Visual attention and language exposure during everyday activities: an at-home study of early word learning using wearable eye trackers

Sara E Schroer[1], Ryan E. Peters[1,2], Alyssa Yarbrough[1], Chen Yu[1]

saraschroer@utexas.edu, ryanpeters@duolingo.com, alyssa.yarbrough@utexas.edu, chen.yu@austin.utexas.edu

[1] Department of Psychology, The University of Texas at Austin, Austin, TX, USA
[2] Learning & Curriculum, Duolingo, Pittsburgh, PA, USA

Abstract

Early language learning relies on statistical regularities that exist across timescales in infants' lives. Two types of these statistical regularities are the routine activities that make up their day, such as mealtime and play, and the real-time repeated behaviors that make up the moment-by-moment dynamics of those routines. These two types of regularities are different in nature and are embedded at two different temporal scales, which led to divergent research in the literature – those who collect long-form recordings and observations of at-home behavior and those who use eye trackers and micro-level analyses to quantify real-time behavior in laboratories. The goal of present paper is to jointly examine and connect the statistical regularities at these two timescales. Towards this goal, we brought wearable eye trackers to English- and Spanish-speaking families' homes to record parent and toddler visual attention during daily routines. We transcribed parent speech during object play and mealtime and coded toddler visual attention during naming moments. We found that parents and toddlers jointly interacted with the unique vocabularies of the two activities. Although naming and attention were more coordinated during object play, mealtime still afforded opportunities for high-quality naming moments. Our results lay the building blocks for connecting these two lines of research and demonstrate the feasibility of at-home data collection with eye trackers.

Keywords: visual attention; parent-toddler interaction; at-home research; language development; wearable eye tracking

Introduction

The environment of a developing child consists of structured information that forms the input for early language learning (Goldstein et al., 2010). To make use of the structure, infants can employ statistical learning (Saffran & Kirkham, 2018) to detect word boundaries, map a label to an object or action, and group their growing vocabulary into meaningful categories.

Infants' daily lives are composed of routines and activities. The constraints and demands of these different routines, such as playing with toys, eating meals, and bath time, create different types of language experiences. There is a growing trend in developmental literature to holistically consider the language learning input infants receive across the day and in different activities – and often this means collecting long-form recordings at home (e.g., Soderstrom & Wittebolle, 2013; Bang et al., 2019; Tamis-LeMonda et al., 2019; Rosemberg et al., 2020). Similar research collects video recordings from head-cameras to specifically study infant's visual experiences over the first few years of life (e.g., Fausey et al., 2016; Clerkin et al., 2017; Long et al., 2021). This work can tell us about the language learning landscape and the statistics in the input at the timescale of hours, days, or weeks. When we take stock of this line of research that has studied language input across routines, we see that the different activities infants can engage in throughout the day are categorized by different amounts and types of parent speech, as well as a subset of concrete nouns and verbs that are unique to that context (e.g., Tamis-LeMonda et al., 2019). Of course, the structure available to infants does not stop at the level of context. The interactions that occur within a routine are also informative cues.

For young children, interactions tend to be scripted and highly predictable, such as in a game of peek-a-boo (Bruner, 1983). Infants will likely encounter different scripts across different contexts, which provide another set of informative cues to scaffold word learning. Within an utterance (or set of utterances) and within an action, there are also predictable statistics to learn. During diaper changing, each step of the interaction is segmented by temporally synchronous speech and actions, such as tickling the feet after putting on pants (Nomikou & Rohlfing, 2011). By tracking these low-level statistics of an interaction, infants can learn the behaviors that make up a routine, creating a familiar script to support the learning of more words. Studying such moment-to-moment dynamics of parent-child interactions has revealed dyadic behaviors that contribute to word learning in real-time. Parents selectivity name objects in response to infant behavior – such as holding or looking the object (e.g., Chang et al., 2016). Naming objects in these moments when infants are engaged with them promotes word learning (Yu & Smith, 2012; Schroer & Yu, in press).

