How everyday visual experience prepares the way for learning object names*

Elizabeth M. Clerkin[1], Elizabeth Hart[1], James M. Rehg[2], Chen Yu[1] & Linda B. Smith[1]

Abstract— **Infants learn their first object names by linking heard names to scenes. A core theoretical problem is how infants select the right referent from cluttered and ambiguous scenes. Here we show how the distributional properties of objects in young infants' visual experiences may help solve this core problem in early word learning. Infant perspective scenes of mealtimes were collected using head cameras worn by 9-month-old infants (147 mealtimes from 8 infants). The frequency distribution of objects was extremely skewed with the most frequent visual objects corresponding to the normatively first learned object names in English.**

I. Introduction

Quine (1) famously imagined a stranger who hears a native say ''gavagai'' and point to a scene. To what does ''gavagai'' refer – a rabbit, the grass, a tree, the rabbit's ears, the fur, the beauty of the whole scene, or even some more complicated but possible meaning such as "never on Tuesday"? Infants are like strangers who do not know the native language and somehow have to solve this word-meaning indeterminacy problem to get a lexical foothold into language learning. There are a number of simplifying assumptions that one can make to solve this problem; the most notable within contemporary theories of early word learning is that infants do not entertain every possible meaning that might be conceptualized by an adult but are biased by their perceptual and conceptual systems to link words to some much smaller set of meanings (2). Many of these proposed biases are about object names, the class of words that refer to categories of concrete things and the class of words that are rapidly acquired by 2 year olds (2). What are the processes that might bias object names over other possible meanings, and can such biases actually solve the problem of how infants break into word learning?

The object names that infants learn very early refer to "basic-level" categories such as "cup", "dog", "chair". A long-term assumption has been that these basic level objects are a perceptually easy starting point, "given" to the learner by a visual system tuned to object recognition (3, 4). In contrast to this idea, theorists of human vision and machine vision do not see visual object recognition as an easy task, but rather as a hard one that requires massive visual experience with the specific categories (5, 6).

Accordingly, we propose a different solution and one that may solve the visual object recognition as well as the referential ambiguity problem: the distributional statistics of objects in infants' visual worlds may select and teach the relevant visual categories for first-learned object names. Although infant visual worlds contain many objects, highly cluttered scenes, and much ambiguity, we propose that their visual world does not contain a uniform frequency distribution of objects. Rather a few object categories are highly frequent, providing visual data that point to a small set of recognizable things that are likely to be named. We further propose that these highly frequent objects are the natural consequence of the distribution of objects in natural scenes. Studies of the frequency of specific object categories in natural scenes reveal an extremely right-skewed distribution (7, 8) in which there are a few very high frequency entities and a large number of low frequency entities. We propose that this distribution characterizes infant-perspective scenes as well.

To capture everyday infant-perspective scenes, we placed head cameras on infants in their homes. Our use of head cameras builds on growing multi-disciplinary efforts directed toward understanding egocentric vision (9, 10). Egocentric views have unique properties because they depend on the wearers moment-to-moment behavior and posture and are thus often highly selective relative to the larger environment (11). As illustrated in Fig. 1, the sink, the father, the woman at the sink, the clock, and the dog are all in the same vicinity as the infant, but they are not in the direct view of the infant and thus not in the infant's head camera image as illustrated by the colored area.

One potential limitation of using head cameras concerns the relation between eye and head direction as head cameras measure head direction, not where gaze is directed. Further, the field of view (FOV) of the camera is less than the FOV of the infant and is particularly limited in the vertical and horizontal directions, so that, in principle, eye-gaze could be outside of the captured image. As discussed in a recent

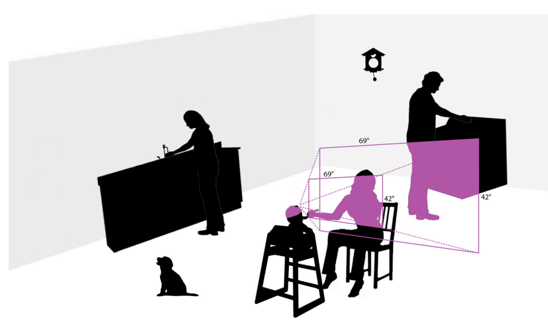


Figure 1. The selective nature of ego-centric views. The field of view indicated in color corresponds to the field of view of the head camera used in this study.

