Received: 21 February 2018 | Revised: 1 February 2019 | Accepted: 13 February 2019

DOI: 10.1111/desc.12816

PAPER

WILEY Developmental Science

Self-generated variability in object images predicts vocabulary growth

Lauren K. Slone | Linda B. Smith | Chen Yu

Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, Indiana

Correspondence
Lauren K. Slone, Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN.
Email: laurenkslone@gmail.com

Funding information
National Science Foundation, Grant/Award Number: BCS1523982 and BCS1730146; Indiana University Bloomington, Grant/Award Number: EAR -- Learning: Brains, Machines and Children ; National Institutes of Health, Grant/Award Number: R01HD074601, R01HD28675 and T32HD007475-22

Abstract

Object names are a major component of early vocabularies and learning object names depends on being able to visually recognize objects in the world. However, the fundamental visual challenge of the moment-to-moment variations in object appearances that learners must resolve has received little attention in word learning research. Here we provide the first evidence that image-level object variability matters and may be the link that connects infant object manipulation to vocabulary development. Using head-mounted eye tracking, the present study objectively measured individual differences in the moment-to-moment variability of visual instances of the same object, from infants' first-person views. Infants who generated more variable visual object images through manual object manipulation at 15 months of age experienced greater vocabulary growth over the next six months. Elucidating infants' everyday visual experiences with objects may constitute a crucial missing link in our understanding of the developmental trajectory of object name learning.

KEYWORDS
eye tracking, infant, object manipulation, variability, vision, word learning

1 | INTRODUCTION

Theorists of human vision distinguish between a distal *object* in the world and the proximal *image* projected by that object at any moment to the retinae. These projected images change continuously–with body movements of the viewer and changes in the physical world. A central question in vision research is how, from such continually changing images, humans robustly recognize the distal object as an entity with stable properties. Despite considerable research, a robust computational mechanism eludes us and we lack a satisfying explanation of how the brain accomplishes object recognition (Pinto, Cox, & DiCarlo, 2008).

Infants' learning of object names also begins with continually changing 2-dimensional retinal images. To learn an object name and generalize the label to varying visual instances of the same object perceived later, infants must start with these variable retinal images, processing them to recognize the distal object first before linking it with a heard word. However, object recognition at the image level has received little attention in the study of early word learning. Instead, researchers have skipped over the problem of variable images, and focused on the mapping problem at a more macro level—how infants map the perceived distal objects to heard words. Here we provide the first evidence that image-level variability matters to early word learning and may be the link that connects infant object manipulation to vocabulary development.

1.1 | Variability can facilitate learning

Most theories of object name learning focus on infants' detection of constancies in mappings between words and objects (e.g., Gleitman & Trueswell, 2018; Smith, Suanda, & Yu, 2014). From this perspective, it might seem reasonable that images with little variability would be good for detecting word-object mappings. This fits with the idea that learning systems may benefit by "starting small"—having access to fewer data early on (Elman, 1993; Nagai, Asada, & Hosoda, 2006). However, if image variability is required

to learn to perceive the same distal object across various viewing conditions (Cadieu & Olshausen, 2008; Földiák, 1991), then learning systems whose visual experiences sample a broad range of possible projected images from the object may acquire more robust object recognition and a stronger visual basis for object name learning (Bambach, Crandall, Smith, & Yu, 2016). This is the hypothesis we test.

This hypothesis is motivated in part by the well-known link between object manipulation and early language learning (e.g., James, Jones, Smith, & Swain, 2014; LeBarton & Iverson, 2016). Object image variation projected on the retinae is ubiquitous, due to the variety of possible poses and positions of objects relative to the viewer, and also to many contextual factors, such as lighting, occlusion and background information. But one kind of image variation, produced by infants' manual actions on objects—holding, moving, stacking—may play a special role in visual object perception. Past research shows that the amount an infant manipulates predicts their visual object recognition (James et al., 2014; Ruff, 1984; Soska, Adolph, & Johnson, 2010) and object name learning (James et al., 2014). Different hypotheses for these relations have been offered; object manipulation may facilitate: selecting and sustaining attention on objects (Rakison & Krogh, 2012; Yu, Smith, Shen, Pereira, & Smith, 2009), deeper encoding of objects (Wilcox, Woods, Chapa, & McCurry, 2007), and integration of multiple object views (James et al., 2014). All of these macro-level explanations may be correct, but all begin with self-generated image variability. Hand actions on objects produce changes in the 2-dimensional images on the retina. We test the specific hypothesis that the *amount of object image variability* generated by the infant's object manipulation varies across individuals and is an important predictor of word learning. That is, it may not be any variability in object images that matters, but rather *self-generated* image variability that is a key predictor of infants' later object name learning.

