Toddlers' Word Learning through Overhearing: Others' Attention Matters

Allison Fitch[1*], Amy M. Lieberman[1], Rhiannon J. Luyster[2] & Sudha Arunachalam[3]

[1]Boston University
[2]Emerson College
[3]New York University

*Corresponding author: 617 Commonwealth Ave., Boston, MA, 02215. afitch@bu.edu

Competing Interests: The authors declare no conflicts of interest with this study.

Abstract

In lab settings, children are able to learn new words from overheard interactions, yet in naturalistic contexts this is often not the case. We investigated the degree to which joint attention within the overheard interaction facilitates overheard learning. Twenty two-year-olds were tested on novel words they had been exposed to in two different overhearing contexts: one in which both interlocutors were attending to the interaction, and one in which one interlocutor was not attending. Participants only learned the new words in the former condition, indicating that they did not learn when joint attention was absent. This finding demonstrates that not all overheard interactions are equally good for word learning--attentive interlocutors are crucial when learning words through overhearing.

Keywords: toddlers, joint attention, overhearing**,** word learning, lexical development, third-party learning

A hallmark of development is the apparent ease with which children learn new words. Much of the word-learning literature has argued that children learn words in contexts that are child-directed (e.g., Rowe, 2008), ostensive (e.g., Horst & Samuelson, 2008), and rich in joint attention cues (e.g., Tomasello & Farrar, 1986). However, some researchers have argued that word learning can also occur through overheard speech—linguistic input that children can access, but is not directed toward them. Overheard speech comprises a significant amount of linguistic input in certain non-western cultures (e.g., Ochs & Shiefflin, 1984) and non-middle class contexts (Sperry, Sperry & Miller, 2018). Although word learning through overhearing has been demonstrated in laboratory studies in children as young as 18 months (e.g., Akhtar et al., 2001; Floor & Akhtar, 2006; Gampe et al., 2012), naturalistic studies have noted that quantity of overheard speech does not significantly predict vocabulary outcomes (Weisleder & Fernald, 2013; Shneidman & Goldin-Meadow, 2012). We propose that--like with directed speech--not all overheard speech is equally conducive to word learning.

One way that potential learning situations vary, whether they involve overheard or directed speech, is in the availability of joint attention cues. For directed interactions, word learning is more likely to occur inside of joint attention than outside it (e.g. Tomasello & Farrar, 1986). It is also possible that children observing an interaction may be sensitive to the presence or absence of joint attention between two interlocutors. In this case, joint attention between the interlocutor and conversational partner would not involve the child, yet the child would observe the shared attention during the interaction. The child's experience of the word form and the referent to map it to will not change if the two interlocutors are engaged in joint attention or not. Still, there is reason to believe that interlocutors who are engaged joint attention may provide a good word learning opportunity. From a theoretical perspective, joint attention reduces

referential ambiguity by providing multiple gaze cues to a target referent and provides evidence that the word learning opportunity is relevant to attend to (e.g., Gergely & Csibra, 2009). There is also empirical evidence to support this claim. Eighteen-month-olds were able to find a hidden object when its location was communicated through a third-party interaction only when that third-party interaction also contained joint attention cues (Grafenhain et al., 2008). In the word learning literature, toddlers (24-32 months) are better word learners in both directed and overheard interactions when those interactions are socially contingent (e.g., O'Doherty et al., 2011; Roseberry et al., 2014), and joint attention is one potential cue to social contingency.

If children rely on joint attention cues when learning words through overhearing, we would expect that interactions where interlocutors are not engaged in joint attention will not support word learning. Our questions were as follows. (1) Can we replicate prior studies to demonstrate that toddlers can learn words from overheard interactions in which both parties are engaged in joint attention? (2) Can these same toddlers learn words from overheard interactions in which one of the parties is not attending, and thus there is no joint attention between the interlocutors? To examine these questions, we measured toddlers' word learning from two overheard interactions that differed in an addressee's visual attention to the interaction, one in which the addressee attended, and one in which she looked down at her notes. If toddlers take the partner's view as their own, as we predict, then they should learn the novel word when the partner is attending to the scene, but not when the partner does not.

