The Effects of Linguistic Context on Visual Attention while Learning Novel Verbs

Matthew J. Valleau and Sudha Arunachalam

1. Introduction

One mechanism understood to be important for the acquisition of verb meanings is *syntactic bootstrapping*, or the use of the linguistic context in which an unfamiliar verb appears as a source of information about its meaning (e.g., Gleitman, 1990; Landau & Gleitman, 1985). Gleitman (1990) provided a useful metaphor for understanding the role of linguistic context: it serves as a "mental zoom lens for fixing on just the interpretation… that the speaker is expressing" (23). This zoom lens metaphor has helped to shape verb learning research, leading to numerous studies asking how linguistic context can help toddlers determine the relevant construal of a visual scene in verb learning tasks.

In the current study, we take this metaphor literally; we ask whether the zoom lens effect manifests in visual attention such that toddlers attend to the scene in front of them differently depending on the linguistic context used to describe it. We present them with novel verbs and manipulate how informative the verb's linguistic context is about verb meaning—that is, how useful the context is for syntactic bootstrapping—and examine toddlers' eye gaze as the visual scene depicting the verb's referent unfolds.

Eye gaze is a common measure in verb learning studies using preferential looking or visual world paradigms. In this literature, this measure is used to assess which of two test scenes toddlers attend to when tested on their knowledge of the verb's meaning. But if the verb learning zoom lens truly shapes attention, we predict that we should see effects of linguistic context on gaze patterns that precede the effects that have been demonstrated at test. Thus, we pursue this question by examining toddlers' visual attention as they view a dynamic visual scene that depicts the referent of a novel verb for the very first time, rather than during a subsequent test phase; we tracked whether toddlers looked to the moving object that was being acted upon during the target event.

The empirical foundation for this experiment is a series of studies documenting that English-acquiring children's verb acquisition is supported by

rich semantic information in the verb's linguistic context (Arunachalam, Syrett, & Chen, 2016; Arunachalam & Waxman, 2010, 2011, 2015; Imai et al., 2008; Syrett, Arunachalam, & Waxman, 2014). In Arunachalam and Waxman (2011, 2015), for example, 24- and 27-month-old toddlers successfully acquired novel transitive verbs in contexts that included full determiner phrases labeling the participants in the event (e.g., "The boy is pilking the balloon"), but failed in contexts in which these were replaced with pronouns (e.g., "He's pilking it"). That is, they performed better when provided with rich semantic information about the event participants than when provided with less informative pronouns. Similarly, Imai et al. (2008) found that with older children, ages 3 and 5 years, more semantically informative contexts supported English learners' verb acquisition better than less informative contexts, although the amount of information required by these older children was less: they performed well when the verb was flanked by pronouns but struggled when both arguments were dropped (e.g., "Look, pilking!").

In these tasks, toddlers were first presented with the novel verb accompanying a scene in which an actor acted on an object, e.g., a boy waved a balloon on a stick. At test, toddlers viewed two new scenes simultaneously; in one, the target scene, the actor performed the familiarized action but on a new object (e.g., waved a rake), while in the distractor scene the actor used the familiarized object, but with a new action (e.g., tapped the balloon). The task thus required them to extend the novel verb to a scene involving a new participant object. Arunachalam and Waxman hypothesized that the full determiner phrases allowed learners to identify the relevant event participants, and to focus their attention on the relation between them. Correct identification of the relation was crucial to allowing toddlers to abstract the novel verb to apply to a new object. For example, if toddlers had incorrectly inferred that the crucial event component to which "pilking" referred was the man's holding of a stick-like object, then both test scenes would have applied.

Although the presence of descriptive determiner phrases labeling a verb's arguments aids acquisition of the verb's meaning, to our knowledge no studies have examined whether this benefit is indeed due to toddlers' identification of the referents of those determiner phrases, and in turn, to the relation between them. Other hypotheses are certainly possible. It could be that full determiner phrases capture learners' interest and result in more overall learning, but not specifically in more attention to the referents of those determiner phrases. It could also be the case that toddlers encode the event identically in both the full determiner phrase and pronoun conditions, but that the pronominal contexts hinder them from extending the verb to apply to a new object.

This prior work thus presents an ideal opportunity to test a literal interpretation of the zoom lens metaphor. Do toddlers hearing novel verbs in utterances with full determiner phrases attend to the visual scene differently than those hearing utterances with pronouns? Or is visual attention (and thus, possibly, encoding) the same across conditions, with the previously observed differences only manifesting at test? To address this question, we analyzed

toddlers' eye gaze as they viewed the scene and attempted to map the novel verb to meaning. Unlike previous work, we assess eye gaze as toddlers view a potential referent for the very first time, as they are in the process of looking out into the world to discover what it might refer to, and mapping it to meaning.

