Socially Motivated Observational Learning: Children Preferentially Learn from Overheard Speech Addressed to Their Own Mothers

Sarah Goodman[1] **Ruthe Foushee**[1]
[1]Department of Psychology, The New School for Social Research, 80 5th Ave, New York, NY, 10011

Abstract

When does overheard speech support early word-learning? The present study provides preliminary evidence that children's closest social partners (here, mothers) structure attention to overheard speech. In a Tseltal Mayan community where overhearing is central to socialization, 75 mother-child dyads participated in overhearing-sessions exposing them to two novel words. In each session, one child's mother was directly addressed ("Mother-Addressed" condition), while the second mother merely observed ("Mother-Unaddressed" condition). Immediately after, both children completed two gaze-based tests of word recognition and learning. Results show that, compared to Mother-Unaddressed children, Mother-Addressed children exhibited (i) greater one-shot recognition of the novel word whose referent was visible during the overhearing-session; and (ii) a word-learning advantage on cross-situational familiarization trials . Insofar as *who* is being spoken to matters, the present study finds that not all overheard input is created equal: rather, results provide preliminary evidence that observational word-learning is socially-structured.

Keywords: overhearing; social learning; attention

Take an ordinary morning rush: two parents are hurrying their children out the door when a glass slips off the table and shatters. One parent looks to the other and declares: '*oh, fork.*' Euphemisms like this one presuppose that overheard speech may inadvertently be taken up by children and, furthermore, point to caregivers' understanding of their potential role in structuring children's attention to overheard speech.

Despite the ubiquity of these amendments, the language-learning literature often assumes that learning proceeds only or primarily through child-directed speech (Fernald, 1984; Gaskins, 2020; Golinkoff et al., 2019; Ramírez-Esparza et al., 2014; Soderstrom, 2007; Tamis-LeMonda et al., 1996, 2001, 2014; Tomasello & Farrar, 1986). Seemingly tailored to maximize uptake, child-directed speech is often prosodically exaggerated, tuned for semantic relevance and complexity, and rich in gaze- and gesture-based referential cues (Fernald, 1984; Tamis-LeMonda et al., 2001; Tomasello & Farrar, 1986). Furthermore, it is contingently learner-responsive: adjustments are continuously fine-tuned in accordance with signs of engagement and comprehension (Clark, 2014). One way to understand these modifications is that they collectively serve to capture, sustain, and direct *attention*, thereby supporting learning (Fernald, 1989; Golinkoff et al., 2015).

Other-directed speech is, by contrast, agnostic to children's attention. Nevertheless, a wave of experimental work suggests that children are indeed capable of learning novel words from other-directed input. These studies often feature two adult experimenters, who, seated across from a child, jointly attend to a novel object while one — pointing and demonstrating — teaches the other its name (Akhtar, 2005; Akhtar et al., 2001; Floor & Akhtar, 2006; Martínez-Sussmann et al., 2011; O'Doherty et al., 2011; Shneidman et al., 2009).

This scaffolding might support attention in a similar manner to child-directed speech: semantic complexity is pared down (Foushee et al., 2016) and referential ambiguity reduced, while ostensive signals and joint attention amplify communicative intent. Under such supportive conditions, children as young as 18 months attend to overheard speech, and subsequently show evidence of learning (Floor & Akhtar, 2006). Insofar as this body of work establishes that other-directed input can support early word-learning, it does so in circumstances for which naturalistic analogues are rare.

Yet in Tseltal Mayan communities and others like them, children appear to develop linguistic competence almost exclusively via overhearing (Brown & Gaskins, 2014; Casillas et al., 2020; De León, 2011; Foushee & Srinivasan, 2024; Rogoff, 2003). Carried on their mothers' backs throughout the day until walking-age, Tseltal children have the opportunity to overhear great volumes of other-directed speech (Brown & Gaskins, 2014; Casillas et al., 2020; Rogoff, 2003). De León's (2011) work establishes this type of observation as a *bona fide* participatory role — akin to speaker, addressee, or bystander — concluding that it affords children structured access to exchanges, which become meaningful sites of learning. Indeed, despite receiving little child-directed input, Tseltal children reach language milestones on a timeline similar to their Western peers (Casillas et al., 2020; Foushee & Srinivasan, 2024). Such findings have primarily been interpreted as evidence of a contextually-bound learning adaptation by children socialized in indigenous contexts where active observation and 'listening in' are central values (*à la* Correa-Chávez and Rogoff, 2009; Rogoff, 2003).