* This research was supported by NSF grant BCS-15233982 and NIH R01HD074601.

1 Psychological and Brain Sciences, Indiana University, Bloomington IN 47405

2 Interactive Computing, Georgia Institute of Technology, Atlanta GA 30332

review of head camera studies with infants (12), there are several facts that mitigate against these limitations. First, in active viewing (not watching screens), infants as well as adults typically turn heads and eyes in the same direction to attend to a visual event (13-16), and sustained visual attention is associated with aligned heads and eyes (17, 18). Further, although eyes often lead heads in directional shifts of visual attention, and although heads undershoot eyes given extreme changes in gaze direction (19), differences in head-eye direction are usually resolved in less than 500 ms in infants (13, 19). Head-mounted eye-tracking studies also show that aligned heads and eyes – that is, fixations to the center of the head-centered image – strongly characterize freely-moving infants viewing behavior (20). The overwhelming predominance of gaze centered within the head camera image (see 20) reduces the likelihood of missed content due to momentary shifts in eye-gaze and the FOV of the camera itself. Thus, if the sample of head camera images is large enough, the observed regularities in content may be assumed to characterize the content of scenes in front of both the heads and eyes.

The participants in the present study were 9 month-old infants. We chose this age in order to focus on the earliest stages of object name learning (21, 22). Further, this is a period considerably prior to the rapid expansion known as the “naming explosion” in early noun learning that occurs around the age of 24 months (2, 22). For the present study, we analyzed the head camera images captured during infant mealtimes. We chose mealtimes as the at-home activity to analyze because it occurs multiples times per day every day and involves a relatively constrained geometry (that aids in the quality and coding of the head-camera images). Further, it seems likely that a large variety of common objects would be in the vicinity and possibly in view (food, dishes, utensils, furniture – the paraphernalia of kitchens and dining tables). Though further research should examine the distribution of visual objects during other activites in infants’ daily lives, the mealtime scenes analyzed in this study ought to be somewhat representative as they include a large variety of situations under this “mealtime” umbrella. Our corpus includes scenes when the infant is in a highchair, when the infant is not in a high chair, when the infant is feeding him/herself, when the infant is being fed by a caregiver, and when the infant is not eating at all but watching others eat or prepare food, do the dishes, etc.

The present study provides the first evidence of the frequency distribution of basic level objects in infants’ every-day visual worlds. From these infant-perspective scenes we found this distribution to be extremely skewed with some objects much more frequent than others and that these highly frequent object categories correspond to object names normatively learned first by infants.

II. Method

A. Participants

The participating infants (n = 8, 3 male) varied in age from 8 ½ to 10 ½ months, with a mean age of 9.2 months.

B. Head Camera

We used a commercial wearable camera (Looxcie) that was easy for parents to use, safe (did not heat up) and very light weight (22 grams). The camera was secured to a hat that was custom fit to the infant so that when the hat was securely placed on the infant, the lens was centered above the nose and did not move. The diagonal field of view (FOV) was 75 degrees, vertical FOV was 42 degrees, and horizontal FOV was 69 degrees, with a 2" to infinity depth of focus. The camera recorded at 30 Hz. The battery life of each camera was approximately two continuous hours. Video was stored on the camera until parents had completed their recording and then was transferred to laboratory computers for storage and processing.

