$, $p = .8$) or gaze proportions to the experimenter ($t(30) = -0.09$, $p = .9$).

The above results suggest that younger children's chance performance on word-learning in the Overhearing Halfalogue procedure used in Experiment 2 cannot be attributed to the halfalogue nature of the overheard speech in that study, as children in Experiment 3 could hear both sides of the dialogue. However, we conducted a final analysis of both experiments' fact-learning data alone, to follow up on the divergent pattern of results in Experiments 1 and 2 versus Experiment 3. The best-fit logit model included age, experimental condition (Overhearing Dialogue or Overhearing Halfalogue), object familiarity, and an interaction

between experimental condition and object familiarity. This model resulted in a significantly better fit than a model without the interaction ($\chi^2(1) = 6.92$, $p = .009$; AIC for model with interaction: 408, AIC for model without interaction: 413). Model coefficients suggest greater accuracy with age ($OR = 4.94$ $[2.85, 8.81]$) and lesser accuracy for novel-object facts ($OR = 0.41$ $[0.20, 0.82]$), an effect attenuated in the Overhearing Dialogue condition, specifically ($OR = 3.61$ $[1.39, 9.57]$ for the interaction of object familiarity and overhearing condition. Thus, children in the Overhearing Dialogue condition tended to outperform children in the Overhearing Halfalogue condition on facts associated with novel objects. Together, our results suggest that although the younger preschoolers in our studies were able to attend to and track the overheard speech enough to learn multiple new facts, they found it challenging to form and retain multiple novel word–object mappings via a short overhearing exposure, even when they were overhearing a dialogue.

2.6 General Discussion

The present studies tested children's ability to acquire novel words and facts from their environments in the absence of external guidance or support. Such tests of children's real-world self-directed learning are a topic of considerable current interest, but are especially underrepresented in the domain of language development, where the role of the adult caregiver directing speech to the child is often emphasized over the role of the child themselves. Our studies compared self-directed learning in a naturalistic context to learning via pedagogical instruction, across a three-year age range. In contrast to previous studies, the overhearing conditions we designed stripped away as many pedagogical cues as possible, providing a stringent test of learning from complex overheard speech. We included multiple novel words and facts, embedded in a variety of sentence frames using the pace and prosody of adult-directed speech. Additionally, we employed a real-world context of overhearing — a nearby phone conversation — that children in our sample frequently experience in their own homes (see Table S2), and which we show to have similar learning potential to an overheard dialogue where both sides are audible.

Extrapolating from the results of previous overhearing experiments — where even toddlers have been found to readily learn words in a overhearing context (Akhtar, 2005; Akhtar et al., 2001; Baldwin, 1991; Floor & Akhtar, 2006; Gampe et al., 2012; Martínez-Sussmann et al., 2011; Shneidman et al., 2009) — we might have expected the preschoolers in our experiments to be just as skilled at learning in an overhearing context as in a pedagogical one. But the overhearing context in our studies was much more demanding than in previous studies, as we aimed to provide a more stringent test of how well children may learn from complex ambient speech in their daily lives. In doing so, we provide a demonstration of children's self-directed learning with a transparent application to the real world.

Taken together, our results show a developmental progression in preschoolers' ability to pick out, map, and remember multiple novel linguistic items outside of a pedagogical interaction (Figure 14). In contrast to the findings of previous studies of overhearing in

more simplified contexts, younger preschoolers (3–4.5 years; $M = 3.8$) were at chance at learning a set of three novel words from overheard speech, though they reliably learned a set of five facts. Their performance for both words and facts improved with age. These younger preschoolers in Experiments 2 and 3 showed a significant learning boost from pedagogical instruction for both types of learning targets, relative to when the novel words and facts had to be learned by overhearing a halfalogue (Experiment 2) or dialogue (Experiment 3). While younger children's word learning did not differ from chance in the overhearing conditions, the older preschoolers in Experiment 1 (4.5–6 years; $M = 5.2$) performed above chance when learning a set of four new words from overhearing, and equivalently to when they were directly taught these words (42% and 41%, respectively), though they were better at learning new facts when these facts were introduced pedagogically (79% mean accuracy versus 64% for overhearing).

Our study endeavored to teach children more novel words and facts — especially in only about one minute of speech — than most previous studies. Even in the Pedagogical condition of Experiment 1, children may have struggled to retain four novel phonological forms that had been introduced so briefly: the overall word learning accuracy for 5-year-olds in Experiment 1 was around 40%, whereas even toddlers will succeed at around 80% when given only one novel word to learn (e.g., Floor & Akhtar, 2006; Gampe et al., 2012; Woodward et al., 1994). Remarkably, even though the younger children in Experiments 2 and 3 did not appear to successfully learn words from overhearing, they were able to learn facts, providing evidence of their ability to independently tune in to overheard speech in a relatively unsupported learning context, sans visual cues from either speaker or addressee.

Across our studies, we found that children reliably learned facts at greater rates than they did words. Children's strong performance on facts in all three experiments, and superior performance for facts corresponding to familiar objects in particular in Experiments 1 and 2, may give us insight into some of the challenges posed by learning from overheard speech more generally. Performance may have been better for facts than words, perhaps because the facts themselves consisted of familiar words, and because the facts afforded more words and familiar concepts to associate with the object description than a single novel word. The child's mapping task could have been further simplified when they were learning a fact about a familiar object, where it would be trivial to identify *which* object they should map the fact to (i.e., they didn't have to look at the objects to know which was the 'dog'). Further, the greater number of memory cues that were present for facts broadly, and for familiar facts in particular, may have made them easier to retrieve at test, relative to their single-word counterparts (Deák & Toney, 2013).

It is also possible that the learning asymmetry between words and facts derived from differences in how children encoded information about individual objects in response to hearing the words and facts used in our study. For example, children could have had more difficulty linking labels to the specific objects in our task due to their understanding of labels as naming *categories*, which could have resulted in coarser encoding of individual category members. Prior work with adults suggests that, because facts express information that is unique to each object — rather than category-level information — they may trigger more

fine-grained representations of individual objects (Lupyan, 2008, 2012). Given that each word as well as each fact in our study was only associated with one object, a further study would be needed to evaluate this hypothesis. For example, to test whether reference to individual items accounted for children's superior performance with facts, a future experiment might test the learning of facts and words that have been associated with categories of objects vs. individual exemplars.

Results from our matching-touch measure additionally suggest development in attentional components of children's self-directed learning skill, from recognizing an opportunity to fill an "information gap" (e.g., information about the novel objects before them; Loewenstein, 1994) to coordinating their attention between potential sources of new information. Our finding across experiments that children's touch behavior was correlated with the order in which the objects were mentioned in the overheard speech suggests that both younger and older preschoolers' manual exploration was influenced in real time by the content of the experimenter's call. Similarly, children's positive scores on our matching-touch measure in Experiments 1 and 2 showed that they were more likely to play with an object as it was being discussed by the experimenter, compared to when another object was being discussed. This behavior, combined with their robust learning of multiple facts, points to children's ability to coordinate their attention between overheard speech and their referential context.

While there was substantial variation in the amount children looked toward the experimenter, children's looking behavior correlated with learning in only one of our three experiments (Experiment 2), contrary to previous results (Martínez-Sussmann et al., 2011; Shneidman et al., 2009). That we didn't find such a correlation reliably might be explained via differences in the structure of the overhearing exposures we used compared to those in previous work. In previous studies, the experimenter's gaze was informative: she looked toward and interacted with the referents of the novel words while the child looked on. In our study, however, the objects were displaced from the experimenter, and she provided only descriptive cues to the referents of the novel words, avoiding looking toward the child or objects. Multiple studies show that toddlers are not only able to use speaker gaze to resolve referential ambiguity, but also actively seek it out (Baldwin, 1991; Vaish et al., 2011), suggesting that our participants' glances to the experimenter may have reflected not only attention, but also their uncertainty and consequent information-seeking (Hembacher & Frank, 2017). In an overhearing experiment testing the impact of joint attention between the overheard adult interlocutors on children's learning, two-year-olds failed to learn a novel word when the addressee was distracted and not looking at the referent objects, which the authors hypothesized reflected children's reliance on the addressee's visual perspective to map the word (Fitch et al., 2020). In a similar context, where objects were labeled without joint attention, toddlers were able to learn new word mappings only with visible focus on the objects by the speaker (Baldwin et al., 1996; Bannard & Tomasello, 2012). In the absence of that cue, toddlers could demonstrate learning in a looking, but not explicit pointing, test (Bannard & Tomasello, 2012). It may be, therefore, that younger children in our study had difficulty establishing word-to-object mappings because the experimenter (and her unseen addressee) did not look toward the objects, but would have been able to show some knowl-

edge of these mappings had we used a more implicit test of learning. Anecdotal evidence that children were looking toward the experimenter at least in part to try and resolve the referential ambiguity of her speech comes from a number of children across experiments who tried to spontaneously engage the experimenter (e.g., one child who, when the experimenter described the *dax*, held the blue object out toward her and asked, "This? You mean this little guy?").

As we mentioned in the Introduction, our findings may speak to a puzzle in the language development literature: while even toddlers have been able to learn words from overhearing in experimental settings, studies consistently find no correlation between the quantity of early overheard input in children's homes between 18 and 30 months, and their vocabulary growth six months to a year later (Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a). We suggest that the reason for the disconnect between toddlers' in-lab overhearing prowess in experimental settings on the one hand, and the lack of a correlation between naturalistic overheard speech and vocabulary growth on the other hand, may lie in the differential learning demands posed by the two types of overheard speech (see Sperry et al., 2019, for a related discussion). As noted in the Introduction, previous experimental studies have tested learning from overheard speech in ways that may have placed lesser demands on children's self-directed learning (see Table E). Compared to the overheard speech presented in previous studies, the overheard speech in children's own homes is liable to bear less resemblance to child-directed speech, to include fewer pedagogical cues, and to include many words that are unfamiliar to the child, rather than a single novel one. Because adult interlocutors "in the wild" will often share knowledge of words that are new to overhearing children, they are unlikely to consistently stress these words, embed them in labeling sentence frames, or supplement them with overt cues to the reference like eye gaze, as has been done in previous studies of learning from overhearing. These differences are likely to make naturalistic overheard speech more complex and difficult for children to learn from, such that they may not even attend to it early in development (Foushee et al., 2016; Kidd et al., 2012, 2014). Even if children do attend to overheard speech, they will often have to use the linguistic context to infer the meaning of an unfamiliar word, which will itself be difficult because the context will often be comprised of other unfamiliar words.

Of course, the complexity of naturalistic overheard speech is only one of the possible explanations for the pattern of results in the literature. As discussed in the Introduction, overheard speech as a category is likely to be much more diverse (including adult speech to other children, sibling productions, etc.), compared to the category of speech that the child receives directly. This makes the lack of correlation between the amount of overheard speech a child receives and their vocabulary growth especially difficult to interpret. In child-directed interactions, words are likely to be easier to hear and to interpret, and to be harder to ignore, by virtue of how adults tailor their input to children (e.g., Yurovsky, 2018). Data from overheard speech is likely to be noisier, and isolating what the child has learned from overheard speech is especially difficult — thus the need for experimental studies like ours and others' to complement observational studies.

Although we found that younger preschoolers did not reliably learn words via overhearing

in our task, we do not wish to imply that children of this age cannot learn language from overheard speech more generally. Our studies focused specifically on the learning of overheard concrete nouns, whose meanings depend heavily on the situational context and would benefit especially from cues like joint attention (Fitch et al., 2020) — imposing significant attentional demands in their absence. Previous work suggests that it may be possible for young children to acquire partial word meanings — falling short of mapping words to their referents — when these meanings can be inferred from the linguistic context, or acquired via passive exposure (as might be the case for aspects of the meanings of verbs, see, e.g., Arunachalam, 2013, 2016; Kline and Snedeker, 2015; Landau and Gleitman, 1985; Messenger et al., 2015; Naigles, 1990; Yuan and Fisher, 2009; and nouns, see e.g., B. Ferguson et al., 2014, 2018; Goodman et al., 2008). Further, even if young learners cannot acquire full word meanings via overhearing, attending to overheard speech may aid learners by increasing their familiarity with a new word form (e.g., learning that "tureen" is a legal English word) and providing information about a new word's semantic domain and context of use. Thus, our data leave open whether young children might construct partial word meanings from overheard speech, paving the way for future learning.

It is also important to note that our conclusions about the utility of overheard speech, and the behaviors associated with learning from overhearing, should be limited to children in this sample, in this context — urban, educated, and child-centered. In contrast to many children across the globe, our participants were likely accustomed to receiving child-directed speech, and to having their attention directed, from infancy (Ochs & Schieffelin, 1984). Regardless of where children are growing up, they need data to learn the language of their community. How children get those data will look different depending on the child-rearing and socialization practices of their community and the availability of the caretakers. Indeed, the contexts in which preschool-aged children come to learn best are partly responsive to their experiences as infants and toddlers, including whether they have had their attention directed and managed by caregivers (e.g., Yu & Smith, 2016) or have spent a large proportion of their time observing third-party interactions among other community members (Gutiérrez & Rogoff, 2003; Mastin & Vogt, 2016). In the domain of vocabulary acquisition, specifically, Mastin and Vogt (2016) found divergent results for the types of engagements that correlated with vocabulary growth for urban versus rural infants in Mozambique, based on what was familiar to them. It is possible, therefore, that we might see earlier or more robust learning from overhearing in children who habitually receive less child-directed speech, who find themselves in joint attentional interactions with adults less frequently, and/or who have more exposure to overheard speech. Indeed, Shneidman and colleagues 2009 found that children who had more practice overhearing at home exhibited distinct patterns of attention during an experimental overhearing exposure, and performed better at test (see also, Correa-Chávez & Rogoff, 2009).

2.7 Conclusion

To conclude, the current experiments make several important contributions to the study of self-directed learning and language development. We show first that preschoolers can learn a substantial amount of linguistic information via naturalistic overheard speech, without their attention being guided by an adult pedagogue. However, their ability to do so is developing during this period, and children's success may depend on the degree to which they need to coordinate attention to the extralinguistic context (as opposed to the speech alone), the availability of referential cues, the child's existing vocabulary, as well as their skill at tracking the speech online and retaining novel phonological forms in memory. While the experimenter in the Pedagogical condition — and likely adults in general when they speak to children — sought to maintain children's attention and reduce referential ambiguity, in overhearing contexts, children must manage their attention themselves, arguably a domain-general learning skill. With respect to the conflict between previous results in the experimental versus correlational overhearing literatures, our study suggests that children may not show evidence of regularly acquiring vocabulary from the overheard speech in their own homes during the first few years of life in part because they are still developing the requisite attentional and linguistic abilities to learn words from overhearing. Future studies are needed to enrich our understanding of the role children themselves play in their own language development, as their self-directed learning abilities evolve.

Chapter 3

Selective Attention Based on Speech Complexity and Learning Rate

Abstract

How, with people talking around them all the time, do children decide what speech to tune into and learn from? Previous studies have shown that children manage their rate of information absorption by selectively attending to stimuli at an intermediate level of complexity — neither too hard, nor too easy. To our knowledge, the present study is the first to investigate the effects of *spoken language complexity* on children's selective attention. Preschoolers (4–6 years) watched a video where the illustration for each page of a children's picture book was displayed alongside a distracting animation. The audio for each page of the picture book was looped such that the story would progress faster if the child looked at the distractor image for an extended period of time — indicating their loss of attention towards the story, and cutting the narration short. The linguistic complexity of the storybook narration was manipulated in two between-subjects conditions, such that children listening to the Simple narration heard largely familiar words, while children in the Complex condition heard multiple words typically acquired later in development. Participants' listening comprehension and word knowledge was tested after the story. While learning did not differ significantly between the two conditions, participants who listened to the Simple narration exhibited numerically greater visual attention to the story illustration, lesser visual attention to the distractor, and longer listening times to the narration itself. Importantly, these indices were significantly related to children's overall learning at test, controlling for condition and age — statistically obviating the possibility that differences in attention owe to superficial differences in the two speech streams. Our results open the possibility that young children may actively direct their attention toward linguistic input that is most appropriate for their current level of cognitive and linguistic development.

> It is the children between five and seven who are the word-lovers. It is they who show a predisposition toward such study... And they may be entirely carried away by their ecstatic, their tireless interest in the parts of speech.
>
> *M. Montessori, 1946*

3.1 Introduction

How do we decide what in the world to pay attention to, and which learning opportunities we should pursue, versus pass up? In adulthood, our choices about where to focus our energies as learners are often conscious, and driven by our own self-assessed level of competence. Inspired to expand our origami skills, or faced with a broken kitchen sink, we might skim YouTube to find channels whose folding or plumbing tutorials fit our levels of folding or plumbing expertise. Similarly, we might challenge ourselves to watch a television show in a foreign language that we want to speak — but stick to soap operas with predictable dialogue, rather than the jargon-filled legal dramas we might watch in our native tongue. College undergraduates, for their part, often enroll in excess courses at the beginning of the semester, then drop the ones that they predict will be either too much work, or redundant with their previous coursework. One way of thinking about what the amateur plumber, second language learner, and college undergraduate are doing in these examples is identifying where their learning will be the most *efficient*. If they select material that is too difficult, they risk experiencing the frustration of there not being enough information to learn much at all. If they stick to material that is too easy, they risk becoming bored to the point of missing what little there might be for them to learn. These examples illustrate the importance of broadly sampling possible sources of information, and of the crucial but counterintuitive learning skill of *giving up* when one's time and cognitive resources could be expended elsewhere with greater reward.

Though likely implicit in early childhood, this process of sampling available sources of information and monitoring one's learning efficiency is arguably especially important for young children, whose cognitive resources and real-world knowledge are limited relative to adults' (Balcomb & Gerken, 2008; Butterfield et al., 1988; Gottlieb et al., 2013; Roebers, 2017). While adults already possess most of the knowledge needed to achieve their everyday goals, children are fairly regularly stumped by the physical world, and are active studies of social world dynamics (Buchsbaum et al., 2012; Piaget, 1954). They are apparently often frustrated by their inability to communicate their inner lives to others, which suggests constraints on their language abilities that are likely to also constrain their day-to-day spoken language comprehension. In the midst of so many learning projects, children would be wise to deploy their attention conservatively — that is, to the problems where they can make the most headway at any given moment. In line with this computational-level goal (Marr, 1982),

rational learning accounts suggest that children's tendency to get bored or 'give up' can be understood as a consequence of optimizing their rate of information absorption (Gerken et al., 2011; Gottlieb et al., 2013; Kidd et al., 2012, 2014). In fact, experimental paradigms in infancy hinge on variants of this observation, such that the duration of infants' sustained attention to experimentally manipulated stimuli and/or infants' boredom across stimulus presentations is regularly used to infer the state of their default expectations of the world (Oakes, 2010). But what does the rational deployment of attention actually look like in practice?

Surveying children's behavior with this question in mind, the myriad demonstrations of children's short attention spans might seem unremarkable, while what suddenly stand out are the contexts where children show seemingly limitless patience for the same learning material. For example, children will often request to be read the same story again and again, even immediately after having just heard it. In our study, we capitalize on this phenomenon to test children's rational attention and learning in an ecologically valid context. In particular, by using naturalistic language stimuli, our study is designed to speak to how children manage their attention in the course of a particularly daunting learning task: developing language. Notably, except in rare contexts, the challenge posed to language learners is not an absence of relevant data (Goldin-Meadow, 2015); instead, children are often surrounded by diverse language sources, from the speech that they receive directly from caregivers, to speech from television and other media, to overheard conversations between family members and strangers, to speech directed to other children or animals, occurring in the home, or in the classroom... In the face of such diversity, how do children decide when to 'tune in,' and when to 'tune out'?

In the service of answering this question, the present work tests whether children's sustained attention to *naturalistic spoken language* is responsive to its complexity — operationalized in terms of words' familiarity, and therefore how difficult the speech as a whole is to process. In contrast to previous research with infants, we use natural language stimuli, which both interests children and carries real information for learning. This expansion on previous methods enables us to investigate how children's attention might be driven by their sense of learning — and how their learning might reflect their attention (Balcomb & Gerken, 2008; Houston-Price & Nakai, 2004). We manipulate speech complexity in our study in a straightforward way, by manipulating the relative age of acquisition (Bonin et al., 2001; Ghyselinck, Lewis, et al., 2004; Izura & Ellis, 2002; Kuperman et al., 2012; Morrison & Ellis, 2000) of the words that a speaker uses to narrate a textless picture book.

To measure children's sustained attention to the speech, we track children's gaze to competing visual targets: the Illustration corresponds to the ongoing narrative, while the Distractor functions as a continuous lure for children's attention. To capture children 'giving up,' we make the end of each trial contingent on children's fixation on the Distractor, such that distracted children listened to the story for overall less time than children who were inferred to be attentive. Finally, to measure learning, we test children's plot and word knowledge using previously established test stimuli (Foushee, Srinivasan, & Xu, unpublished manuscript) following the narration of the story. If we find evidence that children's

auditory attention is responsive to the relative 'complexity' of the speech stream, then we might expect attention to be a gating mechanism for children's learning from the different sources of spoken language in their environments across development. Specifically, children may only learn from more complex sources of language — like overheard conversations between adult caregivers or radio news broadcasts — later in development, when children's linguistic sophistication renders the language less subjectively complex.

In what follows, we briefly review literature relating stimulus complexity, child attention, and learning — in particular as they might inform predictions about language development — before describing the current study in greater detail.

Background

Empirical research in cognitive development is founded on the assumption that infants' attention to experimental stimuli is driven by a principled comparison to what they already know or have encoded (Aslin, 2007; Sokolov, 1966). This research comes out of an older literature, originating with the observation that the strength of organisms' reactions to stimuli decreased with repeated exposures (Fantz, 1964; see Colombo and Mitchell, 2009 for a review). Capitalizing on this observation, experimental paradigms designed for small human organisms expose infants to the same stimulus over and over, until it no longer holds their interest: they have habituated (Fantz, 1964; Oakes, 2010). Researchers use patterns of *dishabituation*, or reawakening of interest, to novel events presented in later trials to make inferences about the structure of infants' knowledge (e.g., Baillargeon et al., 1985; Cohen & Strauss, 1979; Spelke, 1994), based on their sensitivity to experimentally manipulated variation along conceptually relevant dimensions (e.g., category membership, Waxman and Markow, 1995; spatial relation, Hespos and Spelke, 2004; and numerosity, Brannon et al., 2004). An increase in looking time to an array of 4 dots, say, when habituated to an array with 2, suggests that infants are sensitive to number, while continuing to show habituation would suggest that infants cannot tell the two numerosities apart. Thus, one context in which children *disattend* to a stimulus is when it is beyond their capacity or already (over-) familiar (O'Connor, 1962), and one context in which they *attend* is when it presents something interestingly different or new to be learned (see Oakes, 2010, for a review).

Rather than use habituation as a discrete signal of discrimination or preference, we draw inspiration from studies that examine the habituation response *itself* as a meaningful indicator of a learner's ongoing processing of a stimulus (Colombo & Mitchell, 2009; Lovibond, 1969; Maltzman & Mandell, 1968; Orr & Stern, 1970). For example, in infant-controlled procedures, infants' exposure to a stimulus is directly related to their continued visual regard, such that trial durations vary with the interest that infants exhibit (Horowitz et al., 1972). Here, we follow recently revived work in this tradition, which independently defines the complexity of different stimuli — irrespective of experimental participants' knowledge or experience — and measures the duration of participants' attention in response (Caron & Caron, 1969; Kidd et al., 2012, 2014; R. M. Martin, 1975; Thomas, 1965).

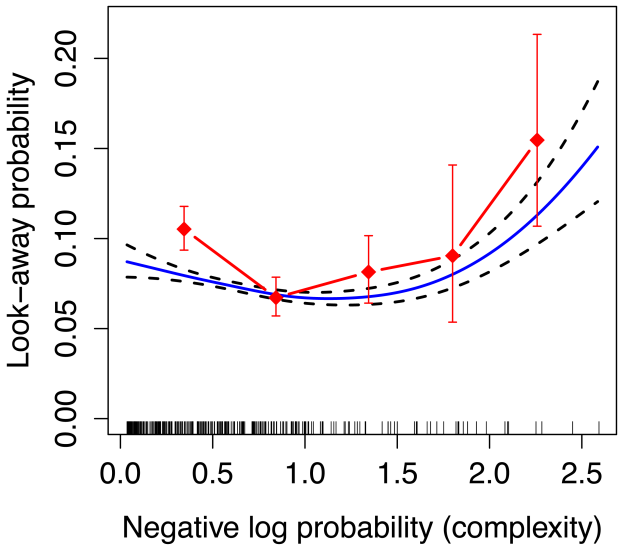


Figure 16: *Figure 3 Reprinted from Kidd, Piantadosi, & Aslin (2012).*

Previous work in developmental psychology suggests that infants' attention is both principled and personalized, such that it is *triggered* by the presence of novel information, but *sustained* by infants' ongoing sense that they are still learning (Gerken et al., 2011; J. T. Hart, 1965; Houston-Price & Nakai, 2004; Hunter & Ames, 1988). In a 2012 study, Kidd and colleagues played simple sequences of visual events for 8-month-old infants, and measured infants' duration of attention in response. They defined the complexity of individual events in terms of their predictability, based on the preceding event sequence: highly predictable events (e.g., event A after the sequence A-A-A-A) represented the lower bound of complexity, while highly unpredictable events (e.g., event B after the sequence A-A-A-A) received high complexity scores. Figure 16 plots what the authors termed the "Goldilocks effect," wherein infants' probability of 'giving up' — here, looking away from the visual display — was lowest for events of *intermediate* complexity. Infants showed the same preference for "just right" auditory stimuli in a 2014 conceptual replication.

In a related study, Gerken and colleagues (2011) exposed 17-month-old infants to artificial grammars that varied in their learnability. There, infants took less time to habituate to stimuli that represented an unlearnable blend of grammatical gender markings than to stimuli whose pattern of morphological markers was principled. More interestingly, infant looking times differentiated two stimuli that were theoretically both learnable. Infants took longer to habituate to the stimuli that was, in the author's terms, *subjectively learnable* (that is, whose pattern same-age infants and adults were able to extract given equivalent exposure in previous studies), compared to stimuli that was *objectively*, but not *subjectively*, learnable. The latter sort of stimuli could be explained by an objectively learnable set of morphological rules, but in fact had not led to learning in previous participant samples. Gerken and colleagues propose a causal relation between learnability and attention, wherein infants im-

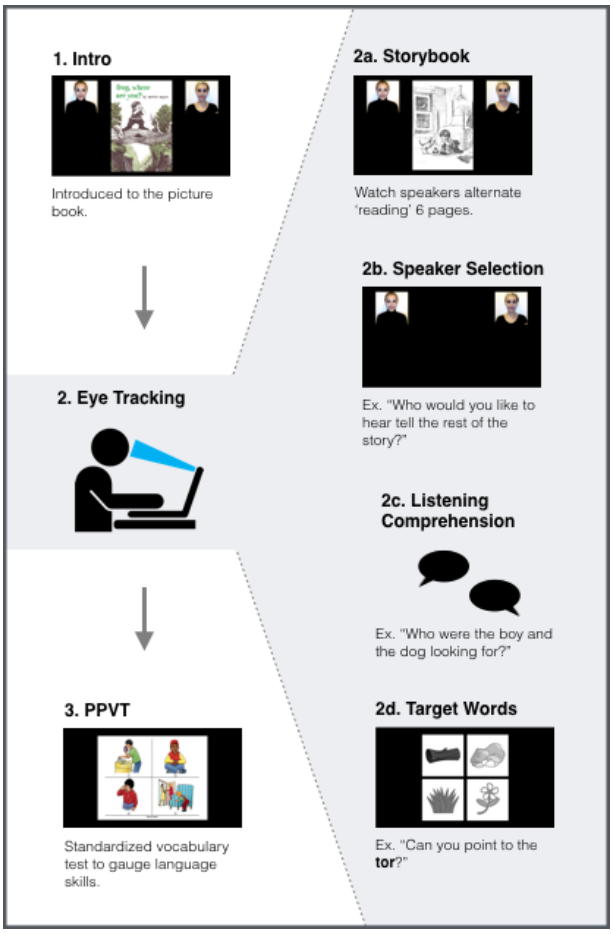
plicitly monitor their own rates of learning, and disattend when their learning rate is below some threshold of efficiency. Such a mechanism could account for the proposed U-shaped function of the "Goldilocks effect" (Figure 16; Kidd et al., 2012, 2014), given that both 'too simple' and 'too complex' stimuli would result in inefficient learning. Importantly, while learning is implicated as the underlying motivator of infants' sustained attention, or probability of 'giving up,' studies of infant attention typically have not directly tested learning within the same experiment.

Such studies provide compelling accounts of learner behavior; but do researchers' observations of infants' rational preference for semi-predictable tone sequences, or learnable toy grammars, generalize beyond the lab? With development comes new opportunities to test children's looking and listening behavior in the face of competing demands on their attention, and with measurable consequences for learning. In particular, language development offers us a real-world test domain for these ideas, as potential sources of language knowledge will naturally vary in how appropriate they are for children's current levels of competence. The work reviewed above hints that children may implicitly monitor the relative complexity or learnability of the language in their environments. If they also budget their attentional resources, then we might expect children to ignore speech that is audible and nearby, but contains too many words they don't understand, or to 'tune in' to a more accessible overheard conversation over one that is less so. This pattern of attention might help account for why only some sources of linguistic input have been shown to be reliably useful for infants' and toddlers' vocabulary development (Hoff & Naigles, 2002). Extending this work to naturalistic spoken language and to older children also has the advantage of enabling us to test learning directly, rather than infer its correlation with children's self-directed attention.

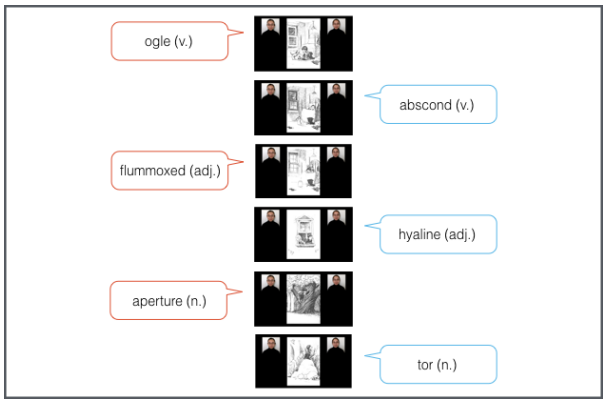
3.2 The Present Study

The present study continues from two prior experiments in this vein. Inspired by 'tandem reading' by teachers and librarians at story hour, these prior experiments presented children across a broad age range (2.5–6.5 years) with videos of two speakers alternating narrating pages of a common textless picture book (see Figure 17). The Simple speaker primarily used words estimated to be familiar to the child, while the Complex speaker used unfamiliar, later-acquired words. In order to test the consequences of the two levels of complexity, each page introduced an unfamiliar target word, embedded in the speech of the narrator for that page. That is, for the children who heard the Simple speaker narrate the first page, the unfamiliar word *ogle* was otherwise surrounded by familiar, early-acquired words. For the children who heard the same page narrated by the Complex speaker, *ogle* co-occurred with words like *companions*, *frolicked*, *attention*, and *amused* (see Table 6).