Rather than exclusively evidencing a culture-specific learning adaptation, this observational 'expertise' may also point to a more general mechanism via which overheard speech becomes effective language-learning input. In virtue of being carried on their mothers' backs, Tseltal infants have near-continuous, privileged access to not just any overheard speech, but "front-row seats" to the conversations of their closest social partners (Foushee & Srinivasan, 2024). A pattern of preferential attention to caregiver-directed speech, then, might help explain how Tseltal children develop linguistic competency without delay, despite receiving little child-directed input (indeed, some of the strongest experimental evidence for Mayan children's facility with observational learning comes from a task in which they learn by overhearing an experimenter instruct their *siblings* in a novel craft (Correa-Chávez & Rogoff, 2009). As such, near-constant physical proximity to caregivers and limited child-directed address may serve to render this more broadly available pathway to word-learning especially visible.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

If children's close social partners drive attention to overheard speech, and attention to speech facilitates word-learning, then children should learn new words more readily from overheard speech addressed to their own mothers, compared to strangers. The current study tests this hypothesis in a Tseltal community where infants are rarely addressed and learn primarily through observation (Brown, 2011; Brown & Gaskins, 2014). In each overhearing-session, a pair of age-matched children observed an interaction that addressed only one of their mothers, while the other mother observed. Immediately after, both children completed two gaze-based tasks testing their recognition and learning of two novel words introduced during the overheard exchange. In finding evidence of learning advantages for Mother-Addressed children, the present study suggests that caregivers play a role in structuring children's attention to overheard speech, vindicating what parents implicitly grasp every time they substitute *'fork'*.

Method

Participants

Participants were 75 children in Chiapas, Mexico, aged 10–41 months ($M = 28.38$, $SD = 7.51$). Each overhearing-session involved a pair of age-matched children and their mothers. In each pair, one child was assigned to the "Mother-Addressed" condition and the other to the "Mother-Unaddressed" condition.

Of the initial sample, two participants were excluded for fussiness, 20 for faulty audio or video, and one due to experimenter error. This resulted in a sample of $N = 52$ for the first task ($n = 31$ Mother-Addressed, $n = 21$ Mother-Unaddressed). On the second task, an additional nine participants were excluded due to faulty audio, and two due to technical issues with stimuli presentation, yielding $N = 41$ ($n = 24$ Mother-Addressed, $n = 17$ Mother-Unaddressed).

Stimuli and Procedure

Exposure Each overhearing-session began with the Mother-Addressed dyad seated in a testing room. The Mother-Unaddressed dyad was instructed to 'wait their turn' at the room's entrance, situated within earshot (Figure 1, left panel). The experimenter entered the room carrying a novel herb in a bucket, and initiated a selling-exchange with the Addressed-Mother. Both children were similar distances from the experimenter to ensure equal opportunity for overhearing. The experimenter haggled with the Addressed-Mother over the novel herb, *nina*, and also introduced a second novel word, *dadin*, whose referent remained unseen.

Test Immediately following the interaction, both children completed two gaze-based tests of word recognition and learning, administered simultaneously in separate testing rooms (see Figure 1, right panel). The tasks began with a brief calibration phase, where an attention-grabbing animation (a sniffing bunny) was serially presented across the top, bottom, and sides of the screen. Children's gaze was continuously video-recorded throughout the test phase, and their initial tracking of these prompts served as calibration points for later manual coding of left–right-looking.

One-Shot Word-Recognition Test. In the first test, children viewed videos of the referent for *nina* — both said and seen during the overhearing session — alongside a distractor, while a female voice instructed them to 'look' (*'ilawil'*) at each image in turn. The next screen, lasting 4 s, showed both the target and distractor simultaneously, while the accompanying audio asked: 'where is the *nina*?' (*banti ay te* nina*?*).