C. Instructions to Parents

Parents were told that we were interested in their infant’s everyday activities, including mealtimes, and were free to choose to record whenever it suited their family’s schedule. Parents recorded a total of 8.5 hours (mean = 1.1, SD = 0.54) of mealtime which yielded a sample of 917,207 frames (mean = 114,651, SD = 57,785). The total number of individual mealtime events in the sample was 147; that is the 8 individual infants had on average 18 different mealtime events. The average duration of mealtime events was 3.5 min (SD = 7.2).

D. Coding of Head Camera Images

The 917,207 (total) frames in the mealtime corpus were sampled at 1/5 Hz (Fig. 2) for coding, which yielded a total of 5,775 coded scenes. Sampling at 1/5 Hz should not be biased in any way to particular objects and appears to be sufficiently dense to capture major regularities. An earlier study of faces in infant head camera images showed that a coarser sampling of scenes at 1/10 Hz yielded the same statistical patterns as a 1/5 Hz sampling. In addition, two different 1/5 Hz samplings over the same set of images with different starting points yielded the same statistical regularities (23).

Quine’s indeterminacy problem applies to coding the objects in the images as any image contains a potentially indeterminate number of “objects” – from the door, the door knob, the screws that hold the door knob, the picture on the wall, the elements in the picture, its frame, and so forth. To determine the psychologically relevant basic level category objects – the objects that speakers of English might name and talk about – we asked a large group of adults (about 500 individuals) to tell us what was in each image. These

Figure 2. Example streams of 15 seconds of continuous recording sampled at 1/5 Hz from 4 different mealtime events.

"coders" were naïve to the purpose of the experiment and were simply asked to label the 5 most obvious objects in the image, providing us with the objects in the scene that were obvious to human perceivers. Each of the sampled scenes was coded by four different coders. Images were coded in sets of 20 sequentially ordered images. Prior to coding the experiment proper, coders were given training on 8 scenes with feedback to implement the following instructions: to exclude body parts but not clothing, to name objects with everyday nouns ("spoon" not "soup spoon" or "silverware."); to label images on objects (for example, the fox on a book page, or the clown on a child's cup); to name foreground but not background objects; and to not repeat names in an image (if there were three cups in view, cup could be provided just once). Coders were asked to supply up to five object names but could supply fewer if the image was sparse (e.g., only a cup in view).

The words supplied by the coders were "cleaned" with respect to the following properties: spelling errors were corrected, plural forms were changed to singular (e.g., dogs to dog), abbreviations were changed to the whole word (e.g., fridge to refrigerator); adjectives were removed (e.g., redbird to bird). From these coder-supplied names for the objects in the scenes, we have three measures of the objects in this corpus of images: Coder Tokens consist of all the object names supplied by coders without regard to image or agreement. Given 4 coders and 5 potential object names offered per image by each coder, the maximum number of Coder Tokens per image was 20. The distribution of object names in the list of all Coder Tokens provides an aggregate measure of the frequency and obviousness (when an object name is repeated by multiple coders within an image) across images. Image Types are an image-based measure -- the unique object names supplied by coders for each image. Focal Types are the count of all object names supplied in an image by at least 2 of the 4 coders. This is thus a more conservative measure of the object types in an image.

E. Object Name Types

The Infant Communicative Developmental Inventory contains 396 words, 172 of which are names for concrete objects (excluding body parts). The words on this inventory are used as a checklist to measure infant receptive and productive vocabulary in infants 8 to 16 months. The words on this inventory were the words in the receptive vocabulary of 50% of infants at 16 months in a large normative study (22). Therefore, these 172 nouns are normatively the first-learned object names in English and comprise our category of First Nouns. The prediction for this study is that the objects named by these nouns will be highly frequent in infant scenes. We compared the frequency of objects named by these First Nouns with two other sets of nouns. The first set consists of common object names typically known by two year olds. This set of Early Nouns consists of 105 concrete object names from the Toddler Communicative Developmental Inventory that were not also on the Infant form. The nouns on the toddler inventory are nouns that were in the productive vocabulary of 50% of 2 ½ year olds in a large normative study (22). This set of Early Nouns serve as a control set of common nouns with an early – but not first – age of acquisition. The key prediction is that objects named by First Nouns will be much more frequent in the infant scenes than objects named by Early Nouns. The second set of comparison nouns consists of the other object names supplied by the naïve adult coders that were not in the First or Early Nouns sets. These Later Nouns consist of the names for the many different kind of things in infant visual experiences.