Our goal was to see whether the utterance of an unfamiliar verb in a linguistic context that is informative as to its meaning guides subsequent event perception. If so, it would suggest that toddlers can encode linguistic information about an unfamiliar verb and then scan the visual world in such a way that helps them to identify that verb's referent. This also permitted us to ask, when toddlers fail to extend a novel verb to apply to a new object at test, as they have in some conditions of Arunachalam and Waxman (2010, 2011), and Imai et al. (2008), is it because they have not assigned the verb the correct meaning at all, or is it because they failed to extend it? If toddlers have correctly encoded the verb's meaning but struggle to extend it to apply to new objects, then we predict that gaze patterns on trials on which they point correctly vs. incorrectly should be the same, as their problem is not at encoding. On the other hand, if failure in studies like these can be traced to difficulty with the identification of the verb's meaning in the first place, then toddlers should show different gaze patterns on trials on which they point ultimately correctly as compared to trials on which they ultimately point incorrectly.

2. Method
2.1. Participants

Forty toddlers (20 males, 20 females; mean age: 33 months, age range: 31 to 36 months) participated and were randomly assigned to one of two between-subject conditions: novel verbs flanked by (1) determiner phrases with an indefinite determiner and noun (Full DP condition) or (2) pronouns (Pronoun condition). We recruited toddlers from the greater Boston, MA area who were reported to be developing typically by their parent and who were acquiring English as a native language with exposure to other languages less than 30% of the time. Twenty-one additional toddlers were excluded from the sample, due to fussiness (2), more than 50% track loss during the session (8), difficulty calibrating eye gaze before beginning the study (1), failure to point at all or to point correctly during training (6), or failure to point on at least two trials at test (4). Although most of the toddlers for whom there was excessive track loss contributed codable pointing responses, and most of those who did not point contributed useable eye-tracking data, we excluded both of these groups so that we could relate eye gaze during the event to performance on the established pointing measure at test.

We focused on toddlers in the latter half of the 3rd year of life because it is roughly the midpoint between the ages investigated in prior work. Arunachalam and Waxman (2010, 2011) studied younger ages, and Imai et al. (2008) studied older ages. We expected our task to be substantially more difficult than Arunachalam and Waxman's because our visual scenes were more similar to

each other due to the need to achieve spatial separation between the actor and object (see below) in all scenes.

Parents completed the MacArthur-Bates Communicative Development Inventory (MCDI) Level III (Fenson et al., 2000). Mean vocabulary production was 69 words in the Full DP condition (range: 39 to 92 words), and 68 words in the Pronoun condition (range: 26 to 95 words).

2.2. Materials

See Table 1. The visual stimuli consisted of videos in which actors performed simple actions on objects (e.g., waving an umbrella). To ensure that we could create non-overlapping Areas of Interest for the agent and the object, we took care to include spatial separation between the actor and a significant portion of the object. We refer to these as the Agent AOI and Object AOI respectively. For example, for the umbrella-waving video, the actor held just the tip of the umbrella handle, thus allowing most of the object, as well as the most interesting and recognizable portion of the object (the canopy) to be spatially separated from the agent. For the scenes we showed at test, on each trial one depicted the same action as seen previously (e.g., waving) but with a new object (e.g., a balloon), while the other depicted a new action but with the familiar object (e.g., tapping the floor with the same umbrella). These scenes also included spatial separation between the agent and object, and were thus visually relatively similar to each other. All stimuli appeared on a white background. The scenes were 6.5" wide and 10.75" tall; during the test phase when two scenes appeared side-by-side, they had 5.75" between them.

The auditory stimuli consisted of sentences produced by a female native speaker of American English in a sound-attenuated recording booth. The linguistic frames used in each condition were like those used in prior studies. Utterances were edited in Praat (Boersma & Weenink, 2011).

Visual and auditory stimuli were synchronized and presented on an eye-tracker monitor, Tobii T60 XL, which records coordinates of gaze at a rate of 60 frames/sec. Sound was presented through the monitor's speakers on either side of the visual display.