Language learning is grounded in routines, but the ways these timescales influence one another is bidirectional. We can borrow Gottlieb's framework on the bidirectional influences in development (2007), to explain the probabilistic

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

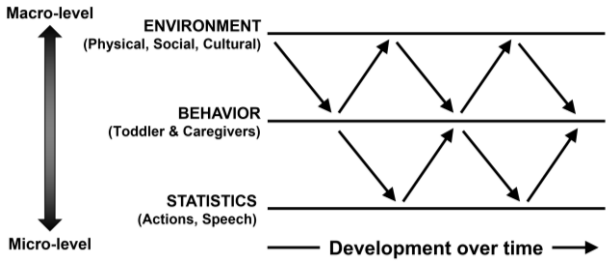

Figure 1: The probabilistic nature of language development over time, inspired by Gottlieb (2007). Consider a mealtime routine of making a sandwich – at the macro-level, the dyad is influenced by environmental factors like being in the kitchen and cultural factors like the type of food they are making. Making a sandwich then follows a script of behaviors such as taking out bread, spreading peanut butter and jelly, and cutting the sandwich into pieces. The dyad may talk to each other as they do this, and toddlers might help their parent. Each step of the script can be segmented into the individual actions and words that form the underlying statistics of the routine at the micro-level.

development of language. The environment of a dyad, including social, cultural, and physical factors, influences their behavior at the macro-level. And the behaviors of caregivers and the infant have an equal impact in selecting and shaping their surrounding environment. The ways the dyad behaves (speech and actions) are structured and patterned throughout the day, making the statistics of the input that infants receive for language learning. In turn, dyadic behaviors are also influenced by the ambient (and each other's) statistics. A re-imagining of Gottlieb's bidirectional influences can be seen in **Figure 1**. Some directions of influence are well-studied, with discovered correlations showing how environment may shape behavior or behavior may shape statistics. Missing from this direction of influence is mechanism – how influences like environment and routines at the macro-level are directly linked to behaviors and statistics at the micro-level. By explicitly measuring micro- and macro-level data types simultaneously, we can better motivate the mechanisms through which environmental influences matter for learning.

The presented data was collected as part of the new FIELD Project, a Family-Infant Eye-tracking and Language Development dataset collected using wearable eye trackers at participants' homes. With the FIELD project, we can study the statistics in toddlers' language learning environments that exist across different timescales, bridging the disconnect between home recordings and in-lab research – providing context for why the findings from each approach matter. How do the real time behaviors we have studied in the lab make up the greater landscape? And how do the macro-level statistics like routines shape the micro-level behaviors? In the FIELD Project, we bring wearable eye trackers into participants' homes to capture the patterns of dyadic behavior across different routines in their natural environment.

The main goal of the paper is to study how the in-the-moment statistics that we know are important for language learning unfold across different routines. We will compare infant's input in two everyday contexts by measuring the words they hear and the objects they see *during* those naming moments. This is the first study we know of using wearable eye trackers at home to tackle this question, with both parents and toddlers wearing the eye trackers. We also tested the feasibility of adding wearable eye tracking to the growing trend of collecting naturalistic, long-form recordings at home. It is a proof-of-concept that dual eye tracking data can be collected from parent-toddler dyads at home and that the data can be processed, coded, and analyzed in a tractable way.

Methods

Participants

The study was conducted in the metropolitan area of a major city in Texas. Families were recruited through word-of-mouth in the Psychology and Neuroscience departments of the university as well as through advertising at the local science museum. Parents were told that we were interested in what children see and hear during their daily lives. Our initial dataset is comprised of 4 families that completed 1-3 recordings each, for a total of 8 recordings.

Toddlers were aged 27- to 31-months old (mean = 29mo, 3 male). The parent wearing the eye tracker (focal parent) was selected by the families and was equally likely to be the mother or father. Other members of the family were allowed to be home during the recordings – often both parents were home, and one toddler also had an older brother present in the recordings. The other three toddlers were only children. Two toddlers were white/Hispanic, one was white, and one was mixed race (white/East Asian). Three of the families spoke more than one language at home (Japanese or Spanish), with varying degrees of language exposure (1 hour/day to 12+ hours/day). Two families spoke only English during the recordings, one family spoke exclusively Spanish, and one family spoke both English and Spanish. All families reported being middle- or upper-middle-class. All participants were neurologically and psychologically typically developing and had no disclosed visual impairments. 2 additional families were excluded due to equipment error (1 white, 1 East Asian).