F. Analytic Approach

Major advances in understanding language acquisition have been made by studying the statistical properties of large corpora of infant-directed speech (assembled from speech to many different children) (24). We take the same approach here with respect to the collected head camera images, combining the meal-time images collected from individual children into one corpus. The analyses, as in corpus analyses of language, are over the frequencies of objects in the visual corpus as a whole and not with respect to the individual participants.

III. Results

The coders provided 745 unique words, or Types, of which 133 were First Nouns, 59 were Early Nouns, and 553 were nouns not on either of these inventories. That is, coders saw a lot of different objects in these mealtime scenes and were not constrained to just supply early learned words. Table I provides a list of the 30 most frequent nouns supplied by the coders for First, Later, and Later Nouns.

We first present (Fig. 3) the frequency distribution of words in the list of Coder Tokens, this the big bag of all words offered by the coders without regard to agreement

TABLE I.

30 Most Frequent Words Per Category		
First Nouns	***Early Nouns***	***Later Nouns***
table	tray	shelf
chair	washing machine	bag
shirt	jar	container curtain
bowl	napkin	cabinet
bottle	knife	lid
spoon	tissue	counter
window	basket	fireplace
cup	sofa[a]	bin
pants	dryer	tablecloth
toy	bench	straw
plate	can	painting
glasses	yogurt	handle
food	bucket	seat
door	sauce	wood
telephone	belt	outlet
couch[a]	walker	cage
picture	grass	mug
box	sandwich	cloth
refrigerator	scarf	cord
book	pretzel	dresser
light	closet	ring
glass	sidewalk	cushion
lamp	soda	brick
fork	ladder	sweatshirt
shoe	popcorn	stool
plant	potato	letters
pillow	stick	railing
sweater	stone	frame
paper	bat	bracelet
blanket	strawberry	

a. Couch and sofa are included on the First and Early Nouns lists respectively because they appear this way (separately) on the Infant and Toddler MCDIs. Other synonyms (e.g., pop and soda) are listed together on the MCDI and were thus collapsed into one word in our coding scheme.

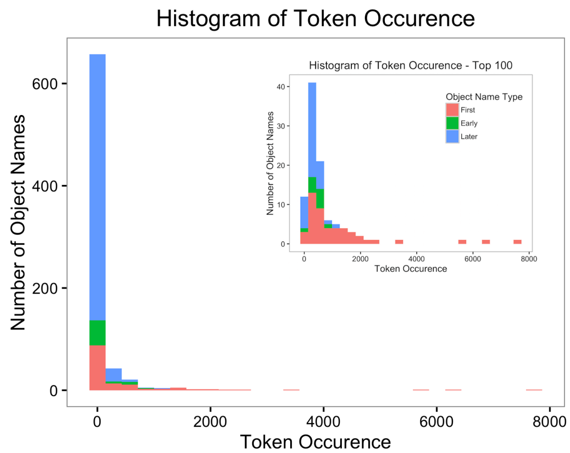


Figure 3. The frequency of object names provided by coders across the image corpus as a whole irrespective of agreement among coders or image (Coder Tokens). Inset is the frequency of object names provided by coders for the 100 object names that occurred most frequently in the corpus, in other words, the tail of the distribution.

among coders nor with respect to number of associated images. This distribution of words reflects the consistency and diversity in the corpus of images as a whole. Frequency of these object names is extremely right-skewed. Moreover, in this big bag of names, the most frequent words are names of objects that are normatively acquired first by learners of English. The long tail is made up of Early Nouns and Later Nouns. The Early Nouns are nouns normatively known by 2 ½ year olds and Later Nouns are also common nouns (see Table 1); however, the object names that coders reported with high frequency in these egocentric mealtime scenes collected from 9 month olds are, quite selectively, first learned object names.