2.3. Procedure

Toddlers played with toys in our reception area while their caregiver completed a consent form and the MCDI. They then entered the testing room, where the toddler sat in a car seat approximately 65 in. from the eye-tracker. The caregiver sat behind and to the side of the toddler and was asked not to talk or interact during the session. One experimenter controlled stimulus presentation from behind a curtain, and another sat beside the toddler to elicit pointing responses. Both experimenters, who were naïve to the study hypotheses, recorded toddlers' direction of pointing.

Table 1. Stimuli from one trial (of six). Images are single frames taken from dynamic scenes, except for the Linguistic Familiarization phase, during which toddlers saw a still frame.

Trial Phase	Video	Audio
Linguistic Familiarization (6 sec)	[still frame]	**Full DP condition:** A girl is gonna *pilk* an umbrella. Let's see. **Pronoun condition:** Let's see what happens here. She's gonna *pilk* it.
Event Familiarization Window 1 (6 sec)		Look, *pilking*!
Event Familiarization Window 2 (6 sec)		Wow!
Test Baseline window (6 sec)	Familiar Object Familiar Action (target)	Now look—they're different!
Test Query window (2 sec)	★	Do you see *pilking*?
Test Response window (12 sec)		Where do you see *pilking*? (experimenter elicits pointing response)

Before the test trials, toddlers participated in a pointing warm-up with no novel words. They were asked to point to familiar characters and actions. Next,

each toddler participated in six novel verb trials, each with three phases: Linguistic Familiarization, Event Familiarization, and Test. See Table 1. Both the Full DP and Pronoun conditions presented the same visual stimuli; the only difference was the utterance they heard during Linguistic Familiarization.

During Linguistic Familiarization, toddlers saw a 6-sec still frame of a live action scene (e.g., a girl holding an umbrella), centered on the screen, and heard either, "A girl is gonna pilk an umbrella. Let's see" (Full DP condition), or "Let's see what happens here. She's gonna pilk it" (Pronoun condition). These linguistic stimuli were nearly identical to Arunachalam and Waxman (2015); recall that in that study toddlers successfully acquired novel verbs in the Full DP condition but failed in the Pronoun condition.

Event Familiarization is the first point in the trial when toddlers see the dynamic scene, and it immediately follows their hearing the utterance. It is therefore the phase during which we were interested in toddlers' gaze patterns. This phase comprised two windows; during the first (6 sec), toddlers saw a dynamic scene in which the actor performed an action with the object (e.g., waving an umbrella). Toddlers heard, "Look, pilking!" After a brief blank screen lasting 0.25 sec, the scene repeated during the second window (6 sec), with the audio, "Wow!"

The Test phase was divided into three windows. First, during Baseline, two new dynamic scenes displayed simultaneously, side-by-side. In the Familiar Action scene, the actor performed the now-familiar action with a new object (e.g., a balloon on a stick); in the Familiar Object scene, she performed a new action on the familiar object (e.g., tapping the umbrella to the floor). Toddlers heard, "Now look—they're different!" as these scenes played. Immediately afterward the Query window began, during which the scenes disappeared, replaced by a centrally placed yellow star, designed to draw toddlers' attention to the center of the screen and allow them to attend to the test query. The test query played: "Do you see pilking?" Finally, in the Response window, the two dynamic scenes reappeared in their original locations and the test query repeated, "Where do you see pilking?"

2.4. Predictions

We were primarily interested in learners' looking patterns during the Event Familiarization phase. During this phase, toddlers in both conditions heard identical linguistic stimuli; what differed was what they had heard previously during Linguistic Familiarization. We predicted that toddlers in the Full DP condition would spend more time looking at the person and object in the scene, as compared to toddlers in the Pronoun condition. We chose one of these, looking to the Object AOI, as our dependent measure, because the objects were in motion more than the people. Because the objects were moving during the Event Familiarization phase, we used dynamic AOIs to track them, noting their locations for each frame of the video and assessing whether toddlers' gaze fell within that location. We also included whether the toddler pointed correctly or

incorrectly at test in our analyses, predicting greater looking to the Object AOI on trials on which toddlers correctly mapped the verbs and extended them to apply to a new object.

Note that we did not examine looking behavior during Linguistic Familiarization because this is when toddlers heard the linguistic stimulus; their gaze during this time could simply be a result of the moment-by-moment unfolding language rather than their encoding of the novel verb (e.g., looking at the umbrella when hearing the word "umbrella"). Instead, we examined eye gaze when toddlers were *not* hearing an informative linguistic stimulus, but were simply being prompted to inspect the scene in search of a referent for the novel verb (e.g., "Look, *pilking*!"). We also did not analyze their gaze during the Test phase, because we elicited pointing responses to provide a link to prior work, and we do not yet know whether toddlers' gaze is affected by being asked to point. Anecdotally, we have observed in our lab that when toddlers are asked to point they shift their gaze away from the scene to which they are pointing; they also may block the eye-tracker with their arm while pointing.