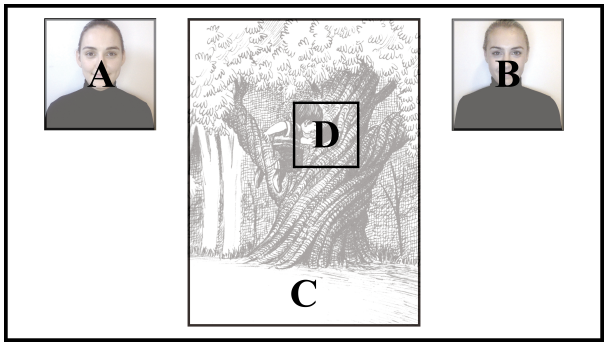
In the preceding two experiments, children's discrimination and preference between complexity levels was assessed via explicit questions about the contrasting speakers (e.g., "Who would you like to hear tell the end of the story?"). Children's explicit responses showed no relation to their age, vocabulary size, or learning from the story. However, one detail of the



(a) *Overview of Previous Experimental Methodology.*



(b) *Target Words Included to Trace Learning from* Simple *versus* Complex *Speaker.*



(c) *Critical Visual Stimulus Regions, Including the Referent of the Target Word [D].*

Figure 17: *Overview of Preceding Experiments (Foushee, Srinivasan, & Xu, unpublished manuscript).*

data suggested that our manipulation of language complexity had not been entirely ineffectual. While children were unable to express an explicit preference for one speaker over the other, analyses of their gaze behavior during the narration suggested that they more readily processed the Simple speech. Specifically, when listening to a storybook page narrated by the Simple, rather than the Complex, speaker, children spent more time looking not just at the story illustration, but at the referents of especially difficult target words embedded in both speakers' narration ($\chi^2(1) = 5.54, p < .05$; see *Supplemental Online Materials* for complete descriptions of these procedures and results).

In the present study, we make two major alterations to our procedure to more precisely target the question of whether children's attention and learning are responsive to speech complexity. First, rather than expose children to stimuli of serially alternating complexity,

we employ a between-subjects design, such that participants hear the entire story narrated at a single complexity level. Second, on the hypothesis that, due to the embedded target words, both speech streams in the previous experiments were excessively complex — and therefore challenging for children to discriminate — we further simplify the Simple speech to amplify the contrast between conditions (see Table 6). We compare this new Simple speech in one condition to the Complex speech from the previous experiment, which we know to be difficult for children of this age. Third, we test a narrower age range, focusing on children for whom we have a strong expectation that one of the levels of complexity will be *too complex*, and the other closer to 'just right.' Together, these changes allow us to make stronger predictions about the patterns of attention and comparative learning that we expect. Finally, rather than use an explicit test of children's preference between speakers or complexity levels, we rely on their visual attention to images that are 'boring' but match the story they could be following, versus to a dynamic animation that is exciting, but does not. In this way, we also bring our method closer to previous paradigms, which manipulated stimulus complexity and measured infants' probability of looking away in response.

Table 6: *Sample Passages at Three Levels of Complexity*

SIMPLE† *all familiar words*	SIMPLE† + *unfamiliar target words*	COMPLEX†‡ + *unfamiliar target words*
Once, a boy and his dog were good friends. They liked to play all day.	Once, a boy and his dog were good friends. They liked to play all day.	Once there lived two companions. They frolicked together.
This night, they were looking at the frog they caught. The boy looked at him from his chair while the dog put his nose in the frog's jar.	This night, they were ogling the frog they caught. The boy ogled him from his chair while the dog put his nose in the frog's jar.	This night, they were ogling the frog they caught. The boy ogled him from his chair while the dog put his nose in the frog's jar.
The frog smiled up at them.	The frog smiled up at them.	Their attention amused him.

† Contrast tested in previous experiments (both speakers introduce unfamiliar target words).
‡ Contrast tested in present study (only COMPLEX speech contains unfamiliar target words).

3.3 Method

Participants

Forty-six preschool-aged children (17 females, $M_{\text{age}} = 4.6$ years, $SD_{\text{age}} = 0.47$ years) who spoke English as their primary language participated in this study.[1] Participants were recruited from preschools and children's museums in the Bay Area, or from a database of interested families maintained by the University of California developmental laboratories. Prior to their study session, participants were randomly assigned to one of two conditions: Simple ($n = 24$, 10 female; 4.1–6.0 years, $M_{\text{age}} = 4.6$, $SD_{\text{age}} = 0.54$) or Complex ($n = 22$, 7 females; 4.0–5.7 years, $M_{\text{age}} = 4.6$, $SD_{\text{age}} = 0.41$). There was no significant difference in age between conditions ($p = .9$). Two additional children participated, but were excluded after another child (1) or teacher (1) intervened on their study session. Children were tested in the lab or a quiet area of the school or museum, and those participating outside of a school setting received a small gift for their participation. When present, caregivers completed a comprehensive language questionnaire and child vocabulary survey during the session.

Stimuli & Procedure

Children sat at a table before a laptop and mounted SMI RED-n eyetracker. To introduce the paradigm, the experimenter first flipped through a hard copy of the textless children's book, *Frog, Where are You?* (Mayer, 1969) to show the child that the book contained pictures, but no text. The experimenter explained that the child was going to see the pictures for the story on the screen, and hear it narrated through their headphones. After calibrating the eyetracker — which recorded the child's gaze throughout the remainder of the experiment — children donned the headphones, and were familiarized with the split-screen display that they would see for the duration of the story.

Familiarization

The first screen of the experiment displayed the Distractor, a black-and-white animation of penguins jumping rope, on the left side of the screen against an otherwise black background. The screen lasted for ten seconds, during which a female voice drew the child's attention to the ongoing animation, and encouraged them to look there "if the story gets boring" (see Appendix J). Next, the cover of the book appeared alongside the Distractor (Figure 18). Both images were displayed for fifteen seconds, during which the voiceover reiterated that the child was going to hear a story, and again directed the child's attention to the Distractor ("Where are you going to look if the story gets boring?"). The familiarization phase ended with a looming fixation cross on a grey background, used to center children's gaze before the onset of the narration — and critical data collection — phase.

[1] Data collection was cut short due to the COVID-19 pandemic.

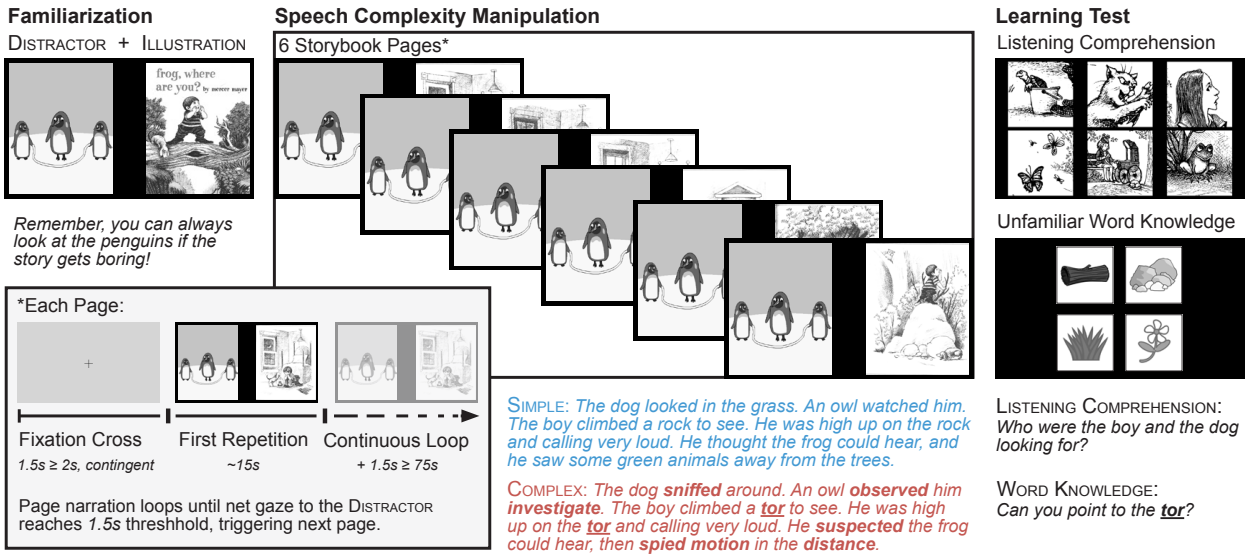


Figure 18: *Schematic of Experimental Eyetracking Procedure.*

Storybook Presentation

The content of the audio differed depending on children's assigned condition (see *Speech Complexity Manipulation*, below, and Appendix J for full script). The visual stimuli did not differ between conditions: for each of the six pages, the distractor animation played continuously on the lefthand side of the screen, while the illustration for that page occupied the right half. A fixation cross appeared before each page to re-establish children's visual attention to the display, ensuring high quality gaze data throughout the duration of the experiment. The storybook presentation lasted between 3.5 and 14 minutes.

Speech complexity manipulation. Depending on the condition to which children were assigned, the split-screen images were accompanied by either Simple or Complex audio narration (Table 6). The speech accompanying each page was matched between conditions on number of syllables (50) and sentences (5), speech rate (approximately 3 syllables/second), and lexical diversity (see *Supplemental Online Materials*). Holding these variables constant, we manipulated speech complexity via the estimated age of acquisition of the content words used in the two conditions. All of the words used by the Simple speaker appeared on the MacArthur-Bates Communicative Development Inventory (M-CDI), a standardized parental report vocabulary measure normed for use with children 18–30 months in age (Fenson et al., 2007). In contrast, each page narrated by the Complex speaker included seven words that do not appear on the M-CDI, and are estimated to be acquired after age 7 (Kuperman et al., 2012). As in our previous experiments, two of the later-acquired words used by the Complex speaker were unfamiliar words that provided a secondary trace of learning tested in the final phase of the study.

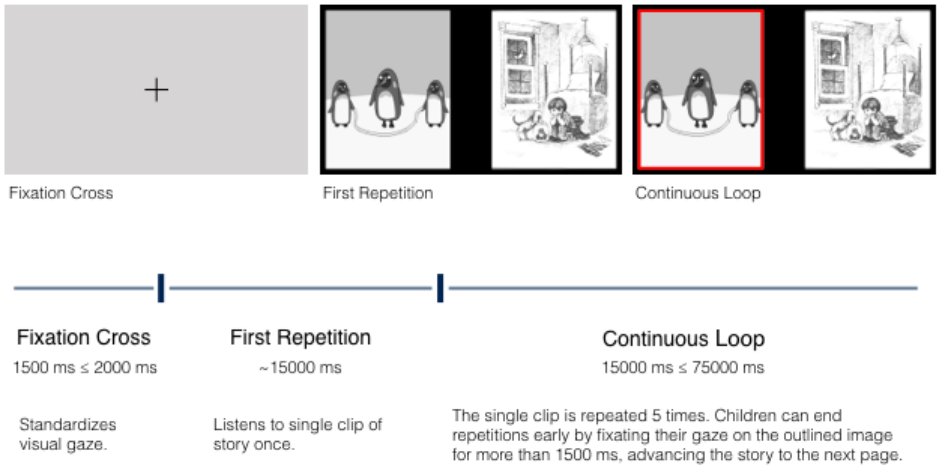


Figure 19: *Timeline of a Single Trial.*

Note. Children could end the looping audio and move on to the next page by switching their attention to the DISTRACTOR.

Unfamiliar target words. On each page, the third and fourth sentences in the Complex audio contained an embedded *unfamiliar target word*, intended to provide an additional trace of learning. The unfamiliar target words replaced semantically related words that occurred in identical sentential contexts in the Simple speaker's narration (e.g., in the Complex condition, the boy looked into an *aperture* in a tree, rather than a "hole"). The set of six unfamiliar target words included two nouns (*aperture*, *tor*), two verbs (*ogle*, *abscond*), and two adjectives (*flummoxed*, *hyaline*), and were estimated to be acquired after age 14, according to the same estimates used to norm the rest of the Complex narration (Kuperman et al., 2012). A vocabulary survey administered to caregivers ensured that these words were indeed novel: 0% of caregivers reported that they were familiar to their children (and many reported having learned them from the study themselves).

Participant-controlled trial duration. The story was divided into six pages, or trials. On each, children were forced to hear the narration for that page of the storybook at least once ($\sim 15s$). After the first repetition of the narration for that page, the same audio narration continued to loop for a total of $75s$, or approximately five further repetitions. Children could cause the story to advance to the next page before hearing all five further audio loops by disattending to the storybook illustration and switching to the GIF, programmed as a 'trigger' region on the eyetracker. Once children had fixated continuously on the distractor GIF for $1.5s$, the story would automatically advance to the next page, starting the cycle once more with a new looping audio narration and accompanying illustration.

Learning Tests

After listening to the entire story, we measured participants' learning via two blocks of test trials. In each block, children demonstrated their recall of the plot and unfamiliar target words by selecting from arrays of images in the same visual style as the preceding story. Each block began with a practice trial to familiarize children with the format of the 'game,' followed by six test trials. Only one child failed to answer the practice trial correctly.

Listening comprehension. Listening comprehension trials tested children's knowledge of events or characters in the story (e.g., "Who were the boy and the dog looking for?"; see Appendix K). Questions were presented in the same order across participants. On each trial, the comprehension question played over a fixation cross against a grey background, after which the display switched to a 3x2 grid of candidate images. All images were by the author–illustrator of *Frog, Where are You?* (Mayer, 1969).

Word learning. While only children in the Complex condition heard the set of unfamiliar target words in listening to the story narration, children in both conditions participated in the word-learning test. On word-learning trials, participants selected from 2x2 grids of black-and-white images in response to a test question (e.g., "Can you point to the one who is *ogling* something?"; see Appendix L). Unlike the listening comprehension trials, word-learning trials required children to generalize the word meaning from its occurrence in the story to unrelated contexts (e.g., from the boy *ogling* the frog in the jar, to a man peering at something through a magnifying glass). Competitor images were selected so that the syntax of the test question did not disambiguate the correct picture, and the correct response for all questions was normed via a sample of undergraduates exposed to the same story narration ($N = 19$).

Coding and Analysis

This experiment provided us with a rich dataset with which to explore the relations between complexity, attention, and learning from naturalistic speech stimuli. In the following subsections, we introduce the primary measures we use, and describe our analysis approach and predictions. Full documentation of our experimental and data processing procedures, raw data, and scripts for analyses can be found at https://osf.io/zsjfb/?view_only=7ae1b045dd774d4db618c2b8735ac148. Study session videos are available on Databrary.org (linked in the preceding repository), for viewing by registered users at the access level permitted by each participating family.

Voluntary Trial Duration

Our first measure of children's attention reflected the duration that participants spent on each page of the storybook, above and beyond (a) the obligatory first listen to the page narration ($\sim 15s$), and (b) the period required to trigger the following page ($1.5s$). This measurement reflects the child's degree of attention in that by default, the narration corresponding to each page played in a continuous loop for up to $1.25min$ ($75s$) following the child's first non-optional listen. The child could trigger the stimulus to move on to the next page *early* (i.e., before hearing any additional repetitions of the page) by fixating for $1.5s$ on the Distractor. *Voluntary trial duration* was calculated by subtracting the duration of the first obligatory page repetition, plus the $1.5s$ trigger duration, from the total duration of the trial. In addition to by-trial values (6 per participant), we sum each child's voluntary trial durations to obtain a single summary value for each participant — we refer to this value as *cumulative listening time*.

Areas of Interest

We defined two critical Areas of Interest (AOIs) during the eye tracking portion of this study: the Illustration (right on-screen image; Figure 19) and the Distractor (left on-screen image). We analyze two metrics of children's visual attention to each of the above AOIs on each trial, as well as two summary metrics of children's visual attention to each AOI over the course of the experiment.

Net dwell time. *Net dwell time* reflects the total time during which a participant's gaze was detectable by the eye tracker and fixated on a given AOI. Thus, this measure combines information about both the length of children's exposure to the story and the distribution of their attention while listening. We analyze these millisecond values for the ILLUSTRATION and DISTRACTOR across the full duration of each trial for each participant, for a total of 12 data points per participant (two AOIs, six pages).[2] As with trial durations, we sum AOI net dwell times (i.e., ILLUSTRATION/DISTRACTOR *total dwell time*), across pages, to obtain single summary values for each child.

Percent dwell time. *Percent dwell time* represents children's AOI dwell times as percentages of their gaze across the entire display, thereby isolating the relative distribution of children's visual attention to the two images, irrespective of their listening duration. For single measures for each child, we use the mean of children's percent dwell time values for each AOI (*mean net dwell time percent*).

Modeling Approach

To analyze by-trial data, we constructed mixed effects models using the lme4 package in R (Bates et al., 2015), with fixed effects for our predictors of interest. Following Barr and colleagues' (2013) recommendations, our default models included random intercepts for each participant and page, and random slopes for page number by participant. In the case that our model of interest failed to converge, we first removed the random slopes for page number and refit the model, followed by the random intercepts for page if this new model also did not converge. Due to the heavy right skew of the distribution of the raw measurements — in terms of milliseconds of visual attention or listening time — we use log-transformed values when modeling attention variables or including them as predictors. In the final section, we use linear models to test the relation between attention and learning by predicting children's accuracy at test (percent correct across listening comprehension or word learning test trials) from summary metrics of their attention (*cumulative listening time, total dwell time*, and *mean dwell time percent*), controlling for condition and age. For all models, we report coefficients or odds ratios and their 95% bootstrapped confidence intervals for fixed effects. To additionally assess significance, we use nested model comparison, including likelihood ratio tests using the anova command in the R car package (Fox & Weisberg, 2019), and Wald tests using the Anova command.

[2] In fact, we collected two summary dwell time metrics for each trial: AOI dwell time during the first mandatory page repetition, and AOI dwell time afterward. While we predicted stronger effects on dwell times in the period after the first listen — once the participant had been exposed, at least, to all the information the page had to offer — these measurements were also much more unwieldy. Zero values were common, and the ratio of lowest to highest non-zero values was significantly greater. Cognizant of already operating with a partial dataset, we opted to combine participants' dwell times on the first and later repetitions of each page in the interest of retaining more data points for analysis.

Before reporting the results of our study, a brief discussion of our underlying linking hypothesis and predictions is in order.

The Linking Hypothesis

Our experimental paradigm hinges on two observations from existing empirical research: First, psycholinguistic experiments with adults suggest that earlier-acquired words — like those that dominate the Simple story narration — are processed more readily than later-acquired words. Words with earlier age of acquisition estimates are associated with lower reaction times in picture naming (e.g., Carroll & White, 1973; A. W. Ellis & Morrison, 1998), word naming and lexical decision tasks (e.g., Butler & Hains, 1979; Gerhand & Barry, 1999; Izura & Ellis, 2002; Morrison & Ellis, 2000), and even semantic tasks like word association (Brysbaert et al., 2000; van Loon-Vervoorn, 1989) and categorization (Ghyselinck, Custers, et al., 2004; Ghyselinck, Lewis, et al., 2004). These data suggest that the Complex narrative should be more effortful to process than the Simple narrative, even for a listener familiar with all of the words.

Second, eyetracking research relies on the assumption that participants' gaze is generally revealing of the current object of their attention — and, more specifically, of their incremental processing during online language comprehension tasks (e.g., Arunachalam, 2016; Golinkoff et al., 1987; see Eckstein et al., 2017 for review). This suggests that when children are actively attending to the speech, the visual illustration of the audio narration will be a highly attractive fixation target. But we also know that children spend most of their time during storybook reading episodes scanning the storybook illustration, *anyway* (e.g., Belfatti, 2012; Evans & Saint-Aubin, 2005). While promising for the ecological validity of our experiment, children's general tendency to attend to stories' illustrations could call into question whether their visual attention toward the illustration in our study is necessarily a meaningful signal of auditory attention. To address this concern, we introduce the DISTRACTOR as a lure for children's attention, reasoning that children should be more likely to fixate on the DISTRACTOR when they are less attentive to the story narration, and therefore less attracted to the illustration currently being described.

Previous samples of children in the relevant age range have shown robust learning of both words and facts from the Simple speech (Foushee, Srinivasan, & Xu, unpublished manuscript), suggesting that if children's attention is at least partly driven by learnability, children in the Simple condition should show greater attention to the speech (longer voluntary trial durations) and greater relative attention to the storybook illustration (longer ILLUSTRATION net dwell times, and greater ILLUSTRATION percent dwell times). More generally, we predict that learning and attention should be linked, such that greater attention to the story — measured via children's sustained attention to the story illustration, rather than the DISTRACTOR — should be associated with greater evidence of learning at test. We test these predictions in the following sections, organized in terms of analyses relating (a) complexity and attention, (b) complexity and learning, and (c) attention and learning.

3.4 Results & Discussion

Complexity & Attention

Our first set of analyses asked whether children's attention varied on the basis of speech complexity. We predicted that children in the Simple condition would listen to the story for greater durations, and as they listened, allot greater visual attention to the Illustration AOI, compared to the Distractor. Table 7 presents median values and 95% bootstrapped confidence intervals for each of our attention metrics between conditions, while Figure 20 compares the distributions of by-participant summary values (i.e., *cumulative listening time*, Illustration *total dwell time*, and Illustration *mean percent dwell time*; see Appendix M for histograms of these measures between conditions).

Table 7: *Attention Metrics by Condition*

		Simple		Complex	
		Mdn	*95% CI*	*Mdn*	*95% CI*
By-*Trial* Metrics	Voluntary Trial Duration (*s*)	7.73	(0.97, 19.90)	3.71	(0.07, 13.30)
	Net Dwell Time (*s*)				
	Illustration	14.76	(11.97, 20.83)	12.66	(7.84, 17.72)
	Distractor	5.42	(3.74, 8.28)	5.64	(3.19, 8.80)
	Percent Dwell Time (%)				
	Illustration	50.5	(40.8, 64.2)	44.4	(26.9, 61.1)
	Distractor	27.9	(17.2, 50.6)	37.0	(21.4, 53.8)
By-*Participant* Metrics	Cumulative Listening Time (*s*)	75.5	(40.7, 109.00)	47.0	(27.50, 68.00)
	Total Dwell Time (*s*)				
	Illustration	98.00	(82.60, 117.60)	80.2	(56.50, 92.70)
	Distractor	41.4	(29.9, 44.7)	39.7	(27.2, 49.5)
	Mean Percent Dwell Time (%)				
	Illustration	50.2	(45.2, 56.9)	41.8	(34.4, 57.5)
	Distractor	30.5	(23.7, 42.7)	37.0	(31.9, 47.0)

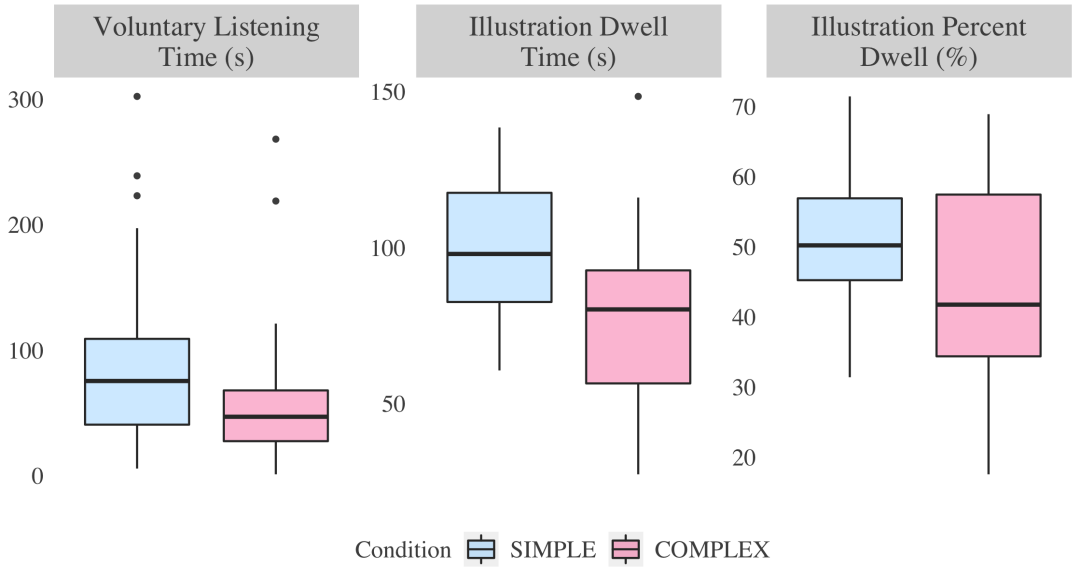


Figure 20: *Summary Metrics of Child Attention between Conditions.*

Note. From left to right: *Cumulative Listening Time*, ILLUSTRATION *Total Dwell Time*, ILLUSTRATION *Mean Percent Dwell Time.*

Voluntary Trial Duration

Voluntary trial durations in our sample ranged from less than $1s$ to $73.9s$, or about $1.2min$ ($Mdn_{\text{duration}} = 5.53s$ $[0.25, 18.10]$). On the median trial, participants heard approximately one half of the narration for the page repeated a second time ($Mdn_{\text{repetitions}} = 0.47$ $[0.12, 1.33]$), only 'timing out' on a total of four trials (three trials from three different four-year-olds in the Simple condition, and one from a four-year-old in the Complex condition). On 29% of trials, participants were apparently already looking at the DISTRACTOR when the second page repetition started, and did so continuously, such that they 'moved on' to the next page as soon as possible (voluntary trial durations of $0s$). On over half (68%) of all trials, children moved on to the next page before hearing a complete second repetition of the page they were currently on (see Figure 21). Comparing median values across conditions suggests that trials tended to be longer in the Simple condition ($Mdn_{\text{Simple}} = 7.73s$ $[0.97, 19.90]$) relative to the Complex condition ($Mdn_{\text{Complex}} = 3.71s$ $[0.07, 13.30]$). However, despite this numeric trend, mixed effects linear models (reported below) suggest that differences in listening times by condition are not reliable.

Table 8 displays the coefficients and confidence intervals for mixed effects linear models fit to the voluntary trial duration data (in *log milliseconds*). Model 1 includes age, con-

Table 8: *Mixed Effects Linear Regressions on Voluntary Trial Duration*

	Dependent variable:			
	Voluntary Trial Duration (*log ms*)			
	(1)	(2)	(3)	(4)
Intercept	10.60^{***}	10.50^{***}	13.70^{***}	12.90^{**}
	(5.19, 16.10)	(5.02, 16.00)	(7.07, 20.20)	(3.11, 22.70)
Condition (COMPLEX)	−0.73	−0.45	−9.50	−4.28
	(−1.83, 0.36)	(−2.06, 1.16)	(−20.60, 1.66)	(−20.90, 12.30)
Page Number	−0.03	0.004	−0.03	0.18
	(−0.20, 0.13)	(−0.23, 0.24)	(−0.20, 0.13)	(−1.89, 2.26)
Age	−0.56	−0.56	−1.22	−1.08
	(−1.72, 0.60)	(−1.72, 0.60)	(−2.63, 0.20)	(−3.19, 1.03)
COMPLEX:Page		−0.08		−1.49
		(−0.42, 0.26)		(−5.01, 2.03)
COMPLEX:Age			1.90	0.83
			(−0.51, 4.30)	(−2.76, 4.42)
Page:Age				−0.04
				(−0.49, 0.41)
COMPLEX:Page:Age				0.31
				(−0.45, 1.07)
Observations	276	276	276	276
Participants	46	46	46	46
SD(Participant)	1.60	1.60	1.57	1.56
Log Likelihood	−667	−667	−664	−665
Akaike Inf. Crit.	1,345	1,349	1,342	1,351
Bayesian Inf. Crit.	1,367	1,374	1,368	1,387

Note: Models include random intercepts for subject. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

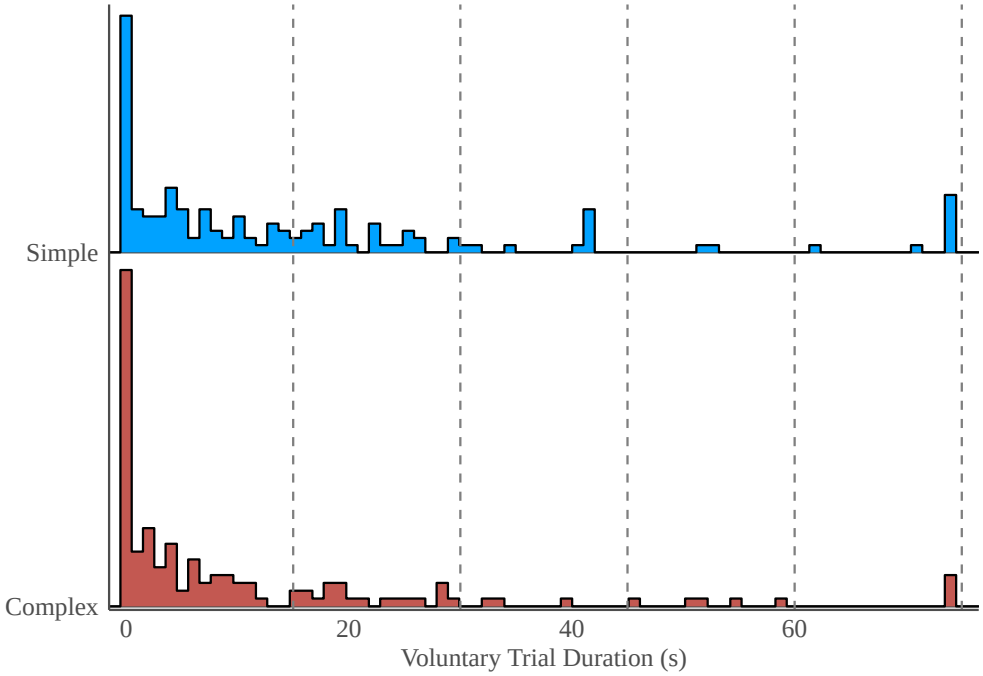


Figure 21: *Histogram of Voluntary Trial Durations by Condition.*

Note. Dashed lines mark approximate page repetition boundaries.

dition, and page number as fixed effects, and random intercepts for participant (models with random slopes for page number and intercepts for page failed to converge). Neither condition (Complex $\beta = -0.73$ $[-1.81, 0.34]$, *Wald* $\chi^2(1) = 1.74$, $p = .19$), page number ($\beta = -0.03$, $[-0.20, 0.13]$, *Wald* $\chi^2(1) = 0.16$, $p = .68$), nor age ($\beta = -0.56$ $[-1.71, 0.59]$, *Wald* $\chi^2(1) = 0.88$, $p = .35$) were significant predictors of children's voluntary trial durations. Models 2–4 test further theoretically motivated interactions among these variables. For example, Model 2 tests the prediction that the relation between page number and listening duration might differ between conditions, such that listening durations for children in the Complex condition might decrease more dramatically than listening durations for children in the Simple condition (*interaction between condition and page number*: $\beta = -0.08$ $[-0.42, 0.26]$, *Wald* $\chi^2(1) = 0.23$, $p = .63$). Model 3 tests the prediction that the relation between age and trial duration might differ between conditions, such that relatively younger children might listen longer in the Simple condition, and relatively older children might listen longer in the Complex condition (*interaction between condition and child age*: $\beta = 1.90$ $[-0.45, 4.25]$, *Wald* $\chi^2(1) = 2.39$, $p = .12$). Finally, Model 4 tests the prediction that listening times for older children in the Complex condition might decrease at a lesser rate across pages than listening time for younger children (*three-way interaction between condition, page number, and age*: $\beta = 0.31$ $[-0.45, 1.06]$, *Wald* $\chi^2(1) = 0.62$, $p = .43$). While none of these

interactions reach significance in our models, we note that the effects are in the predicted directions (refer to model coefficients in Table 8).