Here, successful identification of the target may reflect knowledge of the correct word–referent mapping, but does not require it. The target may also be identified via an in-

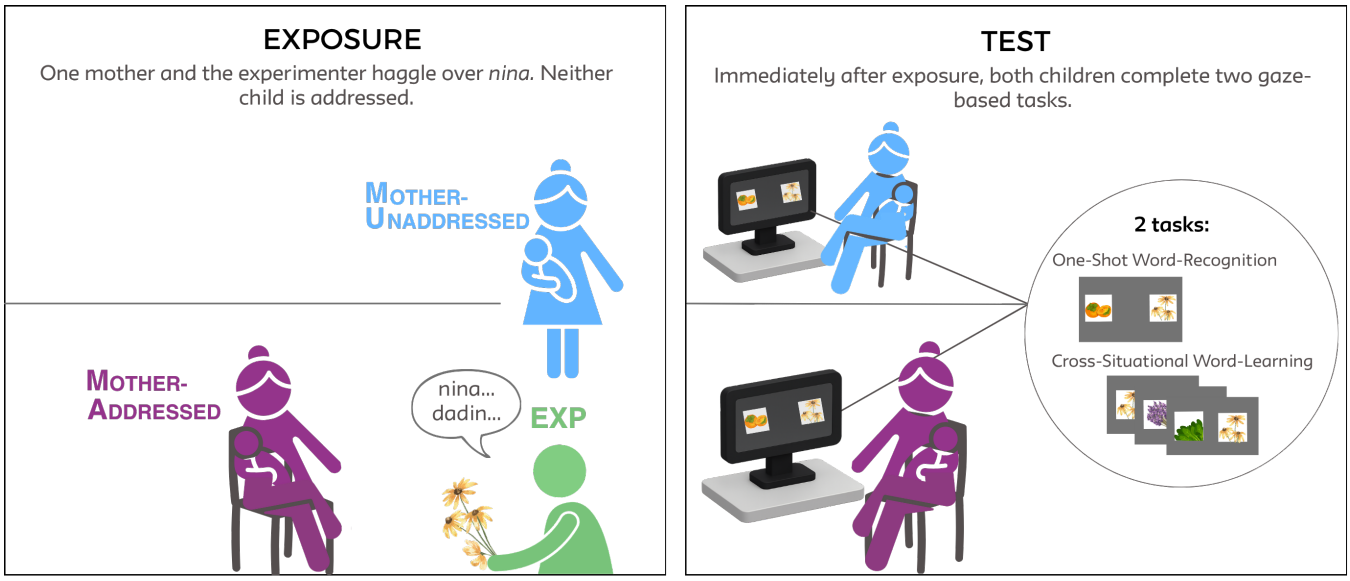

Figure 1: Setup of the overheard interaction (left) and the two subsequent gaze-based learning tasks (right).

ference from the recent co-occurrence of the word and referent during the overhearing-session. Accordingly, this task was intended to capture early sensitivity to word-recognition immediately following minimal exposure, without requiring evidence of stable learning (Fernald et al., 1998).

Cross-Situational Word-Learning Task. Participants then completed a cross-situational word-learning task (Smith & Yu, 2008). Across 28 trials, participants encountered four novel word-referent pairs, which differed in their prior exposure histories: the word *nina* — along with its visible referent — was previously introduced during the overhearing-session, and encountered again in the previous task; *dadin* was uttered during the overhearing-session, but without a visible referent; and *soman* and *tani* were both entirely unfamiliar.

Participants saw, in a fixed order, 24 individually ambiguous familiarization trials followed by four test trials of a similar structure. At the onset of each trial, two images appeared on the screen (see Figure 2), where they were briefly animated for 4*s*. On familiarization trials, the accompanying audio played the two corresponding words 1500 ms apart, in pseudorandom order. On the four test trials — one for each novel word — the audio repeated the same target word twice.

Words were audio-recorded by a female native speaker of Tseltal in a consistent register. Referent images — matched roughly for size and salience — depicted unfamiliar flora determined to be unnameable in this context via consultation with research assistants and caregivers during piloting.

Trials were counterbalanced such that each word-referent pair appeared equally with every other, in both the left and right visual position, and such that the left word was played first on half of trials, and second on the other half — together ensuring that correct mappings could only be inferred via statistical learning across exposures (Smith & Yu, 2008).

This paradigm models two important features of ecological word-learning. First, it captures the naturalistic interdependence of vocabulary knowledge. On early trials, word-referent mappings are ambiguous, but as participants progress, evidence accumulates for the correct mappings via the tracking of statistical co-occurrences. Inferences to one correct word-referent mapping support inferences to all the others, such that the individual ambiguity of each trial may ultimately be extinguished (Smith & Yu, 2008). Everyday disambiguation works similarly: say, after Penn introduces himself, my confidence about which one is Teller skyrockets.