Data collection

Families that expressed interest in the study were provided with a pair of adult-sized sunglasses and a glasses strap to help the toddler get used to wearing oversized glasses. After a week of practicing, researchers visited the family's home with a pair of wearable eye trackers (**Figure 2**).

Families were asked to record data for an hour or until their toddler no longer wanted to wear the eye tracker. Parents were not asked to participate in any specific activity, but were told that we were especially interested in toy play, mealtime, book sharing, and chores. Parents were told they could speak in the language that was the most comfortable and typical for their family. The researchers left the house while the family

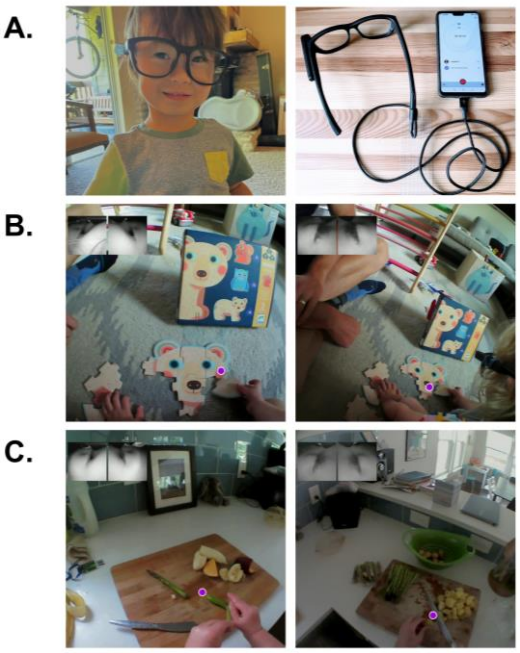

Figure 2: A. Toddler wearing the eye tracker (left) and an overhead shot of the Invisible (right). **B.** Toddler's (left) and parent's (right) view during object play; and **C.** during mealtime. The purple dot indicates gaze.

wore the eye trackers. After data collection, parents filled out a demographic survey.

Toddlers and the focal parent wore Pupil Labs Invisible eye trackers that were attached by a single cord to an Android phone. Toddlers wore a vest with a pocket in the back and the parents wore a running arm band to hold the phone. A glasses strap was used to snuggly hold the glasses onto the toddler's head. Not tethered to a computer, participants were able to move freely and go about their daily lives. Toddlers were tolerant of wearing the eye tracker for long periods of time (a bit more than an hour, which is the battery life of the phone).

The Invisible eye trackers look like a pair of glasses and are equipped with a fish-eye scene camera that captures the view in front of the participant, as well as two infrared cameras that record the participant's eyes. The Pupil Invisible requires no calibration and uses an algorithm for gaze-estimation that is robust to changes in lighting, slippage of the eye tracker, and movement (Tonsen, Baumann, & Dierkes, 2020). The eye trackers also recorded audio.

Measuring language input

Recordings were first annotated to identify the activities the toddlers engaged in. To be coded, activities had to last at least a minute but could also include bouts of "off-task" behavior that lasted less than a minute. Dyads participated in a diversity of activities, including object and non-object play, book sharing, mealtime (which includes making and cleaning up food), chores, screen-use, going for walks, and more. For the present study we will focus on object play and mealtime as both English- and Spanish-speaking families engaged in these object-focused activities.