We next report the image-based distribution of objects: Image Types (Fig. 4a) – the proportion of images in which each unique object was reported to be present by at least one coder, and Focal Types (Fig. 4b) – the more conservative measure, requiring at least two coders to report the presence of the object in the image. These image-based measures again show an extremely right-skewed distribution. The items in the tail of the distribution, the 10 most frequent object names are pervasively present throughout the scenes (M = 27% of images) while the next 10 most frequent object names occur much less frequently (M = 11% of images) and the next 10 after that even less frequently (M = 7% of images). In brief, it is a very small set of visual objects are *repeatedly* present in infant-perspective views of mealtime. By hypothesis, these highly frequent objects in the tail of the frequency distribution may form the restricted class of candidate referents for object names heard during mealtime activities. Consistent with this idea, the 10 most frequent Types (Fig. 4a) and Focal Types (Fig. 4b) were objects named by words that are normatively among the first learned object names by infants learning English.

Fig. 5a shows the proportion of images in which each word occurred as a Type by object name type: First Nouns, Early Nouns, and Later Nouns (553). The mean and median for each object name type are shown. Fig. 5b shows this

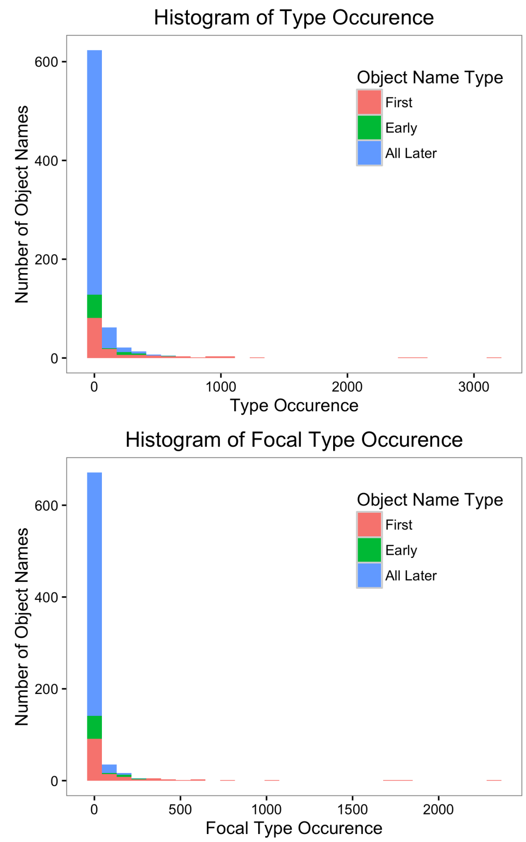
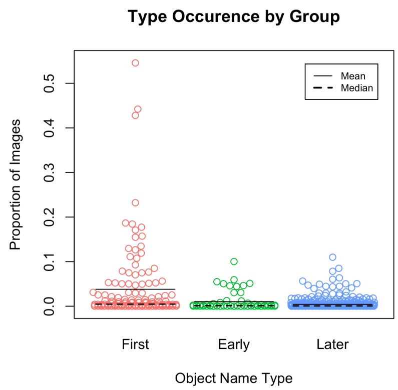


Figure 4. The frequency of object names, measured as unique object names provided by any coder(s) per image (Types) and as unique objects names supplied by at least two coders per image (Focal Types).