Second, the task was included because of its potential to capture the relative advantages of different exposure conditions for word-learning. Familiarity with a word's phonological form may give learners a 'leg-up' if its referent is later encountered (Swingley, 2007). Children's phonological familiarity with two out of the four unfamiliar words — which here, varies as a function of attention during exposure — may differentially afford attentive overhearers a head-start on subsequent word-learning. On this logic, and consistent with research on cross-situational learning showing that knowledge of word–referent mappings begins to emerge before test, we expect any advantages of prior exposure to be most pronounced during early familiarization trials (Smith & Yu, 2008; Trueswell et al., 2013).

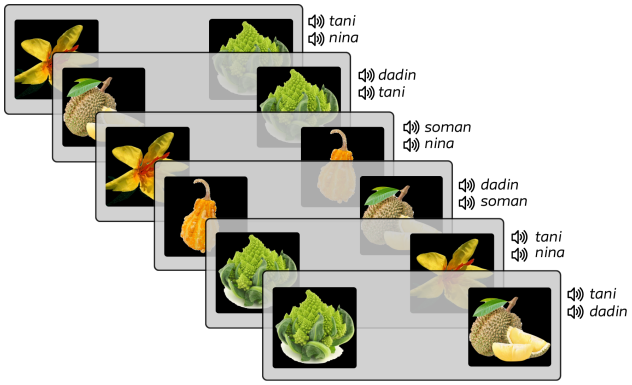

Figure 2: Structure of the cross-situational word-learning familiarization trials (1–6).

Coding and Analysis

Manual Gaze Coding Trained research assistants, blind to the location of the target image, manually coded periods of left- and right-looking from video recordings of participants' gaze-behavior using ELAN (Ver. 7.1; Max Planck Institute for Psycholinguistics, The Language Archive, 2026; Wittenburg et al., 2006). Coding proceeded at the video's frame-level resolution of 30 fps. A left- or right-looking period was defined as beginning when a child's gaze crossed the midline of the screen toward one side, and ended when gaze returned to center, crossed to the opposite side, or moved off-screen.

Preregistered Analyses Following a preregistered plan (osf.io/me6nd), word-recognition on each test component — the one-shot test, cross-situational familiarization trials, and cross-situational test trials — was indexed via target-looking proportion. For each trial, target-looking proportion was calculated as the time spent looking to the target image out of the summed time spent looking at either image during a designated analysis window (detailed by task type, below).

One-Shot Word-Recognition Test. For the one-shot word-recognition test, the analysis window began 367 ms after the onset of the word *nina*, accounting for the latency in children's initiation of saccades in response to audio stimuli (Fernald et al., 1998), and ended 3133 ms later.

We use a two-sample t-test to compare target-looking between conditions, and fit a mixed-effect linear model predicting proportion target-looking on the one-shot test from condition and age (mean-centered), with a random intercept for overhearing-session to account for the paired design. We predicted that children in the Mother-Addressed condition would show significantly higher mean target-looking compared to children in the Mother-Unaddressed condition.

Cross-Situational Word-Learning Familiarization Trials 1–6. Previous studies interested in tracking the trajectory of children's word-referent mappings across familiarization trials typically use a forced-choice paradigm (e.g., Berens et al., 2018; Trueswell et al., 2013). Given the broad age range of our participants, we elected to minimize task demands by instead using gaze to the referent after hearing a word as an implicit, graded measure of knowledge. Each familiarization trial thus contained two analysis windows: Each analysis window began 367 ms after the onset of the target word, and ended at the onset of the next task event (either the onset of the second word or the end of the trial).

We used t-tests to compare mean target-looking proportion across familiarization trials 1–6, and fit a linear mixed-effects model to children's word-by-word target-looking data, with condition, age (mean-centered) and trial number as fixed effects, and random intercepts for overhearing-session and participant. Again, we predicted that Mother-Addressed children would show a significantly higher mean target-looking proportion than Mother-Unaddressed children.

To examine whether any condition advantage was graded by exposure type, we fit a second linear mixed-effects model with word exposure type and its interaction with condition as added predictors, with planned contrasts for each exposure-type. We conducted pairwise comparisons to assess whether any condition advantage differed in size across exposure-types, with the prediction that the advantage for Mother-Addressed children would be greatest for *nina* (both said and heard during exposure), followed by *dadin* (heard only), and then by *soman* and *tani* (entirely novel). Thus, contrasts compared the condition advantage for *nina* vs. all other words; for *dadin* vs. the entirely novel words; and for *soman* vs. *tani*.

