Self-Directed Learning by Preschoolers in a Naturalistic Overhearing Context

Ruthe Foushee,[1] Mahesh Srinivasan,[1] & Fei Xu[1]

[1]University of California, Berkeley

Please address correspondence to:

Ruthe Foushee

Department of Psychology

University of California, Berkeley

Berkeley, CA 94720-1650

Phone: 520-664-6126

Email: foushee@berkeley.edu

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.

Keywords: self-directed learning, overhearing, word learning, child-directed speech, effective input

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 input over others, and selectively explore objects, suggesting that children choose the information they want to receive from early in life (e.g., Gerken, Balcomb, & Minton, 2011; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Kidd, Piantadosi, & Aslin, 2012, 2014; Piantadosi, Kidd, & Aslin, 2014; Sim & Xu, 2017a; 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; 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, McGovern, Yung, & Kidd, 2012; Ruggeri, Gureckis, Markant, Bretzke, & Xu, 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, Goodman & Schulz, 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 a credible source of new linguistic information (Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Luchkina, Sobel, & Morgan, 2018). But what about in the real world, when children have not only to *evaluate* new information from potential sources of learning, they have to recognize learning opportunities and 'tune in' to them over competitors?

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 which 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, Carpenter, Call, Behne, & Moll, 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, we can think 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 how we often think of 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, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Shneidman, Arroyo, Levine, & Goldin-Meadow, 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, Brown, & Levinson, 2019; Shneidman & Goldin-Meadow, 2012; see also Brown, 1998; Cristia, Dupoux, Gurven, & Stieglitz, 2017; de León, 1998; Mastin & Vogt, 2016; Ochs & Schieffelin, 1995; Pye, 1986a, 1986b; Sperry, Sperry, & Miller, 2018; Vogt, Mastin, & Schots, 2015; Weisleder & Fernald, 2013). For example, in one Yucatec Maya community, up to 80% of words that 12-month-olds heard were overheard (Shneidman & Goldin-Meadow, 2012; 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., 2018). 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., 2018).

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, García-Sierra, & Kuhl, 2014; Shneidman & Goldin-Meadow, 2012; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; 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, Jipson, & Callanan, 2001; Baldwin, 1991; Floor & Akhtar, 2006; Fitch, Lieberman, Luyster & Arunachalam, 2019; Gampe, Liebal, & Tomasello, 2012; Martínez-Sussmann, Akhtar, Diesendruck, & Markson, 2011; Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009; for review see Shneidman & Woodward, 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

[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.

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 dense, 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 A1 in the Appendix), children often only needed to learn a single novel word (Akhtar, Jipson, & Callanan, 2001; Floor & Akhtar, 2006; L. A. Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009). This word was repeated as many as nine times was embedded in a small number of explicit labelling or directive sentence frames (Akhtar et al., 2001; Akhtar, 2005; Floor & Akhtar, 2006; Martínez-Sussman 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-Sussman 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 Table A1 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, which we would 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 studies of learning from overhearing, which have focused on children 18 to 30 months in age (see Table A1). Our goal was in part to determine the lower bound with respect to age in which children can learn from overhearing when demands on self-directed learning are high.

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: they 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, Lupyan, Goldstein, & Spivey, 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 this is a common feature, 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 this 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 picked up a 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 task requires that children attend to both the overheard speech and the objects in front of them, suggesting that the project 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.

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; d=.55, at α =0.05), using a one-sample t-test comparing children's learning from overhearing to chance. Power was calculated using the *pwr* package (Champely et al., 2013) in R (R Core Development Team, 2011).

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) = .1, p = .9; Cohen's d = -.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 1. 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 2). 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 a 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 1. Stimuli used in Experiment 1. Novel objects appear in the top row, familiar objects in the bottom row.

Table 1

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*. 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 parents' estimates of their child's exposure to overheard phone calls—can be found 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. Parents 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 feet from the table, where she began “working” on a laptop that had been resting there, surreptitiously starting a timer on her phone for one minute. 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.

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 15 seconds 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 said 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. Parents 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 they 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 repositiory), 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 in two ways: first, we conduct 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* subsection below). Second, we code whether a child gives the same object multiple times in response to different prompts within a test block (e.g., puts the purple object in the hat when asked for both the "pimwit" and the "zav"). In the *Supplemental Online Materials,* we separately analyze the test performance of these children—whose behavior suggests they were considering the entire set of objects on each test trial—and the data from children who never provided the same object twice within a block, consistent with their test trials *not* being independent.

Mixed effects models. We use mixed effect logit models constructed using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Development Core Team, 2008) 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 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). We use likelihood ratio tests and compare models' Akaike Information Criteria (AIC) to select among nested models, updating the null model (only random intercepts for participant) to include predictors in the following order: (1) age, (2) condition, (3) learning target type, (4) the interaction of condition and learning target type. When models with the predictors of interest fail to converge, we refit a null model to the data, excluding random effects, and repeat the process of nested model construction and comparison. Predictors that do not significantly improve fit are dropped. We report 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, 2011) 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), and (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 Core Development Team, 2011). 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 *rmcorr* function in 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 2 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 2) 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 via a third coder, whose 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. 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.