Net Dwell Time

Next, we looked at the distribution of participants' visual attention to our predefined Areas of Interest (AOIs). ILLUSTRATION dwell times ranged from $0s$ (two children, one in each condition) to $62.8s$ ($Mdn_{\text{Illustration}} = 13.72s$ $[10.48, 19.18]$). Given that dwelling continuously on the DISTRACTOR for $1.5s$ — after hearing first narration loop — triggered the end of the trial, DISTRACTOR dwell times occurred in a narrower range than ILLUSTRATION dwell times, as the only way that children could dwell on the DISTRACTOR for fewer than $1.5s$ was by 'timing out' of the trial. Beyond the four trials mentioned above, DISTRACTOR dwell times ranged from $1.5s$ to $24.5s$ ($Mdn_{\text{Distractor}} = 5.49s$ $[3.59, 8.39]$), with greater values reflective of both more time spent on the DISTRACTOR during the first narration loop, and more $< 1.5s$ visits during later loops.

If our manipulation of speech complexity affected children's attention to the content of the story, we expected to see differences between the two conditions in the distribution of children's visual attention to the story-relevant AOI, the ILLUSTRATION, relative to the perceptually salient DISTRACTOR. Specifically, we predicted that, by being better able to follow the story, children in the Simple condition would be more engaged by the image that matched the narration, and therefore attend more to the story illustration (longer ILLUSTRATION dwell times) than children in the Complex condition. We might expect a somewhat opposite pattern of results for attention to the DISTRACTOR, such that children in the Complex condition who are having more trouble following the story might be relatively more likely to be lured away by the distracting animation, thereby ending their trial early.

For a more coherent picture of our dwell time data, we next fit a linear mixed effects model to participants' dwell times (in log-transformed milliseconds), with fixed effects for age, page, condition (Simple or Complex), AOI (ILLUSTRATION or DISTRACTOR), and an interaction between condition and AOI (see Table 9). The model included random intercepts for participants. AOI was a significant predictor of dwell times, suggesting that children in both conditions looked overall longer at the ILLUSTRATION over the course of each page (DISTRACTOR $\beta = -0.99$ $[-1.12, -0.86]$ *Wald* $\chi^2(1) = 306.93$, $p < .001$). While condition was not a reliable predictor of dwell times itself (COMPLEX $\beta = -0.27$ $[-0.42, -0.11]$, *Wald* $\chi^2(1) = 3.77$, $p = .05$), its interaction with AOI was, such that children exhibited a diminished preference for the ILLUSTRATION in the Complex condition, relative to the Simple condition ($\beta = 0.28$ $[0.09, 0.47]$ *Wald* $\chi^2(1) = 8.19$, $p = .004$). Finally, child age was not a significant predictor of dwell times ($\beta = -0.05$ $[-0.18, 0.08]$ *Wald* $\chi^2(1) = 1.03$, $p = .31$).

Table 9: *Mixed Effects Linear Regression on AOI Net Dwell Time*

	Dependent variable:
	Net Dwell Time (*log ms*)
Intercept	9.90^{***} (9.26, 10.50)
Condition (Complex)	-0.27^{***} (-0.43, -0.11)
AOI (Distractor)	-0.99^{***} (-1.12, -0.86)
Page Number	-0.02 (-0.04, 0.01)
Age	-0.051 (-0.19, 0.09)
Complex:Distractor	0.28^{***} (0.09, 0.47)
Observations	546
Participants	46
Pages	6
SD(Participant)	0.15
Log Likelihood	-491
Akaike Inf. Crit.	998
Bayesian Inf. Crit.	1,032

Note: Model includes random intercepts for subject.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Percent Net Dwell Time

Participants spent between 0 and 99.5% of their overall screen dwell time on one of our two AOIs (M= 89.9% [77.1, 95.5]). The Illustration ($M_{\text{Illustration}}$ = 47% [45%, 50%) typically elicited a greater percentage of children's page–long dwell times than did the Distractor ($M_{\text{Distractor}}$ = 35% [33%, 38%]). Table 10 displays the coefficients and confidence intervals for linear mixed effects models fit to the percent net dwell time data for the Illustration (first column) and the Distractor (second column). Models included condition, page number, and age as fixed effects, random intercepts for participant and page, and random slopes for page number by participant. In neither model were condition (Illustration: Complex $\beta = -5.53$ [-13.06, 1.99], *Wald* $\chi^2(1) = 2.12$, $p = .15$; Distractor: Complex $\beta = -0.27$ [-0.42, -0.11], *Wald* $\chi^2(1) = 2.16$, $p = .14$), page number (Illustration: $\beta = -1.12$ [-4.00, 1.76] *Wald* $\chi^2(1) = 0.59$, $p = .44$; Distractor: $\beta = -0.99$ [-1.12, -0.86] *Wald* $\chi^2(1) = 0.30$, $p = .58$), or age (Illustration: $\beta = 0.80$ [-7.21, 8.78] *Wald* $\chi^2(1) = 0.04$, $p = .84$; Distractor: $\beta = -0.05$ [-0.18, 0.08] *Wald* $\chi^2(1) = 1.12$, $p = .29$) significant predictors of participants' percent dwell times.

With this suggestive evidence from children's net dwell time measurements — if not their *percent* net dwell times — that our manipulation of verbal complexity influenced children's

Table 10: *Mixed Effects Linear Regression on AOI Percent Net Dwell Time*

	Dependent variable:	
	Illustration (%)	Distractor (%)
Intercept	57.00*** (17.50, 96.60)	13.00 (−27.90, 53.90)
Condition (Complex)	−6.87 (−14.60, 0.82)	6.28 (−1.86, 14.40)
Page Number	−1.12 (−3.82, 1.57)	−0.52 (−2.23, 1.19)
Age	−0.53 (−8.73, 7.68)	4.61 (−4.07, 13.30)
Observations	648	648
Participants	46	46
Pages	6	6
SD(Participant)	19.99	16.50
SD(Page)	5.33	2.91
Log Likelihood	−1,179	−1,177
Akaike Inf. Crit.	2,375	2,373
Bayesian Inf. Crit.	2,408	2,405

Note: Models include random intercepts for subject and page.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

visual behavior, we next asked whether it had any influence on learning.

Complexity & Learning

To test the relation between complexity and learning in our study, we analyzed the effect of speech complexity on children's performance on the six listening comprehension and word learning test trials following the story presentation. As children in both conditions heard each page at least once, they shared a baseline exposure to the elements of the story plot that were tested in the listening comprehension trials. We ask first whether performance on these trials differed based on whether children heard the Simple or Complex story narration.

Listening Comprehension

As a group, children tended to perform very well on listening comprehension trials. Children in the Simple condition answered an average of 70.4% [59.0%, 80.6%] trials correct, about 4.2 out of 6. Participants in the Complex condition performed slightly worse: 64.4% [55.3%, 72.7%], or about 3.9 trials correct. Performance by question and condition are plotted in Figure 22. A permutation test comparing listening comprehension scores across the two conditions suggests that the difference in mean scores is not significant ($p = .44$).

To further explore the consequences of our speech complexity manipulation, we fit a mixed effects logit model to participants' trial-by-trial listening comprehension data (coded as 0 = *incorrect*, 1 = *correct*). The model included age, trial order, and condition as fixed effects, random intercepts for both participant and trial, and random slopes for trial order by participant. Odds ratios and confidence intervals for this model can be found in Table 11. Age was a significant predictor of accuracy in the model, with older children significantly more likely to respond correctly ($OR = 6.41$ $[2.22, 24.03]$, *Wald* $\chi^2(1) = 10.19$, $p = .001$). Neither trial number ($OR = 1.33$ $[0.65, 2.96]$, *Wald* $\chi^2(1) = 0.32$, $p = .57$) nor assignment to the Complex condition ($OR = 0.72$ $[0.25, 1.97]$, *Wald* $\chi^2(1) = 1.14$, $p = .29$), on the other hand, was reliably related to children's listening comprehension scores.

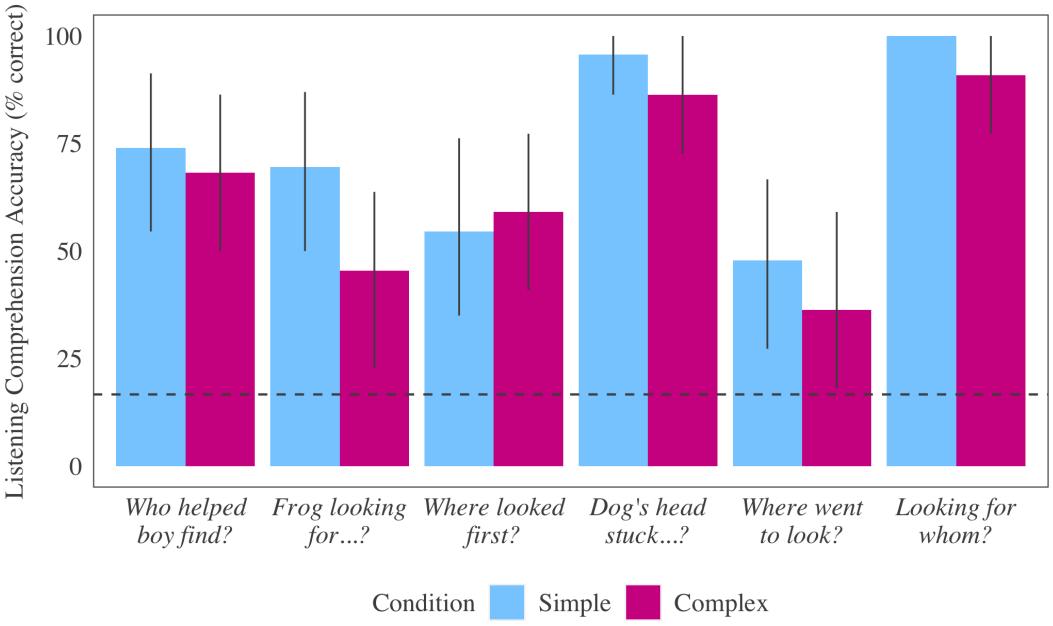


Figure 22: *Listening Comprehension Performance by Item and Condition.*

Note. See Appendix K for test arrays.

Table 11: *Mixed Effects Logistic Regressions on Test Trial Accuracy*

	Dependent variable:	
	Listening Comprehension {0, 1}	Word Learning {0, 1}
Intercept	0.0003 (0.00, 0.07)	0.44 (0.02, 11.13)
Condition (Complex)	0.62 (0.24, 1.52)	1.06 (0.82, 2.77)
Trial Number	1.33 (0.65, 2.96)	0.44 (0.61, 1.33)
Age	6.41** (2.22, 24.03)	1.49*** (0.55, 2.02)
Observations	648	648
Participants	46	46
Trials	6	6
SD(Participant)	1.66	0.45
SD(Participant, Trial Number)	0.10	—
SD(Trial)	1.46	.60
Akaike Inf. Crit.	304	357
Bayesian Inf. Crit.	333	379

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Word Learning

The final block of the study consisted of six trials testing every participants' knowledge of the unfamiliar target words. As only children in the Complex condition actually had opportunity to learn the unfamiliar target words, we predicted chance (25%) word-learning performance for children in the Simple condition. Surprisingly, overall performance was less distinguished between the two conditions than we anticipated (M_{Complex} = 39% [30%, 48%]; M_{Simple} = 32% [23%, 40%]; p = .2). Examining accuracy by word, performance in neither condition exceeded chance (*ogling*: M_{Complex} = 55% [36%, 73%], M_{Simple} = 38% [21%, 58%]; *absconding*: M_{Complex} = 23% [5%, 41%], M_{Simple} = 25% [8%, 42%]; *flummoxed*: M_{Complex} = 55% [32%, 77%], M_{Simple} = 42% [21%, 63%]; *hyaline*: M_{Complex} = 46% [27%, 68%], M_{Simple} = 33% [17%, 54%]; *aperture*: M_{Complex} = 9% [0%, 23%], M_{Simple} = 13% [0%, 25%]; *tor*: M_{Complex} = 50% [27%, 68%], M_{Simple} = 38% [21%, 54%]; all ps > .05, with Bonferroni correction for multiple comparisons).

As when analyzing listening comprehension data, we fit a mixed effects logit model to children's word-learning test trial accuracy, including age, trial order, and condition as fixed effects, and random intercepts for participant and trial (a model including random slopes for trial number failed to converge). In contrast to our listening comprehension results, this model did not show a significant effect of age (OR = 1.05 [0.55, 2.02]; *Wald* $\chi^2(1) = 0.03$, p = .86), trial number (OR = 0.91 [0.61, 1.33]; *Wald* $\chi^2(1) = 0.32$, p = .57), or condition

($OR = 1.49$ $[0.82, 2.77]$; *Wald* $\chi^2(1) = 1.78$, $p = .18$).

We note that our test of word learning was also difficult, as it demanded not only recall, but *generalization* of the newly-introduced word to a novel context and visual referent. Comparing the means plotted by item in Figure 23 suggests that children especially struggled with the word *aperture*, which was both the second-to-last unfamiliar word to be heard by children in the Complex condition, and the second-to-last word to be tested for all children. Closer inspection of children's selections at test reveals a strong preference for a specific test image of a small flower — unfortunately, the only image not used in our prior experiments, as children there, too, had shown a disinclination to pick the muddy hole in the ground that was the generalization target, leading us to replace the array. The flower was selected by 46% of children in the Complex condition (10/22), and 38% of children in the Simple condition (9/24), evidence of its general appeal.

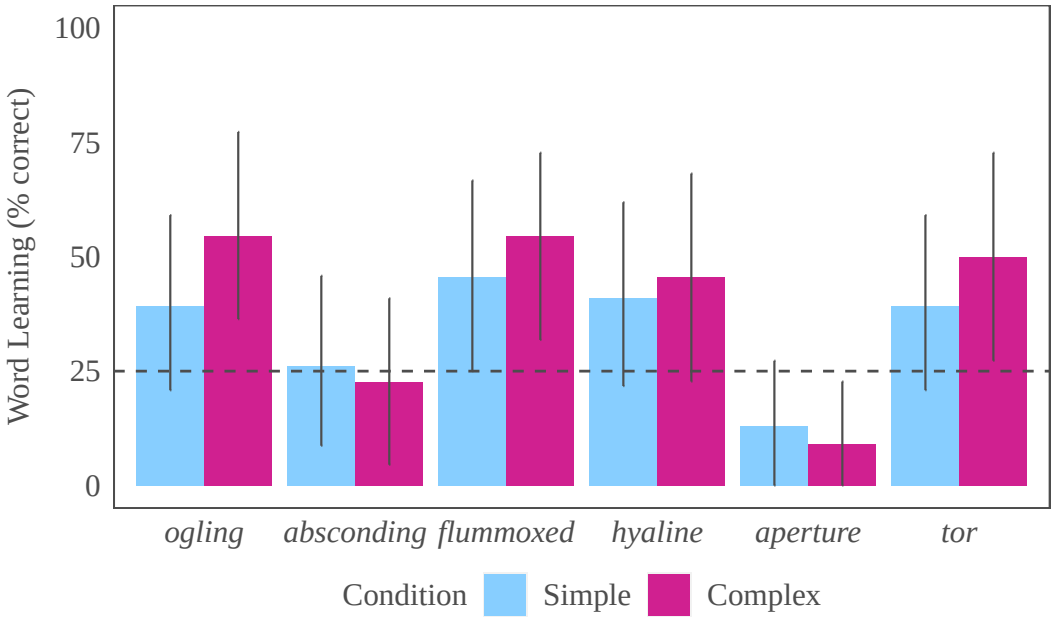


Figure 23: *Word Learning Performance by Word and Condition.*

Note. Order of bars reflects the sequence in which children in the Complex condition heard the words in the story, and the sequence in which children in both conditions were tested (see Appendix L for test arrays).

More generally, these results may speak to *contextual* determinants of complexity. Like Gerken and colleagues (2011), we used stimuli that we had advance reason to believe was well-calibrated for our participants. Specifically, the speech used in the Simple condition led to robust learning of the same unfamiliar words in previous samples of children in this

age range. However, we did not replicate children's word-learning performance from our previous experiments, suggesting a disruptive effect of the GIF: in spite of children's increased exposure to the story itself, the presence of the continuous animation in their periphery may have taxed children's memory or interrupted their encoding of the story.

Attention & Learning

A primary goal of this study was to explore the link between children's capacity to learn from a particular stimulus and the attention that children dedicated to it. For our critical question of the relation between child attention and subsequent learning, we predicted children's learning scores from measures of their online attention to the storybook.

Learning as a Driver of Attention

Thus far, we have primarily been thinking about the *objective* complexity of our stimuli: words in the Simple condition are typically acquired earlier than words in the Complex condition, which also includes regular unfamiliar words. While enabling us to define verbal complexity *a priori*, this absolute scale overlooks the *subjective* complexity of the speech — that is, how children's own language development might dictate the perceived complexity of incoming speech, such that the same objective level might be 'just right' for one child, and too difficult for another. Applied to our study, this suggests that the experience of the speech complexity *within* each condition might be very different for children with larger versus smaller vocabularies. In other words, while the speech in the Simple condition might be pretty close to 'just right' for young 4-year–olds, it might be 'way too easy!' for children approaching 6. Thus, we might expect children at roughly the same stage of development to exhibit distinct patterns of attention depending on their condition: for example, blazing through the story in the Simple condition, but staying for multiple repetitions of each page in the Complex condition — or, conversely, lingering on each page in the Simple condition, but 'moving on' quickly in the Complex condition.

Also up until this point, we have primarily discussed *continuous* data (e.g., dwell times, listening duration) as proxies for children's degree of attention to the story. But we can also think of children in this study as having a choice. Like the toddler — who, at the close of the bedtime story, immediately exclaims, "again!" — children had the choice on each page of listening to the same audio "again!" or immediately moving on to the next page. We know that children frequently made both the choice to hear the page again and the choice to immediately move on. Of the six pages in the story, children in the Simple condition immediately moved on for an average of 1.67 [1.08, 2.29], and children in the Complex condition moved on for an average of 2.45 [1.68, 3.18]. However, we don't yet know what those choices mean; they might be different 'signals' in different conditions.

As discussed in the Introduction, the choice to continue listening (or looking) is very plausibly a sign that the child is still extracting information from the stimuli. The choice to immediately move on, on the other hand, is an ambiguous signal between 'information

already extracted' and 'information overly taxing to extract.' Given that all of the words in the Simple narration were expected to be familiar to children during this period of development, we might expect decisions to move on in the Simple condition to more often reflect saturation with the page content than decisions to move on in the Complex condition. In the Complex condition, while *some* children might move on after learning all they wanted from the current page, we might expect relatively more discontinued trials to come from overwhelm, or poorer processing of the page, relative to the Simple condition. We would especially expect this to be the correct interpretation for 'move on' decisions in the Complex condition for children with smaller English vocabularies.

As we did not collect standardized language assessments in this study, we use children's age as a proxy for their level of language and cognitive development. Our own past research (Foushee, et al., unpublished manuscript) supports this inference; in fact, we've shown it to be a reasonable assumption for two previous samples of children during the same period of development and recruited from the same venues. Specifically, children's age in these previous samples was highly correlated with their raw scores on the Peabody Picture Vocabulary Test (PPVT; $r = .74$, $p < .001$), as well as with their learning of the same set of unfamiliar target words that we embed in the Complex speech of the current study ($r = .39$, $p < .001$).

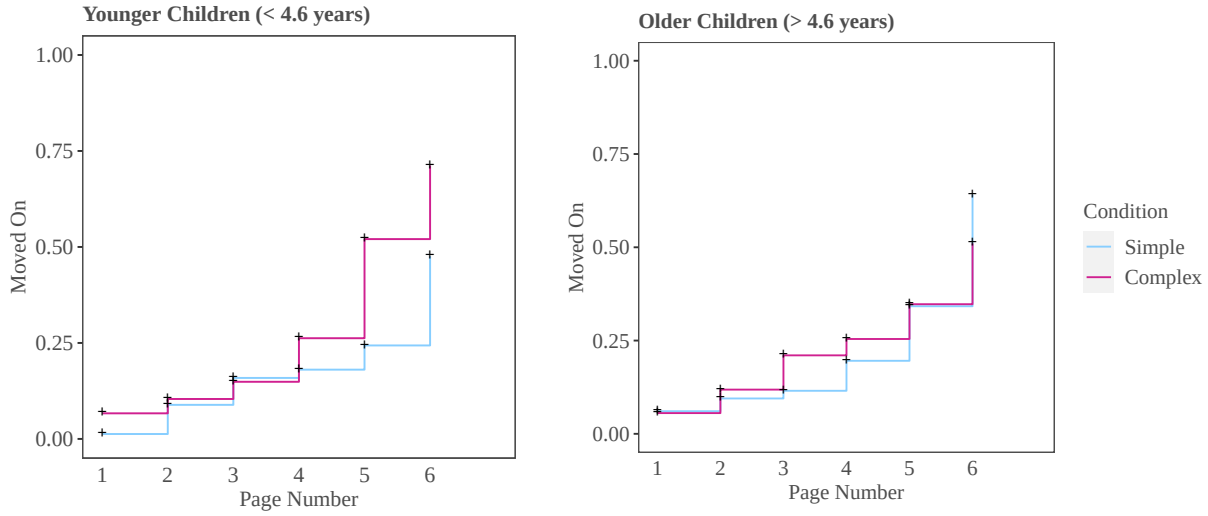


Figure 24: *Survival Analysis of Proportion of Children Having 'Moved On' by Each Trial.*

Note. Left panel shows 'move on' event data for children younger than the median age of our sample; right panel shows data for children in the older half of our sample. Lines are color-coded by condition.

With these predictions in mind, we explore the effect of age and condition on children's page-by-page decisions to move on, versus continue listening. Figure 24 provides an initial visualization of 'move on' events in each condition, by children whose ages fall above and

below the median age of our sample. With trial number (page) along the horizontal axis, and lines color-coded by condition, the height of each point represents the proportion of children who have made the 'move on' decision at some point over the course of the previous trials. The height of the "step" reflects the proportion of children who joined that group in the immediately preceding trial. Impressionistically, it appears that younger children in the Complex condition are more likely to move on than their age-matched peers in the Simple condition, especially as the story wears on. For our younger group (left panel), more children have moved on in the Complex than Simple condition by the last page, but the reverse is true for our older group (right panel), whose listening tendencies appear less distinguished between the two conditions.

Table 12: *Logistic Regression on 'Move On' Decisions*

	Dependent variable:
	'Moving On' {0, 1}
Intercept	0.001 (0.00, 0.16)
Condition (Complex)	31693.07 (3.08, 71507.57)
Page Number	1.08 (0.92, 1.29)
Age	3.55 (1.08, 13.06)
Complex:Age	0.13^{*} (0.01, 0.92)
Observations	276
SD(Participant)	1.15
Log Likelihood	−160
Akaike Inf. Crit.	331
Bayesian Inf. Crit.	353

Note: Model includes random intercepts for subject.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

To systematically test these impressions, we fit a mixed effects logit model to children's binary choice data, classifying trials with voluntary durations of less than $1s$ as decisions to 'move on.' The model included fixed effects for age, condition, page number, and the interaction between age and condition, as well as random intercepts for participants. Odds ratios and confidence intervals for each predictor in the model are presented in Table 12. In the text, we interpret predictors whose bootstrapped 95% confidence intervals suggest that

the direction of the effect is reliable (that is, do not span 1), even if Type II significance tests do not find them significant at $\alpha = .05$.

The results of our model suggest that children were more likely to 'move on' in the Complex condition (OR = 31693.07 [3.08, 71507.57], $\chi^2(1) = 2.66$, $p = .10$), and with greater age ($OR = 3.55$ [1.08, 13.06], $\chi^2(1) = 1.09$, $p = .30$). Contrary to appearances in Figure 24, page number was not a reliable predictor of 'moving on' ($OR = 1.08$ [0.92, 1.29], $\chi^2(1) = 0.89$, $p = .35$). Finally, we were particularly interested in what the interaction term ('Complex:Age' in Table 12) might tell us about how children's own development interacted with our *objective* complexity manipulation to determine the *subjective* complexity of our stimuli. In accordance with our predictions, the model showed a significant interaction between age and condition, such that the increased tendency to move on in the Complex condition was attenuated as children got older ($OR = 0.13$ [0.01, 0.92], $\chi^2(1) = 4.20$, $p = .04$). This is consistent with the idea that the Complex speech was less *subjectively complex* for children whose language development is more advanced, putting even the unfamiliar words closer within reach.

Attention as a Driver of Learning

The notion of 'subjective complexity' relies on the child's sense of their own learning, such that a child's attention to a stimulus is sustained when there is both something to learn, and when there is evidence that it can be learned *by them*. In the next section, we test the potential learning implications of differences in the distribution of children's attention.

Listening comprehension. We first ask whether children's attention during the storybook presentation predicts their performance on the listening comprehension test trials. For each of our attention indices, we fit linear models to children's listening comprehension accuracy (in terms of percentage trials correct), including age and condition as covariates. Model summaries appear in Table 13. Four out of five models showed significant ($p < .05$) relations between children's patterns of attention and their later learning (Model 1: *cumulative trial duration* $\beta = 6.00$ $[0.58, 11.40]$, $F(1) = 4.71$, $p = .035$; Model 2: Illustration *total dwell time* $\beta = 23.20$ $[5.67, 40.70]$, $F(1) = 6.73$, $p = .012$; Model 3: Distractor *total dwell time* $\beta = -11.30$ $[-28.40, 5.76]$, $F(1) = 1.68$, $p = .20$; Model 4: Illustration *mean percent dwell time* $\beta = 0.53$ $[0.04, 1.01]$, $F(1) = 6.73$, $p = .012$; Model 5: Distractor *mean percent dwell time* $\beta = -0.67$ $[-1.11, -0.24]$, $F(1) = 9.10$, $p = .004$).

In the final section, we test the relation between word learning and attention.

Table 13: *Linear Regressions Predicting Listening Comprehension from Attention*

	Dependent Variable:				
	Listening Comprehension (*% correct*)				
	(1)	(2)	(3)	(4)	(5)
Constant	−111.00*	−305.00**	89.20	−62.70	−27.20
	(−203.00, −18.40)	(−518.00, −92.10)	(−109.00, 287.00)	(−130.00, 4.32)	(−86.50, 32.00)
Condition (Complex)					
	−3.36	−0.62	−6.26	−2.62	−2.01
	(−16.10, 9.37)	(−13.50, 12.30)	(−19.10, 6.62)	(−15.50, 10.30)	(−14.20, 10.20)
Age	25.00***	23.70***	21.70**	23.10**	25.90***
	(11.60, 38.40)	(10.70, 36.70)	(7.89, 35.50)	(9.80, 36.40)	(13.10, 38.80)
Attention Metric:					
	Listening Time†	Illustration‡	Distractor‡	Illustration¶	Distractor¶
	(*log ms*)	(*log ms*)	(*log ms*)	(*% time*)	(*% time*)
	6.00*	23.20*	−11.30	0.53*	−0.67**
	(0.58, 11.40)	(5.67, 40.70)	(−28.40, 5.76)	(0.04, 1.01)	(−1.11, −0.24)
Obs.	46	46	46	46	46
R^2	0.29	0.32	0.24	0.28	0.35
Adj. R^2	0.24	0.27	0.18	0.23	0.30
F Statistic ($df = 3; 42$)					
	5.65**	6.50**	4.38**	5.58**	7.50***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

† Participant total across experimental trials

‡ Participant total dwell time across experimental trials

¶ Participant mean across experimental trials

Word learning. As we did for listening comprehension, we fit linear mixed effects models to children's overall test accuracy data — this time restricted to the children in the Complex condition, who had opportunity to learn the words from the story. Model results are displayed in Table 14. Here, attention to the story illustration — both in terms of the total duration of participants' ILLUSTRATION dwell time, and the percentage of their overall dwell time — was uniquely predictive of accuracy at test, controlling for age and condition (Table 14, Model 2: ILLUSTRATION *total dwell time* $\beta = 26.20$ $[4.22, 48.20]$, $F(1) = 5.46$, $p = .031$; Model 4: ILLUSTRATION *mean percent dwell time* $\beta = 0.81$ $[0.28, 1.34]$, $F(1) = 8.91$, $p = .008$). Neither the cumulative duration of children's exposure to the story (Model 1: *cumulative trial duration* $\beta = 3.80$ $[-4.21, 11.80]$, $F(1) = 0.86$, $p = .36$), nor either measure of their attention to the GIF (Model 3: DISTRACTOR *total dwell time* $\beta = -17.50$ $[-38.70, 3.62]$, $F(1) = 2.64$, $p = .12$; Model 5: DISTRACTOR *mean percent dwell time* $\beta = -0.51$ $[-1.18, 0.16]$, $F(1) = 2.22$, $p = .15$) were significantly related to word learning.

3.5 General Discussion

What material or level of complexity best supports children's learning? How do learners decide which problems they are ready to tackle, and which are better saved for later? Our study draws inspiration from classic ideas in developmental psychology (e.g., Berlyne, 1960; Bruner, 1961; Vygotsky et al., 1978) as well as recent formalizations of rational learner behavior to test the implications of general-purpose learning mechanisms in a rich, real-world domain. Rational models suggest that the ideal learner maximizes their efficiency by sampling broadly from the environment, using their own learning rate as a signal of whether they should persist with the same stimulus, or move on (Gerken et al., 2011; Gottlieb et al., 2013; Kidd & Hayden, 2015). Such computational-level descriptions of learner behavior are more and more useful the more they can help us understand real human behavior — but it can be challenging to test their predictions. Indeed, research interested in these questions has typically studied infants' deployment of attention to highly simplified stimuli in controlled laboratory settings (Gerken et al., 2011; Kidd & Hayden, 2015; Kidd et al., 2012, 2014). We saw potential for a more naturalistic test of the relation between complexity, attention, and learning in early storybook reading episodes, where young children are notorious for requesting and re-requesting the same story. We asked whether theories of rational attention — typically modeled and tested under toy conditions — generalize to children's attention to naturalistic language stimuli. If so, then children's continued appetite for hearing the same story might indicate their sense that they are still learning from it.

Having identified an ecologically valid context, we were challenged with how to capture meaningful differences in children's attention (their 'continued appetites'). We 'pit' two black-and-white images against one another: the static illustration for each page was designed to be attractive to children who were "really listening" to the story, while the animated DISTRACTOR was designed to absorb the liberated gaze of children who were no longer actively processing the story content. In part to prevent children from disattending from

Table 14: *Linear Regressions Predicting Word Learning from Participant Attention*

	Word Learning (*% correct*)				
	(1)	(2)	(3)	(4)	(5)
Intercept	5.05	−60.60	256.00	37.80	79.40
	(−134.00, 144.00)	(−194.00, 72.50)	(−19.50, 532.00)	(−55.00, 130.00)	(−36.50, 195.00)
Condition (Complex)					
	−3.36	−0.62	−6.26	−2.62	−2.01
	(−16.10, 9.37)	(−13.50, 12.30)	(−19.10, 6.62)	(−15.50, 10.30)	(−14.20, 10.20)
Age	−0.90	−2.46	−6.88	−6.99	−4.04
	(−24.80, 23.00)	(−23.90, 19.00)	(−30.80, 17.00)	(−27.40, 13.40)	(−27.40, 19.40)
Attention Metric:					
	Listening Time†	Illustration‡	Distractor‡	Illustration¶	Distractor¶
	(*log ms*)	(*log ms*)	(*log ms*)	(*% time*)	(*% time*)
	3.80	26.20*	−17.50	0.81**	−0.51
	(−4.21, 11.80)	(4.22, 48.20)	(−38.70, 3.62)	(0.28, 1.34)	(−1.18, 0.16)
Obs.	21	21	21	21	21
R^2	0.05	0.23	0.13	0.33	0.11
Adj. R^2	−0.06	0.15	0.03	0.26	0.01
F Statistic ($df = 2; 18$)					
	0.44	5.46*	2.64	8.91**	2.22

* $p < 0.05$; **$p < 0.01$; ***$p < 0.001$

† Participant total across experimental trials

‡ Participant total dwell time across experimental trials

¶ Participant mean across experimental trials

the stimuli entirely (recall that the study largely took place in corners of their preschool classrooms and noisy museums), trials advanced after a pre-determined fixation duration on the distractor, such that children heard the story narration for variable lengths of time. Children who spent more time on the DISTRACTOR advanced through the narrative more quickly than children who spent more time on the ILLUSTRATION.