We then transcribed the focal parent's speech directed at the toddler (or a group the toddler was a part of) during object play and mealtime. Parent utterances were considered separate if there was more than 400ms of silence. Spanish transcriptions were completed by a fluent speaker. Naming utterances were identified as when parents named a concrete noun (an object that the toddler could reasonably see, touch, or hold). The noun was then assigned to a category (based on Long et al., 2021): own or social's partner's face, empty hands, or rest of body; people - real (but not wearing the eye trackers); people - drawings or toys; animal – real; animals – drawings or toys; vehicles – real; vehicles – drawings or toys; books; clothing; food; utensils/dishes; cleaning supplies; plants; screens; furniture/appliances; other toys; other small objects; and other large objects. Parts of animals or vehicles (e.g., nose or tire) were categorized as the whole object.

Measuring visual attention

Toddler visual attention was then coded within the onset and offset of naming utterances (**Figure 3**). First, attention was coded at the frame-level (rate of 25 frames/sec) using an in-house program. A coder identified what the toddler was looking at (as indicated by the gaze-estimation from the eye tracker) and assigned it to a category, the same 22 used for categorizing nouns. If toddlers were looking at their own or their parent's hands holding an object, that object was coded as the location of their attention (hence why "empty hand" is a category). The coder was blind to the noun being spoken by the parent.

A coder then watched the toddler's eye tracking video during each naming utterance and determined whether the toddler was looking at a potential referent for the annotated

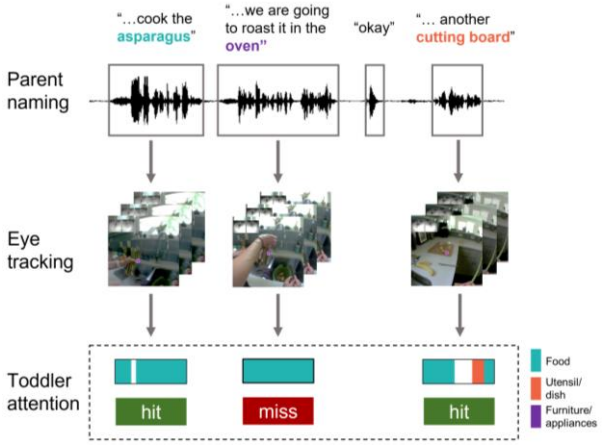

Figure 3: Overview of coding process. Parent utterances and naming moments were first transcribed. Frames of the eye tracking video were then coded to identify toddler attention to the noun categories. Matching colors of naming and attention indicate a category match. Lastly, naming moments were coded as hits or misses.