3. Results

We first excluded trials on which toddlers did not point (11%). For the remaining trials, track loss and frames on which toddlers were looking outside the area of interest were included as data points in addition to looks to the scene. Track loss was comparable across conditions: 14% in the Full DP condition and 18% in the Pronoun condition, and comparable across trials on which toddlers pointed correctly vs. incorrectly: 15% on trials on which they pointed correctly and 18% on trials on which they pointed incorrectly. This suggests that incorrect pointing was not simply due to inattention.

Overall, pointing behavior did not significantly differ across conditions; toddlers in the Full DP condition correctly chose the target on 55% of trials, and those in the Pronoun condition did so on 50% of trials. Chance performance is 50%, as there were only two scenes from which to choose. The number of trials included in each cell was thus comparable across conditions. Though the trend was in the same direction as in Arunachalam and Waxman (2011, 2015), performance in the Full DP condition was poorer than in these studies. However, we expected this would be the case; because we had to create similar spatial separation between the agent and patient entities, the visual scenes were much more similar to each other than in prior work. Further, the effect sizes in the prior studies were already relatively small. Our question of interest here is not, of course, how often toddlers successfully choose the target, but rather how their visual attention differs when they do and when they do not, and, importantly, whether this is related to the linguistic information they hear.

Figure 1 illustrates toddlers' preferences for the Object AOI (e.g., umbrella) over the course of the Event Familiarization Phase, split by condition. Evident in the figure is, first, an overall similarity in how toddlers looked at the visual scene in both conditions. This is unsurprising; after all, toddlers are viewing the

same dynamic scene and dynamic visual features are likely to capture attention similarly. Recall that the linguistic material presented during the Event Familiarization Phase is constant across conditions and in itself provides no information that should affect gaze. However, in Figure 1, a small difference between conditions emerges throughout Window 2. It appears that after toddlers have fully absorbed the visual scene (i.e., in Window 1), they inspect it differently depending on the linguistic context they had previously heard.
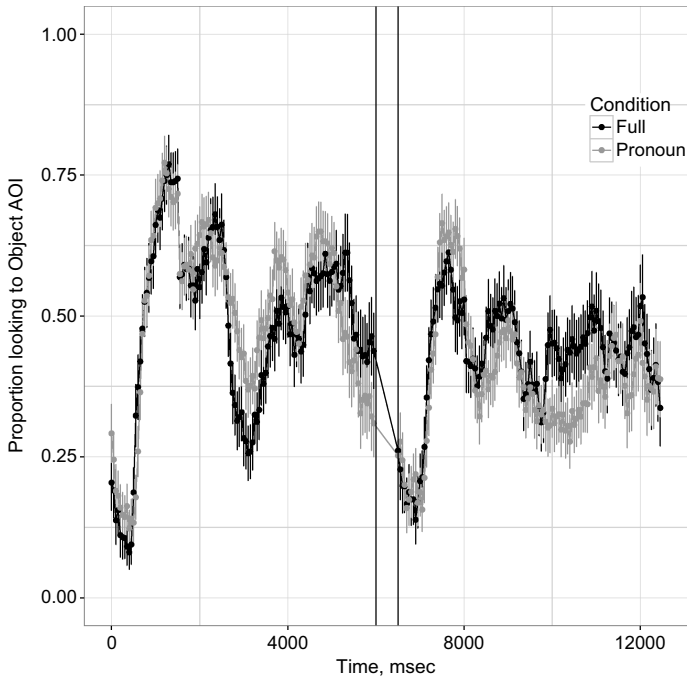

Figure 1. Toddlers' gaze to the Object AOI during the Event Familiarization Phase, by condition. The vertical lines indicate the brief interval between Windows 1 and 2.