Methods

Participants

Twenty healthy, English-speaking toddlers (10 male, 10 female; 21;01-31;22, M = 25;19, SD = 118 days) participated in the study. An additional four toddlers were excluded; three did

not pass the Familiarization trials (see Procedure), and one did not meet the language eligibility criterion (exposure to English at least 70% of the time). The sample size and age range were selected from prior similar studies (Akhtar, 2005; O'Doherty et al., 2011). An a priori power analysis using effect sizes from those studies with a goal of achieving 95% power yielded a sample size of 20. Families were recruited from the greater Boston area. Most participants were White (16/20), one was Black/African-American, and three were two or more races. Of those who reported maternal education (n = 16), all except one had a Bachelor's degree or higher. Ethical approval of all recruitment and study procedures was obtained from the Boston University Institutional Review Board.

Stimuli

Stimuli were three familiar objects (a toy pig, small wooden spoon, and child's drinking cup) and six novel objects. The novel objects were uncommon household objects and toys, such as a citrus squeezer, clapper toy, and silicone oven mitt. Before the study, caregivers were shown photographs of the novel objects and asked if their child was familiar with or had a label for any of the toys. An additional novel object was held in reserve to replace an object indicated as familiar. Novel objects were divided into two sets; the objects within a set were always presented together. During the experiment, the objects were hidden in opaque, blue plastic boxes. During test trials, toddlers were encouraged to place their chosen object into a "penguin," which was a small black wastebasket decorated as a penguin (see Luyster & Lord, 2009; Luyster & Arunachalam, 2018).

Procedure

Families arrived to the laboratory and were asked to provide written informed consent, demographic information, and information about their child's vocabulary (MacArthur-Bates

Communication Development Inventory II Short Form A; Fenson et al., 2000). During this time, toddlers were introduced to the "penguin" and shown that objects (generic playroom toys) could be put into it. No stimulus objects were shown at this time.

Toddlers were then directed to sit at a child-sized desk and chair. The study began with a **Familiarization** trial, whose purpose was to familiarize the toddler with the procedures and ensure that they understood the verbal instructions. The **Familiarization** trial began with an exposure phase. Experimenter 1 (E1) sat across from the toddler and placed three opaque boxes on the desk. E1 told the toddler "I'm going to show you what's in here. Let's see what's in here!" She then opened the first box, removed a familiar object (pig, cup, or spoon) and commented on it twice ("Wow, I like this one! What a cool thing!") before replacing it in the box. This was repeated for each box, moving from right to left. The objects were never labeled, only referred to generically as "this one" or "thing." Once all three familiar objects had been shown, E1 removed the three objects from their boxes, placed them on the toddler's desk, and allowed the toddler to play with them for approximately 30 seconds.