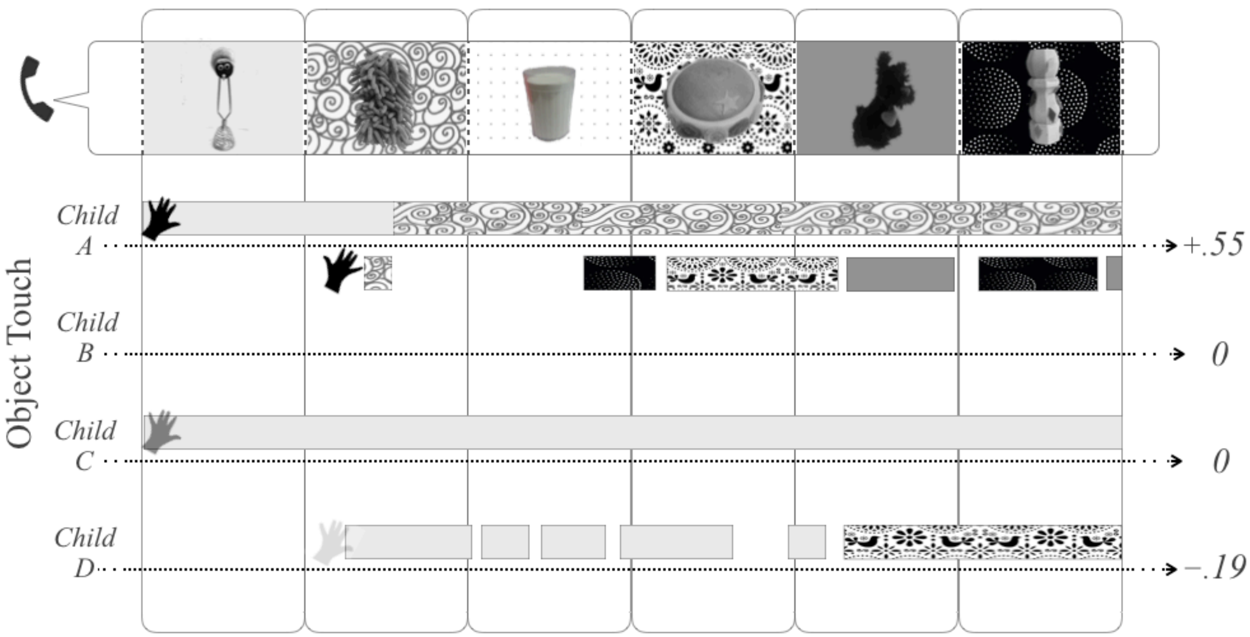

Figure 2. Experiment 1 touch behavior and matching-object touch scores for four participants, including the receiver of the highest (*Child A*) and lowest (*Child D*) scores. Participants' periods of touching each object (horizontal bars for each hand, filled according to which object they were touching) 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 and Discussion

Test performance. Preliminary analyses revealed no effects on 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.