Because we also expected that whether toddlers ultimately pointed correctly at test would be reflected in their earlier gaze patterns, we plotted separately in Figure 2 only trials on which toddlers ultimately pointed correctly. Indeed, the small difference between linguistic conditions is more apparent than in Figure 2. In Figure 3 we depict gaze to the Object aggregated over time within each window, now separated by Window, Condition, and Pointing Response. Notice that mean preference for the Object AOI is similar across all four groups within each Window, with the exception of trials in the Full DP condition on which toddlers pointed correctly. That is, trials on which toddlers both heard the novel verb flanked by content nouns and ultimately pointed correctly at test show a zoom lens effect of more gaze to the Object AOI in Window 2.
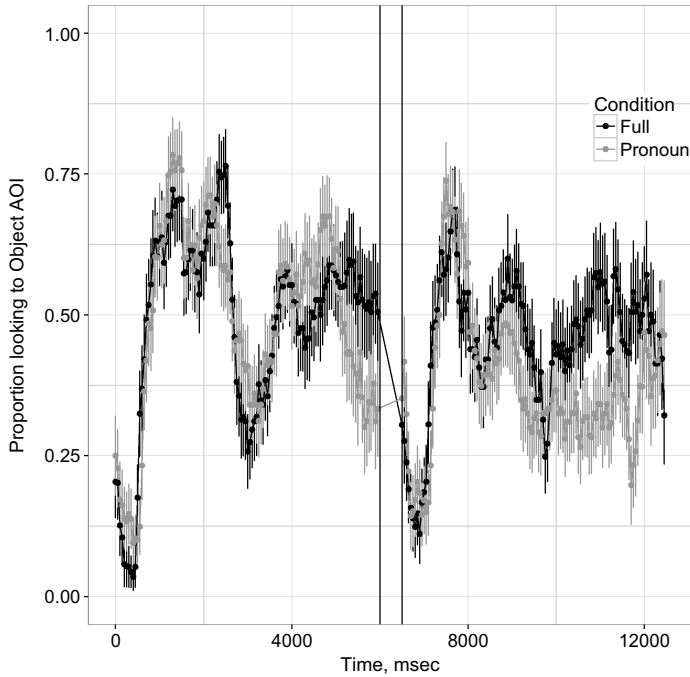
Figure 2. Toddlers' gaze to the Object AOI during the Event Familiarization Phase for trials on which toddlers ultimately pointed correctly at test, by condition.
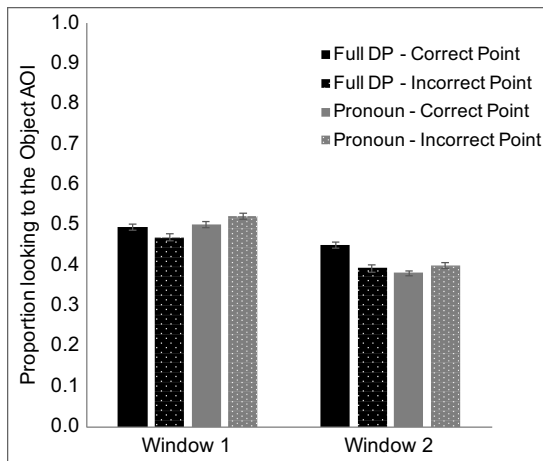

Figure 3. Toddlers' gaze to the Object AOI during the Event Familiarization Phase, by window, condition, and pointing response.

To investigate these patterns statistically, we aggregated the gaze data into 50 ms bins (of 3 frames each), transformed the binned data using an empirical-logit function, and fit the data to a multilevel logistic regression model using maximum likelihood estimation (Barr, 2008). As random effects, we included participant and trial, and as fixed effects, Time (in seconds), Condition (Full DP vs. Pronoun), Pointing Response (correct vs. incorrect), and Event Familiarization Window (1st vs. 2nd). Our critical question is whether Condition plays a role in gaze patterns, but we also expected to see interactions with Pointing Response. A main effect or interactions with Window would indicate that it mattered whether it was the toddler's first or second time viewing the event. We used model comparison to determine the significance of individual factors. We found main effects of Condition, Window, Time, and Pointing Response, as well as the expected interaction between Condition and Pointing Response (all ps < .05). We also found interactions of Window and Time, Window and Pointing Response, Time and Pointing Response, and 3-way interactions between Window, Condition, and Pointing Response; and Window, Time, and Pointing Response (ps < .05). See Table 2 for the parameters for those effects that contributed significantly.

Table 2. Parameter estimates for mixed-effects logistic regression model.