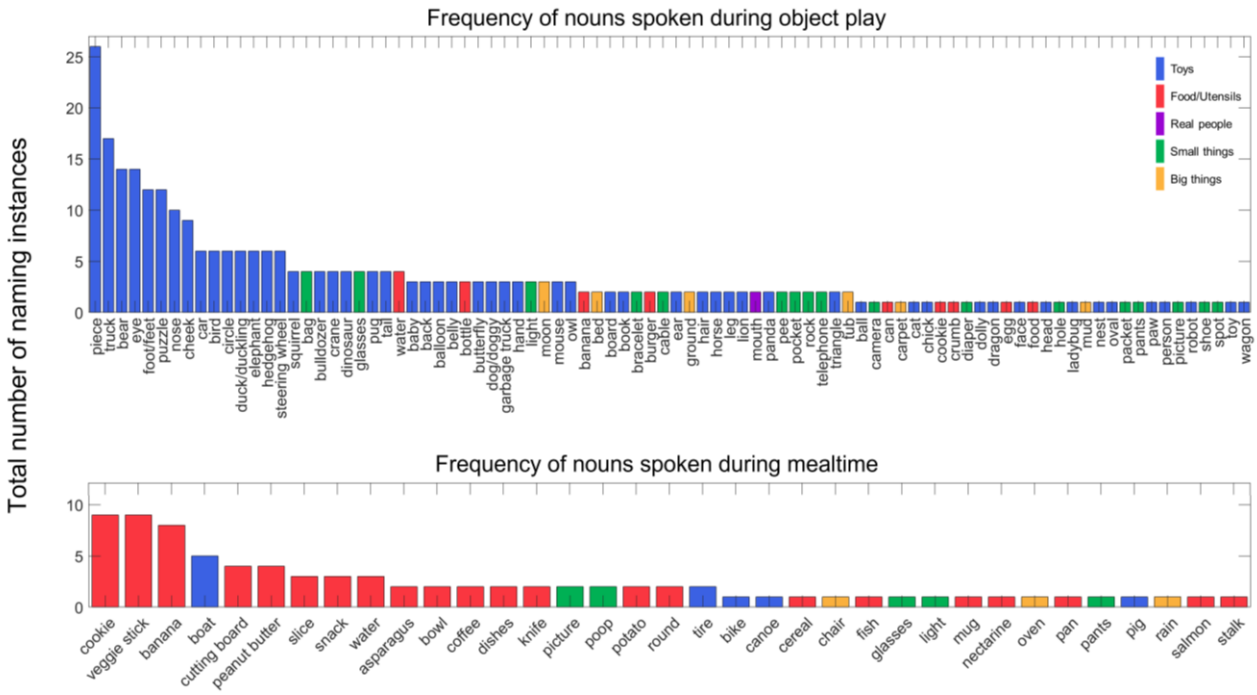

Figure 4: The total number of times each unique noun was spoken during object play (top) and mealtime (bottom). Nouns are assigned into 5 colored categories. All nouns spoken in Spanish were translated to English for the visualization.

nouns (e.g., parent said "car", a potential referent is a picture, toy, or real car). A *hit* naming utterance was defined as the toddler looking at potential referent for at least a single frame of the utterance. All other naming utterances were *misses*.

Analysis plan

The first set of presented analyses considers **language input** during two everyday activities – including how much parents talk and name objects during object play and mealtime, as well as the types of objects they are talking about. Our second set of analyses considers toddlers' **visual attention during naming moments**, specifically the objects they look at across all naming moments. Lastly, we consider the **coordination of naming and attention** by measuring how often naming moments are "hits" during object play and mealtime and if toddlers look at objects in the same category as the referent (matches) or other categories (mismatches) during naming.

Our intention with continuing data collection is to be able to compare the data collected from English- and Spanish-speaking participants to identify both similarities and differences in these families. In the present paper, however, we will focus on a comparison of the two activities and merge the data collected in the two languages.

Results

Across the 8 recordings, we had 5 instances of object play and 3 instances of mealtime. On average, object play bouts lasted 10.80 minutes (range: 6.41-18.87) and mealtime bouts lasted 6.69 minutes (range: 2.46-12.05). Because of the varying length, comparisons between object play and mealtime will use the frequency of behaviors per minute and/or the proportion of total instances in each activity.

Language input during everyday activities

We first measured how much parents talked and named objects during each activity. In **object play**, parents spoke an average of 14.77 times/minute (range: 4.54-21.69) and named objects 6.23 times/minute (range: 1.62-10.30). During play, 42% of all utterances contained a naming event (range: 23-67%). In **mealtime**, parents spoke 11.98 utterances/min (range: 6.97-16.36) and named objects 4.88 times/minute (range: 2.74-7.01). In mealtime, 40% of utterances contained a naming event (range: 39%-43%).

There were no significant differences in frequency of utterances and naming between the two activities (using t-tests, *ps* > 0.507), though this may be due to the small sample size. With more data, we expect the trend of more speech and naming during object play to become significant. This would be in line with previous observations (e.g., Tamis-LeMonda et al., 2019). Nonetheless, we are intrigued by the apparent lack of differences in the *proportion* of utterances that contain naming events and hypothesize that this result would be sustained, even as speech rates diverged.