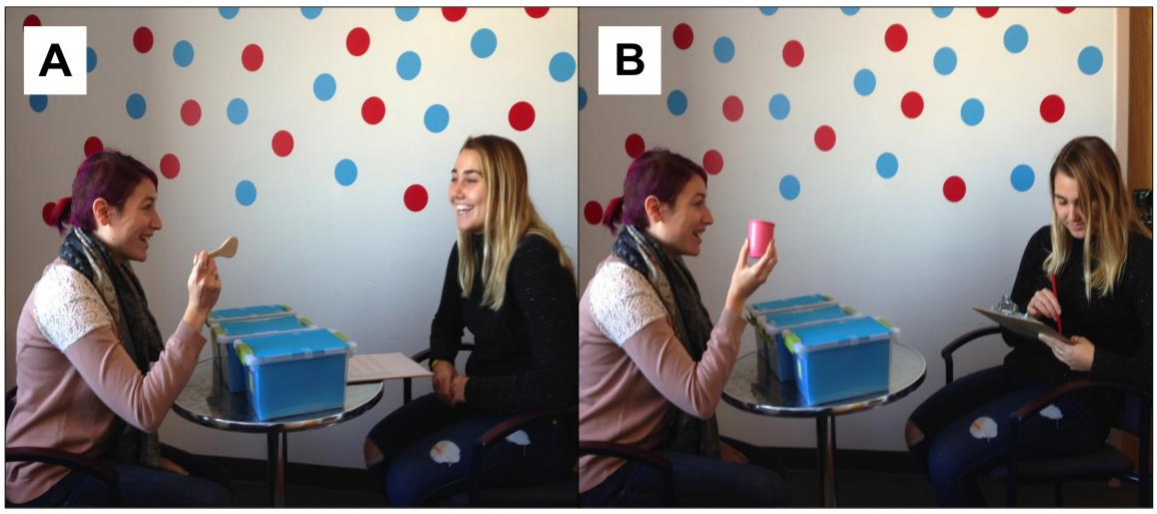
Next, **Familiarization** concluded with a test phase in which the three objects were placed on a tray, and toddlers were prompted for labels and preference. In labeling prompts, toddlers were instructed to: "Get the pig and put it in the penguin." If a toddler was hesitant to place toys in the penguin, the prompt was modified to "…and give it to me," or "…and give it to [caregiver]." Once the toddler responded (pointed to a toy, held it up, put it in the penguin, or gave it to the experimenter or caregiver), the toy was placed back on the tray and the same prompt was repeated for the spoon. We took two correct responses to mean that participants understood the task and instructions. Thus, if the participant responded correctly for both the pig and spoon, E1 moved on to the preference prompt. If the participant provided an incorrect

response or did not respond to one of those two, the cup was prompted. In preference prompts, toddlers were instructed to "Get the one you like and put it in the penguin." Toddlers were prompted up to three times per object. Toddlers' responses were recorded on-line by a second experimenter (E2), who was sitting at a table approximately six feet away, and confirmed off-line by review of video recordings of the session (confirmation was not possible for one participant because of a technological failure).

After familiarization, toddlers participated in two **Overhearing** trials (Figure 1). In these trials, toddlers were told that it was E2's turn to play the game, that it would be their turn next, and that they should stay at their desk while E1 played the game with E2, who was seated at a table perpendicular to the toddler's desk (thus providing a full view to the toddler). Participants were not explicitly told to attend to the interaction, and they were not redirected by the experimenters if they did not attend. Like the Familiarization trial, each Overhearing trial contained two phases: exposure and test. In the exposure phase, E1 presented E2 with the three opaque boxes (opened from right to left), this time containing the novel objects from one of the two object sets. One box contained the target object—the box that contained the target was randomized. If the box contained a non-target object, the procedure was similar to Familiarization. She told E2: "I'm going to show you what's in here. Let's see what's in here!" She then opened the box to reveal a novel object and commented on it twice "Wow, I like this one! What a cool thing!" If the box contained the target-object, the generic language was replaced with a novel label (either *bosa* or *manu*, counterbalanced across conditions). She told E2: "I'm going to show you the [*bosa*]. Let's see the [*bosa*]!" She then opened the box to reveal a novel object and labeled it twice "Wow, I like this [*bosa*]! What a cool [*bosa*]!" Thus, the target object was labeled four times per exposure phase. The exposure phase ended with E1

bringing the three objects over to the toddler, placing them on the desk, and encouraging the toddler to play with the three objects (for approximately 30 seconds). If a toddler had not touched all three objects during this time, E1 held up each object, asked "Did you see all of these?" and gave the toddler one more opportunity to play with all of the objects. This exposure phase was performed twice, such that each object was displayed twice and the target object was labeled a total of eight times. The target object (of two possible objects; one of the objects in the set was never the target) and which box the target was located in was randomly selected using a random number generator.

Figure 1. Schematic of toddler's perspective during the exposure phase in the (A) Attended and (B) Distracted conditions.



The test phase in Overhearing trials was similar to Familiarization. In line with prior work (e.g. Akhtar, 2005), the three objects were placed on a tray and the toddler was given one labeling prompt (for the target object) followed by one preference prompt. In the labeling prompt,

the toddler was instructed to indicate the target item (*bosa* or *manu*) by putting it in the penguin. In the preference prompt, they were again instructed to indicate the item they liked. This prompt served as a control to ensure that toddlers who selected the target item on the labeling prompt were doing so because they had indeed mapped the label to the target item, and were not just drawn to it more. The test phase was only performed once, such that each Overhearing trial contained two exposure phases followed by one test phase.