Cross-Situational Word-Learning Test Trials. For cross-situational test trials, analysis windows began 367 ms after the audio onset of the first occurrence of the target word and ended 3133 ms later.

Mean target-looking proportion was analyzed separately for each of the four words using a one-sample t-test against chance (0.5). Any early advantage for children in the Mother-Addressed condition was predicted to be 'washed out' by test, by which time all participants have had the opportunity to accumulate sufficient evidence of the correct word–referent mappings. Therefore, we predicted that participants in both conditions would demonstrate above-chance target-looking across the four novel items. We fit a linear mixed-effects model predicting target-looking from condition, age, and word, with random intercepts for participant and overhearing-session. By the same logic, we predicted no significant effect of condition or word.

All analyses were conducted in R (Team, 2021). We report means and bootstrapped confidence intervals by condition, along with coefficients and their standard errors for models fit using the `lme4` package (Bates et al., 2015), and likelihood ratio tests conducted using the `anova` command to compare models with and without a given variable.

Results

One-Shot Word Recognition

Children[1] in the Mother-Addressed condition looked significantly more to the target than children in the Mother-Unaddressed condition ($M = .58$, 95% CI [.49, .67] vs. $M = .44$, 95% CI [.33, .54]; $t(39.65) = 2.12$, $p = .041$; see Figure 3), further supported by a linear mixed-effects model fit to the proportion target-looking data (condition $\beta = -0.15$, $SE = 0.07$, $\chi^2(1) = 4.414$, $p = .036$). Age was not a significant predictor ($\beta = 0.003$, $SE = 0.005$, $\chi^2(1) = 0.433$, $p = .51$).

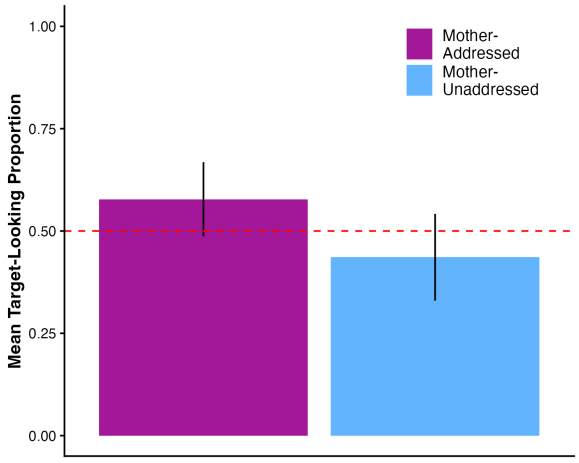

Figure 3: Comparison of mean target-looking proportion during the one-shot test between Mother-Addressed and Mother-Unaddressed participants. Error bars indicate 95% bootstrapped confidence intervals, here and throughout.

Cross-Situational Word-Learning Task

Familiarization Trials 1–6. Children in the Mother-Addressed condition looked at the target significantly longer than children in the Mother-Unaddressed condition ($M = .54$, 95% CI [.49, .60] vs. $M = .45$, 95% CI [.39, .52]), $t(279.42) = 2.12$, $p = .035$ (Figure 4, top panel).[2] A linear mixed-effects model showed a significant effect of condition ($\beta = -0.10$, $SE = 0.05$, $\chi^2(1) = 4.453$, $p = .035$), such that Mother-Addressed children showed significantly greater target-looking than Mother-Unaddressed children across trials 1–6. Age ($\beta = -0.001$, $SE = 0.003$, $\chi^2(1) = 0.261$, $p = .61$) and trial number ($\beta = -0.005$, $SE = 0.012$, $\chi^2(1) = 0.157$, $p = .692$) did not significantly improve model fit.

A second model additionally included word and its interaction with condition as fixed effects, and planned contrasts

[1] As per our preregistered exclusion criteria, data from four participants were excluded from analysis for $< 1/3$ total looking-time during the analysis window, resulting in a final sample of $N = 48$ ($n = 30$ Mother-Addressed, $n = 18$ Mother-Unaddressed).