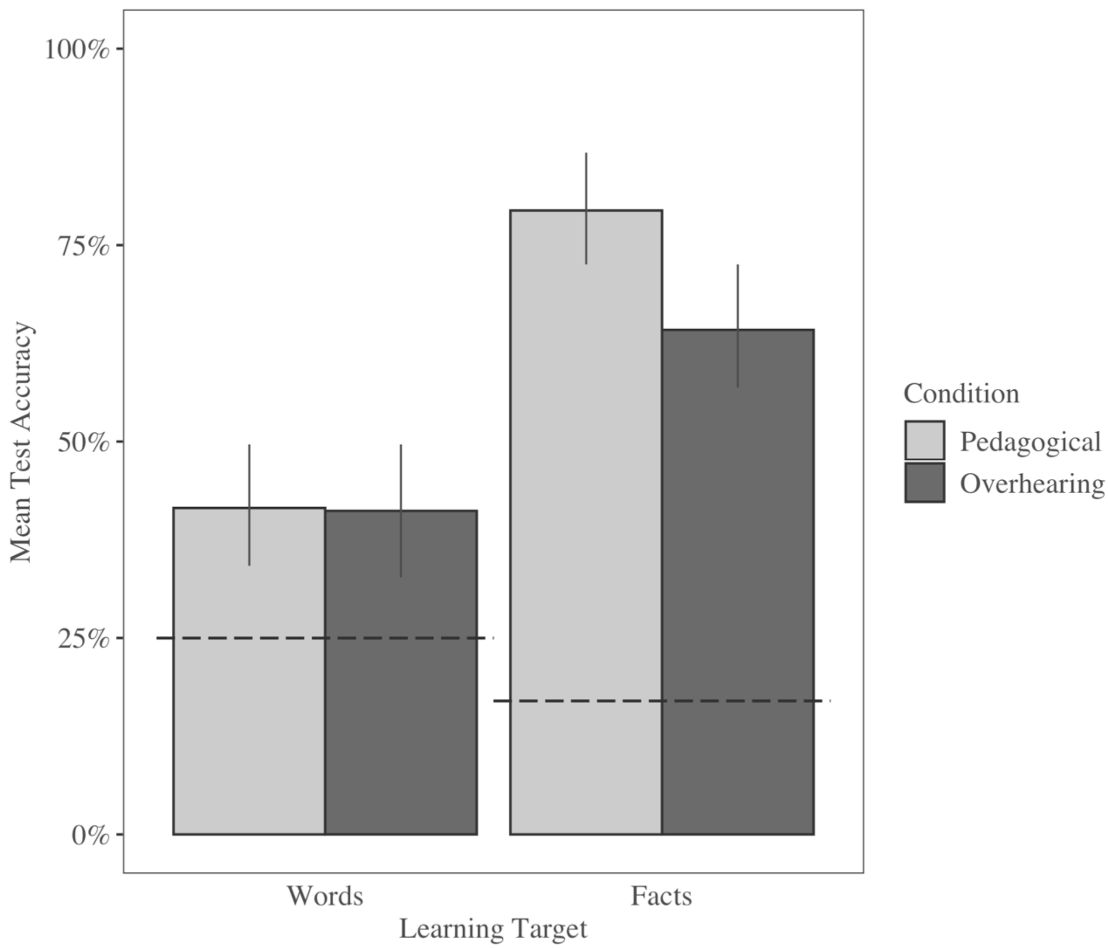

Figure 3. Experiment 1 mean accuracy at test by learning target and condition. Chance for each target type is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

Comparisons to chance. Figure 3 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 $d = .71$; 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 $d = .61$; 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%]; $t(33) = 4.0$, $p < .001$, Cohen's $d = 1.44$; first fact-learning trials: 71% [56%, 85%]; $t(33) = -243$, $p < .001$, Cohen's $d = 1.17$).

Next, we separately analyzed test data from children who gave the same object more than once in a block (suggesting their test trials *are* independent; 38 participants total, 19 in each condition) and those who did not (consistent with avoiding the object they gave previously, and therefore that their trials may be dependent on one another; 30 participants, 15 in each condition). Both groups responded at rates significantly above chance, for both learning target types (see Table S1 in the Supplemental Online Material).

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 $X^2(1) = 81.88$), suggesting that word learning was more difficult than fact learning in both conditions (Overhearing: $t(400) = -5$, $p < 0.001$, Cohen's $d = .47$; Pedagogical: $t(500) = -9$; $p < .001$, Cohen's $d = .83$). 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 $X^2(1) = 7.14$, $p < .01$). 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 $d = .66$), there was no difference in performance between the two conditions for words ($t(70) = .1$, p = .9, Cohen's $d = -.015$). This model resulted in a significantly better fit than the null model with no predictors and only random intercepts ($X^2(3) = 103.33$, $p < .001$; 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 ($X^2(1) = 7.23$, $p < .01$; AIC for model without interaction: 1218.6).

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 $X^2(1) = 0.01$, $p = .92$). 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).

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 ($X^2(1) = 4.9$, $p < .05$; AIC without interaction: 466, AIC with interaction: 463), and also outperformed a model with condition as the sole fixed effect ($X^2(2) = 6.7$, $p < .05$; 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 $X^2(1) = 4.9$, $p < .05$). 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.

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.

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) = -.46 [-.64, -.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, Lombrozo, Griffiths, & Xu, 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 4). 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 $r = .45$ [.10, .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 $r = -.04$ [-.40, .32]; $t(30) = -.20$, 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 4; 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 $r = -.09$ [-.44, .28]; $t(30) = -.05$, p = .6) or their accuracy at test (Pearson's $r = .21$ [-.17, .53]; $t(30) = 1$, p = .3).

Although this result conflicts with those of previous overhearing studies (Martínez-Sussman 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.