The key manipulation of the study was the difference between the two Overhearing trials. In the Attended condition, E1 and E2 maintained eye contact during exposure and E2 had full view of the objects. In contrast, in the Distracted condition, E2 looked down at a clipboard throughout the exposure phase, pretending to take notes and occasionally nodding along with what E1 was saying, thus mimicking common situations in which people are being spoken to while distracted. During this time, E2 did not make eye contact with E1, and could not see the objects being presented by E1. The toddler always had full view of the objects and boxes. We predicted that if toddlers utilized the partner's view in the interaction, they should succeed in the Attended condition and fail in the Distracted condition. If they did not utilize the partner's view, they should succeed in both conditions.

All participants participated in one Overhearing trial in the Attended condition and one Overhearing trial in the Distracted condition. Conditions were counterbalanced for order: half of the participants saw the Attended condition first, the other half saw the Distracted condition first. The set of objects used for each condition was also counterbalanced.

Results

Toddlers who failed to select the correct object on at least two of the labeling prompts during Familiarization were excluded from further analysis (three toddlers, see Participants).

Toddlers' responses during Overhearing test trials were coded as target or non-target and then compared to chance (33% because there were three objects to choose from; see Table 1). If a toddler chose multiple objects, this was coded as a non-target selection, even if one of their selections was the target. Although not responding is technically a fourth choice (and thus changes the level of chance), we conducted these analyses with non-responses (n = 6) coded as non-target choices to be conservative. Excluding those responses yields the same pattern of results, albeit with larger effects.

Our primary planned analyses (comparing response patterns to chance) were one-tailed Bayesian binomial tests for two reasons. First, we chose one-tailed binomial tests to emulate the analyses conducted in previous overhearing studies (e.g., Akhtar, 2005; O'Doherty et al., 2011). Second, we developed separate directional hypotheses for two distinct research questions (Can toddlers learn in the attended condition? Can they learn in the distracted condition?). Additionally, we used Bayesian analyses because we were interested in evidence in favor of both the null (children do not learn the target words) and alternative hypotheses (children learn the target words). However, frequentist *p*-values are also listed in Table 1.

To answer our first question (if toddlers mapped the novel label to the target novel object in the Attended condition), we compared the number of target responses in the Attended condition to chance (.33). Of the 20 participants, 12 (60%) selected the target object, which suggests the alternative hypothesis is eight times more likely than the null hypothesis (BF_{10}= 8.28), and is similar to the success rate in prior studies (60% in the current study, 62.5% in both Akhtar, 2005 and O'Doherty et al., 2011). On the preference prompt, only 8 toddlers selected the target object as the one they liked, which is considered small to moderate evidence (van Doorn et

al., 2019) for the null hypothesis (BF_{10}= .38). Thus, toddlers were not drawn more to the target object (because it was labeled or possibly more salient) than the non-target objects.

Table 1. Target responses by condition and question type.

Condition	Question	Proportion of Target Responses	*p*	95% CI
Attended				
	Labeling	0.60	0.012	(0.39, 1.00)
	Preference	0.40	0.327	(0.22, 1.00)
Distracted				
	Labeling	0.35	0.51	(0.18, 1.00)
	Preference	0.15	0.98	(0.04, 1.00)

To answer our second question (if toddlers mapped the novel label to the target novel object in the Distracted condition), we repeated the same analyses as in the Attended condition. Here, only 7 of the 20 participants selected the target object, which suggests the null hypothesis is more than four times more likely than the alternative (BF_{10}= .24). Toddlers also did not prefer the target on preference prompts, selecting it only 3 times (BF_{10}= .08, strong evidence for the null hypothesis).