Parameter	Estimate	Standard Error	t-value
Intercept	1.90	0.061	31.51
Condition (C)	0.14	0.068	2.01*
Window (W)	0.36	0.053	6.87*
Time (T)	-0.034	0.0047	-7.22*
Pointing Response (P)	-0.18	0.037	-4.86*
W * T	-0.040	0.0072	-5.54*
C * P	-0.15	0.074	-2.03*
W * P	-0.33	0.071	-4.73*
T * P	0.026	0.0050	5.25*
W * T * P	0.027	0.0096	2.83*
C * W * P	-0.29	0.14	-2.083*

The expected main effect of condition indicates that as toddlers watched the event for the first time, their gaze was indeed affected by the linguistic context they had previously heard: those in the Full DP condition preferred to look at the moving object as compared to those in the Pronoun condition. Interestingly, our finding of a main effect of Pointing Response, and of interactions between Condition and Pointing Response, suggest a nuanced relationship between these factors. Inspection of Figure 3 suggests that the effects were driven by toddlers in the Full DP condition on trials on which they pointed correctly. However, in both conditions, pointing performance was close to chance, and although some points to the correct scene may have been reflective of learning, others may have simply been random guesses. The relationship between gaze and pointing performance, then, remains a topic for future work, at least in the case of pronoun contexts.

We next wondered whether the preference for the object AOI in the Full DP condition as compared to the Pronoun condition could be traced to specific features of toddlers' gaze, such as more entrances into the object AOI region—that is, that toddlers keep looking back to the object AOI—or due to longer fixations on the object, as we would expect if they were tracking its movement. We thus tallied both the number of distinct entrances into the object AOI as well as the mean length of looks at the object AOI. However, neither of these measures was found to be indicative. The average number of fixations to the object AOI during Familiarization was 12 in the Full DP condition (correct points, 13; incorrect points, 11.89) and 15 in the Pronoun condition (correct points, 15, incorrect points, 15.78); a mixed-effects model including subject and trial as random factors and condition, pointing response, and their interaction as fixed effects found no significant main effects or interactions. While toddlers in the Full DP condition looked longer on average to the object AOI (correct points, 588.59 ms; incorrect points, 590.55 ms) than those in the Pronoun condition (correct points, 441.30 ms; incorrect points, 355.06 ms), a mixed-effects model including subject and trial as random factors and condition, pointing response, and their interaction as fixed effects found no significant main effects or interactions. These results suggest that although toddlers in the Full DP condition, on trials on which they pointed correctly, preferred the Object AOI overall, this preference did not uniformly manifest either as more distinct entrances into the Object AOI region nor as longer fixations. It may be that other gaze measures are better indicators. Just as latency and overall preference are taken as good measures for static images (e.g., Fernald et al., 2008), we consider the identification of eye gaze measures for dynamic scenes that relate to knowledge or performance to be critical. This is something we are currently investigating (Valleau et al., in prep). Identification of specific dependent measures that instantiate these gaze differences will be fruitful for future investigations of exactly how the zoom lens manifests.

4. General Discussion

In the current study we aimed to consider literally Gleitman's (1990) linguistic "zoom lens" to investigate how toddlers' visual attention is shaped by linguistic context and in turn shapes language acquisition. Building upon an established finding that toddlers' acquisition of novel verb meanings is affected by the linguistic context in which the verb is presented, we asked whether hearing novel verbs in (semantically informative) full determiner phrase contexts vs. (less informative) pronominal contexts affected toddlers' looking patterns as they viewed the referent event for the first time and sought to map the novel verb to meaning.

Our findings indicate that linguistic context indeed matters; toddlers who had heard the novel verb flanked by full noun phrases were more likely to look at the object in the scene than those who had heard pronouns, as reflected in a main effect of linguistic condition in our analysis. Hearing a full determiner

phrase directed toddlers' attention to the object, perhaps allowing them to then focus on the action in which it was engaged. That is, the determiner phrases affected not just toddlers' ultimate representation of the verb's meaning, but also *how they inspected the scene in the first place*. Whether or not toddlers ultimately pointed correctly after this Familiarization phase also mattered; trials on which toddlers pointed correctly also resulted in more looking at the object, as reflected in a main effect of pointing response. We also observed an interaction between condition and pointing; trials in the Full DP condition on which toddlers ultimately pointed correctly were the primary driving force between the observed differences. Interestingly, the difference in attentional patterns manifested in overall looking to the object as a function both of linguistic condition and pointing performance, but it did not manifest in the number or duration of fixations on the object, suggesting that future work must investigate more closely what kind of gaze measures best reflect these differences.

