USING EYE TRACKING TO BETTER UNDERSTAND CHILDREN'S PROCESSING OF EVENTS DURING VERB LEARNING: IS THE FOCUS ON PEOPLE (FACES) OR THEIR ACTIONS (HANDS)?

JANE B. CHILDERS*[1], EMILY WARKENTIN[1], MARISSA YOUNG[1], SNEH LALANI[1], and AKILA GOPALKRISHNAN[1]

*Corresponding Author: jchilder@trinity.edu
[1]Psychology Department, Trinity University, San Antonio, Texas USA

Verbs are central to the syntactic structure of sentences, and thus important for learning one's native language. This study examines how children are visually inspecting events as they hear, and do not hear, a new verb. Specifically, there is evidence that children may focus on the agent of the action (Maguire et al., 2008; Childers et al., 2017) or may prioritize attention to the action being performed (e.g., Behrend, 1990); to date, little evidence is available. This study uses an eye tracker to track participants' looking to the agent (i.e., face) vs. action (i.e., hands) when viewing events linked to a new verb as well as distractor events to better understand children's visual attention patterns. Two 1/2, 3 1/2- and 4 1/2-year-olds saw dynamic target and distractor events and heard new verbs while a Tobii x30 eye tracker recorded fixations to AOIs (head/face, hands). Results show that children in all age groups were able to learn and extend the new verbs to new events at test. Additionally, across age groups, when viewing target events, children increased their looking to the hands (where the action is taking place) as those trials progressed and decreased their looking to the agents' face, which is less informative for learning a new verb's meaning. In contrast, when viewing distractor events, children decreased their looking to hands over trials and maintained their attention to the face. In sum, children's visual attention to agents' faces and hands differed depending on whether the events cooccurred with the new verb. These results are important as this is the first study to show this pattern of visual attention during verb learning, and thus these results help reveal underlying attentional strategies children may use when learning verbs. Additionally, the study of children's acquisition of verbs, which vary across languages, could provide clues to how languages may have changed in response to the learnability of emerging predicate structures.

1. Introduction

Verbs are central to the syntactic structure of sentences. A controversy in this area is whether children focus on the agent of the action or the action being performed, and to what extent they focus on each one. Some studies suggest that young children learn verbs best when a single agent is seen, which could mean that they are attending too much to agents rather than the actions while learning verbs (Maguire, Hirsh-Pasek, Golinkoff & Brandone, 2008; Childers, Paik, Flores, Lai & Dolan, 2017: complex events). In other studies, children seem to be attending to actions (e.g., Scott & Fisher, 2012) or results (e.g.,

Childers, 2011; Behrend, 1990). Additionally, in everyday contexts, children often see events linked to a new verb that are interspersed with distracting events. Yet most laboratory studies of verb learning show children relevant events as they hear the new verb. Thus, the present study includes distracting interleaved events to examine whether participants focus more on agents (faces) or their actions (hands) when learning verbs, and whether this varies depending on event relevancy.

To accurately acquire a verb, learners must solve what Gleitman and Gleitman (1992) described as the "packaging problem", wherein learners must package together only the relevant aspects of a dynamic event and disregard any irrelevant information. Given the difficulty of this problem, learners often use information across events and sentences or engage in cross-situational learning (e.g., Imai & Childers, 2020). Research has shown that the comparison of events can help children learn and generalize verbs (e.g., Childers & Paik, 2009; Haryu, Imai & Okada, 2011; Waxman, Lidz, Braun & Lavin, 2009). Additionally, children can benefit from seeing similar or varied events, depending on the nature of the task and test conditions (Haryu et al., 2011; Imai, Haryu, Okada, Hirsh-Pasek, Golinkoff, Shigematsu, 2008; Childers, Parrish, Olson, Burch, Fung & McIntyre, 2016).

In light of this prior research, it seems clear that young children can glean information across a set of events as they learn verbs, but theories differ in the mechanisms they posit that underlie this cross-situational processing. In statistical learning, observers compare events by noting co-occurrences between specific words and referents (e.g., Yu & Smith, 2007). In structural alignment, observers compare events linked to a new verb over instances by aligning elements across the instances based on their common relational structure (e.g., Gentner & Namy, 2006). Our study tests whether children's looking at target and distracting events during the learning phase differs in terms of their focus on agents (faces) or actions (hands). By including both types of trials, we can ask whether there are general looking preferences (e.g., for faces or hands) or whether children attend to events differently when viewing events linked to verbs vs. distracting events. To our knowledge, no prior study verb has used eye tracking in this way.