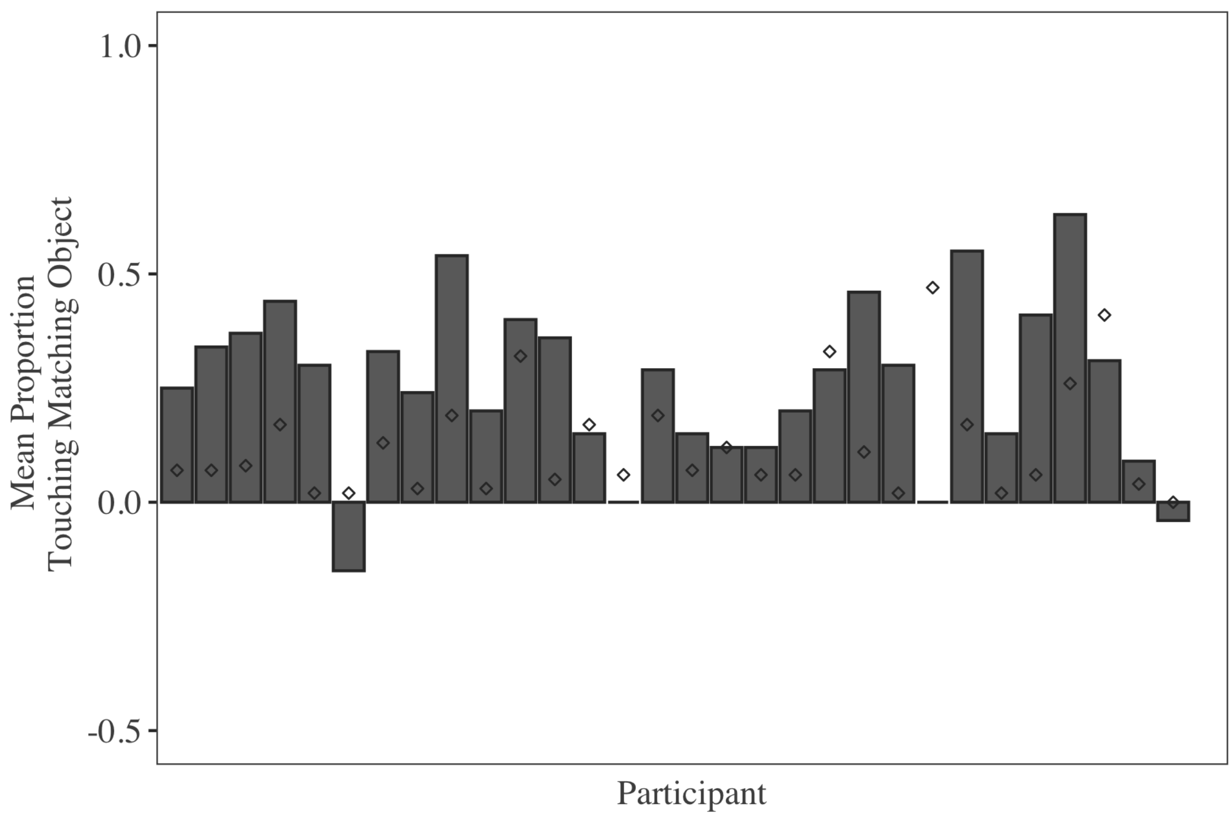

Figure 4. Mean proportion of matching-object touch and proportion gaze to the experimenter for each participant in Experiment 1. Positive values on the matching-touch measure (bars) indicate that the child touched the specific novel object that the experimenter was discussing more often as they was discussing it than when they were not. Overlaid points reflect the proportion of the call each participant spent looking toward the experimenter.

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 A1 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, $t(31) = 5.13$, $p < .001$, $d = 0.90$; Pedagogical: 74% [66%, 82%], $t(31) = 13.14$, $p < .001$, $d = 2.34$). 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, $t(31) = -0.74$, $p = .46$, $d = -0.14$; see Figure 6).

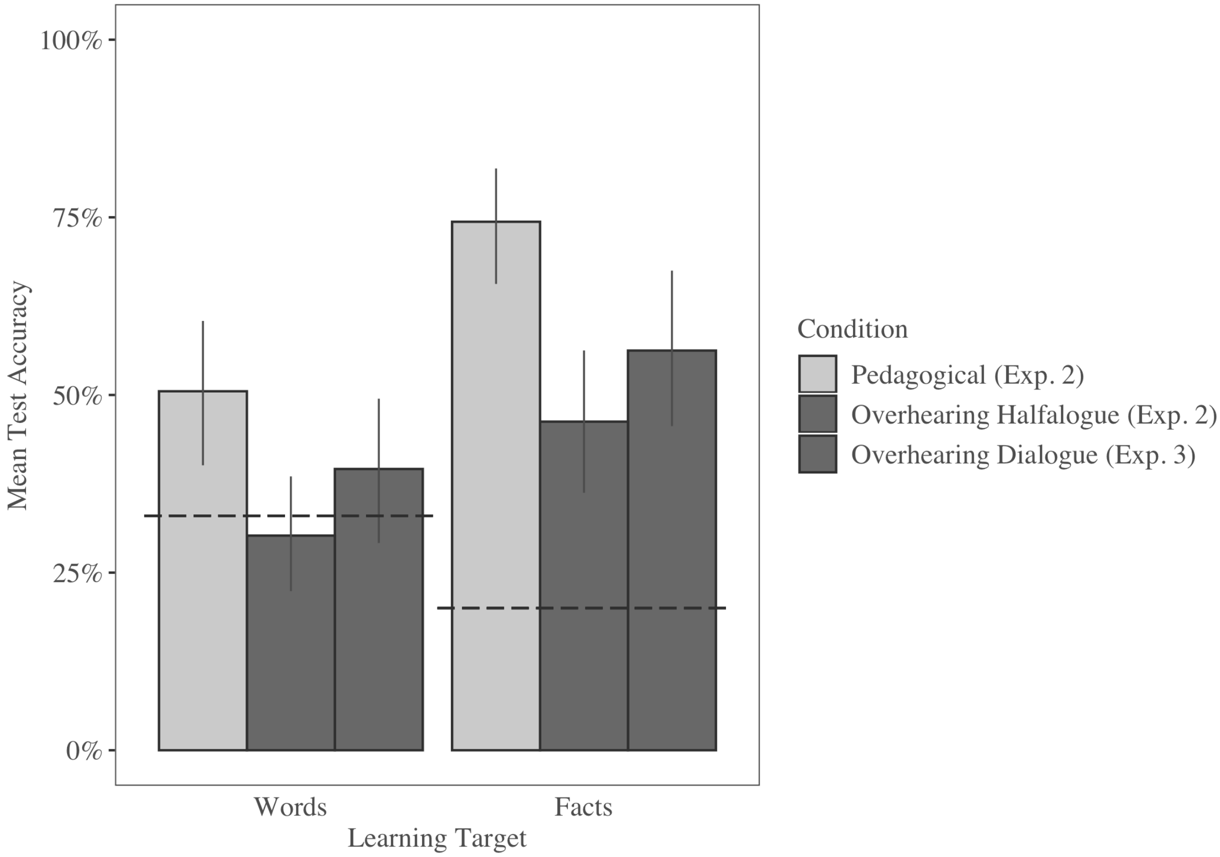

Figure 6. Experiments 2 and 3 mean accuracy at test by learning target and condition. Chance for each target type (0.20 for facts, and 0.33 for words) is indicated with a dashed line, and error bars indicate 95% bootstrapped confidence intervals.