Our findings expand our understanding of how language shapes visual attention in the service of language learning. It is well established that adults will look to plausible upcoming referents in a linguistic context; for example, on hearing that "The boy will eat the…" adults fixate edible objects in a visual scene even prior to hearing the verb's direct object "cake" (Altmann & Kamide, 1999). But 2-year-olds in an adaptation of this paradigm only showed anticipatory looks to the edible object at the verb if they had large productive vocabularies, suggesting that the ability of linguistic input to shape visual attention is reliant on existing language skill (Mani & Huettig, 2012). Similarly, Gampe and Daum (2014) found that 2-year-olds hearing a sentence with a familiar verb (e.g., "I'll show you eating") examined the visual scene in a way that indicated they had predicted how the action would unfold (e.g., looking specifically to food items) only for verbs that were very early acquired. How then can language shape attention when toddlers do not yet have referents for some or all of the linguistic expressions they hear? Our results indicate that linguistic context allows toddlers to search for referents that fit larger pieces of language (e.g., syntactic structures, familiar words), and that this attentional shift can support acquisition of unfamiliar words in the context.

More specifically, this study offers insight into the mechanism underlying verb learning performance in earlier studies using similar experimental paradigms. Imai et al. (2008) and Arunachalam and Waxman (2010, 2011) found that unless the linguistic context was sufficiently rich, young English learners had difficulty acquiring the meanings of novel verbs and extending them to new scenes at test. But what specific aspect of the task did they struggle with in these studies? Imai et al. (2005) demonstrated that the difficulty was not in encoding and remembering the action itself in the absence of a novel verb, but rather pertained to mapping it to the novel lexical item. Our current findings further refine this conclusion, suggesting that it is mapping the verb to meaning during familiarization that poses the challenge, rather than extending the verb to apply to new participant objects at test. This is because we found differences in

attention during familiarization, even before the test scenes appeared. Although Waxman et al. (2009) and Arunachalam et al. (2013) suggested that chance performance in Imai and colleagues' studies was due to children's being "captured" by the participant object during familiarization, leading to difficulty extending the verb to a new object at test, our study reveals that toddlers in fact performed better at test if they had spent more time looking at the participant object during familiarization; it cannot be that the encoding problem toddlers faced was one of being captured by the object at least in an attentional sense. Instead, it may be that toddlers struggled with identifying the appropriate referent of the novel verb in the first place. When they failed to identify the correct referent of the novel verb, they may have been entertaining multiple possibilities, including, for example, 'holding a stick' in addition to 'waving.' The difficulty of establishing the intended meaning of a verb from observation is of course well established (e.g., Gillette et al., 1999); our results suggest that it is indeed this problem, and not one with extending the verb to a new object, that some toddlers found challenging in previous work as well as in our task.

One alternate interpretation of our results warrants discussion. Could it be the case that toddlers did not actively scan the visual scene differently depending on the kind of linguistic information they had heard, but rather that those who happened to fixate on the balloon during the familiarization phase (perhaps driven by the fact that they had heard the word "umbrella" earlier) were more successful at retaining the action associated with the verb and identifying the verb's referent accordingly? We think this is unlikely because it was only in the second presentation of the visual scene—Window 2—that differences became evident. If low-level association of the noun "umbrella" with its referent was driving toddlers' gaze, we would predict that visual attention to the scene would differ from the beginning, just after toddlers heard the noun. Instead, it looks like they initially inspected the visual scene similarly—perhaps as any viewer would regardless of the task (Papafragou, Hulbert, & Trueswell, 2008)—and only afterward focused on that part of the scene that was most relevant for acquiring the verb's meaning. (Unfortunately, then, our findings suggest that toddlers may not be in a good position to acquire verb meanings that refer to fleeting actions on a single hearing, as their attention is not immediately drawn to the part of the visual scene labeling the novel verb; they may instead rely on accrual of information over multiple instances to acquire such verbs. Indeed two recent studies suggest that verbs labeling typically durative actions are more easily acquired than verbs labeling punctual actions (Abbot-Smith, Imai, Durrant, & Nurmsoo, 2016; Horvath, Rescorla, & Arunachalam, forthcoming).) Nonetheless, this interpretation, even if true, is not incompatible with our primary hypothesis, which is that linguistic information can shape the way learners scan a visual scene in ways that are beneficial for language acquisition.

Language can be a powerful source of information to guide visual attention (e.g., Lupyan & Spivey, 2010). But even with adults, few studies have investigated precisely how listeners coordinate processing of a visual scene with linguistic processing and integrate the two streams during language

704

comprehension (e.g., Ferreira, Foucart, & Engelhardt, 2013). Expansion of such lines of research not only in adults but also developmentally will allow us to discover how event apprehension and linguistic processing relate in real-world learning situations in which learners must coordinate their observation of the world with the language they hear. Crucially, such investigations can help us unite studies of language processing with language learning.