We measured children's attention to the speech stimuli via three classes of indices. *Voluntary trial duration* and *cumulative listening time* captured participants' total exposure to the story narration, or the total amount of data to which they had access. While these first measures accurately reflect a child's 'window of opportunity' to learn from the story, they may be a noisy measure of their attention. That is, we know that a child with a voluntary trial duration of 13.5*s heard* the page narration for more than twice as long as a child with a voluntary trial duration of 6*s*: the first child heard every syllable, word, and sentence on the page repeated twice, while the second child heard only the first part of the page repeated before moving on. But — as anyone who has ever been asked to repeat an announcement at the train station knows — *hearing* is different from *listening* (Houston & Bergeson, 2014). Thus, we relied on our AOI metrics for a clearer signal of children's sustained attention to the story, as they capture the subset of children's overall exposure devoted to the image that went along with the story, over the dynamic distractor. Specifically, *net dwell time* and *total net dwell time* captured the duration of children's selective attention to the ILLUSTRATION and the DISTRACTOR, while *percent net dwell time* and *mean percent dwell time* captured the *relative* distribution of children's visual attention between AOIs, irrespective of the magnitude of their dwell time durations.

On the whole, our results suggest that our experiment was engaging for children, and that our manipulation of words' ages of acquisition effectively manipulated the complexity of the speech. Across conditions, children listened to the story for highly variable periods of time (Figure 21), and displayed variable attention to the ILLUSTRATION and DISTRACTOR. While children's voluntary trial durations were numerically greater in the Simple condition ($M_{\text{Simple}} = 7.73$ [0.97, 19.90], $M_{\text{Complex}} = 3.71$ [0.07, 13.30]), mixed effects models predicting trial durations from age, page, and condition suggested that the difference was not significant (Table 8). For a cleaner reflection of children's attention to the speech, we examined children's net dwell times to each AOI: though children in both conditions spent significantly longer on the ILLUSTRATION than on the DISTRACTOR, the *magnitude* of the difference was reliably greater in the Simple condition (Table 9). Combined with the numeric difference in trial durations, this result suggests that Simple participants' voluntary trial durations were longer specifically because of the greater duration of time that they spent fixated on the illustration for the story. This was not a given for two reasons: first, longer voluntary trial durations could in principle come from equivalent durations spent on the ILLUSTRATION, but more looks back and forth from the ILLUSTRATION to the DISTRACTOR. In this scenario, the DISTRACTOR would receive more total dwell time — accumulated across individual visits each too short to meet the trigger threshold — than other, shorter trials. Second, longer trial durations without greater ILLUSTRATION dwell times could theoretically

also have come from children looking away from the display altogether.[3] That none of our models predicting children's attention showed a main effect of condition is unexpected, but ultimately strengthens our results: any effects of complexity that we see in later analyses are unlikely to owe to mere intelligibility, or some other superficial quality of the speech streams that makes one preferable over the other for all child listeners.

Among the motivations for our experiment was the as-yet-unproven link between children's complexity-based attention and their learning. In seeking the earliest evidence for these patterns, previous research necessarily employed highly simplified stimuli, with limited potential for testing learning (Gerken et al., 2011; Kidd et al., 2012, 2014). By virtue of our increased age range and stimulus complexity, we were able to directly test children's learning within the experiment. If children deploy their attention rationally, we expected the duration of their attention to reflect the rate of their learning. That is, children should be least likely to terminate the trial, and most likely to continue dwelling on the illustration, when they are currently learning from the story. In contrast, we expect that children who have already encoded the information for the page on the first repetition will 'move on' immediately. We also expect children who are struggling to learn anything from the audio to move on quickly, albeit for the opposite reason. The suggestive pattern of results that we saw in our analysis of children's trial-by-trial 'move on' decisions (Table 12) is compatible with this interpretation. In particular, the interaction between condition and child age in predicting decisions to 'move on' suggests that the inclination to move on from the page as soon as possible — which is associated with the Complex condition — becomes weaker with age. This is the pattern that we would expect if children's increased age — a proxy for their language development — functionally decreased the (subjective) complexity of the complex speech.

Further evidence that children's attention was reflective of their learning comes from relating summary metrics of their attention to their accuracy at test. Controlling for age and condition, children's listening comprehension scores were significantly related to their listening times and to both absolute and relative measures of their gaze toward the ILLUSTRATION. Examining the data for only the children who actually heard the unfamiliar words in the narration suggests that visual attention toward the story illustration was also significantly related to word-learning accuracy. These results imply that it was not the complexity manipulation *per se* that made children learn more or less from the speech, but at least partly a product of their own attentional investment.

[3]This is an unlikely explanation for our data, as children spent the vast majority (an average of 89.9%) of their total dwell times on one AOI or the other, rather than in or beyond the margins of the display.

3.6 Conclusion

To our knowledge, this study is the first to extend ideas about the interplay between infant learning and attention to natural language stimuli and formal definitions of verbal complexity. Relative to studies of infant language development, the idea that low-level processes of attention to spoken language might continue to mediate language development into the preschool years has received little attention (Houston & Bergeson, 2014). This may be because we think of the preschooler's task in language learning as less basic than the infant's. However, while preschool–aged children are arguably masters of the phonetic inventory of their language and much of its syntax, they are a long way from possessing adult vocabularies. Our study offers a novel contribution to the literature by explicitly testing the relation between learning and self-directed attention to language of varying complexity, suggesting that selective attention may be a gating mechanism for word-learning into the preschool years. Even results relating attention and listening comprehension scores have implications for word-learning, as a new word is going to be more learnable when embedded in a linguistic context that the child otherwise understands (e.g., Sullivan & Barner, 2015).

In linking prior word familiarity with attention in older children, our study tested an implicit form of a skill — namely, selecting what to learn from, and when to give up — that early educators explicitly teach. A teacher of a nearby second-grade classroom in one of our participating schools shared her own method as an example: during reading time, she tells students to turn to a page of a book that interests them, and put up one finger for every word they don't know. If they get to the end of the page without raising a finger, the book is too easy; if they get to the end of the page and are raising their whole hands, the book is too hard. These sorts of heuristics, especially in an artificial domain like reading, may help children make explicit decisions about how to manage their learning time. However, our study adds to the body of evidence suggesting that they already know *something* about whether they will be able to learn from potential sources of information in their environments. This is an important conclusion for more applied fields, where notions of what constitutes 'developmentally appropriate' material are often difficult to cash out (e.g., Antonacci, 2000).

Finally, we were particularly interested in children's sensitivity to naturalistic speech complexity as a means of explaining why certain sources of language input have proven to be more useful for children's learning than others. We hypothesized a critical role of attention, such that simpler, more 'developmentally appropriate' spoken language inputs might more effectively elicit and maintain children's attention. As in other domains where, for example, children track potential informants and select who they want as a teacher (e.g., Pasquini et al., 2007), our results leave open the possibility that children may track the relative difficulty of processing and encoding different sources of information, and preferentially attend to those where their learning will be the most efficient.

Chapter 4

Capturing Qualitative Variability in Early Overhearing Experiences: A Case Study

Abstract

Inspired by qualitative studies typically limited to child-directed speech, we develop a coding scheme designed to characterize *all* utterances accessible to the child in terms of their relative utility for word-learning. We focus in particular on contributors to *referential transparency* as a well-established and meaningful dimension of language learnability in context. These include the spatial positions of the caregiver and child, the caregiver's use of gaze or gesture to illustrate their meaning, the child's visual access to the caregiver or to the referent of the utterance, and the caregiver's use of modified prosody. As a proof of concept, we apply this coding scheme to existing naturalistic video corpora for a single child whose language development is well-documented. We find that both speech directed to the child and overhearable by her are highly variable along the qualitative dimensions we coded, and identify the heterogeneity of overheard speech as a source of noise in previous investigations. While irrelevant as a referential cue, our results suggest caregivers' prosodic modification may play a functional role in marking speech intended for the child — especially given the significant qualitative overlap between overhead and child-directed speech along other dimensions. In spite of the frequent similarity between overheard and child-directed speech, overheard utterances was significantly less associated with child attention. Taken together, our results shed light on how adults and children co-structure the early language environment, and promise to provide similar insights when applied to naturalistic video corpora for children across the world.

> It would be strange, indeed, to equip the child with subtle means for detecting lexical, morphological, and syntactic structures, while leaving her with only the most primitive equipment for learning to become an interactive member of human society. Every linguistic structure that we have explored in child-directed speech takes its meaning in definable communicative contexts.
>
> CROSSLINGUISTIC EVIDENCE FOR THE LANGUAGE-MAKING CAPACITY
> *D. I. Slobin, 1996*

4.1 Introduction

Despite substantial research on (a) differences in the contributions of child-directed versus overheard speech to vocabulary size (Ramírez-Esparza et al., 2014; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a; Weisleder & Fernald, 2013), and (b) the impact of qualitative differences in child-directed speech contexts on vocabulary acquisition (e.g., Cartmill et al., 2013; Hirsh-Pasek et al., 2015; Ramírez-Esparza et al., 2014), to our knowledge there remains no systematic study within these literatures of qualitative differences in *overhearing* contexts as they relate to learning. This is important because overheard speech is common across the world (e.g., Casillas et al., 2019; Sperry et al., 2019), and because overheard speech is likely to be a more heterogeneous category than child-directed speech, such that understanding the range of roles it may play in young children's lives is critical and not straightforward.

In quantitative studies correlating amount of speech with vocabulary size, both child-directed and overheard speech are treated as monolithic. However, speech directed *to* the child is likely to be a much more coherent category than speech *around* her, which might be directed to variable audiences, at variable distances, and with variable relevance to the child. The two categories of speech (child-directed versus overheard) undoubtedly differ in their overall rates of features that we know children can use to solidify word–referent mappings. Mindful of this, our study takes inspiration from previous studies of input *quality*, where researchers unpack the influence of child-directed speech by hand-coding qualitative aspects of naturalistic audio or video recordings, often with the intention of relating that variability to metrics of children's development of language (e.g., Hirsh-Pasek et al., 2015; Ramírez-Esparza et al., 2017; Rowe et al., 2004; Rowe et al., 2016).

In studies using qualitative coding schemes, utterances with the same token count, and even same ratio between types (unique words) and tokens, might be found to differ along some social-contextual dimension that we know is relevant for learning. Previous work analyzing such qualitative diversity has found, for example, that amount of speech not only directed to the child, but specifically one-on-one and in the sing-songy register of so-called

parentese, is predictive of vocabulary growth (Ramírez-Esparza et al., 2014), as is caregivers' tendency to use nouns when the noun's referent is highly salient or easily inferred from context (Cartmill et al., 2013). Notably, fine-grained coding schemes of this nature have historically been applied exclusively to speech that is child-directed, leaving a gap in the extant literature. Here, we develop a coding scheme that will enable us to characterize the full range of linguistic inputs experienced by children across contexts, and to analyze their relative utility for language-learning. We initially apply this coding scheme to an existing naturalistic English-language video corpus, corresponding to a target child whose language development is well-documented. However, our system is designed to be used to capture the richness and latent structure within the early language environments of children across contexts, cultures, and languages.

In Chapter 1, our means of evaluating the relative learnability of speech *around* versus speech *to* the learner employed coarse-grained, text-based metrics computed on large-scale corpora of child- and adult-directed speech. In this chapter and in ongoing cross-linguistic work, we explore the question at a different scale, via fine-grained coding of the learnability of individual child- and adult-directed utterances *in context*. Specifically, we use longitudinal samples of the language environment of a single child to ask five questions regarding overheard language quality and its empirically-grounded support for learning language:

(1) How does the *quality* — in terms of hypothesized utility for language learning — of overheard speech compare to the quality of child-directed speech?

(2) How does the qualitative *variability* of overheard speech compare to the qualitative variability of child-directed speech?

(3) In a naturalistic context, how *distinguished* are overheard and child-directed speech?

(4) How does the quality of child-directed and overheard speech change as the child matures?

(5) What aspects of speech quality are associated with child attention?

Guided by the prior work reviewed in the Introduction, we focus on *referential ambiguity* as a meaningful and well-studied dimension of individual utterances that is reliably associated with learning.

4.2 Method

Sample Selection

We selected videos (Databrary.org) and transcripts (https://phonbank.talkbank.org/browser/index.php?url=Eng-NA/Providence/) from the Providence corpus to explore qualitative differences in varieties of adult speech. The Providence corpus was collected by Demuth and colleagues (2006) as part of a longitudinal study of phonological development, and documents the early language development of six children, approximately one year in age at the time of enrollment. Data collection for the children in the corpus began at the onset of children's first words, after which they were videotaped in their homes for one hour every two weeks, for up to three more years. Our only prerequisite in selecting videos to analyze was that there be at least two adults present during the recording, and that it include multiple adult–adult conversational turns. This ended up being highly constraining, as videos for five out of the six children largely recorded single adult–child dyads, effectively narrowing our sample from six children to one.

The recordings that we ultimately analyzed for this case study, then, represent hour-long samples of naturalistic speech from the home of a single child, Naima, across the first three years of her life. Naima is one of the most densely sampled children in the corpus: she and her family contributed an impressive 88 sessions total, spanning the time just before Naima's first birthday (00;11;27), to a couple months before her fourth (03;10;10). Of these videos, 12 met our criteria for inclusion.

Procedure

We used Datavyu (Datavyu Team, 2014) to code the sample of videos. Transcripts of the relevant sessions were downloaded from the CHILDES database, and used to populate time-locked coding cells, organized by speaker. Two coders were responsible for coding eleven out of the twelve transcripts. Coders were responsible for alternating videos with respect to the age of the child, so that potential inconsistencies in coding were not confounded with child age. As the codes hinged on an understanding of the pragmatic context of the utterances, coders watched each video in full before coding the utterances. Coders also used this initial viewing to annotate coded dimensions that typically spanned multiple utterances, including the context of the interaction and the physical position of the child. In the critical coding pass, coders entered values for each of the qualitative dimensions described below, for all adult utterances in the recording. Ambiguity in the application of the codes was rare by design, as any dimensions triggering disagreement in previous adult coders were dropped before finalizing the scheme. When uncertainty did arise, the primary coder(s) and first author reviewed the video to reach a decision. In cases where the dimension could not be coded (i.e., where the utterance was inaudible, or the speaker or child were out of frame), the code for that utterance was marked as 'na.'

Coding Scheme

Speech Audience

We used a combination of pragmatic cues to code the audience to whom each utterance was directed, including: (1) the content of the utterance, (2) the surrounding linguistic context, (3) the gaze of the speaker, (4) the focus of attention of the scene participants, and (5) the physical positions of the speakers, combined with the relative volume or force of the utterance. The audience of the utterance was coded as 'target child,' 'adult,' or 'phone.' Utterances receiving the latter two codes were classified as *overheard speech*. Our method of classifying overheard speech differs from most previous studies in that it is coded on a by-utterance basis, rather than generally across segments of speech (e.g., Weisleder & Fernald, 2013), and in that it includes adult phone conversations that take place when the child is within earshot (*contra* Shneidman et al., 2013).

We next coded a set of six qualitative features for each utterance individually. The coding scheme is based on evidence for qualitative dimensions of spoken language associated with heightened child attention at Naima's age, and/or cues that children can reliably use to resolve referential ambiguity and learn new words (e.g., Cartmill et al., 2013; Cooper & Aslin, 1990; Golinkoff et al., 2015; Golinkoff & Hirsh-Pasek, 2006).

Here & Now Reference

Coders indicated whether the utterance described or referred to the current environment. Utterances referring to the "here and now" are argued to make the task of word-learning easier (e.g., R. Ellis & Wells, 1977), particularly early in the course of acquisition. "Here and now" coding reflects the intuition that if "coffee" is an unfamiliar word, it will be easier to learn when Naima's family is in the kitchen and her mother says, "Yum, Daddy's drinking coffee," than when Naima and her mother are in the living room when her father comes home, and Naima's mother says, "Daddy was shopping, he was looking for coffee." One prominent view in the literature emphasizes the information available in the syntax of an utterance (Gleitman et al., 2005). If a sentence is about the immediate context of the utterance, the child can use the relational structure implied by syntax to parse the scene and infer the meanings of new embedded words (Hoff & Naigles, 2002; Naigles, 1990). More generally, assuming that speech refers to the "here and now" is a sensible starting hypothesis for a learner, with the implication that learning will be enhanced when that assumption is met (Mervis, 1983; Shatz, 1978).

When coding utterances about the "here and now," coders further distinguished between utterances where the child was visibly attending to the relevant part of the scene, and utterances where she was not.

Referential Gesture

"Referential gesture" was coded as present when an utterance was accompanied by non-verbal cues to its reference (Baldwin et al., 1996; Booth et al., 2008; Brooks & Meltzoff, 2008; Frank et al., 2013; León, 1999; Slobin, 1985). Referential gesture occurred in a variety of forms: when Naima's mother leans in to pull a fine thread off Naima's tongue and says, "You had a hair in your mouth," and lifts the hair before Naima's eyes, but also when Naima's mother points at the laundry basket in conversation with Naima's father, or looks toward the fridge when discussing dinner plans, or even when she mimes sleeping when whispering about a nap. Thus, referential gesture coding considered a more expansive *locus of reference* (R. Ellis & Wells, 1977), and captured distinct information from the "here and now" code.

Child Gaze Toward Speaker

Caregiver's referential gestures might be lost on Naima if she were not attending to the speaker. Thus, we additionally coded Naima's visual attention to the speaker (Bakeman & Adamson, 2019; Grassmann et al., 2015).

Sing-song Prosody

This code captured whether the utterance had the cadence or exaggerated prosody typical of infant-directed speech (Fernald & Kuhl, 1987; Saint-Georges et al., 2013; Snow & Ferguson, 1977; Soderstrom, 2007), and best reflected pitch *variability*. Previous work suggests that this dimension attracts and maintains infants' attention, resulting in enhanced learning of associations between, e.g., visual and auditory stimuli (Cooper & Aslin, 1990; Kaplan et al., 1996; Ma et al., 2011), or of mappings between sound and meaning (Graf Estes & Hurley, 2013).

In addition to the above binary features, we analyzed three continuous measures of speech quality, auditory clarity, morphological complexity, and utterance length.

Auditory Clarity

We rated the auditory *clarity* of the utterance (e.g., Fernald & Simon, 1984), from 0 (inaudible) to 3 (clear). Of course, clarity for Naima may be different than for the coder viewing the tape. However, likely because the original study (Demuth et al., 2006) targeted phonological development, the camera placement was always designed to optimize the recording of Naima's productions. Therefore, the recording audio may provide a more accurate reflection of the child's own auditory experience than if our data had come from a study with a different intent. We note that some dimensions could still be coded, even for "inaudible" utterances, as in the case of inaudible speech on the phone.

Morphosyntactic Complexity

To capture trends in structural complexity, we used the pre-existing annotations of morpheme and token counts for each utterance to analyze *utterance length*, as well as to compute a measure of "morphological complexity" that increased with the ratio of morphemes to tokens (MacWhinney, 2008). For example, the utterance, "Oh baby, sorry," with three morphemes and three tokens, receives a morphological complexity score of 1, while the utterance "Why're you growling," with five morphemes and three words, receives a score of 1.67.

4.3 Results & Discussion

Summary data for all dimensions can be found in Appendix N. To assess the reliability of our coding scheme, an independent research assistant coded two especially dense thirty-minute segments of videos from the first and fourth quartiles of our age range. Agreement was typically high ('Here and Now': 97%, Referential Gesture: 100%, Sing-song Prosody: 70%, Speech Type: 97%, Child Gaze toward Speaker: 78%).

Distribution of Utterances

Both speaker and child were visible for 56% of all utterances, enabling complete coding for 3,801 utterances. In an additional 16% of utterances (1,075 total), the child, but not the speaker, was in frame, enabling coding of child position and gaze, but not referential gesture. We include all coded utterances in our analyses; scripts for accessing transcripts from the CHILDES database (childes-db; Sanchez et al., 2018), populating Datavyu coding spreadsheets, and all data analyses can be found at https://osf.io/hy5z2/.

Speech Context and Child Position

Utterances occurred most frequently in contexts coded as "play time" (71.2% of utterances), followed by "meal time" (22.6%), "bath time" (3.8%), and "bed time" (2.4%). An average of 2 contexts occurred in each video. The child was typically seated in a high chair (31.7% of utterances) or standing (26.7% of utterances). There was insufficient variability in early videos to support further analyses of the relation between the child's physical position and her language environment.

Table 15: *Examples of Qualitative Overlap in Child-Directed and Overheard Speech*

Speech Type	+ Qualitative Features	− Qualitative Features
Child-directed	MOT: Mmmm we're eating our supper!" CHILD: Mmmm" MOT: Here it is...yummy! Here's another bite... mmm! Thank you!	MOT: Remember, what was I reading, remember when I was trying to try out the backpack to carry you? I had to read the directions. CHILD: Why? MOT: Directions explain how to use or fix something.
Overheard	MOT: I packed you a towel and diaper and all that. FAT: Oh, good. MOT: I mean I'm just gonna go do that then come right back.	MOT: Hello? Hi! That's all right. I called them right at five and she told me what was on the regular menu but she didn't have the specials yet...

Speech Type

Despite selecting videos based on the presence of overheard speech, the majority (85.3%; 5,643 utterances) of utterances were child-directed. Overheard speech accounted for 12.4% of all utterances (773 utterances between adults, and 102 utterances over the phone). A remaining 2.3% (184 utterances) were uncodeable, all due to sufficiently poor audio quality that the original researchers had not been able to transcribe the content of the speech.

The infrequency of overheard speech may accurately reflect the statistics of the child's environment, or may reflect the original study's focus on the child's verbal *production*, leading Naima's mother to engage in more speech-eliciting behaviors during recordings than she might otherwise. Speech was also not equally distributed across caregivers: Naima's mother accounts for 71.2% of all utterances (64.0% or 4,310 utterances in child-directed speech and 8.3% or 558 utterances in overheard speech), while Naima's father accounts for 23.5% (19.8% or 1,333 child-directed, and 4.6% or 312 overheard). We collapse across utterances from both caregivers in our analyses, and structure the results below according to our primary research questions.

Is overheard speech less *learnable*?

Below, we compare child-directed and overheard speech along the qualitative dimensions designed to capture the *referential transparency* of each utterance in context, as well as caregivers' structural simplification of their speech.

Table 16: *Mean Values and Differences in Means (Child-directed − Overheard Speech)*

Dimension	Child-directed	Overheard	Difference[†]
About the 'Here & Now'	0.34 (0.32, 0.37)	0.28 (0.25, 0.30)	0.34^{***}
Child Looking at Referent	0.26 (0.25, 0.28)	0.02 (0.01, 0.02)	0.25^{***}
Child Looking at Speaker	0.86 (0.85, 0.88)	0.44 (0.36, 0.52)	0.43^{***}
Referential Gesture	0.77 (0.75, 0.79)	0.09 (0.05, 0.13)	0.23^{***}
Sing-song Prosody	0.99 (0.99, 1.00)	0.11 (0.07, 0.17)	0.88^{***}
Speech Clarity	2.83 (2.81, 2.85)	2.64 (2.52, 2.75)	0.19^{***}
Morphological Complexity	1.21 (1.20, 1.22)	1.30 (1.24, 1.35)	-0.09^{***}
Utterance Length	4.59 (4.45, 4.74)	6.48 (5.72, 7.26)	1.90^{***}

[†] Observed difference in means (Child-directed − Overhead)
[***] $p < 0.001$, via exact permutation test comparing observed difference in means to empirical null distribution.

All coded features of caregivers' utterances in context were reliably different between speech directed to Naima and speech that Naima could overhear (all ps $< .001$; see Table 16). Interestingly, overheard utterances were not uncommonly about the "here and now" ($M = 0.28$), though very infrequently combined with a referential gesture that the child could use to identify that this was the case ($M = 0.02$). The absence of referential gesture may partly explain why Naima rarely gazed toward the referent in overheard speech, even when the utterance was about the here and now (Figure 25). Alternatively, the disparity between speech types in how regularly Naima looks at the co-present referent might indicate that Naima's parents talk about "here and now" objects *because* Naima is looking at them. Overheard speech was also typically rated high for clarity ($M = 2.64$), suggesting that even when parents were speaking with one another or on the phone, they maintained proximity to Naima. This is consistent with Naima's tendency to look at the overheard speaker ($M = 0.44$), which we might expect to be reduced if the speaker were further away.

Child-directed and overheard speech were also reliably distinguished along structural dimensions, although variably so. Child-directed utterances were consistently shorter ($M_{\text{tokens}} = 4.59$) than overheard utterances ($M_{\text{tokens}} = 6.48$). However, the two often overlapped along our measure of morphological complexity, suggesting that while caregivers tended toward shorter utterances, they did not refrain from inflecting the words they used.

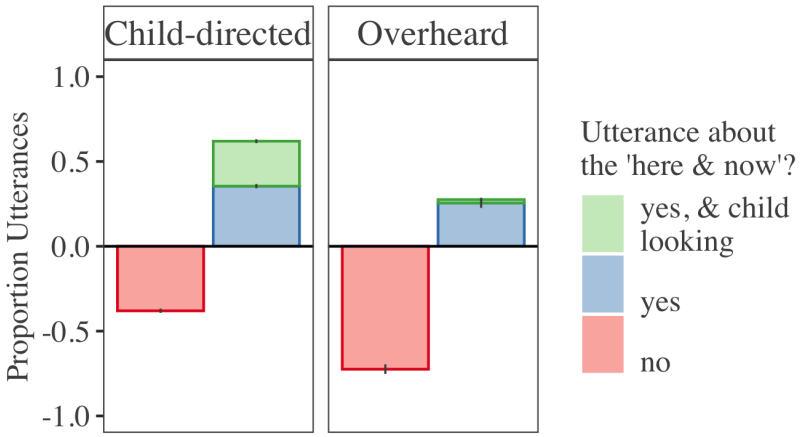


Figure 25: *Semantic Accessibility in Child-directed and Overheard Speech.*

Differences in what we coded as "sing-song prosody" are the most dramatic in our data. "Sing-song prosody" characterizes almost all child-directed utterances ($M = 0.98$), and 11% of overheard utterances. The frequency of exaggerated pitch variation in speech directed *to* Naima is not surprising, as the videos start when Naima is still an infant (Cristia, 2013; Soderstrom, 2007; Spinelli et al., 2017). However, Naima's parents' prosodic modification when addressing *each other* is unexpected, and may speak to the competing demands they experience as caregivers of a small child. For example, Naima's videos reveal various motivations for one of her caregivers to be in sustained physical proximity to her (e.g., to feed her in her high chair, or to prevent her from climbing precarious furniture, breaking something, or dissolving into tears). This means that, for much of the day, if Naima's parents also need to have a conversation, Naima will be present for it (and potentially experiencing a reduction in the attention she so recently enjoyed). Thus, adults in such contexts may be driven to "multi-task" in their speech production, using word meanings and syntax to transmit their *messages* to their partners, and using melodic prosody to signal their continued care and awareness to their infants. Consistent with this interpretation of caregivers' verbal behavior, "sing-song" adult-directed utterances (e.g., "Daddy what's today's date, is it the twenty-first?") in our data were equivalent to unmodulated adult-directed utterances in terms of length ($M_{\text{tokens}} = 5.59$) and morphological complexity ($M = 1.28$). We return to caregivers' instrumental use of prosody as a signal in the General Discussion.

Is overheard speech more *variable*?

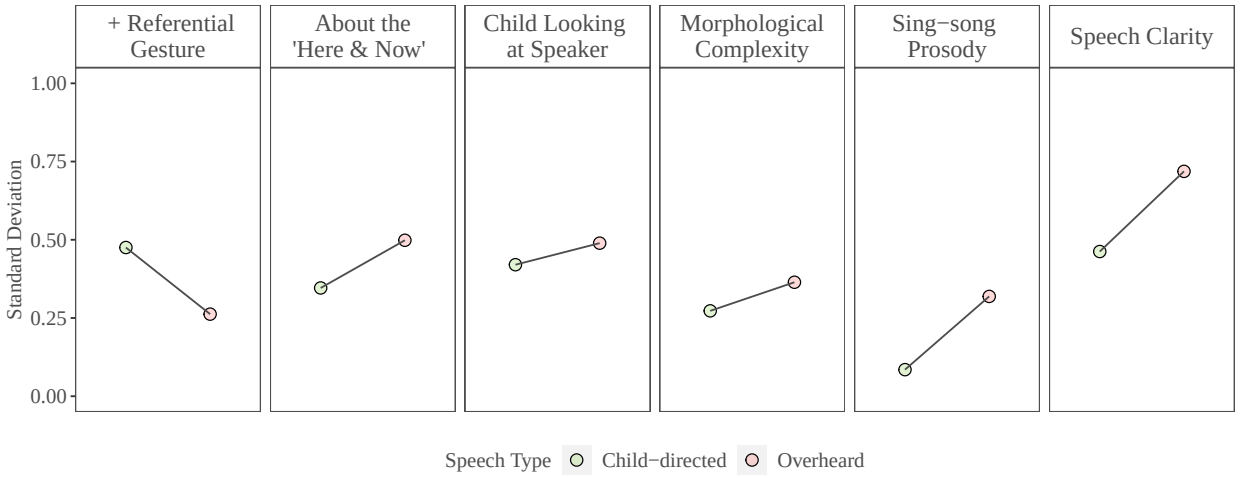


Figure 26: *Standard Deviations for Qualities in CDS and OHS.*

We predicted that overheard utterances would comprise a more heterogeneous category than child-directed utterances. We explore this prediction in two ways. Figure 26 plots the standard deviations for each qualitative variable by speech type, controlling for age. Panels where the point on the righthand side is higher than the point on the lefthand side suggest greater variability along that particular dimension within the set of overheard utterances.

For a better sense of the reliability of this difference — especially in light of the difference in the size of the two datasets — Figure 27 plots the frequency distributions of each binary feature in 1,000 bootstrapped samples of each dataset, and Figure 28 does the same along the continuous dimensions we coded. The width of each distribution gives a sense of the reliability of our frequency estimate, based on our dataset, while its horizontal position gives a sense of the overall rate or value range of that feature. Together, these analyses suggest that overheard and child-directed speech are reliably differentiated in their prosodic modification, tendency to describe the current environment, and correspondence with the current target of the child's visual attention — in terms of both an utterance's referent and its speaker. However, they are more frequently similar in their co-occurrence with referential gesture, clarity, and utterance length. Importantly, Figure 27 suggests that neither speech type is entirely predictable in its degree of referential transparency.

Figure 27: *Binary Feature Frequency in Resampled Distributions of Utterances.*

Are child-directed and overheard speech reliably distinguished?

Our third question concerned how *distinguishable* overheard speech is from child-directed speech in a naturalistic context. Again, we tested this in two ways.

We first fit a logit model to the data, using our coded variables to predict the type of the speech (coded as *overheard speech* = 0, *child-directed speech* = 1) to which each utterance belonged. The model included age, "here and now" reference (a categorical variable with three levels: "no," "yes, but not looking at the referent" and "yes *and* the child's gaze is on the referent"), referential gesture (0, 1), whether the child was looking at the speaker (0, 1), "sing-song prosody" (0 = absent, 1 = present), clarity (rating 0–3), and morphological complexity (computed values 0–3).