[2] Based on exclusion criteria, 63 of 246 possible early familiarization trials were excluded from analysis due to $< 1/3$ total looking-time, leaving 183 trials from 39 participants (1–6 trials per participant).

for exposure type. Condition remained a significant predictor ($\beta = -0.11$, $SE = 0.05$, $\chi^2(1) = 4.5$, $p = .034$). No condition-word interactions reached significance ($\chi^2(3) = 0.734$, $p = .865$): *nina* vs. *dadin*, *soman*, and *tani* ($\beta = 0.02$, $SE = 0.02$); *dadin* vs. *soman* and *tani* ($\beta = -0.02$, $SE = 0.04$); *soman* vs. *tani* ($\beta = -0.005$, $SE = 0.06$), indicating that the advantage for Mother-Addressed children did not significantly vary by exposure-type.

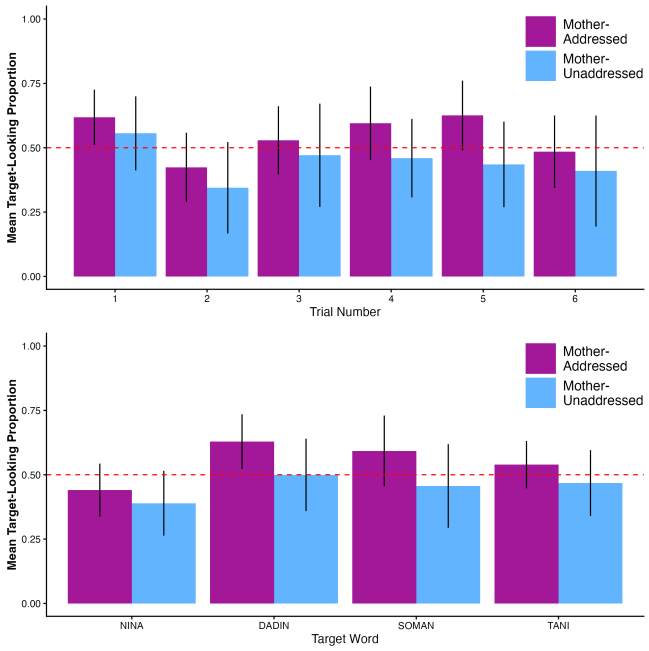

Figure 4: Mean target-looking proportion during cross-situational familiarization trials 1–6. *Top*: Mean target-looking proportion by trial. *Bottom*: Mean target-looking proportion by word.

Test Trials One-sample t-tests revealed that target-looking did not differ significantly from chance for any of the four words: *nina* ($M = .50$, 95% CI [.34, .66], $t(15) = -0.002$, $p = .998$); *dadin* ($M = .53$, 95% CI [.38, .67], $t(17) = 0.38$, $p = .707$); *tani* ($M = .58$, 95% CI [.41, .75], $t(15) = 0.96$, $p = .350$); *soman* ($M = .42$, 95% CI [.26, .58], $t(17) = -1.05$, $p = .308$). After applying exclusion criteria, the Mother-Unaddressed group was underpowered for between-group comparisons.[3]

Exploratory Analyses of Task Engagement

Given the exclusion of trials with $< 1/3$ total looking time, and participants with $< 3/4$ usable familiarization trials, we compared total looking time between conditions to examine whether exclusion patterns reflected broader differences in task engagement. On the one-shot test, Mother-Addressed children demonstrated significantly greater total looking time ($M = 4491.61$ ms, 95% CI [4063, 4920]) compared to Mother-Unaddressed children ($M = 3184.63$ ms, 95% CI [2513, 3856]), $t(33.17) = 3.42$, $p = .002$. A linear model predicting total looking time from condition and age, with a random intercept for overhearing session, confirmed the significance of this pattern ($\beta = -453.52$, $SE = 181.98$, $F(1, 33.42) = 6.21$, $p = .018$; see Figure 5).

This difference in total looking time may partially account for the differential exclusion rates between conditions, and raises the interesting possibility that attentional engagement during exposure extended beyond the overhearing session itself, perhaps influencing engagement with subsequent tasks.

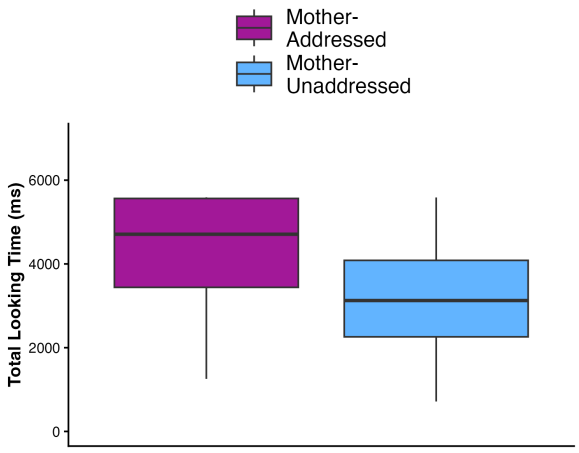

Figure 5: Distribution of total looking time during the one-shot task by condition.