same information for Focal Types, our more conservative measure. For statistical analyses, we used planned comparisons, comparing objects named by First Nouns to Early Nouns, and to Later Nouns. Object name was the random variable in these analyses. The prediction in each case is that objects named by First Nouns will be more frequent in the images than objects named by Early Nouns or by Later Nouns. All p-values were corrected for multiple comparisons using the Bonferroni-Holm adjustment. First Nouns occurred as Types more frequently than Early Nouns, $t(164.24) = 3.64$, $p < 0.001$ and Later Nouns, $t(133.08) = 4.76$, $p < 0 .0001$. Finally, First Nouns occurred as Focal Types more frequently than Early Nouns, $t(148.18) = 3.53$, $p < 0.001$, and Later Nouns, $t(132.30) = 4.2787$, $p < 0.0001$. Because of the non-normality of this data, nonparametric Mann–Whitney–Wilcoxon tests were also conducted and confirm the findings of the t-tests. First Nouns occurred as Types more frequently than Early Nouns, $U = 5145$, $p < 0.001$ and Later Nouns, $U = 54610.5$, $p < 0.0001$. Finally, First Nouns occurred as Focal Types more frequently than Early Nouns, $U = 5281.5$, $p < .001$, and Later Nouns, $U = 55433$, $p < .0001$.

In sum, by all measures, the objects that are pervasively prevalent in 9-month-old infants' egocentric views of mealtime correspond to the object names that will be acquired first. It is notable that most of these words on the First Nouns list are not (normatively) known by 9 months.



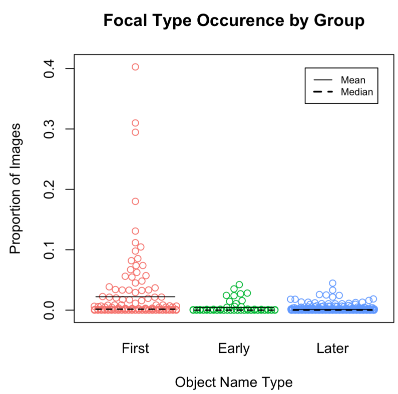


Figure 5. The proportion of images in the corpus in which each individual object name occurred, measured as unique object names provided by any coder(s) per image (Types) and as unique objects names supplied by at least two coders per image (Focal Types). The mean and median for each object name type is included.

That is, the high frequency visual objects at 9 months correspond to the words normatively learned *several months later*.

IV. General Discussion

The results show that during one everyday activity, mealtime, many different objects are present in the infant-perspective scenes but that a very small number of object categories are repeatedly present. The frequency distribution of objects in these scenes, like many natural frequency distributions (25, 26) is extremely right-skewed with a few highly frequent objects being very frequent and with a very large number of infrequent objects. The results also indicate that the highly frequent objects in these egocentric scenes correspond to the object names normatively learned first by infants. These head camera images, collected from 9-month-old infants, capture visual experiences at the earliest stages of receptive lexical learning (22). Our findings indicate the objects whose names infants learn first are highly prevalent – day in and day out – in infant views. This visual prevalence of a small number of high-frequency objects may train the visual recognition system on these objects and draw attention to them when they are named, helping to solve the referential ambiguity problem. The finding that objects that are frequent in the lives of infants have early learned names is not surprising; however, the theoretically significant finding of the present study concerns the frequency distribution of objects in scenes and the selective high frequency of just a few objects – those whose names are learned very early.

The learning that underlies infants' acquisition of their first object names is likely slow and incremental (27), emergent in the aggregation of learning instances over many days, weeks, and months. Incremental, statistical, word-referent learning that emerges by linking heard words to elements in cluttered, noisy visual scenes poses many well-recognized challenges (28). The main idea behind the present study is that these challenges may be resolved (at least in part) by the distributional properties of objects in egocentric scenes. Learning aggregates over current and previous experiences and requires the learner to connect those experiences in memory which is difficult for infants to do even over very short delays (28). The extremely right-skewed frequency distribution of objects in scenes may play a significant role in solving this problem: If the same few object categories are frequently present day in and day out, the problems of attending to, remembering, and aggregating information across their multiple and varied appearances may be substantially reduced. In brief, these pervasively and repeatedly present visual objects may be well recognized (and generalized by infants) and provide a bootstrap into word learning. The potential referents for the young learner may not be all the possible meanings adults can entertain, nor all the referents of the names that they will know when they are two years old, but a small set of referents that are frequently present, thereby reducing referential ambiguity even given highly cluttered visual scenes.