We then compared the types of objects being labeled by parents (**Figure 4**). Across all recordings, there were 88 unique nouns spoken in the 53.99 minutes of **object play** and 35 unique nouns spoken in the 20.07 minutes of **mealtime** – yielding a comparable frequency of unique nouns/minute at the corpus level (1.63 in play and 1.74 in mealtime). There were, however, marked differences in the types of objects

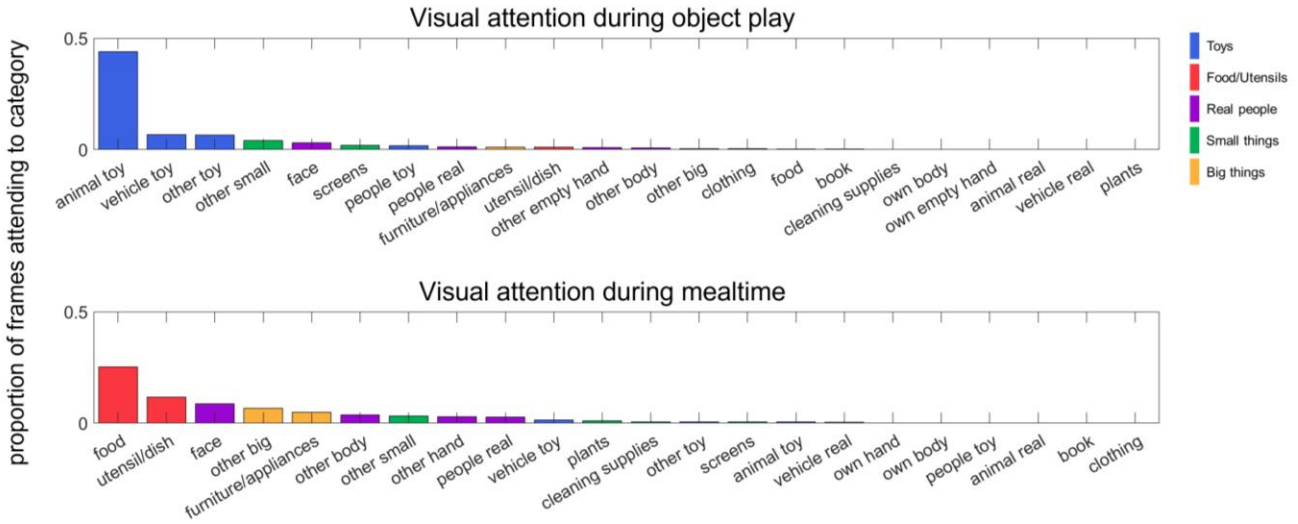

Figure 5: Toddler attention to objects was coded for every frame within each naming utterance. Shown are the proportion of frames infants looked at the noun categories during object play (top) and mealtime (bottom). Nouns are assigned into 5 colored categories.

parents labeled. The majority of naming instances during **object play** were of toys and the majority of naming instances during **mealtime** were of food and various utensils/dishes. For ease of visualization, we grouped the 22 types of nouns into 5 categories (toys, food/utensils, real people, small things, and big things). As in Tamis-LeMonda et al. (2019), we found evidence of "unique vocabularies" in the types of objects parents talk about during play and mealtime – but do toddlers distribute their attention to the same vocabulary?

Visual attention during naming moments

To look for a unique vocabulary in visual attention, we then visualized what objects the toddlers look at during naming instances in object play and mealtime. To do so, we calculated the proportion of total frames in each activity that toddlers gazed at the 22 categories of nouns (**Figure 5**). Just as with naming itself, we found that infants predominantly attended to the activity-specific objects.

Coordination of naming and visual attention

To test the coordination of parent naming and toddler attention, we calculated the proportion of naming instances that are "hits", when infants are attending to a potential referent for any amount of time during a naming utterance.

In **object play**, hits occurred an average of 3.62 times/minute (range: 0.62-7.25), with 56% of naming events counting as hits (range: 0.38-0.75). This proportion of hits is in line with a recent in-lab finding that infants are equally likely to be attending to the correct or an incorrect referent during naming (Yu et al., 2021). In **mealtime**, a different pattern emerged. Hits occurred 1.52 times/minute (range 0.81-2.16), with only 35% of naming events counting as hits (range: 0.17-0.58). There was a moderate association between hit/misses and activity type, in that object play is more likely to have hits (and mealtime more likely to have misses; Yule's Q = -0.362).

To learn more about hits and misses during these two everyday activities, we then looked at the proportion of each naming instance toddlers looked at the matching category (e.g., parent names "elephant" and the toddler is looking at an elephant or other animal), at a mismatching category (e.g., toddler is looking at a toy car), at their parent's face, at empty hands (toddlers' or parents'), or nothing (**Figure 6**).