One possibility is that toddlers were less attentive to the interaction in the Distracted condition and that this explains the difference in target responding. To determine if this was the case, we coded the video recordings for participant behavior during the exposure phases of the Overhearing trials (as in Luyster & Arunachalam, 2018). In the Attended condition, 8 toddlers got up from their seat during one or both exposure phases. Of those, only 1 toddler was not looking at the experimenters during at least one presentation of the target object. In the Distracted condition, 7 toddlers got up from their seat during one or both exposure phases. Similarly, only 1 toddler (not the same one as in the Attended condition) was not looking at the

experimenters during at least one presentation of the target object. Thus, there is no evidence that toddlers in the Distracted condition were less attentive than those in the Attended condition.

Finally, we looked at age and vocabulary size (MCDI receptive language score) as potential predictors of performance. In the Attended condition, a binomial logistic regression on target selection using age and receptive vocabulary as predictors was significant [$X^2(2) = 7.39$, $p = .025$, Nagelkerke $R^2 = .417$]. Both age [Wald $X^2(1) = 4.17$, $p = .042$] and receptive vocabulary [Wald $X^2(1) = 4.04$, $p = .045$] were unique predictors of selecting the target object. However, the model was nonsignificant in the Distracted condition [$X^2(2) = .28$, $p = .869$, Nagelkerke $R^2 = .019$]. Neither age [Wald $X^2(1) = .08$, $p = .777$] nor receptive vocabulary [Wald $X^2(1) = .013$, $p = .910$] were significant predictors of target selection.

Discussion

Our goal was to determine the role of joint attention in word learning through overheard interactions. Whereas previous studies of joint attention have typically included the child as one of the interlocutors, we examined whether toddlers are sensitive to the shared attention of others. We demonstrated that toddlers only learned novel words when the learner they were observing was attending to the scene, providing new evidence that toddlers who overhear a teaching interaction still use joint attention cues to support learning. This was despite the fact that the toddlers themselves had equal visual access to the scene across conditions, suggesting that mere visual access was insufficient. Further, age and vocabulary size predicted success on Attended trials only; thus, learning by overhearing attended interactions may become a more useful strategy as toddlers age and develop better language skills.

While joint attention is clearly an important factor for word learning in directed interactions, future work will need to disentangle why it is important for learning through

overhearing. One possibility is perspective-taking. This hypothesis, proposed by Moore (2007), posits that toddlers learn through third-party interactions by taking the perspective of the conversational partner. Here, the toddler observed that the partner did not look at the target object and thus lacked the information to correctly map the label to that object. Thus, taking that partner's perspective was uninformative. A related explanation is gaze-following. Participants who followed E2's gaze would have attended to the target object in the Attended condition and the clipboard in the Distracted condition. An alternative that is compatible with perspective-taking and gaze-following concerns the role of social contingency. The lack of joint attention between conversational partners creates a social interaction that is not reciprocal, a feature that is key to early word learning even in directed interactions (e.g., O'Doherty et al., 2011; Strouse et al., 2017). A third explanation is that the conversational partner's attentional state serves as a cue about the status of E1's information. It is possible that toddlers did not think the information was relevant to learn in the Distracted condition because E2 was not attending. However, E1 communicated ostensively to E2, and toddlers attended equally to the interaction as a whole in the Distracted condition and the Attended condition. Thus, we do not think this is a fitting interpretation for our data.

Another opportunity for future research concerns the role of age on third-party word learning. While our data suggest that age improved toddlers' word learning in the Attended condition only, children beyond our age range may have stronger mechanisms for third-party word learning. Foushee & Xu (2016) demonstrated that at 4.5 years of age, children successfully learned words through an overheard phone conversation (which necessarily lacks joint attention between interlocutors), and their ability to do so improved from 4.5 to 6 years. It is therefore likely that children at this age would successfully learn in our Distracted condition, although

conversational dynamics differ between in-person and telephone contexts (Emberson, Lupyan, Goldstein & Spivey, 2010), and preschoolers remain sensitive to other aspects of addressee behavior in in-person interactions (Tolins, Namiranian, Akhtar, & Fox Tree, 2017).