Repetition and diversity have both been shown to support lexical development (29). For example, the most frequent words in a language show marked advantages in many aspects of word learning (30). Highly frequent visual objects also have been shown to attract infant attention to and encourage visual learning about those objects (31, 32). However, diversity is also generally helpful to learning. The contextual diversity of individual words predicts both age of acquisition and the speed of adult judgments in lexical processing tasks (33). A still open theoretical question is whether there is some optimal mix of repetition and diversity.

This question of the relative benefits of consistency versus diversity in the training set has been subject to many experimental investigations of human learning more generally (34-36). Many studies indicate that the diversity of training instances increases generalization, but both theory and evidence suggest that for novices, consistency and repetition may be more important (34, 37, 38). Training sets with a uniform distribution of instances are the standard in these experimental studies and thus their generalizability to training sets with power-law distributions may not be warranted. However, the power-law distribution itself provides a kind of "balance" between consistency and diversity. That is, the high frequency objects provide consistency and the low frequency objects provide diversity. A paper on the role of power-law distributions in visual object recognition proposed that the extremely skewed distribution of visual instances and categories in the learning environment had computational benefits (7). That is, the power-law distribution of objects in the world may make learning easier because learning about the vast number of rare objects borrows strength from the very few high-frequency instances with which the learning system breaks the initial learning barrier. In this way, the consistency of the very few high frequency visual objects may be essential both for solving the visual object recognition problem and the word learning problem.

Because Quine's indeterminacy problem (and the problem infants must solve to break into word learning) requires linking seen objects to heard words, future research should focus on the frequency with which objects are named by adults in infants' auditory environments. However, this study demonstrates the critical role that visual experience with objects may play in helping infants to learn language as they are only beginning to learn their first words. In sum, infants may be able to break into object name learning because of the distributional statistics of objects in egocentric visual experience. The potential referents for first object names may be the relatively small set of well-known visual objects highly frequent in the visual environment.

Acknowledgment

We thank the families who participated in the study and the members of the Cognitive Development Lab, especially Ariel La, Swapnaa Jayaraman, Caitlin Fausey, Jaclyn Berning, and Amanda Essex.

References

[1] Quine WV. Word and object. Cambridge: Technology Press of the Massachusetts Institute of Technology; 1960. 294

[2] Bloom P. How children learn the meanings of words: MIT press Cambridge, MA; 2000.

[3] Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P. Basic objects in natural categories. Cognitive psychology. 1976;8(3):382-439.

[4] Gentner D. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Center for the Study of Reading Technical Report; no 257. 1982.

[5] Kourtzi Z, Connor CE. Neural representations for object perception: structure, category, and adaptive coding. Annual review of neuroscience. 2011;34:45-67.

[6] Pinto N, Cox DD, DiCarlo JJ. Why is real-world visual object recognition hard? PLoS Comput Biol. 2008;4(1):e27.

[7] Salakhutdinov R, Torralba A, Tenenbaum J, editors. Learning to share visual appearance for multiclass object detection. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on; 2011: IEEE.

[8] Yuen J, Russell B, Liu C, Torralba A, editors. Labelme video: Building a video database with human annotations. Computer Vision, 2009 IEEE 12th International Conference on; 2009: IEEE.

[9] Fathi A, Ren X, Rehg JM, editors. Learning to recognize objects in egocentric activities. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On; 2011: IEEE.

[10] Pirsiavash H, Ramanan D, editors. Detecting activities of daily living in first-person camera views. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on; 2012: IEEE.

[11] Foulsham T, Walker E, Kingstone A. The where, what and when of gaze allocation in the lab and the natural environment. Vision research. 2011;51(17):1920-31.

[12] Smith LB, Yu C, Yoshida H, Fausey CM. Contributions of head-mounted cameras to studying the visual environments of infants and young children. Journal of Cognition and Development. 2015;16(3):407-19.