References

Abbot-Smith, Kirsten, Imai, Mutsumi, Durrant, Samantha, & Nurmsoo, Erika. (2016). The role of timing and prototypical causality on how preschoolers fast-map novel verb meanings. *First Language*. In press.

Altmann, Gerry T. M., & Kamide, Yuki. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264.

Arunachalam, Sudha, Leddon, Erin M., Song, Hyun-joo, Lee, Yoonha, & Waxman, Sandra R. (2013). Doing more with less: Verb learning in Korean-acquiring 24-month-olds. *Language Acquisition*, *20*(4), 292-304.

Arunachalam, Sudha, Syrett, Kristen, & Chen, Yongxiang. (2016). Lexical disambiguation in verb learning: evidence from the conjoined-subject intransitive frame in English and Mandarin Chinese. *Frontiers in Psychology*, *7*.

Arunachalam, Sudha, & Waxman, Sandra R. (2010). Specifying the role of linguistic information in verb learning. In: Katie Franich, Kate Iserman, & Lauren Keil (Eds.), *Proceedings of the 34th annual Boston University Conference on Language Development* (pp. 11-21)*.* Somerville, MA: Cascadilla Press.

Arunachalam, Sudha, & Waxman, Sandra R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development, 7,* 169-184.

Arunachalam, Sudha, & Waxman, Sandra R. (2015). Let's see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, *22*(2), 117-131.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457-474.

Boersma, Paul & Weenink, David. (2011). Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Fenson, Larry, Pethick, Steve, Renda, Connie, Cox, Jeffrey L., Dale, Philip S., & Reznick, J. Steven. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*(1), 95-115. doi:10.1017/S0142716400001053

Fernald, Anne, Zangl, Renate, Portillo, Ana Luz, & Marchman, Virginia A. (2008). Looking while listening: Using eye movements to monitor spoken language. In: Irina A. Sekerina, Eva M. Fernandez, & Harald Clahsen (Eds.), *Developmental Psycholinguistics: On-line Methods in Children's Language Processing* (pp. 113-132). Amsterdam: John Benjamins.

Ferreira, Fernanda, Foucart, Alice, & Engelhardt, Paul E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69,* 165-182.

Gampe, Anja and Daum, Moritz M. (2014), Productive Verbs Facilitate Action Prediction in Toddlers. *Infancy, 19,* 301–325. doi: 10.1111/infa.12047

Gillette, Jane, Gleitman, Henry, Gleitman, Lila, & Lederer, Anne. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176. doi:10.1016/S0010-0277(99)00036-0

Gleitman, Lila R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1)*,* 3-55. doi:10.1207/s15327817la0101_2

Horvath, Sabrina, Rescorla, Leslie, & Arunachalam, Sudha. (Forthcoming). Acquiring a verb lexicon: Semantic features of toddlers' early vocabularies. In: Kristen Syrett & Sudha Arunachalam (Eds.), *Semantics in Language Acquisition* (Trends in Language Acquisition Research). Amsterdam: John Benjamins.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development*, *76*(2), 340-355.

Imai, Mutsumi, Li, Lianjing, Haryu, Etsuko, Okada, Hiroyuki, Hirsh-Pasek, Kathy, Golinkoff, Roberta M., & Shigematsu, Jun. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79,* 979-1000.

Landau, Barbara, & Gleitman, Lila R. (1985). *Language and Experience: Evidence from the Blind Child.* Cambridge, MA: Harvard University Press.

Lupyan Gary, & Spivey, Michael J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS ONE, 5(7),* e11452

Mani, Nivedita, & Huettig, Falk. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843.

Papafragou, Anna, Hulbert, Justin, & Trueswell, John C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155-184.

Syrett, Kristen, Arunachalam, Sudha, & Waxman, Sandra R. (2014). Slowly but surely: Adverbs support verb learning in 2-year-olds. *Language Learning and Development, 10,* 263-278.

Valleau, Matthew J., Golinkoff, Roberta M., Hirsh-Pasek, Kathy, Konishi, Haruka, & Arunachalam, Sudha. (in prep). Receptive verb knowledge in toddlers: An eye-tracking study.

Waxman, Sandra R., Lidz, Jeffrey L., Braun, Irena E., & Lavin, Tracy. (2009). Twenty-four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67-95.

Proceedings of the 41st annual Boston University Conference on Language Development

edited by Maria LaMendola
and Jennifer Scott

Cascadilla Press   Somerville, MA   2017