The present research is related to a set of studies with 2 ½- to 4 ½-year-old children shown the same events as in the present study (without an eye tracker) (Howard et al., 2019). Children as young as 2 ½ years were able to extend the new verb, demonstrating an ability to distinguish between target and distractor events. However, without an eye tracker, we only have indirect evidence of children's ability to ignore distracting events.

We hypothesized that children would look more to the hands as the hands AOI (area of interest) is larger (in our stimuli and in everyday life) than the face AOI, and the hands are moving. However, two prior studies demonstrate that children can also be overly attentive to the agent in an event, performing fewer verb extensions in events with multiple agents (Maguire et al., 2008), particularly when events were more complex (Childers et al., 2016). Thus, we asked whether children focused more on the face/head or the hands while seeing events in the learning phase, whether this differed depending on whether the event was linked to a new verb (target) or was not (distractor), and whether children could extend the new verbs at test.

1.2. Method

1. Participants

Twenty-four 2 1⁄2 -year-olds (*M*age= 2;8; range: 2;0-2;11), thirty-one 3 1⁄2 -year-olds (*M*age= 3;4; range: 3;0-3;11) and twenty-one 4 ½-year-olds (*M*age= 4;4; range: 4;0-4;10) participated in this study; 40 girls, 36 boys. Children were excluded if exposure to English was <80%, if teachers reported a speech delay, for equipment failure (17) or child refusal (1).

2. Materials and Design

Video stimuli were created with 3 target events showing a single causative action and 2 distinct distractor actions for each of two novel verbs. Test scenes showed a new target scene and a new distractor action. A three second black screen appeared between each event.

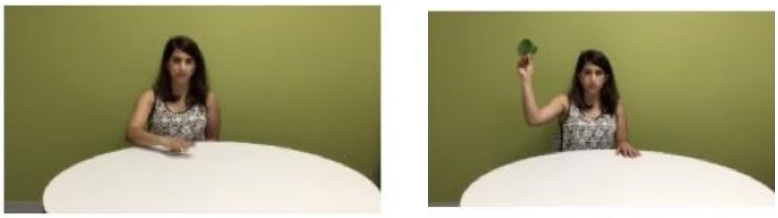
For example, in one set, an actor picks up a natural object so that it sticks to an open hand (picking up a leaf, a stick and a rock in the target events, see Fig. 1, left), and waves a leaf around and twirls a stick on the table using her finger in the two distractor events (see Fig. 1, right). At test, children saw her pick up an object using an open hand (correct) or move an object from the center to the sides of the table (incorrect) (see Fig. 2).

Four sets of events were constructed and two of these were shown to each participant so that they could learn two verbs. Different children saw different sets to minimize the influence of a single set of events on the results. We also created three orders of the events: Target first (TTDDT), Distractor first (DTTDT) and Alternating (TDTDT), and children were assigned randomly to one of these three orders (T=Target, D=Distractor).

3. Experimental Set-up

Children sat in front of a 21-inch flat screen video monitor; a Tobii X30 eye tracker device was at the bottom of the monitor connected to a laptop. A webcam recorded the children's pointing responses. The distance between

the table holding the monitor and participant was 16 inches. The eye tracker used a corneal reflectance tracking technique to measure eye movements. The experimenter calibrated the Tobii x30 eye-tracker using the Tobii 5-point calibration stimuli for infants; the software used was Tobii Studio.



"She's going to snarf it. She's snarfing it. She snarfed it!" "Look what she can do!"

Fig 1: Learning phase example Target (L) and Distractor event (R).

Test Trial 1
"Which one is snarfing it? Can you show me? Can you point to it?

Fig 2: Test trial example.

4. Procedure

Two experimenters were present: one produced sentences from a script while controlling the eye tracker, the other coded children's pointing. Participants saw two warm-up trials and were asked to point to a familiar object out of a pair (e.g. "Can you point to the grapes?"). Then each child was shown two sets of events, one at a time and heard two novel verbs: gorp and snarf (one verb for each set). In each set, children saw 3 target events and 2 distractor events in one of three orders while hearing the new verbs. During the target events, they heard the novel verb three times ("She is going to ___ it. She is ___ing it. She ___ed it."). During the distractor events, they heard non-labelling speech ("Oh, look what she's doing."). At test, they saw a split screen with two different events while hearing "Now it's your turn to find <verb>ing", and were asked to point, "Point to <verb>ing. Can you point to the one who's <verb>ing?"(see Fig. 1). In a second test trial, the same videos were shown on the opposite sides of the screen "You get one more turn to play the game. Can you point to <verb>ing? Which one is <verb>ing?").

The correct side of the screen was counterbalanced. This process was repeated for a second verb. Interrater reliability calculated between the second and a third coder, both from video, showed 94% agreement with Cohen's kappa = .88 (almost perfect agreement).