Odds ratios and 95% confidence intervals for this model are shown in Table 17. Whether the speech was about the "here and now" was a significant predictor of child-directed status (OR = 1.80 [0.79, 4.10]; $\chi^2(2) = 19$, $p < .001$), especially when the child was currently looking at the referent (OR = 8.88 [3.09, 27.00]). The child's concurrent gaze toward the speaker was also associated with child-directed speech status (OR = 2.42 [1.16, 5.10]; $\chi^2(1) = 6$, $p = .018$). Of all variables measured, prosody was the most reliable predictor of child-directed speech status (OR = 369.76 [190.68, 769.50]; $\chi^2(1) = 521$, $p = .001$). Finally, neither

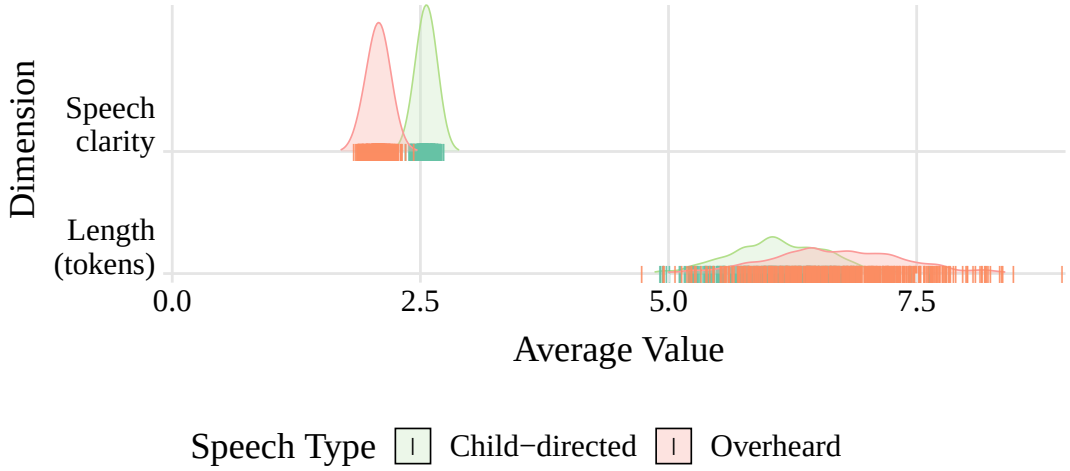


Figure 28: *Average Continuous Feature Values in Resampled Distributions of Utterances.*

referential gesture ($\chi^2(1) = 1$, $p = .250$), speech clarity ($\chi^2(1) = 0$, $p = .740$), utterance length ($\chi^2(1) = 1$, $p = .385$), morphological complexity ($\chi^2(1) = 1$, $p = .359$), nor age ($\chi^2(1) = 4$, $p = .058$) were reliable predictors of whether the utterance was child-directed.

To further evaluate the distinguishability of child-directed and overheard speech, we conducted a linear discriminant analysis using the MASS library in R (Ripley et al., 2013), with a uniform prior on whether each data point was child-directed or overheard. We used only the coded speech variables that were not contingent on the child's own attention or behavior (that is, we included acoustic, semantic, and morphosyntactic variables, but not whether the child was looking at the speaker or referent). The loadings for each variable in the single linear discriminant function appear in the first column of Table 18. Echoing previous results, "sing-song prosody" was almost entirely responsible for distinguishing child-directed from overheard speech, with reference to the "here and now" serving as a very distant second. Referential gesture, speech clarity, utterance length, and morphological complexity did little to contribute to the between-group variance captured by the function.

Interestingly, child-directed speech was better identified than overheard speech, which bears directly on our hypothesis of greater within-class variance for speech that could be *overheard* by the child relative to speech directed *to* her. To assess the accuracy of the linear discriminant, we withheld 25% of the raw data as a test set. The function accurately classified 89% of the overheard utterances in our test set, and 99% of child-directed utterances, with an overall error rate of less than 1% (0.91%). Removing "sing-song prosody" from the function (loadings shown in second column of Table 18) and repeating the procedure with the same training and test data illustrates the critical contribution of prosody to distinguishing speech intended for the child in this household and age range. Without information about prosody, only 66% of overheard and 77% of child-directed utterances were accurately classified, with an increased error rate of 24%.

Table 17: *Logit Model Predicting Child-directed versus Overheard Utterance Status*

	Dependent variable:	
	Child-Directed {0, 1}	
Constant	0.03	(0.002, 0.40)
Here & Now (Child Looking at Referent)	8.88***	(3.09, 27.00)
Here & Now (Not Looking at Referent)	1.80	(0.79, 4.10)
Referential Gesture	1.77	(0.68, 5.00)
Child Looking at Speaker	2.42*	(1.16, 5.10)
Sing-song Prosody	369.76***	(190.68, 769.50)
Speech Clarity	1.11	(0.59, 2.00)
Morphological Complexity	0.59	(0.21, 1.80)
Utterance Length	0.97	(0.89, 1.00)
Age	1.06	(0.10, 1.10)
Observations		2,225
Log Likelihood		−167
Akaike Inf. Crit.		355

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 18: *Linear Discriminant Functions Classifying Utterances as Child-directed or Overheard*

Variable	+ Prosody Loading	− Prosody Loading
About the 'Here & Now'	0.19	1.57
Referential Gesture	0.07	0.86
Sing-Song Prosody	7.91	—
Speech Clarity	0.001	0.25
Morphological Complexity	−0.06	−0.98
Utterance Length	−0.002	−0.08

Does overheard speech change with child age?

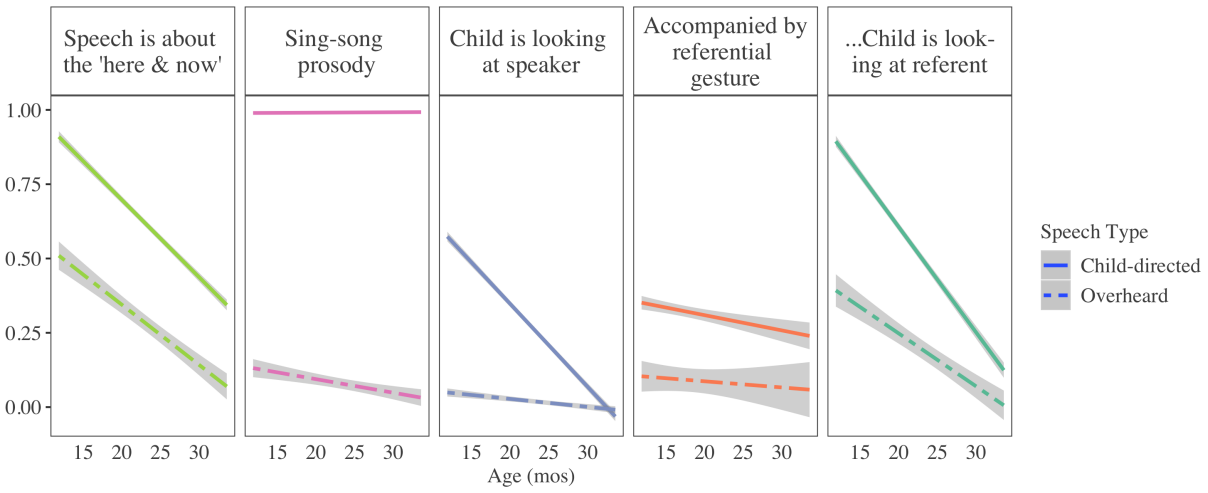


Figure 29: *Feature Frequency across Child Age.*

In child-directed speech, referential cues typically associated with caregiver modification like speech about the "here and now" ($r = -.5$, $[-.53, -.47]$, $p = .01$), utterance length ($r = .07$, $[.05, .10]$, $p < .001$) and referential gesture ($r = -.05$, $[-.1, -.01]$, $p = .05$) were correlated with child age, as was the likelihood that the child was looking at the speaker as they were talking ($r = -.57$, $[-.6, -.54]$, $p = .01$). Remarkably, qualitative *overheard* features also showed correlations with age. As in child-directed speech, caregiver talk about the "here and now" was negatively correlated with age ($r = -.43$, $[-.55, -.29]$, $p = .01$), along with the child's tendency to be looking at a speaker as they talked ($r = -.33$, $[-.46, -.18]$, $p = .01$). In contrast to the interpretable pattern of increasing utterance length in child-directed speech, in overheard speech, utterance length was negatively correlated with age ($r = -.14$, $[-.19, -.08]$). We speculate that this may reflect caregivers conducting fewer full-fledged conversations in Naima's vicinity, and exchanging more brief, functional utterances, more frequently interrupted by their now-verbal daughter. Finally, referential gesture in overheard speech was not correlated with child age, and in neither speech type was caregivers' prosody or morphological complexity related to the age of the child. This is surprising, as previous work suggests that caregivers' exaggerated prosody decreases as the child matures (Bornstein et al., 1992; Cooper & Aslin, 1990), while morphological complexity increases (Ervin-Tripp, 1978; Huttenlocher et al., 2007; Sherrod et al., 1977). We speculate that our result might be a reflection of (a) Naima's age in the study, and/or (b) Naima's caregivers' awareness that they were being recorded, which might have caused them to exaggerate the child-directed features of their speech. To further explore correlations among contextual features of the learning environment and age, please see Appendix O.

Table 19: *Logit Models Predicting Binary Features from Child Age*

	Child-directed			Overheard		
	Constant	Age		Constant	Age	
About the 'Here & Now'	38.95	0.88	(0.87, 0.88)	3.72	0.90	(0.88, 0.92)
Child Looking at Referent	11.49	0.84	(0.83, 0.85)	1.19	0.78	(0.64, 0.88)
Child Looking at Speaker	54.47	0.84	(0.83, 0.85)	2.93	0.88	(0.86, 0.91)
Referential Gesture	0.73	0.98	(0.96, 0.99)	0.16	0.97	(0.90, 1.04)
Sing-song Prosody	79.4	1.02	(0.98, 1.05)	0.33	0.94	(0.91, 0.97)

Finally, we fit models to the overheard and child-directed data for each binary speech quality, with age as the sole predictor. Exponentiated coefficients and confidence intervals for the effect of age are shown in Table 19.

Does overheard speech attract children's attention?

To better understand relations between speech qualities and child attention, we created a new, "child attention" variable that indexed whether the child was looking at either the speaker or the referent of an utterance. We fit another logit model to the by-utterance data, including all other speech qualities as predictors (see exponentiated coefficients and 95% confidence intervals in Table 20). "Sing-song prosody" was highly predictive of child attention (OR = 5.87 [4.34, 7.99], $\chi^2(1) = 146$, $p = .001$), as was "here and now" reference (OR = 3.83 [3.14, 4.68], $\chi^2(1) = 173$, $p = .001$). Speech clarity was also associated with child attention (OR = .71 [1.48, 1.98], $\chi^2(1) = 54$, $p < .001$), again possibly speaking to the role of proximity in eliciting or following the child's attention. Interestingly, age was negatively related to child attention (OR = 0.90 [0.89, 0.91], $\chi^2(1) = 352$, $p = .001$). This may likewise reflect increased independence and distance from her caregivers, or even increased capacity to distribute her attention, such that she can comprehend her caregivers' meaning without needing to look at them or the scene. Indeed, if Naima's prior gaze meant that she was seeking the referent of her caregivers' utterance, she will need to do so less with greater word knowledge.

Our structural variables were the only measures *not* reliably associated with Naima's visual attention (utterance length: OR = 1.01 [0.99, 1.04], $\chi^2(1) = 1$, $p = .23$; morphological complexity: OR = 1.17 [0.84, 1.64], $\chi^2(1) = 1$, $p = .36$). At face value, this result might appear to cast doubt on our premise of complexity as a key driver of child attention and learning. However, we suspect it might say more about the sensitivity of this measure of complexity. If nothing else, that our calculation of "morphological complexity" showed no relation to Naima's age in child-directed speech — especially during this critical period

of linguistic development — suggests that it may be ill-suited to capture the meaningful variation in language structure that we would expect to influence attention.

Table 20: *Logit Model Predicting Child Attention from Qualitative Dimensions of Speech*

	Dependent variable:
	Child Attention {0, 1}
Constant	0.33** (0.17, 0.64)
About the 'Here & Now'	3.83*** (3.14, 4.68)
Sing-song Prosody	5.87*** (4.34, 7.99)
Speech Clarity	1.71*** (1.48, 1.98)
Utterance Length	1.01 (0.99, 1.04)
Morphological Complexity	1.17 (0.84, 1.64)
Age	0.90*** (0.89, 0.91)
Observations	3,605
Log Likelihood	−1,485
Akaike Inf. Crit.	2,985

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
† Composite variable combining codes for child gaze toward the speaker and/or the referent.

4.4 General Discussion

Theories of early learning suggest that children's attention is motivated by an ongoing sense that they are making sense of incoming data (e.g., Balcomb & Gerken, 2008; Gerken et al., 2011; Houston-Price & Nakai, 2004; Hunter & Ames, 1988; Hunter et al., 1983). Our study analyzed the degree to which different sources of spoken language in a child's daily environment might support that sense, focusing especially on support for learning new words. We homed in on *referential transparency* as a demonstrably important dimension of language learning contexts that could be coded from video, and took a case study approach, capitalizing on longitudinal video recordings documenting the language environment of a single child — Naima, from the Providence corpus (Demuth et al., 2006). We analyzed over six thousand utterances spanning the first two years of Naima's life, when cues to words' meanings are argued to be especially critical for language development (e.g., Cartmill et al., 2013).

Our study is rare in considering the quality or learnability of *all* of the speech in the language learner's environment, including adult conversations that take place when the child is nearby, and even caregiver phone calls ('halfalogues'; Emberson et al., 2010). By applying

the same qualitative coding scheme to caregiver utterances coded as 'child-directed' versus 'overheard,' we find that greater referential transparency characterizes the set of utterances spoken to Naima directly. Child-directed utterances were more frequently about Naima's immediate context, rather than the past, future, or another place, and more frequently coincided with her current focus of attention. Child-directed utterances were also more frequently accompanied by physical behaviors like pointing and pantomime, which Naima could use to infer her caregivers' communicative intentions.

That the child-directed speech in our study appears highly supportive of word-learning is concordant with findings in other samples that the amount of child-directed speech that children receive during this period of development is predictive of their medium-term growth in vocabulary size (e.g., Huttenlocher et al., 2010; Ramírez-Esparza et al., 2017; Rowe, 2012; Shneidman et al., 2013; Shneidman and Goldin-Meadow, 2012a; Weisleder and Fernald, 2013; see Hoff, 2006 for a review). These studies typically limit their analyses to language addressed directly to children, rather than considering the range of language sources that young children experience over the course of a day. However, the rare studies that also analyze speech addressed to others consistently find no correlation between the amount of overheard speech regularly available to a child, and that child's level of language development (Ramírez-Esparza et al., 2017; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012a; Weisleder & Fernald, 2013), inviting researchers to conclude that children "do not readily make use of overheard input when learning words in naturalistic situations" (Shneidman et al., 2013, p. 7).

Our study was partly motivated by a potential measurement issue in these prior investigations: namely, that the heterogeneity of the overheard speech category might introduce a significant amount of noise into such correlational measures of learning. Not only are overheard utterances likely to be highly diverse, but we reason that child-directed utterances are likely to represent a significantly more homogeneous category in precisely the contexts where child-directed speech is typical — and typically marked. We see this study as confirming the hypothesis that overheard speech represents a less coherent category than speech directed to children; however, our data also reveal significant *within*-category variability for child-directed speech. These results are consistent with claims by language development researchers that input *quantity* (i.e., the total number of words the child hears) predicts language outcomes by virtue of input *quality* (i.e., dimensions of learnability like those we code here; Cartmill et al., 2013; B. Hart and Risley, 2003; Rowe et al., 2017). The idea is roughly that greater 'quantity' means a greater number of samples from the frequency distributions in Figures 27 and 28, and with those samples, greater opportunities for individual high-quality learning episodes.

Our fine-grained coding of the learning opportunities afforded by the overheard speech within a single child's home suggests that high-quality exposures to new words are not limited to child-directed utterances; however, they may be less likely to co-occur with the child's current focus of attention when overheard. This observation suggests new avenues of research: for example, how might children's attention be conditioned by the relative frequency and quality of the child-directed versus overheard speech in their daily environments?

That both child-directed and overheard speech were highly variable suggests a functional role for prosodic modification in discriminating two language sources that might be less naturally distinguished than previously assumed. Indeed, prosody's decisive role in identifying speech as intended for the child was borne out in our analyses, where classification error by a linear discriminant function skyrocketed when information from prosody was removed (Section 4.3). Caregivers' prosodic modification is especially interesting in light of our study's focus on *referential transparency*. In contrast to eye gaze or pointing (gestures we explicitly coded as "referential"), variable pitch does not in *itself* provide a disambiguating cue to reference. That is, while your mother's gesture to your father's coffee cup might help you infer the meaning of /kafi/, her melodic pronunciation does not. Learners may make use of caregivers' non–adult-like prosody not to decrypt language itself, as has been suggested in previous literatures. Instead, we hypothesize that prosody may be understood as a learned cue to highly transparent language data, and a self-reinforcing cue, as our data suggest that infants' attention has a higher probability of being rewarded when an utterance is in fact intended for them. Here again, Figures 27 and 28 provide a useful illustration of this point: prosody may mark an utterance as coming from the *green* distributions, which offer greater promise for learning — motivating simultaneously children's selective attention to child-directed speech and inattention to ambient overheard speech. This perspective is consistent with evidence that infants whose caregivers do not habitually acoustically exaggerate their speech show weaker or absent preferences and learning benefits from hearing exaggerated infant-directed speech in the lab (see e.g., Cristia, 2013; Soderstrom, 2007, for reviews). It is also reminiscent of evolutionary accounts of infant-directed song as a way for caregivers to signal attentional investment to their infants from afar (Mehr & Krasnow, 2017).

Nonetheless, we note that these results also come with a caveat, as our assessment of caregivers' "sing-song prosody" was highly impressionistic, potentially leading the code to reflect something like "child-directed register," rather than prosodic modulation, *per se*. In support of this hypothesis is the relatively low agreement between our initial coding and an independent reliability coder (70%) — though the fact that both parties also identified "sing-song prosody" in utterances coded as "overheard" suggests that they were not basing their acoustic assessment entirely on intended audience. Our coding of caregivers' prosody was also notably independent of Naima's age, despite the well-documented observation that caregivers typically reduce their acoustic exaggeration as children mature (e.g., Henning et al., 2005; Smith & Trainor, 2008). While it is possible that Naima's caregivers persisted in exaggerated 'baby talk' for the entirety of our study, it is also possible that they gradually reduced their exaggeration, but that this continuous trend was obscured by our binary "sing-song" code. To address these concerns, ongoing work further grounds our scheme in objective proxies for theoretically important variables. For example, to capture caregivers' prosodic modification, we use variability in the first formant of clips of speech, quantified via acoustic analysis software, rather than subjective coding of the "sing-song" quality of caregivers' utterances (Cristia, 2013).

4.5 Conclusion

Even in the absence of information about individual children's language outcomes, qualitative coding schemes like ours provide valuable vocabularies with which to describe early language environments, which are in turn useful for generating hypotheses and making contact with more humanistic fields like anthropology. Relative to child-directed speech, the unknowns of overheard speech are remarkably basic: *how variable are a child's overhearing experiences over the course of a day*, and *how does both the scale and quality of that variation compare across ages, versus across households, versus across cultures?*

Evidence for claims about cross-cultural differences in linguistic and child-rearing practices have typically taken the form of ethnographies (e.g., de León, 1998; Heath, 1983; Ochs, 1982; Schieffelin, 1990; Ward, 1971), which provide rich descriptions of community customs and beliefs, but make systematic comparisons between contexts difficult. We cannot build theories about the mechanisms underlying language development without a sense of how universal versus idiosyncratic the language environments that developmental scientists typically study are (Frank et al., 2017; Lieven, 1994; Ochs, 1990). It is difficult to understand how children transition to acquiring language in classroom contexts without understanding how the overheard input there — between the teacher and another student, or among nearby peers — compares to the overheard input before schooling. Likewise, we can develop better hypotheses about how young children's attention is organized if we can find patterns among features of non–child-directed contexts, and understand how those environments vary in their support for child participation, observation, and apprenticeship (Rogoff et al., 2003).

Language development research that continues to be focused on the impacts of child-directed speech may be missing nuances in how different environments are organized to support children's entry into the adult speech community (Leon, 1998; Ochs, 1990; Vogt et al., 2015). To this end, we are currently applying the scheme developed and analyzed here to naturalistic video corpora in Mandarin and Spanish. Coding of the Forrester (Forrester, 2002), Tong (Deng et al., 2018; Xiangjun & Yip, 2018) and Llinàs-Ojea corpora (Llinàs-Grau & Ojea Lopez, 2000) in CHILDES (MacWhinney, 2000) is ongoing, as is recruitment of speakers and novel video sources to capture the linguistic landscapes of learners across the world. In providing a common vocabulary with which to describe diverse milieux, we aim to bring the psychological and anthropological literatures into contact, such that theories of language development can be tested against the full range of children's linguistic lives.

Chapter 5

Ongoing Work and Future Directions

As in many fields, empirical methods in language development research have been dictated by what researchers expected to find, on the basis of their theoretic commitments and disciplinary training. As a consequence, empirical results whose methods derive from distinct academic perspectives are often difficult to put into contact, distributed as they are across methodologies, such that apparently conflicting conclusions are confounded with investigative approach. The forms of data from, say, ethnographies and laboratory experiments would be literally difficult to reconcile even if researchers from the respective fields sought to describe the same constructs, just with different tools — but they largely don't. Instead, the variables that are critical to one field's story of how (whether) linguistic experience drives linguistic development are often invariable or uncollected by the other. Two sources of ongoing work aim to address this problem, extending the ideas discussed in this dissertation to new contexts and populations, and developing new methods to bridge and translate these valuable bodies of evidence.

5.1 'Learning to Learn' in Language Development

As discussed in the Introduction, there is something of a disjuncture between the primary demographic source of our data on language acquisition — families that are white, Western, and wealthy — and the fact that we admire language acquisition precisely for its robustness across human contexts. In particular, by studying the efficacy of features of the early language environment in primarily Western settings, theories of how children learn language are arguably starting to assume that there is a single pathway to language, one that involves frequent adult speech to children even before children can meaningfully reply. Yet a parallel intellectual history in anthropology challenges this assumption by documenting wide global variation in child-directed speech customs, which nonetheless result in apparently similar timetables for language development. This opens the possibility of multiple routes to language-learning. More specifically, it suggests that while infants in Western contexts may have adapted to adults verbally engaging and maintaining their attention, in-

fants raised in contexts with little child-directed speech might develop distinct patterns of attention and skills at 'listening in' on speech around them in order to learn. Gathering data to bear on this hypothesis is critical for our understanding of what drives language development, and, more generally, of the degree to which learning strategies represent environmental adaptations, versus universal mechanisms. This question is impossible to answer if we restrict ourselves to a narrow subset of the world's learners, using metrics tailored to the developmental progression observed in their environments. In ongoing work, we apply the qualitative coding scheme developed in Chapter 4 to naturalistic video data from diverse contexts. In the following section, we design tests specific to the developmental context of infants in an indigenous society in southern Mexico, where prelinguistic infants experience the world from a sling on their mothers' backs.

5.2 Language Socialization in Tseltal Maya Infants

While ethnographies of cultures like the Tseltal Maya in Chiapas, Mexico report that infants and young children (1) are rarely addressed, and (2) must learn language through overhearing (see Lieven, 1994, for a review), other studies suggest that the value of child-directed speech remains even in these contexts. For example, Shneidman and Goldin-Meadow (2012a) found no correlation between the amount of speech Yucatec Maya infants overheard at 13 months and their vocabularies in the next year, but did find a correlation between vocabulary and the amount of *child-directed* speech they had received. However, as discussed in preceding chapter, measurements of 'overheard speech' in this and other studies represent a heterogeneous category of language input that is defined solely in the negative: that is, as any and all speech which is not child-directed. "Overheard speech" might therefore be directed to an adult, or to a different child; it might be more or less linguistically complex than the speech that the child typically receives; its auditory quality will vary with the position of the speaker, as will the availability of cues the overhearing child can use to understand what the speaker is talking about... This intense variability across the category of overheard speech makes it difficult to draw conclusions about exactly what might typically make it difficult to learn from — or what makes child-directed speech so beneficial. Several details of the socialization context of Tseltal Maya infants motivate the present study as an ideal way to test the 'best case scenario' for learning from overhearing.

First, based on observations by multiple generations of researchers (e.g., P. Brown, 2008; Abarbanell, p.c.), as well as our own data, Tseltal infants are truly almost never directly addressed during their first year of life. Only in later infancy — after they begin walking, and are no longer continuously attached to their mothers — do infants regularly become the direct recipients of speech. The child-directed speech that young children receive increases in quantity and complexity from this point (Shneidman & Goldin-Meadow, 2012a, 2012b).

Second, Tseltal infants tied to their mothers' backs find themselves in a naturally arising overhearing context that almost resembles laboratory experiments in its degree of control. While infants may not be spoken to, they are also almost never put down, or even passed to siblings, ensuring them front-row seats to the entirety of their mothers' social interactions.

Certainly when peeking over her shoulder, and possibly even when tucked beneath the hem of her shawl, an infant's privileged vantage is accompanied by high-quality audio and at least some access to their mother's focus of attention, which will often suggest the referent of her and others' speech.

Third, Tenejapa infants are exposed to several distinctive social speech signals that are never directed to them, yet share many of the features thought to make infant-directed utterances especially conducive to learning. In Tenejapa, greetings among adult community members are highly ritualized, occurring at high volume and with a unique prosodic contour. The specific greeting term depends on the sex, relative age, and sometimes familial relation or authority of the "greetee:" for example, a mother would typically greet a female peer with a call of *kantsil*, while an older woman would be greeted with *metik* (for audio recordings of these exchanges, please see *Supplemental Online Materials*). Here, these high-frequency items provide culturally-specific stimuli to test language that could *only* have been learned by overhearing, as such greetings would never be addressed to infants directly.

It is critical to test infants' language knowledge using an implicit measure as early in development as possible. Previous work suggests the effects of lower quantities of child-directed speech exposure compound as children age (Weisleder & Fernald, 2013). If this trend also holds in high-overhearing environments, then we expect testing infants as early in development as possible to provide the least biased comparison between word knowledge gleaned from the low- versus high-overhearing environments. Not only that, but there remains a question of the degree to which our existing metrics of vocabulary might be biased toward children in child-centered contexts.

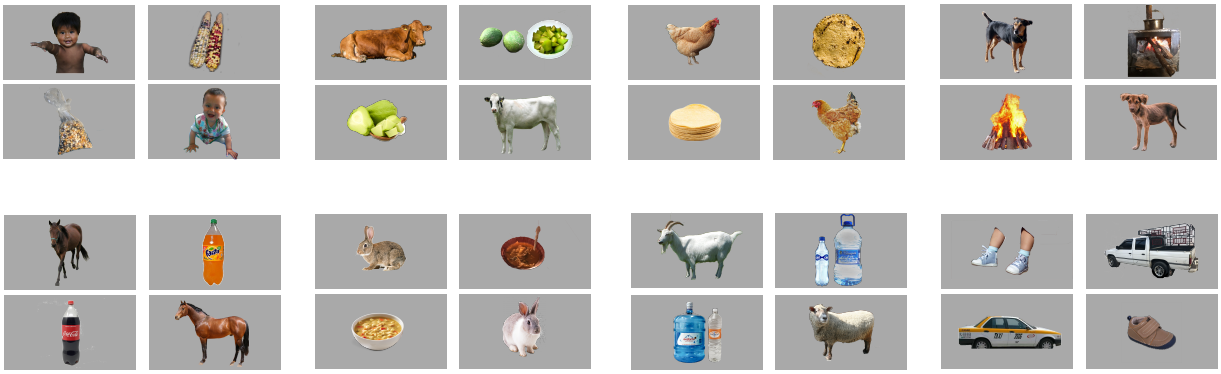


Figure 30: *Adapted Stimuli for Testing Early Knowledge of Common Nouns in Paired-Picture Trials.*

The language socialization literature emphasizes how the learner — her strategies, verbal behavior, and attention — adapt to the values and demands of her social environment (P. Brown, 2011; P. Brown & Gaskins, 2014; Rogoff et al., 2003; Schieffelin & Ochs, 1983). This might mean that explicitly testing receptive vocabulary (e.g., Shneidman & Goldin-Meadow, 2012a) may be less natural for children from language socialization contexts where they are

infrequently engaged directly, and are instead expected to observe. Likewise, inferring vocabulary from spontaneous language *production* (Mastin & Vogt, 2016) might systematically underestimate the word knowledge of children who are not encouraged to talk. Implicit measures of word knowledge evades these issues.

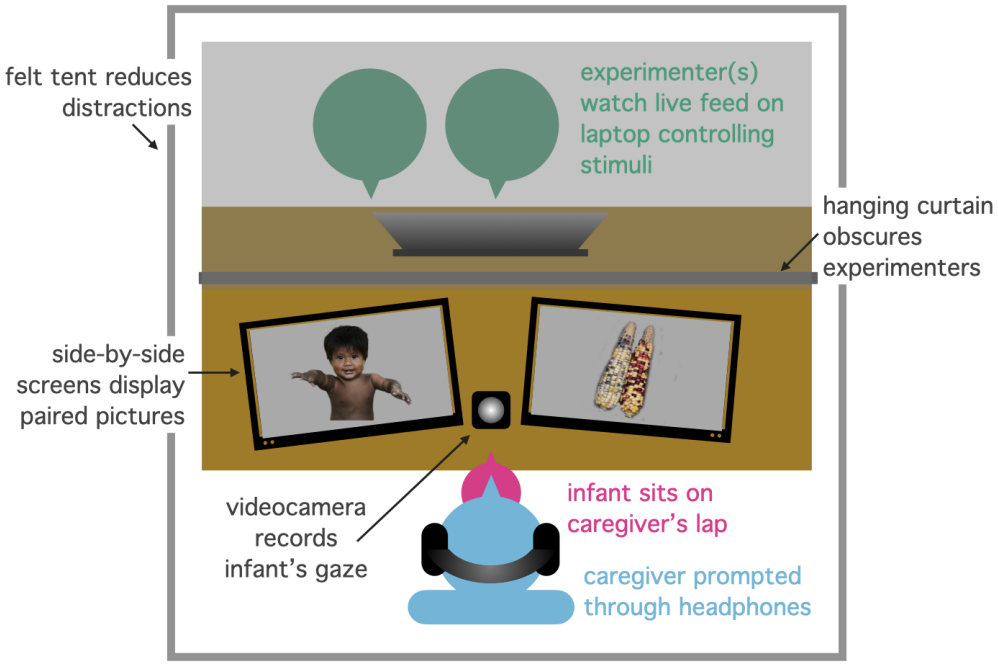


Figure 31: *Experimental Setup for Paired-Picture Trials.*

Responsive to these concerns, in ongoing work we adapt and extend a previously-established experimental method (Bergelson & Swingley, 2012), which uses relative looking time to pairs of picture stimuli to infer infants' earliest associations between word forms and referents. In addition to adapting the original method for a field context (where bringing and powering an eyetracker would be infeasible), we develop context-specific stimuli to test infants' implicit knowledge of common nouns (Figure 30) and of a set of honorifics regularly used as greetings among adult community members (Figure 32).