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 = 0.27, 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 $d = .69$).

Next, we separately analyzed the test data from children who gave the same object more than once in a block ($n = 30$, 16 participants in the Pedagogical condition, and 14 in the Overhearing condition) and those who did not ($n = 34$ participants, 16 in the Pedagogical condition, and 18 in the Overhearing condition). Both groups showed the same pattern of success, responding above chance on fact trials in both conditions, and on word trials only in the Pedagogical condition, and there was no difference in accuracy between them (see Table S4 in the *Supplemental Online Materials*).

Mixed effects models. Models with condition (Pedagogical or Overhearing), learning target (words or facts), 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 ($X^2(2) = 51.62$, $p < .0001$; AIC for model with condition and target type: 888.78, AIC for null model: 936.40). A model which additionally included an interaction between condition and learning target did not result in a significantly better fit ($X^2(1) = 1.35$, $p = .25$; AIC with interaction: 889.43), suggesting that the impact of condition did not differ substantially by learning target, as it had in Experiment 1. 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]; $X^2(1) = 20.21$, $p < .0001$), 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], $X^2(1) = 31.84$, $p < .0001$).

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; $X^2(1) = 6.2$, $p = .01$). Children's odds of accuracy were lower in the Overhearing condition overall (OR = .59 [.28, 1.23], Wald $X^2(1) = 15.5$, $p < .001$), and children were especially bad at learning a fact associated with an unfamiliar object through overhearing (interaction OR = .27, [0.12, 0.83]; Wald $X^2(1) = 6.2$, $p < .05$). 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. Finally, children's odds of accuracy improved significantly with age (OR = 2.32 [1.37, 4.01]; Wald $X^2(1) = 6.3$, $p = .01$).

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

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 ($r_{rm}(27) = -.67$ [-.84, -.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 5) 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 $r = .01$ [-.38, .40]; $t(20) = .05$, p = 1), nor were they correlated with age (Pearson's $r = .03$ [-.36, .42]; $t(20) = .2$, p = .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.

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 $r = .33$ [-.07, .63]; $t(20) = 2$, p = .1), nor their age (Pearson's $r = -.08$ [-.45, .32]; $t(20) = -.4$, p = .7). These results suggest that in our experiment, merely looking frequently toward the experimenter may not be a good indicator that the speech is relevant.

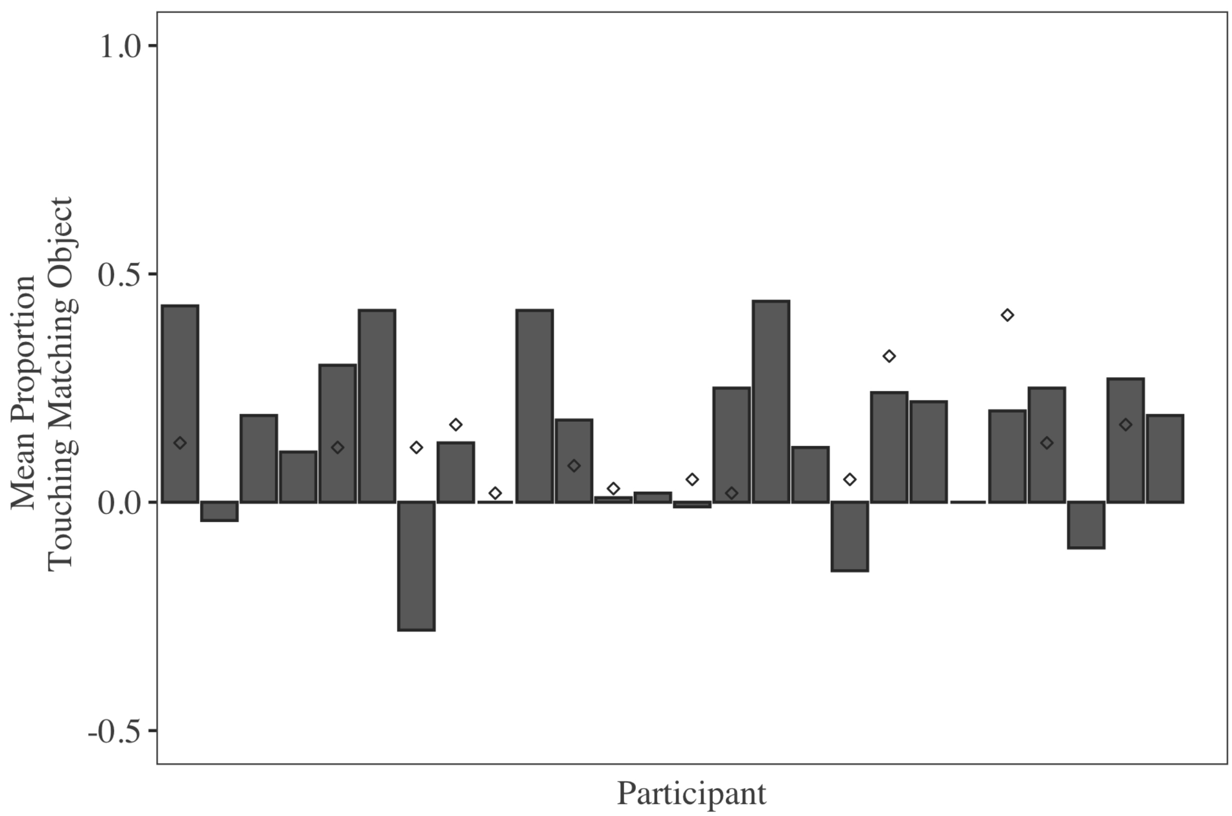

Figure 5. Matching-object touch and gaze toward experimenter by participant in Experiment 2. Bars represent mean proportion of matching-object touch by participant; points indicate the proportion of the overheard call that children looked toward the experimenter.