Although there was an association between hits and play, toddler attention during hits and misses did not appear to differ in object play versus mealtime. **During hits**, toddlers are most likely to look at the category matching the labeled object, which includes time spent looking at a potential referent (play=0.62, mealtime = 0.49 of utterance), and spend little time attending to category mismatches (play=0.20, meal =0.27). **During misses**, toddlers are most likely to look at objects from a mismatching category (play = 0.39, mealtime = 0.40). Crucially, misses are not "near hits" as toddlers do not spend a large proportion of time attending to objects in the same category as the labeled referent (e.g., during a miss the toddler is more likely to be looking at a car than a giraffe while parent says "elephant") (play = 0.17, mealtime = 0.06).

Using a linear mixed effects model, we confirmed that the proportion of time attending to a category match could only be predicted by whether the utterance was a hit/miss, not which activity the utterance came from ($\beta = 0.389$, $p < 0.001$, Cohen's d = -1.29). Similarly, time attending to a mismatching category was only significantly predicted by hit/misses (β = -0.173, $p < 0.001$, Cohen's d = 0.49). Both models included a random effect of subject and were better than a null model with the random effect only (using a Chi-Square difference test, $ps < 0.001$).

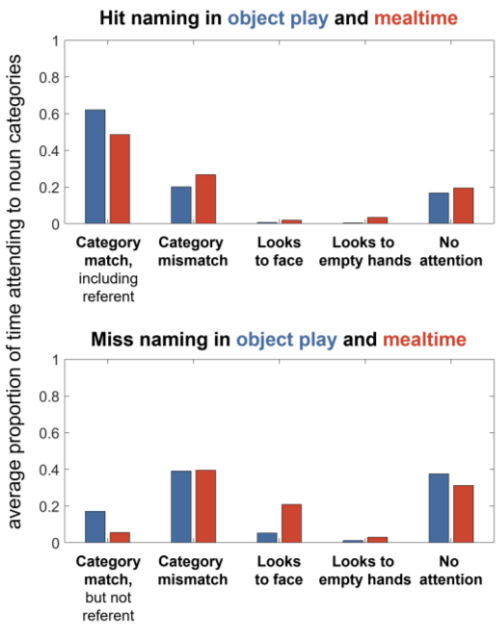

Figure 6: Proportion of a naming utterance that toddlers attended to matching and mismatching categories when the naming utterance was a hit (top) or a miss (bottom).

Taken together, our results suggests that even though the amount of parent speech, and especially the amount of hits, varies between these everyday activities, play and mealtime still afford equally good word learning opportunities.

Discussion

We present the first results from the FIELD Project, an at-home dataset with parent-toddler dyads wearing eye trackers while going about their daily lives. To our knowledge, a study of this nature has yet to be conducted. We observed differences in the objects parents talk about during two everyday activities and found that toddlers visually attend to the same types of objects their parents talk about, demonstrating that the unique vocabularies of daily routines extend to the visual environment as well. Although "hit" naming moments were more likely to occur in object play, the quality of hit naming events was the same in object play and mealtime. Dyads may be utilizing differing techniques to create learning moments in mealtime, a context that is only beginning to be studied (e.g., Clerkin et al., 2017).

The presented dataset does have limitations. Although the micro-level data we collect is dense, we have a small sample and few observations. It is still unclear how our findings may generalize to a larger, more diverse sample of families – as well as other everyday activities. Nonetheless, we were able to replicate earlier studies (e.g., Tamis-LeMonda et al., 2019) by observing more speech and naming in object play than in mealtime, as well as unique vocabularies in each activity. Additionally, we replicated behaviors observed during lab experiments. Chiefly, that toddlers will look at objects during naming moments, but only in half of cases is the attended object a potential referent (Yu et al., 2021). By measuring behavior at both the macro- and micro-levels, we can test for similarities in behavior across diverse contexts. When we zoom in to the micro-level of an interaction and a single instance of naming, we see constants. Although the events surrounding naming may differ across routines and contexts, the behavior itself may unfold in the same way.