Figure 32: *Example Trial Testing Knowledge of Tseltal Honorifics for Greeting an* Older Man *versus a* Younger Woman.

Note. Target greeting was 'tatik' (Older Man).

5.3 Conclusion

Our research with infants growing up in Tenejapa, Chiapas is designed to shed light on the tension between ethnographies on the one hand, which emphasize that all children come to speak the language of their communities, and quantitative studies, which reach broad conclusions about the primacy of child-directed speech via analyses of children from largely Western, child-centric households. Evidence that child-directed speech is important in the contexts in which it is received is only condemnatory of learning by overhearing if we assume that learners are static, and learning mechanisms universal. In fact, the idea of 'learning to learn' has been central within the active learning framework since at least Bruner (1961). It may also be the case that our typical measures of language development — which reflect and reinforce ways of thinking about language as an acquisition object — are suited to pick up on precisely the components of language knowledge that adult-guided, child-directed interactions promote. By collecting quantitative data in a series of carefully designed and context-specific experiments that are sensitive to the socialization environment of Tseltal infants and their mothers, we hope to be contributing to the mutual intelligibility of qualitative ethnographic and psycholinguistic accounts. By testing culturally specific greeting routines at this young age, we also hope to be expanding our (testable) notion of what constitutes language knowledge to include familiarity with linguistic *practice*, in addition to vocabulary. Regardless of how children perform, the highly controlled nature of Tseltal infants' naturalistic overhearing context means that this work will help us better understand what it is about different linguistic inputs and socialization environments that makes them more or less supportive of language-learning, as well as how (whether) young learners adapt to the affordances of their particular language environments.

Conclusion

In contrast to the majority of current rhetoric around language development, this dissertation adopts a view of the language-learning child as an active, or self-directed, language learner. Of the manifold skills and strategies of the self-directed learner (Bruner, 1961; Chi, 2009; Gureckis & Markant, 2012), the dissertation focused on children's strategic attention allocation and independent information-gathering in explaining a puzzle in the extant literature. Namely:

(1) Across studies, the amount of child-directed — but not overheard — speech in children's early environments predicts their later vocabularies.

   (a) This remains true even in contexts where infants are rarely spoken to directly, and where overhearing instead predominates.

(2) In well-controlled lab studies, children as young as 18 months of age can learn a new object label through overhearing.

(3) All children ultimately become competent speakers of their native tongues.

A constellation of experimental and observational studies sought evidence for an intuitive solution, leveraging the idea from the active learning framework that children rationally allocate their attention to stimuli based on its complexity or learnability. We hypothesized that overheard speech is likely to be highly linguistically complex, such that young rational learners may initially ignore it. However, with greater language knowledge, the difference between the subjective complexity of the speech that children receive in interactions with adults and the overheard speech available for them to 'tune in' to in their environments will shrink, such that older children may be able to reliably benefit from overheard language as a source of input for word-learning.

The studies in Chapters 1–4 suggest that our solution may be on the right track:

Chapter 1 established that the speech that children receive directly is reliably simpler along multiple learning-relevant dimensions, relative to typical adult-directed speech. This remains true in aggregate through at least the first four years of development, accounting for the period in which previous studies have failed to uncover a correlation between overheard speech quantity and child vocabulary growth.

Chapter 2 built on previous experimental methods to test preschool-aged children's ability to learn from naturalistic overheard speech. In contrast to previous studies, our experimental overheard speech was adult-directed and contained *multiple* learning targets, including both novel nouns and facts. While all children reliably learned facts composed of familiar words, only older preschoolers (4.5–6 years) reliably learned the set of novel words through overhearing; younger preschoolers (3–4.5 years) apparently struggled. Analyses of children's play and gaze during the overheard phone call suggested that older children were better able to coordinate their attention between the speech and the referential context — though even younger children showed evidence of attention to the overheard speech.

Chapter 3 suggested that children's attention to naturalistic language stimuli is responsive to its complexity, and to children's ability to learn from it. There, complexity and age interacted, such that older children were significantly less likely than younger children to disattend to the naturalistic spoken language *only* when it was highly complex.

Chapter 4 introduced contextual dimensions that inform the difficulty of learning from an utterance into our conceptualization of language complexity. In a case study of the complexity landscape surrounding a single child learner, child-directed and overheard utterances exhibited overlapping distributions of multiple learning-relevant features, including reference to the 'here and now' and even exaggerated prosody. Nevertheless, referential transparency was significantly higher in child-directed utterances, which were also more likely to be prosodically marked and associated with the child's attention — suggesting a role for learning even of the cue structure of the early language environment.

Chapter 5 described ongoing work extending the idea of the learner as adapted to their environment by testing early language knowledge in infants exclusively exposed to (high quality) overheard speech. If the preceding chapters in the dissertation sought to expand our notions of how language knowledge is acquired (i.e., by overhearing, in addition to being taught), then our work with Tseltal Maya infants is partly aimed at expanding our notions of what counts as legitimate language knowledge. Specifically, we capitalize on Tseltal's rich system of honorifics to test infants' knowledge of the social language conventions used in the speech community that they are in the process of joining.

Ultimately, our approach to this particular puzzle should be understood more generally as demonstrating the potential in a research program at the intersection of active learning and language development — especially one with an eye toward ecologically valid demonstrations of children's abilities. Diverse empirical questions lie at this intersection. As an example, children use newly-encountered words long before they have adult-like semantics for them, refining their reference via further observations and feedback from their interlocutors. Can we see children's productions — which garner informative feedback, whether confusion, explicit corrections, or acceptance — as akin to hypothesis-testing? Are children

more likely to produce a word whose semantics they have inconsistent evidence for when there is a knowledgeable adult around?

As the work in this dissertation hopefully reveals, reframing the child as an "active" language learner can introduce novel explanations for phenomena in the development of language. At the same time, using language as a test domain for formal accounts of rational learning can provide researchers with complex learning tasks that not only make sense to children, but are truly informative of how children navigate the complexity within their own lives.

Bibliography

Abel, J., & Babel, M. (2016). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*.

Adi-Bensaid, L., Ben-David, A., & Tubul-Lavy, G. (2015). Content words in hebrew child-directed speech. *Infant behavior and development*, *40*, 231–241.

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430. https://doi.org/10.1111/1467-8624.00287

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of child language*, *42*(2), 239–273.

Antonacci, P. A. (2000). Reading in the zone of proximal development: Mediating literacy development in beginner readers through guided reading. *Reading Horizons: A Journal of Literacy and Language Arts*, *41*.

Aquino, J. M., & Arnell, K. M. (2007). Attention and the processing of emotional words: Dissociating effects of arousal. *Psychonomic Bulletin & Review*, *14*(3), 430–435.

Arunachalam, S. (2013). Two-year-olds can begin to acquire verb meanings in socially impoverished contexts. *Cognition*, *129*(3), 569–573. https://doi.org/10.1016/j.cognition.2013.08.021

Arunachalam, S. (2016). A new experimental paradigm to study children's processing of their parent's unscripted language input. *Journal of Memory and Language*, *88*, 104–116. https://doi.org/10.1016/j.jml.2016.02.001

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53.

Baars, B. J. (1986). *The cognitive revolution in psychology*. Guilford Press.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(MAR), 456. https://doi.org/10.3389/fpsyg.2017.00456

Bakeman, R., & Adamson, L. B. (2019). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *55*(4), 1278–1289.

Balcomb, F. K., & Gerken, L. A. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, *11*(5), 750–760. https://doi.org/10.1111/j.1467-7687.2008.00725.x

Baldwin, D. A. (1991). Infants' Contribution to the Achievement of Joint Reference. *Child Development*, *62*(5), 874–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, *67*(6), 3135–3153.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance*, *10*(3), 340.

Balota, D. A., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., Neely, J., Nelson, D., Simpson, G., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Bannard, C., & Tomasello, M. (2012). Can We Dissociate Contingency Learning from Social Learning in Word Acquisition by 24-Month-Olds? *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0049881

Barnes, S., Gutfreund, M., Satterly, D., & Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of Child Language*, *10*(1), 65–84. https://doi.org/10.1017/S0305000900005146

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., B., B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113.

Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 3–28). Springer.

Beckage, N. M., Smith, L., & Hills, T. (2011). Small Worlds and Semantic Network Growth in Typical and Late Talkers (M. Perc, Ed.). *PLoS ONE*, *6*(5), e19348. https://doi.org/10.1371/journal.pone.0019348

Begus, K., Gliga, T., & Southgate, V. (2014). Infants Learn What They Want to Learn: Responding to Infant Pointing Leads to Superior Learning. *PLoS ONE*, *9*(10), e108817. https://doi.org/10.1371/journal.pone.0108817

Belfatti, M. A. (2012). *CONTESTING NONFICTION : FOURTH GRADERS MAKING SENSE OF WORDS AND IMAGES IN SCIENCE INFORMATION BOOK DISCUSSIONS in Education Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor o.*

Belsky, J., Goode, M. K., & Most, R. K. (1980). Maternal stimulation and infant exploratory competence: Cross-sectional, correlational, and experimental analyses. *Child development*, 1168–1178.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. McGraw-Hill Book Company.

Blewitt, P. (1983). Dog versus collie: Vocabulary in speech to young children. *Developmental Psychology*, *19*(4), 602–609. https://doi.org/10.1037/0012-1649.19.4.602

Bohannon, J. N., Stine, E. L., & Ritzenberg, D. (1982). The "fine-tuning" hypothesis of adult speech to children: Effects of experience and feedback. *Bulletin of the Psychonomic Society*, *19*(4), 201–204.

Bonawitz, E. B., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. E. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of-acquisition and word frequency in the lexical decision task: Further evidence. *Current Psychology of Cognition*, *20*(6), 401–443.

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, *4*(3), 179–202.

Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pecheux, M.-G., Lamour, M., Toda, S., Azuma, H., Ogino, M., & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, france, japan, and the united states. *Developmental Psychology*, *28*(4), 593.

Braginsky, M., Sanchez, A., Yurovsky, D., MacDonald, K., & Meylan, S. (2019). Childes-r. https://github.com/langcog/childesr

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163–197.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57.

Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition*, *93*(2), B59–B68.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274–291.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, 31–44

Brent, M. R. & Siskind, J. M (2001). The role of exposure to isolated words in early vocabulary development. Cognition, 81, 31-44.

Broen, P. A. (1972). The verbal environment of the language-learning child. asha monographs, no. 17.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31–43. https://doi.org/10.1080/15248372.2013.833923

Broesch, T. L., & Bryant, G. A. (2018). Fathers' infant-directed speech in a small-scale society. *Child Development*, *89*(2), e29–e41. https://doi.org/10.1111/cdev.12768

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*(1), 207.

Brown, P. (1998). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics*, *36*(4), 713–753.

Brown, P. (2008). Conversational Structure and Language Acquisition: The Role of Repetition in Tzeltal. *Journal of Linguistic Anthropology*, *8*(2), 197–221. https://doi.org/10.1525/jlin.1998.8.2.197

Brown, P. (2011). The cultural organization of attention. *The handbook of language socialization* (pp. 29–55). Wiley Online Library.

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. *Cambridge handbook of linguistic anthropology* (pp. 187–226). Cambridge University Press.

Brown, R. (1973). *A first language: The early stages*. Harvard U. Press.

Brownell, R. (2000). Expressive one-word picture vocabulary test.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand english word meanings. *Behavior research methods*, *49*(4), 1520–1523.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215–226.

Buchsbaum, D., Seiver, E., Bridgers, S., & Gopnik, A. (2012). Learning about causes from people and about people as causes: Probabilistic models and social causal reasoning. *Advances in child development and behavior* (pp. 125–160). Elsevier.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*(2), 68–76.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*(5), 654.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Caron, R. F., & Caron, A. J. (1969). Degree of stimulus complexity and habituation of visual fixation in infants. *Psychonomic Science*, *14*(2), 78–79.

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.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, *25*(1), 85–95.

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 of the United States of America*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early Language Experience in a Tseltal Mayan Village. *Child Development*. https://doi.org/10.1111/cdev.13349

Champely, S. (2014). *Basic functions for power analysis*. R Foundation for Statistical Computing. Vienna, Austria
Package 'pwr'.

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

Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*(1), 29–39. https://doi.org/10.1037/0012-1649.19.1.29

Choi, S. (2000). Caregiver input in english and korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, *27*(1), 69–96.

Chomsky, N. (1957). Syntactic Structures, 117. https://doi.org/10.1515/9783110218329

Chomsky, N. (1959). A review of bf skinner's verbal behavior. *Language*, *35*(1), 26–58.

Clark, A., & Lappin, S. (2013). Complexity in Language Acquisition. *Topics in Cognitive Science*.

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*(1), 1–23. https://doi.org/10.1075/gest.11.1.01cla

Clark, H., Herbert, & Brennan, S. E. (1991). Grounding in communication.

Clark, H. H. (2014). How to talk with children.

Clark, H. H., & Brennan, S. E. (2004). Grounding in communication. *Perspectives on socially shared cognition.* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child development*, 419–424.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of learning and memory*, *92*(2), 225–234.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349. https://doi.org/10.1016/j.cognition.2011.03.003

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

Correa-Chávez, M., & Rogoff, B. (2009). Children's Attention to Interactions Directed to Others: Guatemalan Mayan and European American Patterns. *Developmental Psychology*, *45*(3), 630–641. https://doi.org/10.1037/a0014144

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Linguistics and Language Compass*. https://doi.org/10.1111/lnc3.12015

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2020). A thorough evaluation of the language environment analysis (lena) system. *Behavior Research Methods*, 1–20.

Cross, T. G. (1977). Mothers' speech adjustments: The contribution of selected child listener variables.

Cross, T. G., & Morris, J. E. (1980). Linguistic feedback and maternal speech: Comparisons of mothers addressing infants, one-year-olds and two-year-olds. *First Language*, *1*(2), 98–121.

Datavyu Team. (2014). *Datavyu: A Video Coding Tool. Databrary Project, New York University*. Databrary Project. New York University.

de León, L. (1998). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2).

Deák, G. O., & Toney, A. J. (2013). Young children's fast mapping and generalization of words, facts, and pictograms. *Journal of Experimental Child Psychology*, *115*(2), 273–296. https://doi.org/10.1016/j.jecp.2013.02.004

de Carvalho, A., Babineau, M., Trueswell, J. C., Waxman, S. R., & Christophe, A. (2019). Studying the real-time interpretation of novel noun and verb meanings in young children. *Frontiers in Psychology*, *10*, 274. https://doi.org/10.3389/fpsyg.2019.00274

Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 415–432.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–174. https://doi.org/10.1177/00238309060490020201

Deng, X., Mai, Z., & Yip, V. (2018). An aspectual account of ba and bei constructions in child mandarin. *First Language*, *38*(3), 243–262.

Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: The role of child directed speech and joint attention. *Journal of Neurolinguistics*, *17*(2-3), 121–145.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000). Santa Barbara corpus of spoken American English, Parts 1–4
Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000–2005. Santa Barbara corpus of spoken American English, Parts 1–4. Philadelphia: Linguistic Data Consortium.

Dunn, L. M., & Dunn, L. M. (1981). Peabody picture vocabulary test-revised.

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

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental cognitive neuroscience*, *25*, 69–91.

Eimas. (1971). Infant speech discrimination. *Science*.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 515.

Ellis, R., & Wells, G. (1977). Enabling Factors in Adult-Child Discourse, 46–62.

Ellwood-Lowe, M., Foushee, R., & Srinivasan, M. (2020). What causes the word gap? financial concerns may systematically suppress child-directed speech. https://doi.org/10.31234/osf.io/byp4k

Emberson, L. L., Lupyan, G., Goldstein, M. H., & Spivey, M. J. (2010). Overheard cell-phone conversations. *Psychological Science*, *21*(10), 1383–1388. https://doi.org/10.1177/0956797610382126

Ervin-Tripp, S. (1978). Some features of early child-adult dialogues. *Language in Society*, *7*(3), 357–373.

Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: Evidence from eye movement monitoring. *Psychological Science*, *16*(11), 913–920. https://doi.org/10.1111/j.1467-9280.2005.01636.x

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670.

Farran, L. K., Lee, C. C., Yoo, H., & Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: An initial study. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0151518

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *Macarthur-bates communicative development inventories: User's guide and technical manual (2nd ed.)* (2nd ed.). Baltimore, MD: Brookes.

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15- and 19-month-olds. *Cognition*, *131*(1), 139–146. https://doi.org/10.1016/j.cognition.2013.12.014

Ferguson, B., Graf, E., & Waxman, S. R. (2018). When Veps Cry: Two-Year-Olds Efficiently Learn Novel Words from Linguistic Contexts Alone. *Language Learning and Development*, *14*(1), 1–12. https://doi.org/10.1080/15475441.2017.1311260

Ferguson, C. A. (1964). Baby Talk in Six Languages. *American Anthropologist*, *66*(6), 103–114. https://doi.org/10.1525/aa.1964.66.suppl{_}3.02a00060

Ferguson, C. A. (1977). Baby talk as a simplified register. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 209–235). Cambridge University Press.

Ferguson, C. A. (1997). Arabic baby talk. *Structuralist studies in arabic linguistics* (pp. 179–187). Brill.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. *The origins and growth of communication*, 5–29.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental science*, *9*(3), F33–F40.

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

Fernald, A., & Morikawa, H. (1993). Common Themes and Cultural Variations in Japanese and American Mothers' Speech to Infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.1111/j.1467-8624.1993.tb02933.x

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

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of child language*, *16*(3), 477–501.

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.

Fischer, K. (2016). *Designing speech for a recipient: The roles of partner modeling, alignment and feedback in so-called 'simplified registers'* (Vol. 270). John Benjamins Publishing Company.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child development*, *67*(6), 3192–3218.

Fitch, A., Lieberman, A. M., Luyster, R. J., & Arunachalam, S. (2020). Toddlers' word learning through overhearing: Others' attention matters. *Journal of Experimental Child Psychology*, *193*, 104793. https://doi.org/10.1016/j.jecp.2019.104793

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*(3), 327–339. https://doi.org/10.1207/s15327078in0903_4

Forrester, M. A. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood*, *9*(3), 255–276.

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847.

Foushee, R., Griffiths, T. L., & Srinivasan, M. (2016). Lexical Complexity of Child-Directed and Overheard Speech : Implications for Learning. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 1697–1702.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*(1), 35–53. https://doi.org/10.1080/01638539909545049

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., et al. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435.

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.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2015). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*
Frank, M. C., Braginsky, M., Yurovsky, D., Marchman, V. A. (under revision). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language.

Fraser, C., & Roberts, N. (1975). Mothers' speech to children of four different ages. *Journal of psycholinguistic research*, *4*(1), 9–16
CDS over time.

Furrow, D., Nelson, K., & Benedict, H. (1979). Mothers' speech to children and syntactic development: Some simple relationships. *Journal of Child Language*, *6*(3), 423–442. https://doi.org/10.1017/S0305000900002464

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*(1), 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language*, *32*(3), 385–397. https://doi.org/10.1177/0142723711433584

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. Basic Books.

Genovese, G., Spinelli, M., Romero Lauro, L. J., Aureli, T., Castelletti, G., & Fasolo, M. (2020). Infant-directed speech as a simplified but not simple register: A longitudinal study of lexical and syntactic features. *Journal of Child Language*, *47*(1), 22–44. https://doi.org/10.1017/S0305000919000643

Gerhand, S., & Barry, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & cognition*, *27*(4), 592–602.

Gerken, L. A., 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. https://doi.org/10.1111/j.1467-7687.2011.01046.x

Ghyselinck, M., Custers, R., & Brysbaert, M. (2004). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 550.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta psychologica*, *115*(1), 43–67.

Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, *71*(1), 105–110.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395–427. https://doi.org/10.3758/BF03201693

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). *Hard Words* (tech. rep. No. 1).

Goldin-Meadow, S. (2015). Studying the mechanisms of language learning by varying the learning environment and the learner. *Language, cognition and neuroscience*, *30*(8), 899–911.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. https://doi.org/10.1017/S030500090001271X

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current directions in psychological science*, *15*(1), 30–33.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. *Child Development*, *90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Goodman, J. C., McDonough, L., & Brown, N. B. (2008). The role of semantic context and memory in the acquisition of novel nouns. *Child Development*, *69*(5), 1330–1344. https://doi.org/10.1111/j.1467-8624.1998.tb06215.x

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1200. https://doi.org/10.3389/fpsyg.2015.01200

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology*, *24*(1), 14.

Guevara-Rukoz, A., Cristia, A., Ludusan, B., Thiollière, R., Martin, A., Mazuka, R., & Dupoux, E. (2018). Are words easier to learn from infant-than adult-directed speech? a quantitative corpus-based investigation. *Cognitive science*, *42*(5), 1586–1617.

Gureckis, T. M., & Markant, D. B. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, *7*(5), 464–481. https://doi.org/10.1177/1745691612454304

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural Ways of Learning: Individual Traits or Repertoires of Practice. *Educational Researcher*, *32*(5), 19–25. https://doi.org/10.3102/0013189X032005019

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208.

Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. cambridge university Press.

Hembacher, E., & Frank, M. C. (2017). Children's social referencing reflects sensitivity to graded uncertainty. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 495–500.

Henning, A., Striano, T., & Lieven, E. V. M. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, *28*(4), 519–536. https://doi.org/10.1016/j.infbeh.2005.06.001

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*, 453–456. https://doi.org/10.1038/nature02634

Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition*, *143*, 87–92. https://doi.org/10.1016/j.cognition.2015.06.009

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Holzrichter, A. S., & Meier, R. P. (2000). Child-directed signing in american sign language. *Language acquisition by eye*, 25–40.

Horowitz, F. D., Paden, L., Bhana, K., & Self, P. (1972). An infant-control procedure for studying infant visual fixations.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Houston, D. M., & Bergeson, T. R. (2014). Hearing versus listening: Attention to speech and its role in language acquisition in deaf infants with cochlear implants. *Lingua*, *139*, 10–25. https://doi.org/10.1016/j.lingua.2013.08.001

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341–348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental psychology*, *27*(2), 236.

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, *43*(5), 1062–1083.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive psychology*, *61*(4), 343–365.

Imbir, K. K., Spustek, T., & Żygierewicz, J. (2016). Effects of valence and origin of emotions in word processing evidenced by event related potential correlates in a lexical decision task. *Frontiers in psychology*, *7*, 271.

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicologica*, 245–281.

Jackson-Maldonado, D., Thal, D. J., Marchman, V. A., Newton, T., Fenson, L., & Conboy, B. T. (2003). Macarthur inventarios del desarrollo de habilidades comunicativas: User's guide and technical manual.

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 323–335.

Kaplan, P. S., Danko, C. M., Cejka, A. M., & Everhart, K. D. (2015). Maternal depression and the learning-promoting effects of infant-directed speech: Roles of maternal sensitivity, depression diagnosis, and speech acoustic cues. *Infant Behavior and Development*, *41*, 52–63.

Kaplan, P. S., Jung, P. C., Ryther, J. S., & Zarlengo-Strouse, P. (1996). Infant-directed versus adult-directed speech as signals for faces. *Developmental Psychology*, *32*(5), 880.

Kempe, V. (2009). Child-directed speech prosody in adolescents: Relationship to 2d: 4d, empathy, and attitudes towards children. *Personality and individual differences*, *47*(6), 610–615.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current directions in psychological science*, *7*(2), 46–49.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kline, M., & Snedeker, J. (2015). 2-Year-Olds Use Syntax To Infer Actor Intentions in a Rational - Action Paradigm. *Cognitive Science*, 1135–1140.

Ko, E.-S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua*, *122*(8), 841–857.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Kuchinke, L., Jacobs, A. M., Grubich, C., Vo, M. L.-H., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: An fmri study. *Neuroimage*, *28*(4), 1022–1032.

Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). *Age-of-acquisition ratings for 30, 000 {English} words*.

Lam, C., & Kitamura, C. (2009). Infant-directed speech to infants with a simulated hearing loss. *The Journal of the Acoustical Society of America*, *125*(4), 2533–2533.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press.

Leon, L. D. (1998). The Emergent Participant: Interactive Patterns in the Socialization of Tzotzil (Mayan) Infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161. https://doi.org/10.1525/jlin.1998.8.2.131

León, L. D. (1999). Verbs in Tzotzil (Mayan) early syntactic development. *International Journal of Bilingualism*, *3*(2), 219–239.

Lieven, E. V. (1994). Crosslinguistic and crosscultural aspects of language addressed to children.

Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically*. IAP. http://www.abebooks.co.uk/servlet/SearchResults?sts=t%7B%5C&%7Dtn=rethinking+language,+mind+,+and+world+dialogically

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of mandarin maternal speech to preverbal infants and five-year-old children: A

longitudinal study. *Journal of Child Language*, *36*(4), 909–922. https://doi.org/10.1017/S030500090800929X

Llinàs-Grau, M., & Ojea Lopez, A. I. (2000). The llinasojea corpus.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The spoken bnc2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3), 319–344.

Lovibond, S. (1969). Habituation of the orienting response to multiple stimulus sequences. *Psychophysiology*, *5*(4), 435–439.

Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science*, *21*(6). https://doi.org/10.1111/desc.12663

Lupyan, G. (2008). From Chair to "Chair": A Representational Shift Account of Object Labeling Effects on Memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369. https://doi.org/10.1037/0096-3445.137.2.348

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*(MAR), 54. https://doi.org/10.3389/fpsyg.2012.00054

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

MacWhinney, B. (2000). *The Database* (3rd ed., Vol. 2). Lawrence Erlbaum Associates
MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. *Corpora in Language Acquisition Research: History, Methods, Perspectives*, *6*, 165–197. http://purl.org/net/MacWhinney-08.pdf

Maltzman, I., & Mandell, M. P. (1968). The orienting reflex as a predictor of learning and performance. *Journal of Experimental Research in Personality*, *3*(2), 99–106.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*(6619), 813–815. https://doi.org/10.1038/385813a0

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Martin, R. M. (1975). Effects of familiar and complex stimuli on infant attention. *Developmental Psychology*, *11*(2), 178.

Martínez-Sussmann, C., Akhtar, N., Diesendruck, G., & Markson, L. (2011). Orienting to third-party conversations. *Journal of Child Language*, *38*(2), 273–296. https://www

.cambridge.org/core/product/identifier/S0305000909990274/type/journal%7B%5C_%7Darticle

Masisi, L., Nelwamondo, V., & Marwala, T. (2008). The use of entropy to measure structural diversity. *2008 IEEE International Conference on Computational Cybernetics*, 41–45.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235–264. https://doi.org/10.1017/S0305000915000148

Mayer, M. (1969). *Frog, where are you?* Dial Press.

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769–785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684.

Mervis, C. B. (1983). Acquisition of a lexicon. *Contemporary Educational Psychology*, *8*(3), 210–236. https://doi.org/10.1016/0361-476X(83)90015-2

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning Verb Syntax via Listening: New Evidence From 22-Month-Olds. *Language Learning and Development*, *11*(4), 356–368. https://doi.org/10.1080/15475441.2014.978331

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A. L., De Schryver, M., De Winne, J., & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–177. https://doi.org/10.3758/s13428-012-0243-8

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, *91*(2), 167–180. https://doi.org/10.1348/000712600161763

Murray, A. D., Johnson, J., & Peters, J. (1990). Fine-tuning of utterance length to preverbal infants: Effects on later language development. *Journal of Child Language*, *17*(3), 511–525. https://doi.org/10.1017/S0305000900010862

Naigles, L. (1990). Children Use Syntax To Learn Verb Meanings. *Journal of Child Language*, *17*(2), 357–374. https://doi.org/10.1017/S0305000900013817

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion*, *9*(5), 609.

Navarrette, R. (2014). A toddler, a cupcake, and a mob of critics. https://www.cnn.com/2014/03/20/opinion/navarrette-boy-cupcake/index.html

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of child Language*, *16*(1), 55–68.

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158–1173. https://doi.org/10.1017/S0305000915000446

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. *Talking to children: Language input and interaction*, (JANUARY 1977), 109–150.

Niwano, K., & Sugai, K. (2002). Intonation contour of japanese maternal infant-directed speech and infant vocal response. *The Japanese Journal of Special Education*, *39*(6), 59–68.

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization. *Culture theory: Essays on mind, self and emotion*, 276–320.

Ochs, E., & Schieffelin, B. (1995). The Impact of Language Socialization on Grammatical Development. *The Handbook of Child Language*, 73–94.

Ochs, E. (1982). Talking to children in Western Samoa. *Language in Society*, *11*, 77–104.

Ochs, E. (1990). Cultural universals in the acquisition of language. *Papers and Reports on Child Language Development*, *29*, 1–19.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-Party Social Interaction and Word Learning From Video. *Child Development*, *82*(3), 902–915. https://doi.org/10.1111/j.1467-8624.2011.01579.x

Orr, W. C., & Stern, J. A. (1970). Effect of stimulus information on habituation rate. *Psychophysiology* (p. 625). Cambridge University Press.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive science*, *42*(6), 1974–1999.

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2015). Young children's self-directed information gathering on touchscreens. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 1835–1840. http://mindmodeling.org/cogsci2015/papers/0318/paper0318.pdf

Partridge, E., Mcgovern, M. G., Yung, A., & Kidd, C. (2012). Young Children's Self-Directed Information Gathering on Touchscreens.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental psychology*, *43*(5), 1216.

Peters, A. M. (1987). The role of imitation in the developing syntax of a blind child. *Text*, *7*, 289–311

Peters, A. (1987). The role of imitation in the developing syntax of a blind child.

Text, 7, 289–311.Wilson, B., & Peters, A. M. (1988). What are you cookin' on a hot?: Movement constraints in the speech of a three-year-old blind child. Language, 64, 249–273.

Phillips, J. R. (1973). Syntax and vocabulary of mothers' speech to young children: Age and sex comparisons. *Child development*, 182–185.

Piaget, J. (1954). *The construction of reality in the child*. Routledge. https://doi.org/10.4324/9781315009650

Piantadosi, S. T., Kidd, C., & Aslin, R. (2014). Rich analysis and rational models: Inferring individual behavior from infant looking data. *Developmental Science*, *17*(3), 321–337. https://doi.org/10.1111/desc.12083

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190. https://doi.org/10.1017/s0140525x04000056

Pine, J. M. (1994). Referential style and maternal directiveness: Different measures yield different results. *Applied Psycholinguistics*, *15*(2), 135–148.

Pye, C. (1986a). An Ethnography of Mayan Speech to Children. *Working Papers in Child Language*, *1*, 30–58.

Pye, C. (1986b). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100. https://doi.org/10.1017/S0305000900000313

R Development Core Team. (2020). *A Language and Environment for Statistical Computing* (Vol. 2). R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Rafferty, A. N., & Griffiths, T. L. (2010). Optimal Language Learning: {T}he Importance of Starting Representative. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 2069–2074.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*(JUN), 1008. https://doi.org/10.3389/fpsyg.2017.01008

Raneri, D., Von Holzen, K., Newman, R., & Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, 1–13. https://doi.org/10.1017/S030500091900093X

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition*, *178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Ratner, N. B., & Pye, C. (1984). Higher pitch in bt is not universal: Acoustic evidence from quiche mayan. *Journal of child language*, *11*(3), 515–522.

Reed, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Learning on hold: Cell phones sidetrack parent-child interactions. *Developmental Psychology*, *53*(8), 1428–1436. https://doi.org/10.1037/dev0000292

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental psychology*, *55*(8), 1656.

Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*, 215–219.