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,[3] 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 do the work of word-learning. Suggestive evidence comes from multiple corners: 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, Hirsh-Pasek, & Golinkoff, 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, Lupyan, Goldstein, & Spivey, 2010). In the context of our study, this latter finding might actually 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, Hirsh-Pasek, & Golinkoff, 2014). This perspective predicts deprecated 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).

[3] 91%, 92%, and 88% of parents in Experiments 1, 2, and 3, respectively, reported talking regularly on the phone around their children; see Table S2 in the *Supplementary Online Materials.*

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, Raczaszek-Leonardi, & Tylen, 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, Namiranian, Akhtar, & Tree, 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, 1991; 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, at least in a study where both overheard interlocutors were visible, two-year-olds learned a novel word when the overheard addressee was attentive and following along, but not when they were 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 co-present 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 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 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). 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, Hirsh-Pasek, & Golinkoff, 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 ($t(70) = -.2$, $p = .8$).

Procedure. The Overhearing Dialogue procedure for Experiment 3 differed from the Overhearing Halfalogue 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. 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). 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 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 $d = .20$).

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 $d = .24$), and above chance (20%) on facts (62% [44%, 78%]; $t(31) = -223$, $p < .001$, Cohen's $d = .86$). The fourteen participants who gave the same object multiple times within a test block – who can thus be thought of as having treated trials independently – exhibited overall chance performance on words, but above chance performance on facts. Children (*n* = 18) who never gave a repeat object within a test block performed above chance on both learning target types (see Table S6 in the *Supplementary Online Materials*).

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 ($X^2(2) = 24$, $p < .001$; 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 ($X^2(1) = 13$, $p < .001$; AIC for model without type: 468). Children's odds of accuracy increased as they got older (OR = 3.68 [1.79, 8.00], Wald's $X^2(1) = 13$, $p < .001$), 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's $X^2(1) = 12$, $p < .001$).

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). That is, while age was a significant predictor of fact accuracy (OR = 8.93 [2.60, 40.30], Wald's $X^2(1) = 11$, $p < .001$), adding object familiarity (familiar versus novel) to the model did not significantly improve fit ($X^2(1) = 1.7$, $p = .19$, 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 be mapped to the correct referent—a task children needed to do on their own in the Overhearing, but not Pedagogical, conditions. If this explanation is valid, children's equivalent performance on familiar and novel object facts in Experiment 3 might reveal that processing the overheard dialogue was indeed less taxing for children than processing overheard halfalogues.

Behavioral proxies of attention. Composite 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 ($r_{rm}(71) = -.46$ [-.62, -.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; $t(50) = .3$, $p = .8$). 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 = .37 [-.01, .65], t(30) = 2, p = .05) or age (Pearson's r = .33 [-.05, .63], t(30) = 2, p = .09).