General Discussion

In each study session, two children overheard a conversation that introduced them to two novel words; one child's mother was the addressee, while the other child's mother only observed. On two subsequent gaze-based tasks, the children whose mothers were addressed demonstrated advantages in recognizing the two novel words, relative to their counterparts whose mothers were mere observers. First, immediately after exposure, Mother-Addressed children demonstrated greater one-shot recognition of the overheard novel word whose referent was visible during the overheard conversation. Second, they demonstrated an advantage on early cross-situational familiarization trials, where differences in attention to the overheard interaction were most likely to show up as differences in target-looking (Trueswell et al., 2013). These findings, therefore, provide some initial support for the idea that overheard speech involving caregivers elicits greater attention from children (here, relative to overheard speech involving *other* children's caregivers).

It has been pointed out that these advantages for Mother-Addressed children may be confounded by the heightened personal relevance of (m)other-directed speech. Yet, rather

[3] In accordance with preregistered criteria, participants had to contribute data on $\geq 3/4$ of all familiarization trials to be included in analyses of test trial performance. This criterion severely reduced our sample for test-trial analyses, to $N = 18$ ($n = 13$ Mother-Addressed, $n = 5$ Mother-Unaddressed).

than constituting an alternative interpretation, this proposed confound is part of the very effect the study is designed to capture: The hypothesis that caregivers structure children's attention to overheard speech is motivated, in part, by the suspicion that caregiver-addressed speech is treated by children as comparatively relevant.

Indeed, work on early social cognition has shown that infants preferentially track the interactants of those closest to them from the earliest months of life and, by the end of their first year, draw inferences from these exchanges, which in turn guide their own social preferences (McMahon & Isik, 2023; Powell & Spelke, 2018). Further, a series of recent studies show that this preference is especially strong for one's caregivers: 12-month-old infants preferentially approached speakers they observed interacting with their parents, but did not generalize the same preference to interactants of unfamiliar adults (Thomas et al., 2022). The present results suggest that caregivers may similarly structure their children's attention to overheard speech — whether via a heightened perception of personal relevance, or otherwise — even when they are mere addresses.

Limitations

The early advantages observed for Mother-Addressed children did not translate into reliable evidence of learning on the final four cross-situational word-learning test trials. Other attempts to adapt cross-situational word-learning paradigms to non-Western 'field' contexts have similarly produced null results at test: In one study, participants showed chance-level performance at test not only on a cross-situational word-learning task, but also on an unambiguous word-learning task (i.e., where familiarization trials presented just one image and its corresponding word Mulak et al., 2021). These results were interpreted as reflecting population differences in task-construal, rather than a lack of learning (Mulak et al., 2021).

Although cross-situational word-learning is assumed to proceed via the automatic tracking of co-occurrences, successful performance at test additionally depends on interpreting the test-phase as a meaningful opportunity to demonstrate acquired word knowledge. When these assumptions are not shared, participants may show chance-level performance at test even in the presence of learning (Mulak et al., 2021).

Future Directions

Although the present study was conducted in a Tseltal Mayan community where children learn primarily through observation, the hypothesis that caregivers structure children's attention to overheard speech need not be culture-specific. Replicating this study in a Western context — where the present task paradigm has yielded robust evidence of learning — would help determine whether the present findings extend beyond contexts where observation is a primary route to language learning. Such a replication would also afford the opportunity to control for the physical proximity of both children to the addressee.

Further, attempts to adapt traditional laboratory-based paradigms to field settings remain valuable, but require continuous fine-tuning for ecological sensitivity. To this end, embedding target words in non-ostensive carrier phrases on cross-situational word-learning trials may be a particularly important amendment in the present context, where isolated word presentations are, in the relative absence of child-directed speech, likely to be rare. Further, establishing that possession of a word's phonological form via overhearing facilitates subsequent word-learning may require not only a socially-structured overhearing opportunity, but a correspondingly 'social' testing context (notably, human-delivered speech during initial piloting yielded improved performance on cross-situational test trials, compared to the computer-based audio used here).