Ringler, N. M. (1981). The development of language and how adults talk to children. *Infant Mental Health Journal*, *2*(2), 71–83.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran R.*, *538*.

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001

Roeper, T. (2013).

Rogoff, B. et al. (2003). *The cultural nature of human development*. Oxford university press.

Rollins, P. R. (2003). Caregiver contingent comments and subsequent vocabulary comprehension. *Applied Psycholinguistics*, *24*, 221–234

Rollins, P. R., (2003) Caregiver contingent comments and subsequent vocabulary Comprehension. Applied Psycholinguistics. 24, 221-234.

Rollins, P. R., & Trautman, C. H. (2006). Child-centered behaviors of caregivers with 12-month-old infants: Associations with passive joint engagement and later language. *Journal of Applied Psycholinguistics*, *27*, 447–463

Trautman, C.H., & Rollins, P.R. (2006) Child-centered behaviors of caregivers with 12-month-old infants: Associations with passive joint engagement and later language. Journal of Applied Psycholinguistics. 27, 447-463 Rollins, P.R., & Trautman, C.H. Caregiver Input before Joint Attention: The Role of Multimodal Input. Presented to International Congress For the Study of Child Language (IASCL): Montreal, July 2011.

Rollins, P. R., & Trautman, C. H. (2011). Caregiver input before joint attention: The role of multimodal input. *Presented to International Congress for the Study of Child Language*.

Rosa, E. C., Finch, K. H., Bergeson, M., & Arnold, J. E. (2013). The effects of addressee attention on prosodic prominence. *Language and Cognitive Processes*, *3798*(April), 1–9. https://doi.org/10.1080/01690965.2013.772213

Rose, Y., & MacWhinney, B. (2014). The phonbank project.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype Me! Socially Contingent Interactions Help Toddlers Learn Language (2013/09/23). *Child Development*, *85*(3), 956–970. https://doi.org/10.1111/cdev.12166

Rovee, C. K., & Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *Journal of Experimental Child Psychology*, *8*(1), 33–39. https://doi.org/10.1016/0022-0965(69)90025-3

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Rowe, M. L., Coker, D., & Pan, B. A. (2004). A comparison of fathers' and mothers' talk to toddlers in low-income families. *Social development*, *13*(2), 278–291.

Rowe, M. L., Denmark, N., Harden, B. J., & Stapleton, L. M. (2016). The role of parent education and parenting knowledge in children's language and literacy skills among white, black, and latino families. *Infant and Child Development*, *25*(2), 198–220.

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*, 162–179.

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159–2173. https://doi.org/10.1037/dev0000240

Ruggeri, A., Markant, D. B., Gureckis, T. M., Bretzke, M., & Xu, F. (2019). Memory enhancements from active control of learning emerge across development. *Cognition*, *186*, 82–94. https://doi.org/10.1016/j.cognition.2019.01.010

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M. C., & Cohen, D. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLoS ONE*, *8*(10), 1–17. https://doi.org/10.1371/journal.pone.0078103

Sanchez, A., Meylan, S., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2018). *childes-db: a flexible and reproducible interface to the Child Language Data Exchange System*. psyarxiv.com/93mwx

Saxton, M. (2009). The inevitability of child directed speech. https://doi.org/10.1057/9780230240780_4

Saylor, M. M., & Ganea, P. A. (2018). *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms*. https://doi.org/10.1007/978-3-319-77182-3

Schieffelin, B. B. (1990). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge: Cambridge University Press

1990. The give and take of everyday life: language socialization of Kaluli children. Cambridge: Cambridge University Press.

Schieffelin, B. B., & Ochs, E. (1983). *A Cultural Perspective on the Transition from Prelinguistics to Linguistic Communication*. 77.

Schieffelin, B. B., & Ochs, E. (1987). *Language Socialization across Cultures*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620898

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

Schulz, L. E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389. https://doi.org/10.1016/j.tics.2012.06.004

Schulz, L. E., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers Engage in More Exploratory Play When Evidence Is Confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Scott, K., Sakkalou, E., Ellis-Davies, K., Hilbrink, E., Hahn, U., & Gattis, M. (2013). Infant contributions to join attention predict vocabulary development. In M. Knauff, M. Pauen, I. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 3384–3389).

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423.

Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive psychology*, *10*(3), 271–301.

Sherrod, K. B., Friedman, S., Crawley, S., Drake, D., & Devieux, J. (1977). Maternal language to prelinguistic infants: Syntactic aspects. *Child Development*, 1662–1665.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686. https://doi.org/10.1017/S0305000912000141

Shneidman, L. A., Buresh, J. S., Shimpi, P. M., Knight-Schwarz, J., & Woodward, A. L. (2009). Social Experience, Social Attention and Word Learning in an Overhearing Paradigm. *Language Learning and Development*, *5*(4), 266–281. https://doi.org/10.1080/15475440903001115

Shneidman, L. A., Gaskins, S., & Woodward, A. (2016). Child-directed teaching and social learning at 18 months of age: Evidence from Yucatec Mayan and US infants. *Developmental Science*, *19*(3), 372–381. https://doi.org/10.1111/desc.12318

Shneidman, L. A., & Goldin-Meadow, S. (2012a). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Shneidman, L. A., & Goldin-Meadow, S. (2012b). Mayan and u.s. caregivers simplify speech to children. In A. Biller, E. Chung, & A. Kimball (Eds.), *Proceedings of the 36th annual boston university conference on language development* (pp. 536–544). Somerville, MA: Cascadilla Press.

Sim, Z. L., & Xu, F. (2014). Acquiring Inductive Constraints from Self-Generated Evidence. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 1431–1436.

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2- and 3-year-old children. *Developmental Psychology*, *53*(4), 642–651. https://doi.org/10.1037/dev0000278

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

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-018-0422-4

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Volume 1: The data*. Psychology Press.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Snow, C. E. (1972). Mothers' speech to children learning language. *Child development*, 549–565.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of child language*, *4*(1), 1–22.

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children*. Cambridge University Press.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Sokolov, E. N. (1966). Orienting reflex as information regulator. In A. Leontiev, A. Ludia, & A. Smirnov (Eds.), *Psychological Research in the U.S.S.R.* Moscow: Progress Publishers.

Solomon, O. (2011). Rethinking baby talk. *The handbook of language socialization*, 121–149.

Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*(1-3), 431–445.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Reexamining the Verbal Environments of Children From Different Socioeconomic Backgrounds. *Child Development*, *90*(4), 1303–1318. https://doi.org/10.1111/cdev.13072

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Spinelli, M., Fasolo, M., Tagini, A., Zampini, L., Suttora, C., Zanchi, P., & Salerni, N. (2015). Linguistic and prosodic aspects of child-directed speech: The role of maternal child-rearing experiences. *European Journal of Developmental Psychology*, *5629*(September), 1–14. https://doi.org/10.1080/17405629.2015.1080159

Srinivasan, M., & Barner, D. (2013). The amelia bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition*, *128*(3), 431–450.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*. https://doi.org/10.1038/srep46730

Sullivan, J., & Barner, D. (2015). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental Science*, *19*(1)

Sullivan, J., and Barner, D. (2015). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. Developmental science.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific english words. *Child development*, *89*(4), 1247–1267.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, *72*(3), 748–767.

Theakston, A. L., Lieven, E., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*, 127–152.

Thomas, H. (1965). Visual-fixation responses of infants to stimuli of varying complexity. *Child Development*, 629–638.

Tolins, J., & Fox Tree, J. E. (2016). Overhearers Use Addressee Backchannels in Dialog Comprehension. *Cognitive Science*, *40*(6), 1412–1434. https://doi.org/10.1111/cogs.12278

Tolins, J., Namiranian, N., Akhtar, N., & Fox Tree, J. E. (2017). The role of addressee backchannels and conversational grounding in vicarious word learning in four-year-olds. *First Language*, *37*(6), 648–671. https://doi.org/10.1177/0142723717727407

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

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

Tomasello, M., & Kruger, A. C. (1992). Joint Attention On Actions Acquiring Verbs In Ostensive And Non-Ostensive Contexts. *Journal of Child Language*, *19*(2), 311–333. https://doi.org/10.1017/S0305000900011430

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195. https://doi.org/10.1111/1467-9280.00240

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340.

Vaish, A., Demir, Ö. E., & Baldwin, D. (2011). Thirteen- and 18-month-old infants recognize when they need referential information. *Social Development*, *20*(3), 431–449. https://doi.org/10.1111/j.1467-9507.2010.00601.x

VanDam, M. (2016). Vandam2 homebank corpus.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). Homebank: An online repository of daylong child-centered audio recordings. *Seminars in speech and language*, *37*(02), 128–142.

van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden* (Vol. 115). Swets; Zeitlinger.

Vogt, P., Mastin, J. D., & Schots, D. M. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands,

and rural and urban Mozambique. *First Language*, *35*(4-5), 341–358. https://doi.org/10.1177/0142723715596647

Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://books.google.se/books?id=RxjjUefze%7B%5C_%7DoC

Ward, M. (1971). *Them children: A study in language*. New York: Holt, Rinehart, Winston.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145
doi: 10.1177/0956797613488145

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49–63.

Whitman, W. (1855). *Leaves of grass*.

Wilson, B., & Peters, A. M. (1988). What are you cookin' on a hot?: Movement constraints in the speech of a three-year-old blind child. *Language*, *64*, 249–273.

Wong Fillmore, L. (1982). Instructional language as linguistic input: Second language learning in classrooms. *Communicating in the classroom*, 283–296.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid Word Learning in 13- and 18-Month-Olds. *Developmental Psychology*, *30*(4), 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Wundt, W. (1897). *Outlines of psychology* [Translated by Judd, C. H.]. New York.

Xiangjun, D., & Yip, V. (2018). A multimedia corpus of child mandarin: The tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69–92.

Xu, F. (2019). Towards a Rational Constructivist Theory of Cognitive Development. *Psychological Review*, *126*(6), 841–864. https://doi.org/10.1037/rev0000153

Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, *201*, 104280.

Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., & Wang, Z. (2016). Effects of valence and arousal on emotional word processing are modulated by concreteness: Behavioral and erp evidence from a lexical decision task. *International Journal of Psychophysiology*, *110*, 231–242. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2016.07.499

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening: Research article. *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x

Yurovsky, D. (2017). A communicative approach to early word learning. *New Ideas in Psychology*. https://doi.org/10.1016/j.newideapsych.2017.09.001

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79. https://doi.org/10.1016/j.newideapsych.2017.09.001

Yurovsky, D., Doyle, G., & Frank, M. C. (2016a). Linguistic input is tuned to children's developmental level.

Yurovsky, D., Doyle, G., & Frank, M. C. (2016b). Linguistic input is tuned to children's developmental level. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Appendix

A Eliciting Child- and Adult-Directed Explanations Online

In this study, we're interested in how people speak to one another. You will be asked to think about different situations, or ideas you would like to communicate to a particular person, and to write exactly what you think you would naturally say to them

Please take a moment to imagine [your child/your best friend] in front of you.

Now, imagine having to explain how to eat a new type of food they've never had before. Please write what you would say, below.

You should write as if you were talking to [your child/your best friend] directly, and try to make your speech as natural and close to what you would actually say as possible.

Now, imagine having to instruct them to get the first aid kit for you, and to call for help. Please write what you would say, below.

You should write as if you were talking to [your child/your best friend] directly, and try to make your speech as natural and close to what you would actually say as possible.

Now, imagine having to explain to them how you wash clothes or do laundry. Please write what you would say, below.

You should write as if you were talking to [your child/your best friend] directly, and try to make your speech as natural and close to what you would actually say as possible.

B Age of Acquisition Estimates: M-CDI vs. Kuperman

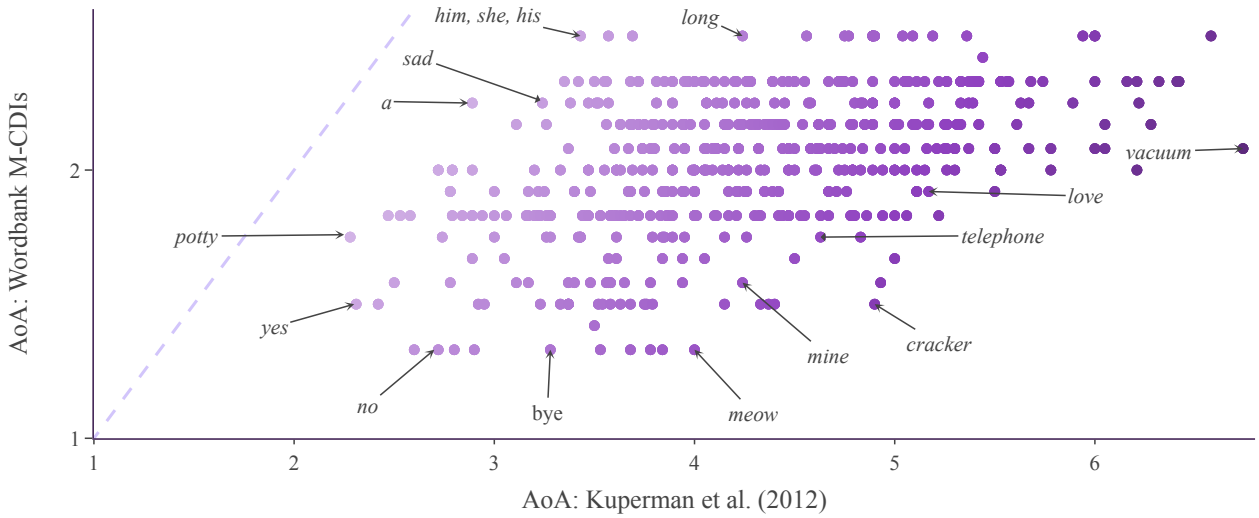


Figure 33: *Age of Acquisition Norms: Kuperman (2012) Adult Ratings vs. MacArthur-Bates Communicative Development Inventory Vocabulary Reports.*

Note. M-CDI estimate represents the month in age when word is reportedly produced by ≥50% of the English-acquiring children whose administrations are archived on wordbank.stanford.edu.

C Data Coverage for Complexity Metrics by Corpus

Table 21: *Tokens by Complexity Metric Across Corpora*

	Concreteness	**AoA**	**Valence**	**Frequency**	**M-CDI**
Child-directed					
Childes	11,955,336	11,414,837	11,243,398	5,072,092	11,497,187
VanDam	35,045	29,424	26,334	11,099	32,971
Manchester	1,349,932	1,293,159	1,277,545	558,453	1,298,276
All CDS	11,990,381	11,444,261	11,269,732	5,083,191	11,530,158
Adult-directed					
Childes	1,512,075	1,439,679	1,412,848	616,759	1,452,454
VanDam	12,905	10,698	9,511	3,937	12,053
Santa Barbara	205,976	154,030	136,928	54,160	175,862
BNC	7,507,640	6,230,899	5,414,582	1,717,135	6,689,738
All ADS	9,238,596	7,835,306	6,973,869	2,391,991	8,330,107

D ADS – CDS Difference in Means Across Corpora

	Concreteness			**AoA**			**Valence**		
Age	**12–24**	**24–36**	**36–48**	**12–24**	**24–36**	**36–48**	**12–24**	**24–36**	**36–48**
Childes	−0.153	−0.078	−0.054	0.199	0.091	0.081	0.017	0.049	0.044
VanDam	−0.077	−0.002	0.022	0.086	−0.022	−0.032	−0.041	−0.009	−0.014
SBC	−0.266	−0.190	−0.167	0.505	0.397	0.387	−0.147	−0.115	−0.121
BNC	−0.361	−0.286	−0.262	0.417	0.309	0.299	−0.074	−0.042	−0.047
All ADS	−0.321	−0.245	−0.221	0.374	0.266	0.256	−0.052	−0.020	−0.025

	Log Freq.			Lexical Comp.			Entropy		
Age	**12–24**	**24–36**	**36–48**	**12–24**	**24–36**	**36–48**	**12–24**	**24–36**	**36–48**
Childes	0.056	−0.055	−0.086	0.111	0.080	0.065	0.507	0.210	0.251
VanDam	−0.804	−0.915	−0.946	0.281	0.249	0.234	−0.367	−0.664	−0.623
SBC	−0.555	−0.665	−0.697	0.454	0.423	0.408	0.881	0.584	0.625
BNC	−0.148	−0.259	−0.290	0.452	0.420	0.406	−0.379	−0.676	−0.635
All ADS	−0.122	−0.233	−0.264	0.390	0.358	0.344	−0.068	−0.365	−0.324

E Summary of Previous Overhearing Experiments

Study	Age	Learning Target	Word Repetitions	Sentence Frame	Child-directed Context Cues	Other Notes
Akhtar et al., 2001	25 & 30 mos	object label	9 total (3 trials of 3 repetitions)	"I'm going to show you the *toma*. Let's find the *toma*. I'll show you the *toma*."	E smiles or gasps at object, engages in joint attention with C, passes object to C	25-month-olds did not demonstrate robust learning of action label
		action label		"Now I'm going to *meek* [character's name]. Let's *meek* [character's name]. I'll show you how to *meek* [character's name]."	E demonstrates action, smiles or gasps, hands character to C to perform action	
Akhtar, 2005	25 & 30 mos	object label	9 total (3 trials of 3 repetitions)	"I'm going to show you the *toma*. Want to see the *toma*? I'm going to show you the *toma*."	E gazes to object, engage in joint attention with C	distractor toy present

Floor and Akhtar, 2006	18 mos	object label	9 total (3 trials of 3 repetitions)	"I'm going to show you the *toma*. Want to see the *toma*? I'm going to show you the *toma*."	E plays a warm-up round of a finding game with child	
Shneidman et al., 2009	20 mos	object label	9 total (3 trials of 3 repetitions)	"Look at the *blicket*! Look at the *blicket*! Look at the *blicket*!"	E uses child-directed speech style, engages in joint attention with C, passes object to C to place down chute	
Martínez-Sussmann et al., 2011	27 mos	object label	9 total (3 trials of 3 repetitions)	"I'm going to show you the one that's in here. It's a *teebu*. Do you want to see the one that's in here? It's a *teebu*. I'll show you the one that's in here. It's a *teebu*."	E begins experiment with familiarization phase with child	distractor toy present

		fact		"I'm gonna show the one my mom gave me. Wanna see the one my mom gave me? I'll show you the one my mom gave me."	E smiles or gasps at object, engages in joint attention with C, passes object to C, who performs action	fact-learning was not robust
		fact + object label		"I'm gonna show you the one my *teebu* gave me. Wanna see the one my *teebu* gave me? I'll show you the one my *teebu* gave me."		
Gampe et al., 2012	18 mos	object label	9 total (3 trials of 3 repetitions	"I'm going to show you the [label]. Do you want to see the [label]? I'll show you the [label]." "Here the [label] goes in. But where is the [label]? I'll get the [label]"	E engages in joint attention with C	Study 2 used a music game

O'Doherty et al., 2011	30 mos	object label	9 total (3 trials of 3 repetitions)	"I'm going to show you the *toma*. Let's see the *toma*. I'm going to find the *toma*"	E gazes to object, engages in joint attention with C, demonstrates action, C imitates	learning only when C handed object

Note. E = Experimenter, C = Confederate.

F Experiment 1 Overhearing Condition Experimenter Script

Hi, how are you?

I'm good, thanks! Yeah, I'm at [Location]. I just brought some fun new toys in to play with. I brought a dog, a toma, a pimwit, a white cup, a zav, and a fep!

Do you know what a pimwit is? I brought one today. It is a purple pimwit. It's springy with a face. The purple pimwit is my sister's favorite. I really like the purple pimwit, too.

I also brought a fep. This fep is blue and tickly and you can put your fingers inside. Have you ever played with a fep? I got this blue fep in Disneyland. This fep is very fun.

Yeah, I like playing with dolls and toys like cups, too. I brought in a white toy cup that I play with my dolls with. It's a nice cup. This cup is full of milk. I have had this white cup for two years.

Um, yeah I just got a new green toma. The toma is a circle-shape, and it even lights up if you press on it! The toma only lights up if you press on the green star, though. My uncle gave the toma to me. I really like playing with the toma.

I brought a fuzzy dog in too. It's a black dog. This dog has a heart around its neck. I bring this dog to school. It looks like a dog I want to have as a pet.

What? Oh yeah, the last thing I brought was a zav. It's a yellow zav and it has a bunch of stickers in all different colors on it. You can take the stickers on and off the zav. I found this zav in the garden. I like this zav best.

Ok I'm going to go back and play now with the green circle toma from my uncle, the fuzzy dog I bring to school, the pimwit with the googly eyes that my sister loves, the blue fep I got from Disneyland, the white cup I've had for two years, and the yellow zav I found in the garden.

Bye! (*'hangs up' phone.*)

[To child:] Hi, [Child's Name]! Are you ready to play a game with me? Alright!

G Experiment 1 Pedagogical Condition Experimenter Script

Hi, [Child's Name]! I brought some fun new toys in to play with. I brought a dog, a toma, a pimwit, a white cup, a zav, and a fep! Do you know what a pimwit is? I brought a purple pimwit today. (*Lifts pimwit.*) It's springy with a face. This purple pimwit is my sister's favorite. I really like the purple pimwit, too. (*Sets down purple pimwit.*)

I also brought a fep. (*Lifts fep.*) This fep is blue and tickly and you can put your fingers inside (*demonstrates*). Have you ever played with a fep? I got this blue fep in Disneyland. This fep is very fun. (*Sets down fep.*)

I like playing with white cups too. (*Lifts cup.*) This cup I brought in is a white toy cup that I play with my dolls with. It's a nice cup. This cup is full of milk. I've had this cup for two years. (*Sets down cup.*)

I also just got a new green toma! (*Lifts toma.*) This toma is a circle-shape, and it even lights up if you press on it! (*demonstrates*). The toma only lights up if you press on the green star, though. My uncle gave this toma to me. I really like playing with the toma. (*Sets down toma.*)

I brought a fuzzy dog in too. (*Lifts dog.*) It's a black dog. This dog has a heart around its neck. I bring this dog to school. It looks like a dog I want to have as a pet. (*Sets down dog.*)

The last thing I brought was a zav. (*Lifts zav.*) It's a yellow zav and it has a bunch of stickers in all different colors on it. You can take the stickers on and off this zav (*demonstrates*). I found this zav in the garden. I like this zav best. (*Sets down zav.*)

Ok, [Child's Name], are you ready to play a game with the (*pointing*) green circle toma from my uncle, the fuzzy dog I bring to school, the pimwit with the googly eyes that my sister loves, the blue fep I got from Disneyland, the white cup I've had for two years, and the yellow zav I found in the garden.

Let's do it!

H Experiment 3 Experimenter–Caller Script

EXPERIMENTER
Hi, how are you?

CALLER
Doing alright, and you?

EXPERIMENTER
I'm good, thanks! Yeah, I'm at [Berkeley/Bay Area Discovery Museum/the preschool]. I just brought some fun new toys in to play with. I brought a dog, a pimwit, a white cup, a zav, and a fep!

CALLER
Whoa, cool, I've never heard of some of those things.

EXPERIMENTER
Do you know what a pimwit is?

CALLER
No...

EXPERIMENTER
I brought one today. It is a purple pimwit. It's springy with a face. The purple pimwit is my sister's favorite. I really like the purple pimwit, too.

CALLER
I bet! What else?

EXPERIMENTER
I also brought a fep. This fep is blue and tickly and you can put your fingers inside. Have you ever played with a fep?

CALLER
No!

EXPERIMENTER
I got this blue fep in Disneyland. This fep is very fun.

CALLER
It sounds like it, but I think I like playing things like house and tea party even better.

EXPERIMENTER
Yeah, I like playing with dolls and toys like cups, too. I brought in a white toy cup that I play with my dolls with. It's a nice cup. This cup is full of milk. I have had this white cup for two years.

CALLER
Yeah, anything else?

EXPERIMENTER
I brought a fuzzy dog in too. It's a black dog. This dog has a heart around its neck. I bring this dog to school. It looks like a dog I want to have as a pet.

CALLER
Aww - I wanna see! And what about that other thing?

EXPERIMENTER
What? Oh yeah, the last thing I brought was a zav. It's a yellow zav and it has a bunch of stickers in all different colors on it. You can take the stickers on and off the zav. I found this zav in the garden. I like this zav best.

CALLER
Wow, cool.

EXPERIMENTER
Ok I'm going to go back and play now with the fuzzy dog I bring to school, (*mmhm*) the pimwit with the googly eyes that my sister loves, (*mmhm*) the blue fep I got from Disneyland, (*mmhm*) the white cup I've had for two years, (*mmhm*) and the yellow zav I found in the garden.

CALLER
Ok, have a good time!

EXPERIMENTER
Bye! (*hangs up phone.*)
[To child:] Hi, [Child's name]! Are you ready to play a game with me? Alright!
(*Comes to sit in chair across from child, and lines toys up in front of child. Reaches down to containers on floor below chair, and lifts one onto the table.*)

I Time-Aligned Object Touch Plots by Child in Experiments 1–3

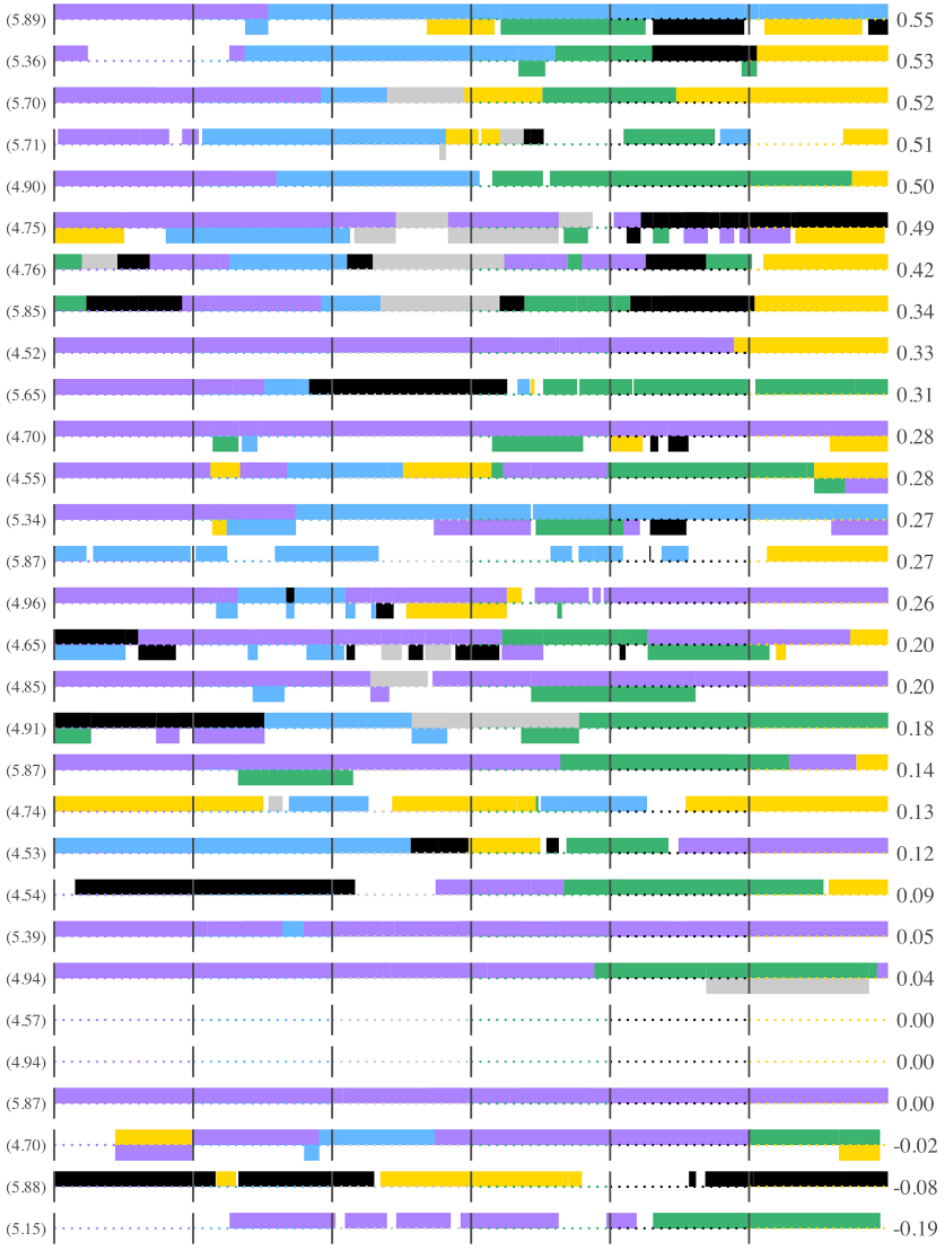


Figure 34: *Touch Behavior for Individual Children in Experiment 1.*

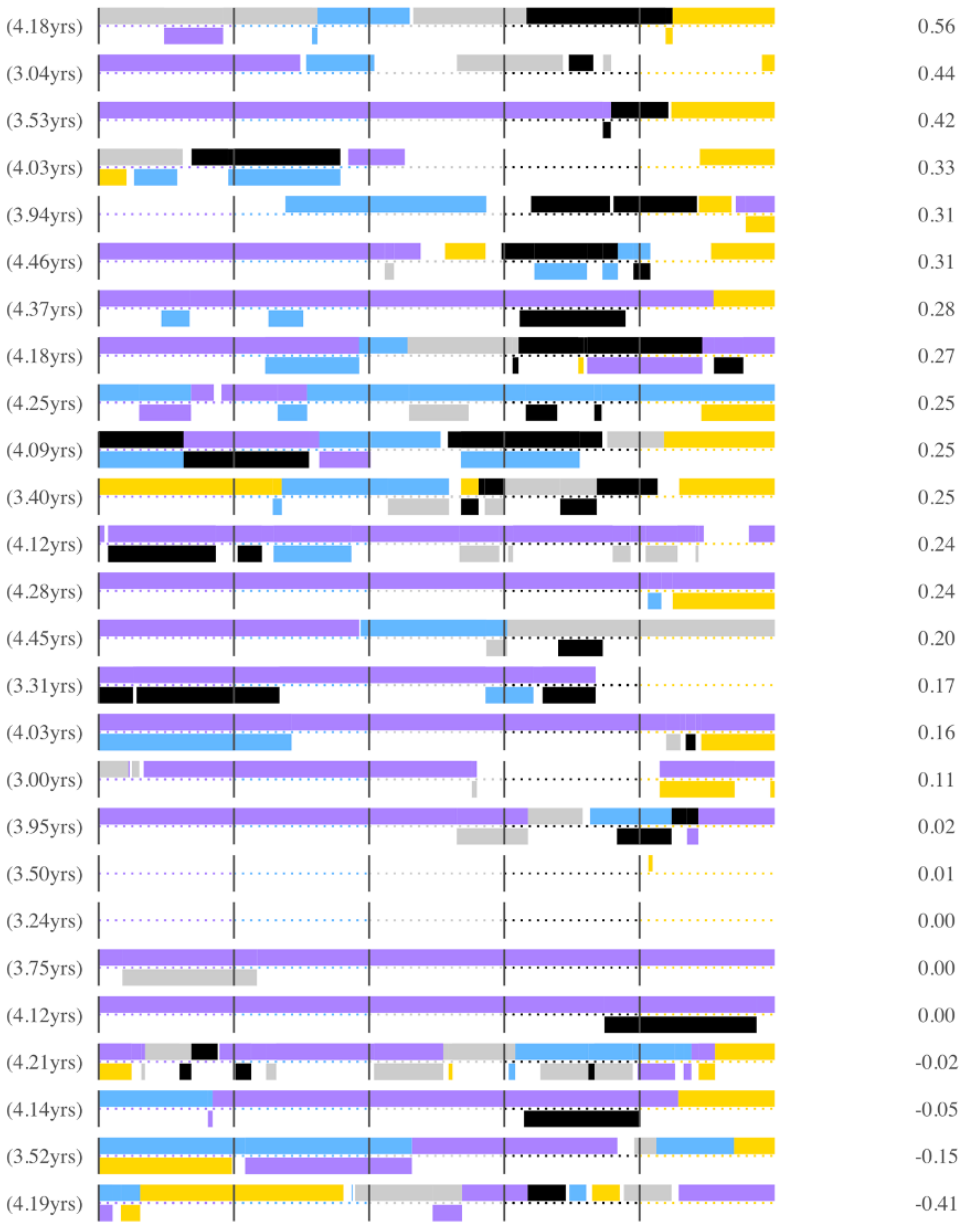


Figure 35: *Touch Behavior for Individual Children in Experiment 2.*

Figure 36: *Touch Behavior for Individual Children in Experiment 3.*

J Storybook Habituation Paradigm

INTRODUCTION *Experimenter sits child in front of laptop and helps child put on headphones.* Today we are going to listen to a story called, *Frog Where are You?* After the story, we're going to play a couple quick games, and that's it! Are you ready?