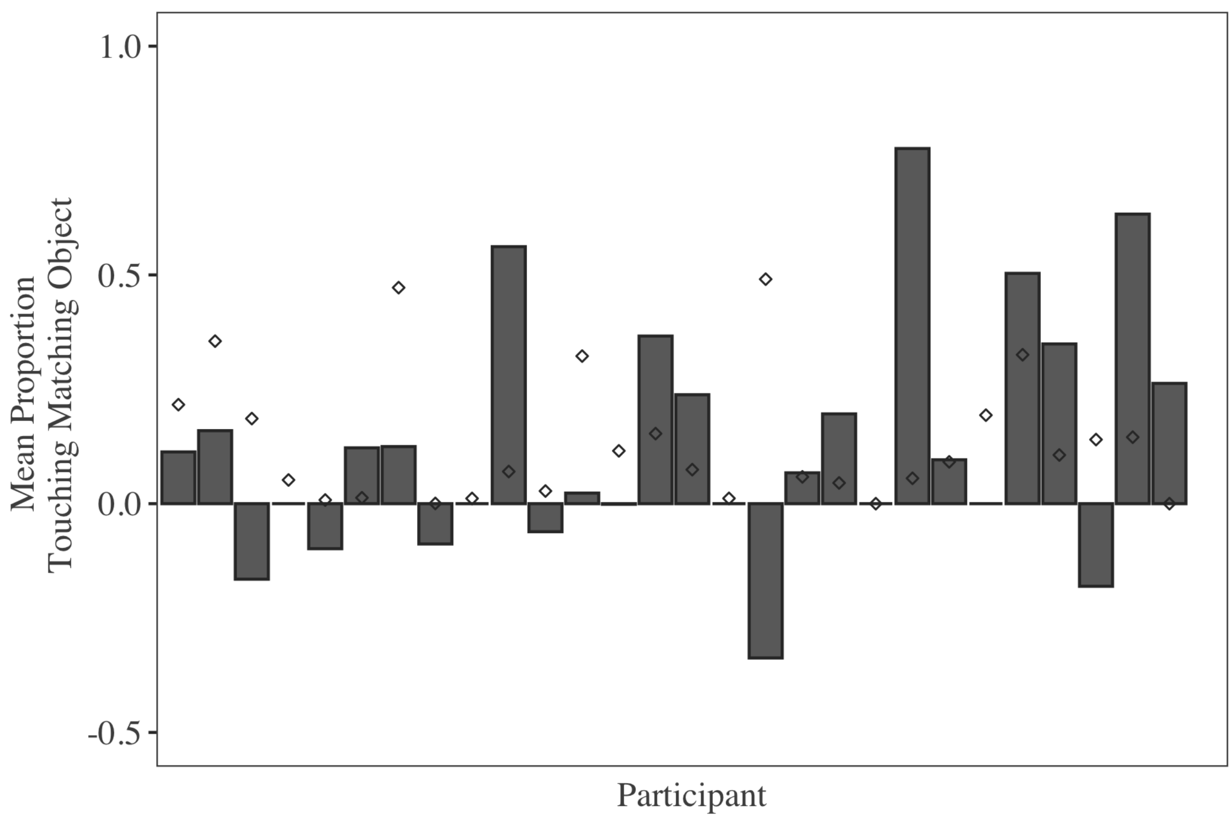

Figure 6. Matching object touch and gaze proportion by participant in Experiment 3. Magnitude of computed matching-object touch score is shown in bars, proportion of phone call in which child gazed at experimenter indicated with points.

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* = .29 [-.10, .60]; *t*(30) = 2, *p* = .1), nor their test performance (Pearson's *r* = -.14 [-.48, .25]; *t* (30) = -.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) = .2, *p* = .8, Cohen's *d* = -.05) or facts (*t*(60) = .5, *p* = .6, Cohen's *d* = -.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 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 ($X^2(1) = 2.73$, *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) = .3, *p* = .8) or gaze proportions to the experimenter (*t*(30) = -.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 ($X^2(1) = 6.92$, 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 = .41 [.20, .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.

Overall evidence for whether access to caller backchannels impacts children's online processing in this task is somewhat mixed. While there was no significant difference between the two experimental overhearing conditions on either of our attentional measures, children in the Overhearing Dialogue condition did not reliably receive positive scores, a distinction difficult to interpret both because of the small sample size and because children's motivation not to touch the objects at all is ambiguous. More promising evidence comes from the attenuation of the difference between familiar object and novel object facts, which we found when analyzing data from Experiment 3 alone and together with Experiment 2. This trend might suggest that—even in the absence of joint attention between overheard speakers—children overhearing a dialogue were able to be more effective and timely at coordinating their attention between the speech and the referential context, such that they incurred less of a cost when mapping facts to unknown 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.

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. As we discussed in the Introduction, such tests of children's real-world self-directed learning are a topic of considerable current interest, but 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 be largely equivalent in this task 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. In contrast to the findings of previous studies of overhearing in more simplified contexts, younger preschoolers (3–4.5 years) 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, Markman, & Fitzsimmons, 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 involved only 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 would likely have been further simplified in the case where they were learning a fact about a familiar object, where it would be trivial to identify *which* object they should map the new fact to, as its noun is already familiar (i.e., they didn't have to look at the objects to know which was the 'dog'). Drawing on studies comparing children's fast mapping of different novel linguistic targets, we hypothesize that the greater number of memory cues, including familiar words, present in facts broadly, and in familiar facts in particular, likely made them easier to retrieve at test, relative to their single-word counterparts (Deák & Toney, 2013).

It is also possible that the learning asymmetry derived from differences in how information about the individual objects were encoded in memory in response to the words versus facts tested in our study. From this perspective, the disadvantage we saw for linking labels to the specific objects in our task might be due to children's understanding of labels as naming *categories,* resulting in coarser encoding of individual category members. Prior work with adults suggests that in expressing information that was unique to each object—rather than category-level information—facts might have triggered more fine-grained representations of the individual objects (Lupyan, 2008; 2012). However, as each word as well as each fact was only associated with one object, a further study would be needed to evaluate this hypothesis. To test whether reference to individual items accounted for children's superior performance on overhearing facts, a future experiment might test learning of facts associated with categories of objects, and category labels that corresponded to multiple exemplars in the task (e.g., more than one *pimwit*), as well as labels that were unique to individual exemplars (e.g., a proper noun for a single *pimwit*). As it stands, our speaks to the contributions of both online language processing and memory in learning from overheard speech, and opens questions for future research.

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; Lowenstein, 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. That said, the younger children in Experiments 2 and 3 had significantly lower matching-touch scores than did their older counterparts in Experiment 1, suggesting that difficulty with tracking overheard speech online may help explain younger children's chance-level performance in learning new words via overhearing. Ultimately, however, our findings cannot adjudicate between the role of this age-related development in attentional coordination, and age-related changes in memory in explaining the difference between older and younger children's ability to learn from overhearing.

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-Sussman 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, Demir, & Baldwin, 2011), suggesting that our participants' glances to the experimenter may have reflected not only attention, but also their uncertainty and consequent information-seeking (Hembacher, deMayo, & 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 knowledge 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 & Goldin-Meadow, 2012; Shneidman et al., 2013). 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., 2018, 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 A1). 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, Griffiths, & Srinivasan, 2016; Kidd, Piantadosi, & Aslin, 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., Landau & Gleitman, 1985; Kline & Snedeker, 2015; Messenger, Yuan, & Fisher, 2015; Naigles, 1990; Arunachalam, 2013; Arunachalam, 2016; Yuan & Fisher, 2009; and nouns, see e.g., Ferguson et al., 2014, 2018; Goodman, McDonough, & Brown, 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 childrearing 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 (Mastin & Vogt, 2016; Gutiérrez & Rogoff, 2003). 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).