Videos of the overhearing-sessions are currently being coded for behavioral indices of attention to the overheard interaction, including children's gaze allocation across the experimenter, caregivers, and the referent of the novel word. Characterizing these attentional patterns across overhearing-sessions may reveal differences in children's engagement with, and access to, overheard speech, accounting for some of the variability in learning outcomes.

Further, the naturalistic nature of the haggling episodes introduced substantial variability between overhearing-sessions, from variability in the contingency between interlocutors (e.g., in the form of joint attention and responsive-turn taking), to variability in caregiver affect, and how frequently, if at all, the Addressed-Mother produced the novel words herself. If this constellation of factors sounds familiar, it is perhaps from the extensive literature delineating the conditions under which *child*-directed speech supports language-learning (Fernald, 1984; Gaskins, 2020; Golinkoff et al., 2019; Ramírez-Esparza et al., 2014; Tamis-LeMonda et al., 1996, 2001, 2014; Tomasello & Farrar, 1986). Overheard input, on the other hand, is often treated by the language learning literature as relatively monolithic (Foushee et al., 2023). Here, by taking seriously the possibility that children's close social partners structure their attention to, and learning from, overheard speech, we find evidence that children show greater recognition of overheard novel words when their own mothers — rather than strangers — are the addressee. These results invite us to consider that observational learning opportunities — and not just pedagogical ones — may constitute meaningful instances of social learning for children, structured by their closest social partners.

Acknowledgments

Many thanks to participating families in Tenejapa; to Petrona Girón López and Miko Eugenio for assistance with data collection; to Alicia Kim, Saina Aryanejad and Omer Raz for coding; and to Travis Stackow, Lorraine Afflitto, and Sofiy Inck for feedback.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199–209.

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *convergence*, *12*(1), 2.

Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*(7), 1132–1136.

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

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

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835.

Clark, H. H. (2014). How to talk with children. In *Language in interaction* (pp. 333–352). John Benjamins.

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.

De León, L. (2011). Language socialization and multiparty participation frameworks. *The Handbook of Language Socialization*, 81–111.

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

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

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

Foushee, R., Horton, G., & Srinivasan, M. (2023). Qualitative variability in early overhearing experiences. *PsyArXiv*.

Foushee, R., Griffiths, T., & Srinivasan, M. (2016). Lexical complexity of child-directed and overheard speech: Implications for learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *38*.

Foushee, R., & Srinivasan, M. (2024). Infants who are rarely spoken to nevertheless understand many words. *Proceedings of the National Academy of Sciences*, *121*(23), e2311425121.

Gaskins, S. (2020). Cultural perspectives on infant-caregiver interaction. *Roots of Human Sociality*, 279–298.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344.

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.

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

Max Planck Institute for Psycholinguistics, The Language Archive. (2026). *ELAN* (Version 7.1) [[Computer software]]. Nijmegen.

McMahon, E., & Isik, L. (2023). Seeing social interactions. *Trends in Cognitive Sciences*, *27*(12), 1165–1179.

Mulak, K. E., Sarvasy, H. S., Tuninetti, A., & Escudero, P. (2021). Word learning in the field: Adapting a laboratory-based task for testing in remote Papua New Guinea. *PLoS One*, *16*(9), e0257393.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-party social interaction and word learning from video. *Child Development*, *82*(3), 902–915.

Powell, L. J., & Spelke, E. S. (2018). Third-party preferences for imitators in preverbal infants. *Open Mind*, *2*(2), 61–71.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891.

Rogoff, B. (2003). *The cultural nature of human development*. Oxford university press.

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.

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

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

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, *43*(2), 454.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748–767.

Tamis-LeMonda, C. S., Bornstein, M. H., Baumwell, L., & Melstein Damast, A. (1996). Responsive parenting in the second year: Specific influences on children's language and play. *Early Development and Parenting: An International Journal of Research and Practice*, *5*(4), 173–183.

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126.

Team, R. C. (2021). R: A language and environment for statistical computing. published online 2020. *Supplemental Information References S*, *1*, 371–78.

Thomas, A. J., Saxe, R., & Spelke, E. S. (2022). Infants infer potential social partners by observing the interactions of their parent with unknown others. *Proceedings of the National Academy of Sciences*, *119*(32), e2121390119.

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

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

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006, May). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the fifth international conference on language resources and evaluation (LREC'06)*. European Language Resources Association (ELRA).