1.3 Results

1.Pointing results

A univariate ANOVA with Age group (3: 2s, 3s, 4s) and Order (3: Target first, Distractor first, Alternating) as BS factors, dv= proportion trials correct (number correct/total trials with responses), showed a main effect of Age group, $F(2, 75) = 3.94$, $p = .024$, np2= .11, and an Age group by Order interaction, $F(4, 75) = 3.04$, p=.023, np2= .15. Given the significant interaction, we split the data by age group and repeated the univariate ANOVA within each age group, following up with one sample t-tests to compare responses to chance.

No significant effects of Order emerged in the 2 ½-year-olds and 4 ½-year-olds data. There was a significant effect of Order only in the 3 ½-year-old group, with Order, $F(2, 30) = 4.25$, $p < .05$, np2= .23. Because this result was only found in one age group, it suggests that order of the events did not exert a major effect on children's responses.

One sample t-tests showed that children in all three age groups exceeded chance. Specifically, 2 ½-year-old children's responses exceeded chance, $t(23) = 3.40$, $p = .002$, as did 3 ½-year-old children's, $t(30) = 4.60$, $p < .001$, and 4 ½-year-old children's, $t(20) = 12.21$, $p < .001$. An independent samples t-test comparing 4½–year-olds' to 3½-year-olds' responses was significant, $t(50) = -2.25$, $p = .029$ (see Fig. 2).

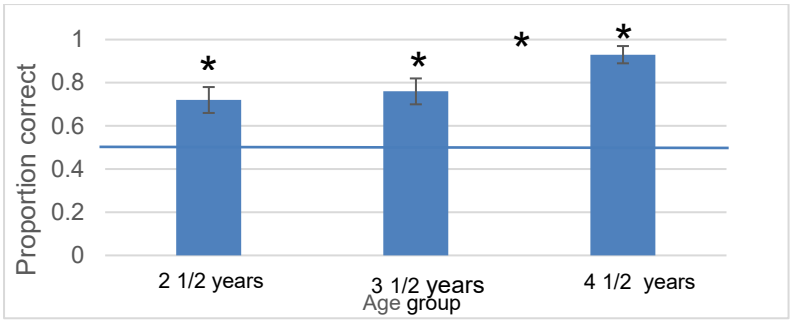


Fig. 2: Pointing results, *$ps < .05$, error bars show SEM, blue line represents chance.

2. Eye tracking results: Looking to the face vs. the hands

A repeated measures ANOVA was computed with Age group (3: 2 ½, 3 ½, 4 ½ years) as BS factor and Trial type (2: target, distractor), Trial number (2: first, last) and AOI (2: face, hands) as WS factors; dv= total fixation duration (with zeros). There was a significant main effect of Trial type, Trial number, and AOI. There were also 3 two-way interactions, and a 3 way interaction of Trial type x Trial number x AOI, $F(1, 67)=27.36$, $p< .001$.

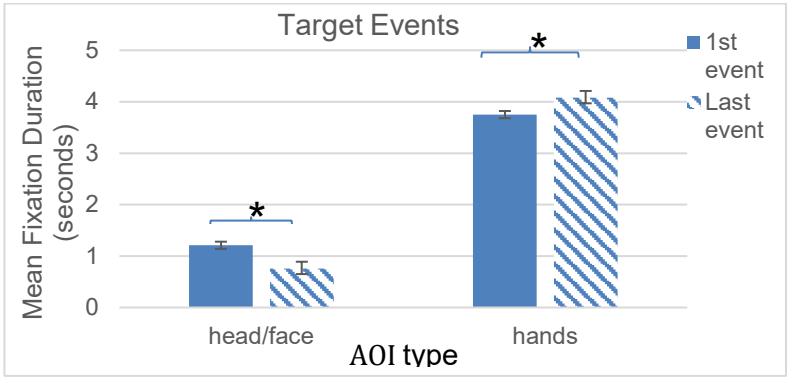


Fig. 3. Target Events: Graph shows mean total fixation duration by Trial (first, last) and AOI type (face, hands), *$ps< .05$.

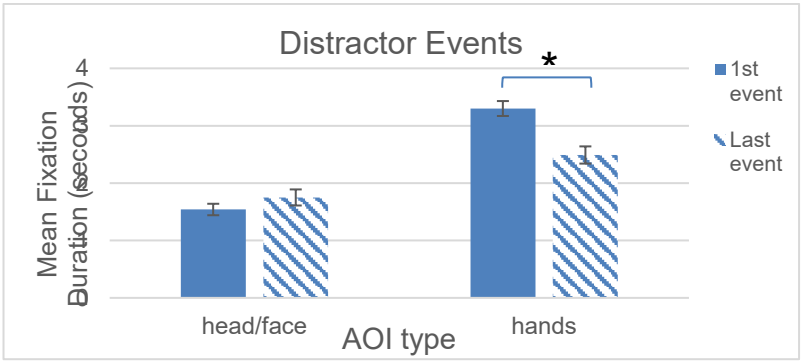


Fig. 4. Distractor Events: Graph shows mean total fixation duration by Trial (first, last) and AOI type (face, hands), *$p< .05$.