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 grow.

References

R Core Development Team. (2011). R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria.

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.

Arunachalam, S. (2013). Two-year-olds can begin to acquire verb meanings in socially impoverished contexts. *Cognition*, *129*(3), 569–573.

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. doi:10.1016/j.jml.2016.02.001

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2017.00456

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890.

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.

Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds?. *PloS One*, *7*(11).

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

Brown, P. (1998). 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

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early language experience in a Tseltal Mayan village. *Child Development*.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., … De Rosario, M. H. (2013). Package `pwr'.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition.* (pp. 127–149). American Psychological Association.

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.

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

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.

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.

De León, L. (1998). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161.

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

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.

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.

Fitch, A., Lieberman, A. M., Luyster, R. J., & Arunachalam, S. (2020). Toddlers' word learning through overhearing: Others' attention matters. *Journal of Experimental Child Psychology*, 104793.

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

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

Fox, J., Friendly, G. G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., ... & Suggests, M. A. S. S. (2007). The car package. *R Foundation for Statistical Computing*.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*(1), 35–53.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.

Fusaroli, R., Raczaszek-Leonardi, J., & Tylen, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, http:// dx.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.

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid "labouring in vain" by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–979. https://doi.org/10.1111/j.1467-7687.2011.01046.x

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., 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.

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. http://dx.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.

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.

Hembacher, E., Benjamin Demayo, & Frank, M. C. (2017). Children's social referencing reflects sensitivity to graded uncertainty. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS ONE*, *7*(5), e36399. 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. In *Proceedings of the 37th Annual Meeting of the Cognitive Science*

Society.

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.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277.

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

Linell (2009). Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making. IAP.

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), e12663.

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.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

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.

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development: The Official Journal of the Society for Language Development*, *11*(4), 356–368. https://doi.org/10.1080/15475441.2014.978331

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(August 1989), 357–374. https://doi.org/10.1017/S0305000900013817

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.

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2012). Young children's self-directed information gathering on touchscreens. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.

Piaget, J. (1954). *The construction of reality in the child*. Routledge.

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.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pye, C. (1986). An ethnography of Mayan speech to children. *Working Papers in Child Language*, *1*, 30–58.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85-100.

Reed, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Learning on hold: Cell phones sidetrack parent-child interactions. *Developmental Psychology, 53*(8), 1428.

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

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.

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

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

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

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.

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.

Shneidman, L., Gaskins, S., & Woodward, A. (2015). Child-directed teaching and social learning at 18 months of age: evidence from Yucatec Mayan and US infants. *Developmental Science*, *19*(3), 372-381.

Shneidman, L. A., & Goldin-Meadow, S. (2012). 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., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning?. *Psychological Bulletin*, *142*(1), 1.

Sim, Z. L., & Xu, F. (2014). Acquiring inductive constraints from self-generated evidence. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Sim, Z. & Xu, F. (2017a) Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology, 35,* 596-608.

Sim, Z. & Xu, F. (2017b) Learning higher-order generalizations through free play: Evidence from two- and three-year-old children. *Developmental Psychology 53,* 642-651.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501-532.

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

Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science*, *40*(6), 1412-1434.

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.

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.

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.

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

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841-864.

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

Yuan, S., & Fisher, C. (2009). “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*(5), 619-626.

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology, 50*, 73-79.

Acknowledgments

We thank the children and families at the Harold E. Jones Child Study Center, Lawrence Hall of Science, and Bay Area Discovery Museum for participating in our research, and Tory Rose Full, Stella Lim, Mona Sterling, Chloé Thorbrogger, Shannon Dailey, Melis Muradoglu, Marly Santora, Mika Braun, Meg Bishop, Minh-Thy Nguyen, Phyllis Lun, Harmonie Strohl, Jacqueline Nguyen, Allison Fong, Gabriela Horton, Luvy Grimaud-Vegas, and Lisa Branum for their assistance with data collection. A subset of these data were presented at the 38th Annual Meeting of the Cognitive Science Society (Foushee & Xu, 2016). This research was supported by a National Science Foundation Graduate Research Fellowship to RF (DGE-1752814), awards from the National Science Foundation (SBE-16302040) and James S. McDonnell Foundation to MS, and an award from the National Science Foundation to FX (SMA-1640816).

Appendix

Table A1

Summary of Previous Overhearing Experiments

Study	Age Range	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	
		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	25-month-olds did not demonstrate robust learning of action label
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 & 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-Sussman 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*]."	E engages in joint attention with C	Study 2 used a music game
				"Here the [*label*] goes in. But where is the [*label*]? I'll get the [*label*]"		
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.

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!

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!

Experiment 3 Experimenter/Caller Script.

EXP: Hi, how are you?

CALLER: *Doing alright, and you?*

EXP: 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.*

EXP: Do you know what a pimwit is?

CALLER: *No…*

EXP: 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?*

EXP: 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!*

EXP: 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.*

EXP: 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?*

EXP: 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?*

EXP: 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.*

EXP: 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!*

EXP: 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.)