[13] Yoshida H, Smith LB. What's in view for toddlers? Using a head camera to study visual experience. Infancy. 2008;13(3):229-48.

[14] Bloch H, Carchon I. On the onset of eye-head coordination in infants. Behavioural brain research. 1992;49(1):85-90.

[15] Daniel BM, Lee DN. Development of looking with head and eyes. Journal of Experimental Child Psychology. 1990;50(2):200-16.

[16] Schmitow C, Stenberg G. What aspects of others' behaviors do infants attend to in live situations? Infant Behavior and Development. 2015;40:173-82.

[17] Pereira AF, Smith LB, Yu C. A bottom-up view of toddler word learning. Psychonomic bulletin & review. 2014;21(1):178-85.

[18] Ruff HA, Lawson KR. Development of sustained, focused attention in young children during free play. Developmental psychology. 1990;26(1):85.

[19] Schmitow C, Stenberg G, Billard A, Von Hofsten C. Using a head-mounted camera to infer attention direction. International Journal of Behavioral Development. 2013;37(5):468-74.

[20] Bambach S, Crandall DJ, Yu C, editors. Understanding embodied visual attention in child-parent interaction. Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on; 2013: IEEE.

[21] Bergelson E, Swingley D. At 6–9 months, human infants know the meanings of many common nouns. Proceedings of the National Academy of Sciences. 2012;109(9):3253-8.

[22] Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, et al. Variability in early communicative development. Monographs of the society for research in child development. 1994:i-185.

[23] Jayaraman S, Fausey CM, Smith LB. The faces in infant-perspective scenes change over the first year of life. PloS one. 2015;10(5):e0123780.

[24] MacWhinney B. The CHILDES Project: Tools for analyzing talk, Vol 2: The database. 2000.

[25] Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661-703.

[26] Kello CT, Brown GD, Ferrer-i-Cancho R, Holden JG, Linkenkaer-Hansen K, Rhodes T, et al. Scaling laws in cognitive sciences. Trends in cognitive sciences. 2010;14(5):223-32.

[27] Roy BC, Frank MC, DeCamp P, Miller M, Roy D. Predicting the birth of a spoken word. Proceedings of the National Academy of Sciences. 2015;112(41):12663-8.

[28] Smith LB, Suanda SH, Yu C. The unrealized promise of infant statistical word–referent learning. Trends in cognitive sciences. 2014;18(5):251-8.

[29] Hoff E, Naigles L. How children use input to acquire a lexicon. Child development. 2002;73(2):418-33.

[30] Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive psychology. 2001;42(4):317-67.

[31] Kovack-Lesh KA, Horst JS, Oakes LM. The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. Infancy. 2008;13(4):285-307.

[32] Kovack-Lesh KA, McMurray B, Oakes LM. Four-month-old infants' visual investigation of cats and dogs: Relations with pet experience and attentional strategy. Developmental psychology. 2014;50(2):402.

[33] Hills TT, Maouene J, Riordan B, Smith LB. The associative structure of language: Contextual diversity in early word learning. Journal of memory and language. 2010;63(3):259-73.

[34] Carvalho PF, Goldstone RL. Putting category learning in order: category structure and temporal arrangement affect the benefit of interleaved over blocked study. Memory & cognition. 2014;42(3):481-95.

[35] Carvalho PF, Goldstone RL. The benefits of interleaved and blocked study: different tasks benefit from different schedules of study. Psychonomic bulletin & review. 2015;22(1):281-8.

[36] Vlach HA, Sandhofer CM. Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. Child development. 2012;83(4):1137-44.

[37] Gentner D. Bootstrapping the mind: Analogical processes and symbol systems. Cognitive Science. 2010;34(5):752-75.

[38] Goodman JC, Dale PS, Li P. Does frequency count? Parental input and the acquisition of vocabulary. Journal of child language. 2008;35(03):515-31.