Target Trials: A repeated measures ANOVA examining looking during the Target trials revealed a main effect of AOI, $F(1, 69)= 580.62$, $p< .001$, and a Trial number x AOI interaction, $F(1, 69)= 30.14$, $p< .001$. Pairwise

comparisons with Sidak corrections showed that, in both the First and Last Target events, children looked longer at the Hands than the Face, *p*s< .05. Importantly though, across trials, looking to the Hands increased, *p*= .003, while looking to the Face significantly decreased, *p*< .001 (see Fig. 3).

Distractor trials: A repeated measures ANOVA revealed a main effect of Trial number, $F(1, 69)= 26.96$, $p< .001$, AOI, $F(1, 69)= 53.91$, $p< .001$, and a Trial number x AOI interaction, $F(1, 69)= 14.10$, $p< .001$. Pairwise comparisons with Sidak corrections showed that children looked longer at the Hands than the Face in the first distractor event, $p< .001$, and in the second distractor event, $p=.004$. However, across trials, a different pattern was seen with children's looking to the Hands decreasing, $p< .001$, while looking to the Face was maintained (see Fig. 4).

1.4 Discussion

Children learning verbs often see other intervening events as they are learning verbs (e.g., seeing a stirring event while learning the verb 'chop' in the kitchen), which need to be processed differently than events linked to the target verb. Our results provide experimental evidence for differential visual processing of relevant vs. distractor events as events are experienced. Specifically, across age groups, when viewing relevant events, children increase their looking to the hand region (actions) over trials and decrease their looking to the agents' face, which is less informative for learning a verb's meaning. In contrast, when viewing distracting events, children decrease their looking to hands over trials and maintain their attention to the head. These results add to the body of research showing that children can compare events during verb learning, and that comparisons help them extend new verbs (e.g., Childers et al., 2016; Scott & Fisher, 2012; Waxman et al., 2009). Specifically, they provide evidence that children are adjusting their visual fixations differently as they see relevant events linked to a new verb as opposed to irrelevant events, suggesting that they are strategic in how they visually inspect events.

Learning new verbs is important to learning one's native language. Our study suggests that by 2 ½ years, children have developed visual strategies for inspecting events that should help them attend to events appropriately when seeing relevant events and hearing verbs (attending more to what the hands are doing than the face) and perhaps help them ignore distracting information (as they focus on faces and not hands) when they see irrelevant events. These are exciting new findings that reveal what mental mechanisms could underlie early verb learning. The study of children's acquisition of verbs, which vary across languages, could suggest ways languages may have changed in response to the learnability of emerging predicate structures.

Acknowledgements

Funding for this research was provided by the National Institute of Health (2R15 HD044447), Murchison Scholars Fund, and Trinity University. We thank Victoria Bell, Sophia Arriazola, Aidan Burke, Priscilla Tovar-Perez, Hannah Bortz, Katy Capps, and Sophia Rodriguez for data collection and coding.

References

Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development*, *61*, 681-696.

Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language, 31(1),* 3-22.

Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36(1)*, 201–224.

Childers, J. B., Parrish, R., Olson, C. V., Burch, C., Fung, G., & McIntyre, K. P. (2016). Early verb learning: How do children learn how to compare events? *Journal of Cognition and Development, 17(1)*, 41–66.

Childers, J. B., Paik, J. H., Flores, M., Lai, G., & Dolan, M. (2017). Does variability across events affect verb learning in English, Mandarin, and Korean? *Cognitive Science, 41,* 808-830.

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15(6)*, 297–301.

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science, 1(1),* 31-35.

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action based verb extension. *Child Development, 82(2),* 674-686.

Howard, T. J., Porter, B. M., Childers, J. B. (2019). Can young children ignore irrelevant events, or subevents, during verb learning? Journal of Cognition and Development, 20(3), 411-432.

Imai, M., & Childers, J. B. (2020). Learning individual verbs and the verb system: When are multiple examples helpful? In J. B. Childers (Ed.), *Language and Concept Acquisition from Infancy Through Childhood* (pp. 131-158). Springer, Cham.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79(4),* 979-1000.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11(4),* 628–634.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122(2),* 163–180.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology, 59(1),* 67–95.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18(5),* 414–420.