Reddotincenterofscreen

EYETRACKER CALIBRATION But first before we get started, I have a mini game for you to play. Do you see this red dot right here? It's going to move around the screen like a fairy. I need you to follow it with your eyes. Can you do that for me? Great! Are you ready? (Press space bar) Where'd the dot go?!

STORYBOOK // GIF + ILLUSTRATION

PREVIEW // GIF ONLY
Oo! Look at that! There will always be these jumping penguins while I'm telling the story, so if you get bored you can always watch them.

STORY INTRODUCTION // GIF + TITLE PAGE
Now we're going to hear a story called Frog, Where Are You? Remember, you can always look at the penguins if the story gets boring.

MANIPULATION // GIF + TITLE PAGE
Where are you going to look if the story gets boring?

STORY NARRATION X6 // GIF + TITLE PAGE

ENDING // TWO-PAGE STORYBOOK
The boy and the dog looked all over for the frog --- then they found him! With his whole family!

LISTENING COMPREHENSION // 2x3 GRIDS

LISTENING COMPREHENSION INSTRUCTIONS // SMILEY
Next, we're going to play a quick game. There are going to be 6 pictures on the screen and I'm just going to ask you to point to one of them. Here's a

practice round.

PRACTICE // FIXATION CROSS
Q: Where do owls live?
A: Trees (2)

TEST ARRAY

PAGE 1 // FIXATION CROSS
That was amazing! Let's try a couple more.
Q: Who helped the boy find the frog?
A: The dog (3)

TEST ARRAY

PAGE 2 // FIXATION CROSS
Q: Who was the frog looking for when he left the boy's room?
A: His mom and dad (1)

TEST ARRAY

PAGE 3 // FIXATION CROSS
Q: Where did the boy and the dog look for the frog first?
A: The boot (5)

TEST ARRAY

PAGE 4 // FIXATION CROSS
Q: What did the dog get his head stuck in?
A: A jar (4)

TEST ARRAY

Page 5 // FIXATION CROSS
Q: Where did the boy and the dog go to look for the frog?
A: The woods (2)

TEST ARRAY

Page 6 // FIXATION CROSS
Q: Who were the boy and the dog looking for?
A: The frog (6)

TEST ARRAY

WORD LEARNING // 2x2 GRIDS

WORD LEARNING INSTRUCTIONS // SMILEY
Great work! Now we are going to play another game, just like we did earlier. There's going to be 4 pictures on the screen and I'm just going to ask you to point to one of them. Let's do a practice round.

PRACTICE // FIXATION CROSS
Can you point to the dog?

TEST ARRAY

WL: OGLING // FIXATION CROSS
Good job! Let's try a couple more.
Q: Can you point to the person who is ogling something?

TEST ARRAY

WL: ABSCONDING // FIXATION CROSS
Q: Can you point to the person who is absconding?

TEST ARRAY WL: FLUMMOXED // FIXATION CROSS
Q: Can you point to the face that looks flummoxed?

TEST ARRAY

WL: HYALINE // FIXATION CROSS
Can you point to the cup that is hyaline?

TEST ARRAY

WL: APERTURE // FIXATION CROSS
Can you point to the aperture?

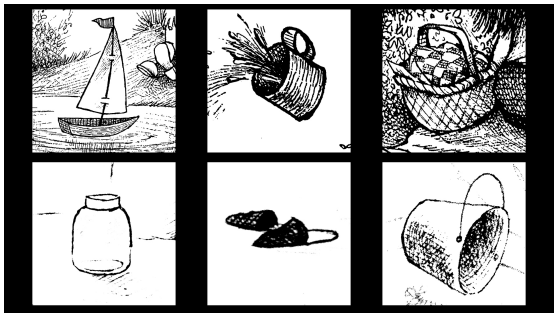
TEST ARRAY

WL: TOR // FIXATION CROSS
Can you point to the tor?

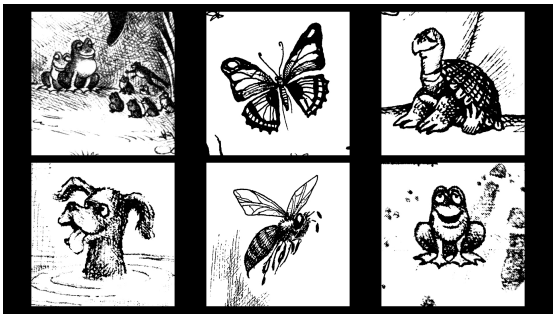
TEST ARRAY

K Listening Comprehension Test Arrays

(a) *Who helped the boy find the frog? (the dog; top-right)*

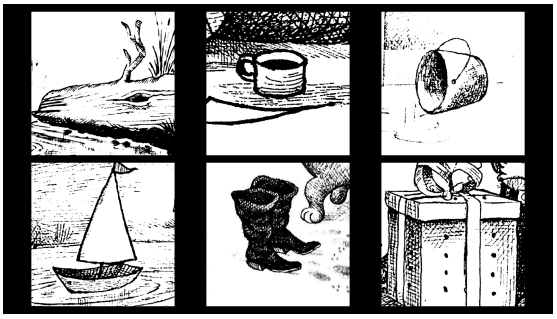


(d) *What did the dog get his head stuck in? (the jar; bottom-left)*

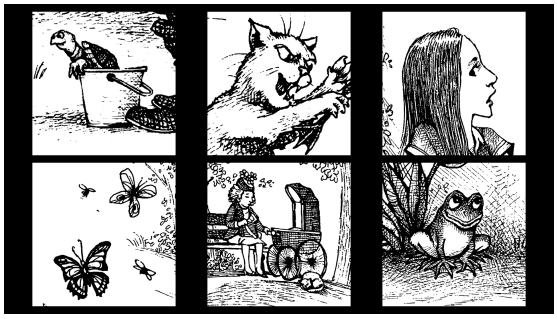


(b) *Who was the frog looking for when he left the boys room? (his family!; top-left*

(e) *Where did the boy and the dog go to look for the frog? (the forest; top-center)*



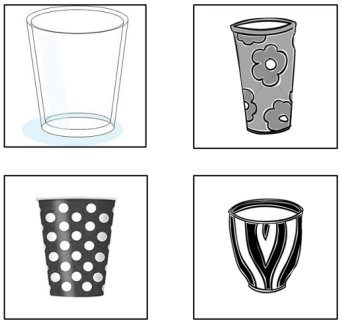
(c) *Where did the boy and the dog look for the frog first? (the boots; bottom-center)*



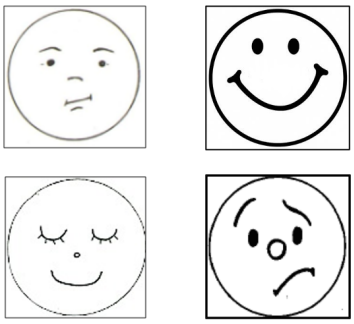
(f) *Who were the boy and the dog looking for? (the frog; bottom-right)*

Note: Training array used to familiarize participants with the task not pictured.

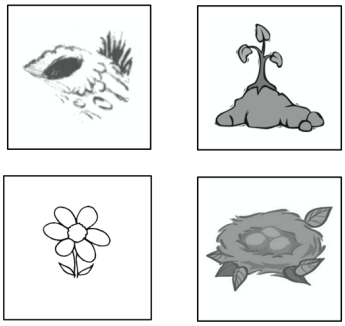
L Word Learning Test Arrays



(a) *Can you point to the cup that is hyaline? (top-left)*

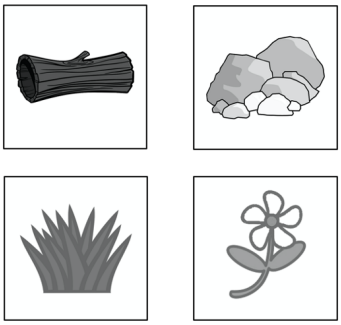


(d) *Can you point to the face that looks flummoxed? (bottom-right)*

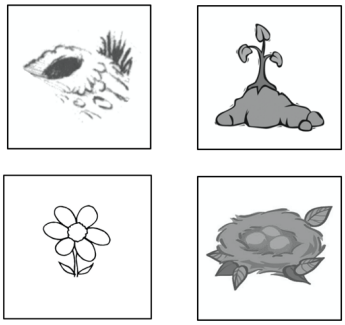


(b) *Can you point to the aperture? (top-left)*

(e) *Can you find the person who is absconding? (top-right)*



(c) *Can you find the tor? (top-right)*



(f) *Can you point to the one who is ogling something? (bottom-left)*

M Distributions of by-Participant Summary Attention Metrics

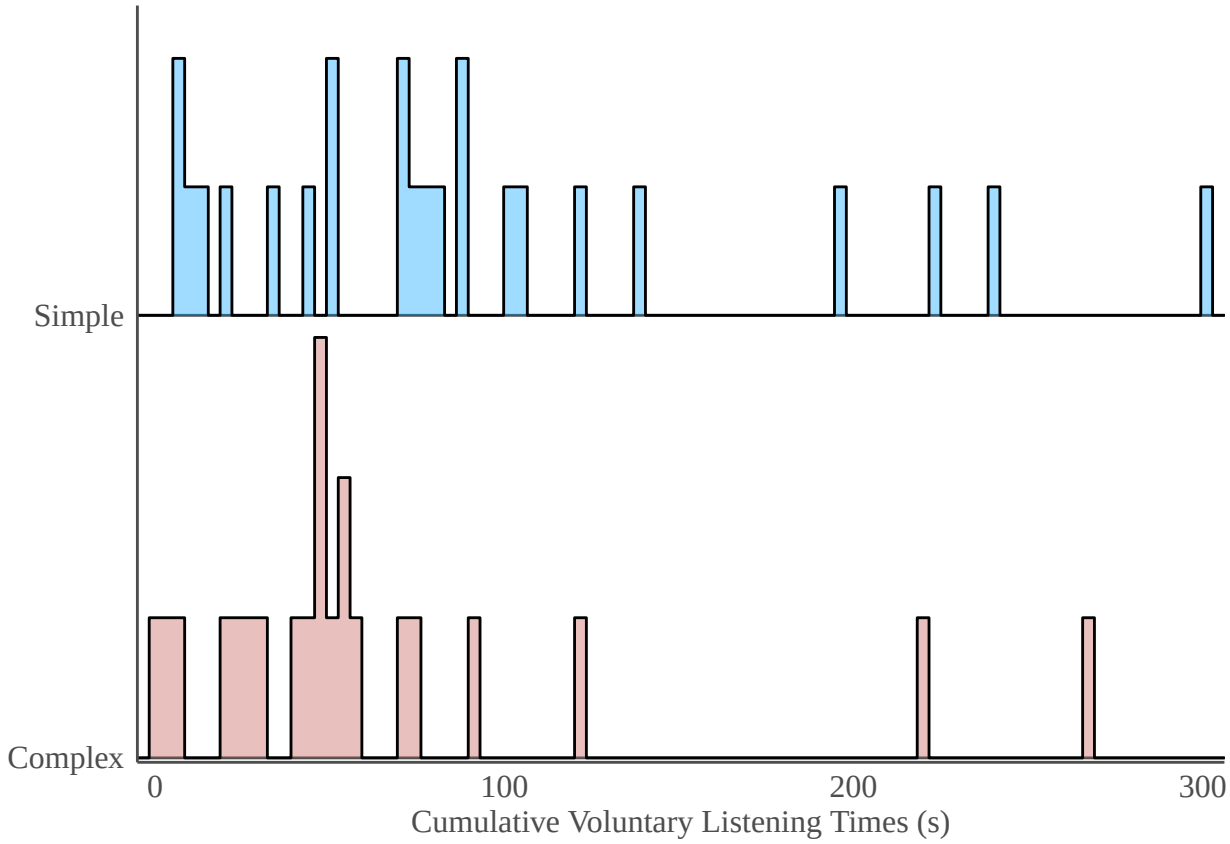


Figure 39: *Histogram of Cumulative Listening Times by Condition.*

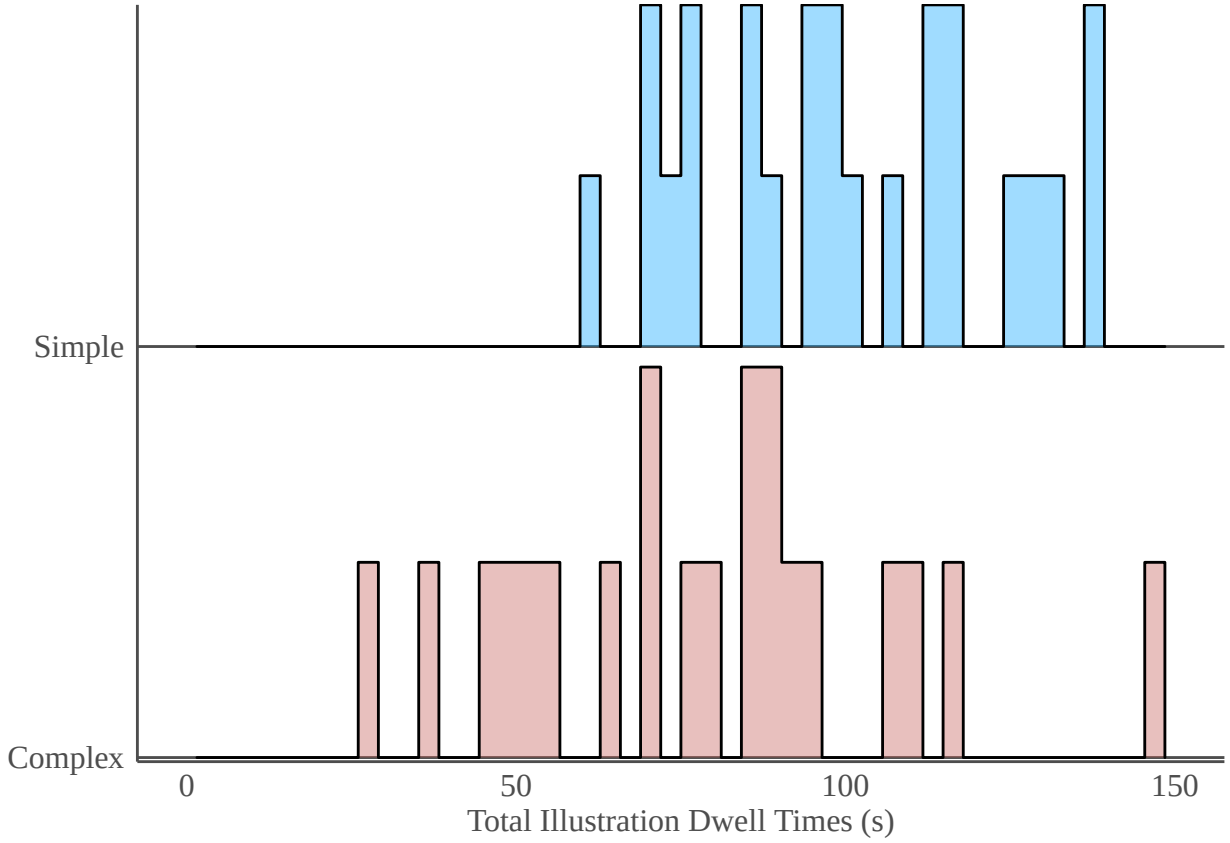


Figure 40: *Histogram of Total* ILLUSTRATION *Dwell Times by Condition.*

Note. Bin size = 3s.

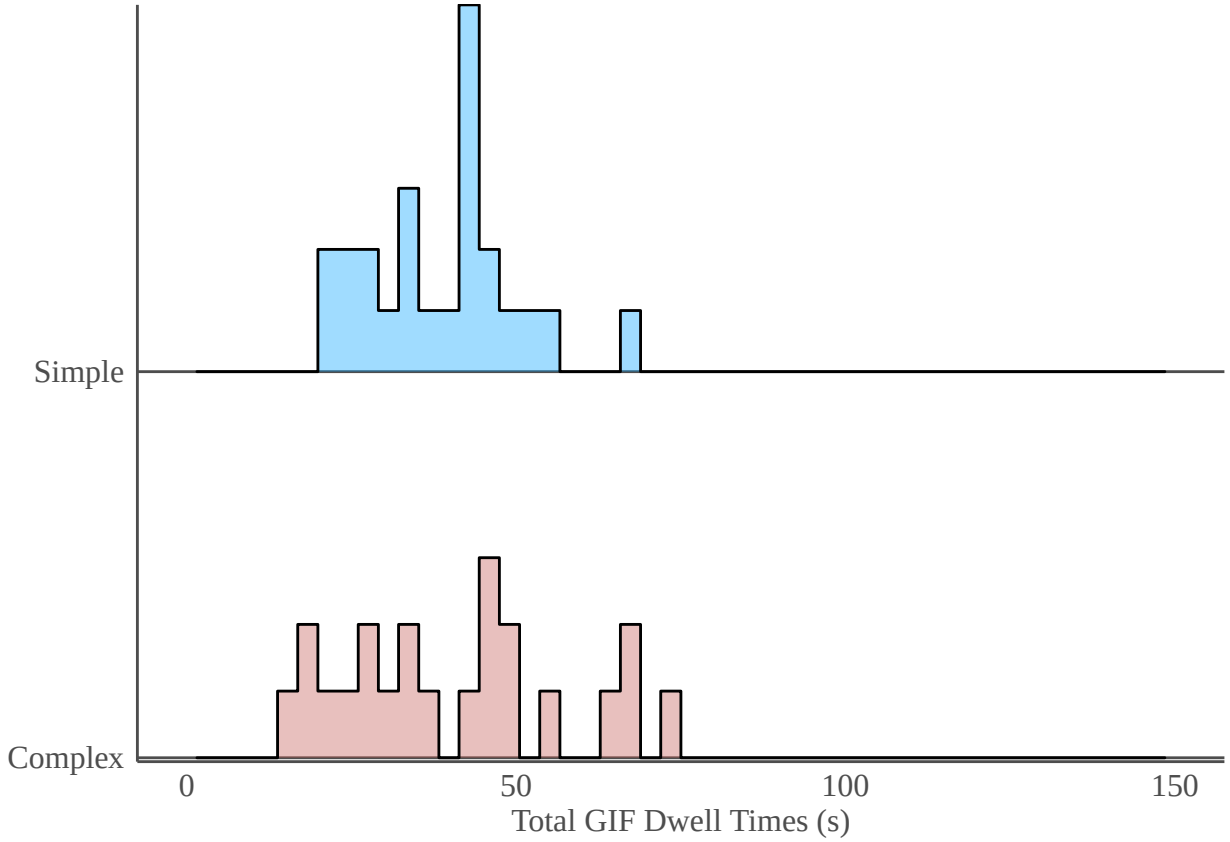


Figure 41: *Histogram of Total* DISTRACTOR *Dwell Times by Condition.*

Note. Bin size = 3s.

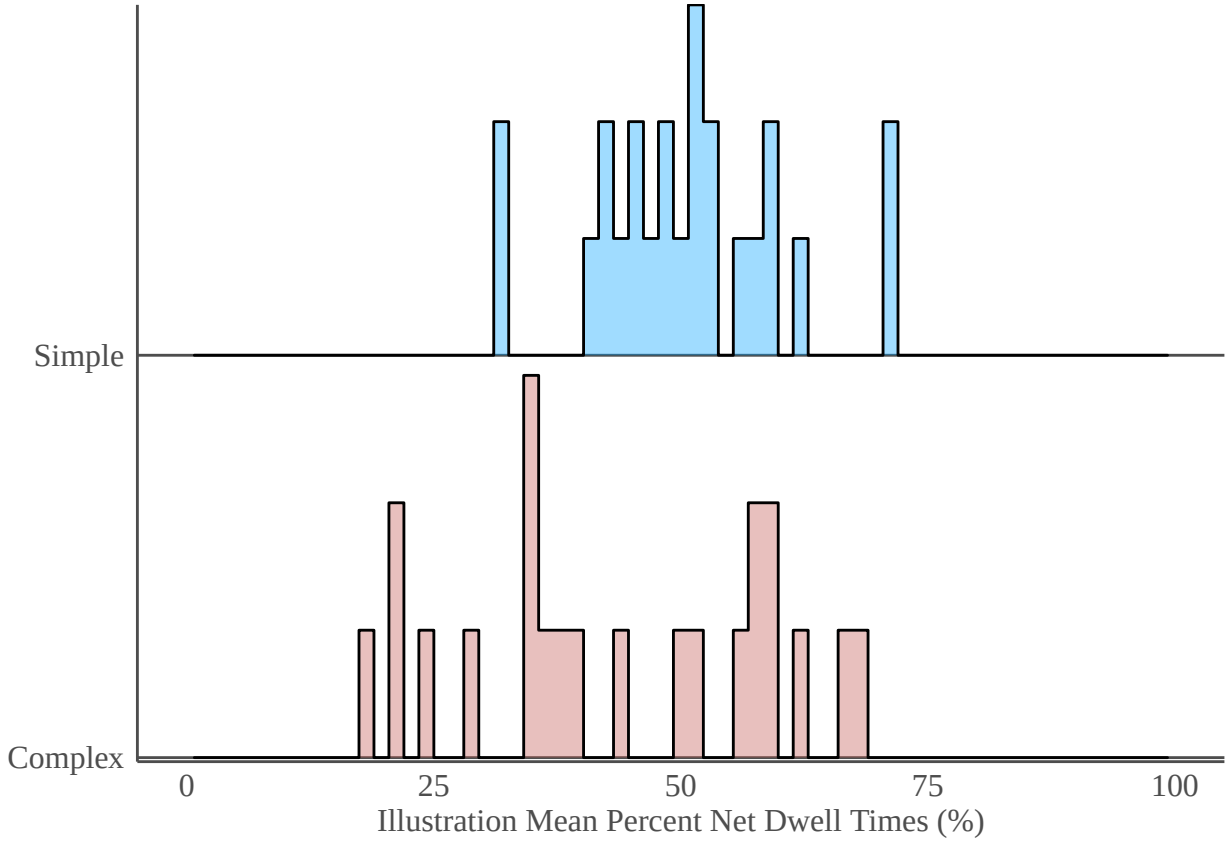


Figure 42: *Histogram of* ILLUSTRATION *Net Dwell Time Mean Percents by Condition.*

Note. Bin size = 1.5%.

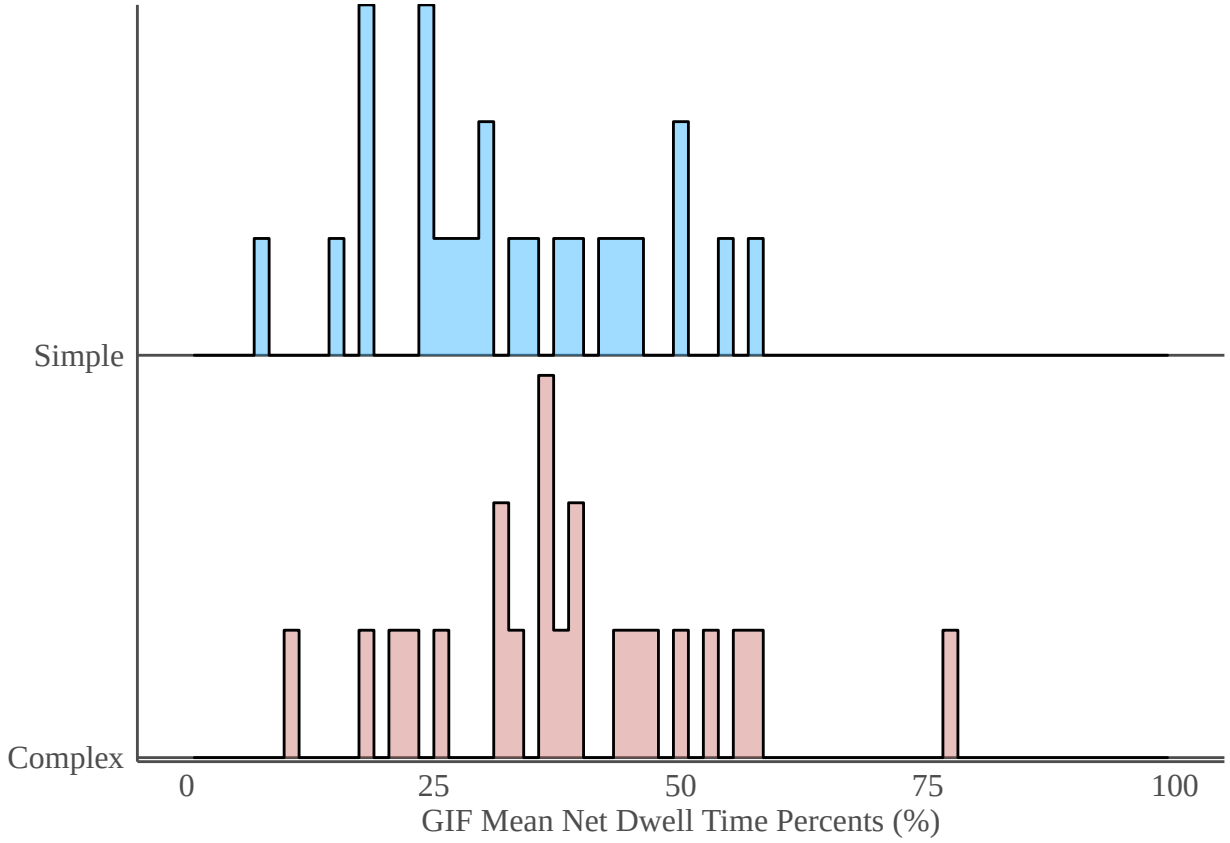


Figure 43: *Histogram of* DISTRACTOR *Net Dwell Time Mean Percents by Condition.*

Note. Bin size = 1.5%.

N Summary of Primary Coded Variables

Variable	**N**	**Mean**	**SD**	**Min**	**Max**
Age	6,733	23.00	8.40	30.00	34.00
Tokens	6,73	4.90	4.00	7	55
Morphemes	6,514	6.20	5.00	8	65
Here & Now	6,532	0.57	0.49	0	1
Referential Gesture	2,685	0.31	0.46	0	1
Looking @ Speaker	3,655	0.56	0.50	0	1
Sing-Song Prosody	6,545	0.86	0.34	0	1
Speech Clarity	6,711	2.50	0.76	0	3
Morphological Complexity	6,514	1.20	0.28	1.30	3.00
Child-Directed	6,733	0.84	0.37	1	1
Looking @ Referent	6,733	0.22	0.42	0	1
Play Context	6,733	0.75	0.43	0	1
Child Standing	6,733	0.25	0.43	0	1
Child Held	6,733	0.012	0.11	0	1

O Correlations in the Language Environment

Table 23: *Correlations Among Qualitative Variables in Speech Directed to Naima*

	Age (mos)	Play	Stand-ing	Tokens	Morphs	Morph. Comp.	Here & Now	Ref. Ges-ture	Sing-song	Look @ Speaker
Age (mos)										
Play	-0.09***									
Standing	-0.16***	0.43***								
No. Tokens	0.14***	0.04	0.01							
No. Morphs	0.12***	0.04	0.00	0.97***						
Morph. Comp.	-0.02	0.05*	-0.03	0.11***	0.30***					
Here & Now	-0.50***	0.10***	0.13***	-0.13***	-0.12***	0.05*				
Gesture	-0.05*	-0.05*	0.04	0.05*	0.05*	0.02	0.24***			
Prosody	0.03	-0.04	-0.05*	-0.01	-0.01	-0.01	-0.02	0.01		
Look@Spkr	-0.57***	0.09***	0.04	-0.10***	-0.08***	0.06**	0.38***	0.04	0.01	
Clarity	-0.54***	0.08***	0.08***	-0.03	-0.02	0.07**	0.40***	0.04	-0.03	0.43***

Table 24: *Correlations Among Qualitative Variables in Speech Around Naima*

	Age (mos)	Play Context	Standing	Tokens	Morphs	Morph. Comp.	Here & Now	Ref. Ges-ture	Sing-song	Look Speaker
Age (mos)										
Play Context	-0.09***									
Child Standing	-0.16***	0.43***								
No. Tokens	0.14***	0.04	0.01							
No. Morphemes	0.12***	0.04	0.00	0.97***						
Morph. Complexity	-0.02	0.05*	-0.03	0.11***	0.30***					
Here & Now	-0.50***	0.10***	0.13***	-0.13***	-0.12***	0.05*				
Referential Gesture	-0.05*	-0.05*	0.04	0.05*	0.05*	0.02	0.24***			
Sing-song Prosody	0.03	-0.04	-0.05*	-0.01	-0.01	-0.01	-0.02	0.01		
Looking @ Speaker	-0.57***	0.09***	0.04	-0.10***	-0.08***	0.06**	0.38***	0.04	0.01	
Speech Clarity	-0.54***	0.08***	0.08***	-0.03	-0.02	0.07**	0.40***	0.04	-0.03	0.43***

P Qualitative Aspects of Overhearing Context Codeable from Video

Category	Type	Code	Description
SEMANTIC	−/+	here_now	*Is the speech about the 'here and now,' or de-contextualized?*
	0–3	adultness	(opposite of 'babiness')
VISUAL	−/+	speaker	*Is the child looking at the speaker?*
ATTENTION	−/+	referent	*Is the child looking at speaker is referring to?*
	0–3	clutter	*How cluttered is the scene?*
REFERENTIAL	−/+	gaze	*Is the speaker looking at what they're talking about?*
	−/+	gesture	*Is the speaker gesturing, demonstrating, or pointing?*
AUDIENCE	=	target child	*To whom is the utterance directed?*
	=	other child(ren)	
	=	adult(s)	
	=	phone	
AUDIO	−/+	sing-song	*Is the speaker using exaggerated child-directed intonation?*
	0–3	auditory clarity	*How clear is the utterance?*
	0–3	proximity	*How near is the speaker?*
	−/+	dialogue	*Does the child have access to addressee backchannels?*

	0–3	noise	(auditory equivalent of clutter) *How much competition is there for the child's auditory attention?*
SOURCE	=	live	*Where is the speech coming from?*
	=	tv	
	=	tablet	
	=	radio	
	=	phone	
CHILD	=	supine	*How is the child positioned?*
POSITION	=	pront	
	=	crawling	
	=	sitting_low	
	=	sitting_high	
	=	held_hip	
	=	held_front	
	=	held_back	
	=	standing	

CODE TYPES:

[+/−]	Binary Feature
[=]	Variable with Mutually Exclusive Values
[0–3]	Subjective Rating Scale