An important contribution of this paper is demonstrating that rich, long-form-style dyadic eye tracking data *can* be collected from parents and toddlers at home. Currently, at-home developmental research is often conducted using audio recorders (like a LENA device), head cameras, or traditional video recorders that are either mounted onto a tripod or carried around by a researcher. Each of these methods has it strengths, but none allow researchers to study the micro-level coordination of dyadic gaze and object manipulation. And while head cameras can contribute to our understanding of the visual environment of young children, they do not provide information on how children actually distribute their attention. In addition to yielding gaze-estimation data, the wearable eye trackers are cameras that necessarily move with toddlers and parents from room to room, eliminating the need for researcher presence during data collection. The eye trackers also provide headcam data that can be used for computer vision research and collect high-quality audio recordings that can be used for speech analyses. Although audio recorders and some head cameras can collect hours of data during a single recording (e.g., LENA devices record up to 16 hours; Bergelson, Casillas, et al., 2019), the eye trackers we used have a battery life of 1 to 2 hours, which is similar to the duration of many at-home studies that still require researchers to set up the recording equipment (e.g., Tamis-LeMonda et al., 2019). Using wearable eye trackers allows for a rich dataset to be collected during at-home research and will provide major insights into how language learning unfolds in infants' daily lives.

Moving forward we plan to expand the FIELD dataset with the goals of (1) better representing the demographics of our community and (2) comparing English- and Spanish-speaking families. With continued data collection we will be able to capture more instances of play, mealtime, book sharing, and chores, as well as learn of other routines that make up toddlers' lives. We also plan on annotating our data in additional ways, including coding parents' gaze, the objects being held, and properties of the participants' field-of-view (e.g., the size of objects, how many potential referents are in view, and low-level features like saliency). With FIELD, our ultimate goal is to provide the bridge that connects at-home and in-lab research by identifying mechanisms and learning processes.

Just as structured information in routines and behaviors scaffolds the early learning of words, studying the patterns of information in toddlers' daily environments will scaffold our understanding of the probabilistic development of language and the bidirectional influences driving its progress.

Acknowledgments

This work was supported by NIH R01HD074601 and R01HD093792 to CY. SES was supported by the NSF GRFP (DGE-1610403) and NIH T32HD007475. We thank the Developmental Intelligence Lab at UT Austin for their support in data collection and coding.

References

Bang, J. Y., Munévar, M., Marchman, V. A., & Fernald, A. (2019). Everyday Language Learning Environments. *OSF Preregistration*. https://doi.org/10.17605/OSF.IO/YHND3

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental science*, 22(1), e12724.

Bruner, J. S. (1983). *Child's talk: Learning to use words*. Norton.

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*,41(5–8).

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160055.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, 152, 101-107.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in cognitive sciences*, 14(6), 249-258.

Gottlieb, G. (2007). Probabilistic epigenesis. *Developmental science*, 10(1), 1-11.

Long, B., Kachergis, G., & Bhatt, N., & Frank, M. C. (2021). Characterizing the object categories two children see and interact with in a dense dataset of naturalistic visual experience. *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

Nomikou, I., & Rohlfing, K. J. (2011). Language does something: body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*, 3(2), 113-128.

Rosemberg, C. R., Alam, F., Audisio, C. P., Ramirez, M. L., Garber, L., & Migdalek, M. J. (2020). Nouns and verbs in the linguistic environment of Argentinian toddlers: Socioeconomic and context-related differences. *First Language*, 40(2), 192-217.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, 69, 181-203.

Schroer, S. E., & Yu, C. (in press). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science.*

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, 8(11), e80646.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child development*, 90(6), 2135-2152.

Tonsen, M., Baumann, C. K., & Dierkes, K. (2020). A High-Level Description and Performance Evaluation of Pupil Invisible. *arXiv preprint arXiv:2009.00508.*

Yu, C., & Smith, L.B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244-262.

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, 118(52).