A final open question is whether the joint attention context is important solely for language learning, or if it generalizes to a broader learning mechanism. Toddlers are also able to learn about object locations and functions through third-party interactions (e.g. Grafenhain et al., 2009; Herold & Akhtar, 2008). Some studies demonstrate that addressee behavior affects success on these tasks (Grafenhain et al., 2009; Matheson, Moore, & Akhtar, 2013), suggesting that social cues are important for general learning. However, in the conditions that lack social cues, those tasks show success by toddlers at similar ages as the toddlers in our study, who failed in the Distracted condition (Matheson et al., 2013; Phillips, Seston & Kelemen, 2012).

Our findings add to a growing body of literature suggesting that not all overheard interactions are equally supportive of word learning. Specifically, a critical feature for learning from an observed interaction is joint attention between conversational partners—a feature that may be present for some, but not all overheard conversations in the home (e.g., those that occur on the telephone). Studies on both live and prerecorded interactions suggest that the addressee's behavior—such as whether they hold the object (O'Doherty et al., 2011) or agree or disagree with the speaker (Tolins et al., 2017)—affects learning. These are ecologically valid cues that determine the quality of overheard interactions (but not the quantity of overheard speech). It is important for future research to further explore if and how the quality of overheard interactions affects language outcomes.

References

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

Emberson, L. L., Lupyan, G., Goldstein, M. H., & Spivey, M. J. (2010). Overheard cell-phone conversations: When less speech is more distracting. *Psychological Science*, *21*(10), 1383-1388.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., & Dale, P. S. (2000). Short-form version of the MacArthur Communicative Development Inventories Short-form versions of the MacArthur Communicative Development Inventories, *21*, 95–116.

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

Foushee, R., & Xu, F. (2016) Active overhearing: Development in preschoolers' skill at "listening in" to naturalistic overheard speech. *Proceedings of the Cognitive Science Society 2016,* 2177-2182.

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

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). One-year-olds' understanding of nonverbal gestures directed to a third person. *Cognitive Development*, *24*(1), 23–33. http://doi.org/10.1016/j.cogdev.2008.10.001

Herold, K. H., & Akhtar, N. (2008). Imitative learning from a third-party interaction: Relations with self-recognition and perspective-taking. *Journal of Experimental Child Psychology*, *101*(2), 114–123. http://doi.org/10.1016/j.jecp.2008.05.004.Imitative

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–257. http://doi.org/10.1080/15250000701795598

Luyster, R. J., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology*, *45*(6), 1774–1786. http://doi.org/10.1037/a0016223

Luyster, R. J., & Arunachalam, S. (2018). Brief Report: Learning Language Through Overhearing in Children with ASD. *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-018-3672-0

Matheson, H., Moore, C., & Akhtar, N. (2013). The development of social learning in interactive and observational contexts. *Journal of Experimental Child Psychology*, *114*(2), 161-172.

Moore, C. (2007). Understanding self and others in the second year. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations.* (pp. 43–65). New York, NY: Guilford Press.

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. http://doi.org/10.1111/j.1467-8624.2011.01579.x

Ochs, E., & Schieffelin, B. B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In *Culture theory: Essays on mind, self, and emotion* (pp. 276–320). Cambridge: Cambridge University Press. http://doi.org/10.1073/pnas.0807041106

Phillips, B., Seston, R., & Kelemen, D. (2012). Learning about tool categories via eavesdropping. *Child Development*, *83*(6), 2057-2072.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*, 185–205. http://doi.org/10.1017/S0305000907008343

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*, 659–673.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2018). Reexamining the Verbal Environments of Children From Different Socioeconomic Backgrounds. *Child Development*. http://doi.org/10.1111/cdev.13072

Strouse, G. A., Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology*, *166*, 310–326.

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. http://doi.org/10.1177/0142723717727407

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463.

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., … Wagenmakers, E. (2019). The JASP Guidelines for Conducting and Reporting a Bayesian Analysis. https://doi.org/10.31234/osf.io/yqxfr

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*, 2143–2152.

Acknowledgements: We are grateful to Kendra Hennegen, the research staff at [REDACTED], and the children and families who participated.

Funding: This work was supported by the National Institutes of Health [GRANT NUMBERS REDACTED]; and the National Science Foundation [REDACTED].