1.2 | Approach

The proximal image of an in-view object varies in many ways: contrast, visual size (with proximity to the viewer), color, as well as image information about shape that changes with object rotations, occlusions, and head and eye movements. In this first study of object image variability and word learning, we focused on this shape-related variability. Infants wore a head-mounted eye tracker as they played with objects. We used a single algorithmic measure, mask orientation (MO), to capture the frame-by-frame image variability of objects on which infants fixated their gaze: MO is the orientation of the most elongated axis of the *object pixels* from the head-camera image as an approximation of the visual information attended by infants. Critically, this is *not* a measure of the object's real-world orientation, nor does it relate in any direct way to the intrinsic shape properties of the distal object. Figure 1a illustrates 2-dimensional images captured by a head-camera worn by an infant. As illustrated in Figure 1b, MO is the orientation of the most elongated axis of whatever object pixels were in view, and will vary with direction of viewing and partial occlusion. Thus, MO offers an objective way to measure moment-to-moment variability of visual projections of the same object.

Participants were 15 months old, an age of expansive growth in object manipulation skills (Lockman, 2000). We measured object

RESEARCH HIGHLIGHTS

- Individual differences in the variability of visual instances of objects were objectively measured from infants' first-person views using head-mounted eye tracking during naturalistic toy play.
- Infants who generated more variable visual object images through manual object manipulation at 15 months experienced greater vocabulary growth over the next 6 months.
- Object image variability generated by parent manipulation and other characteristics of the play session were unrelated to infant vocabulary growth.
- This is the first evidence that image-level object variability matters and may be the link that connects infant object manipulation to vocabulary development.

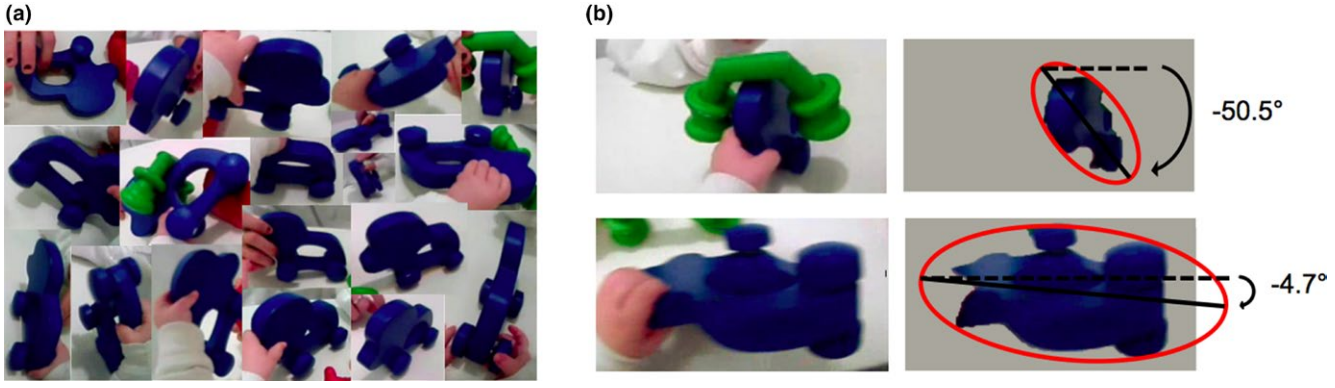

FIGURE 1 Object images and mask orientation coding. (a) Sample object images captured by an infant's head-mounted camera while the infant looked at one of the objects. (b) Left: sample cropped scene images; right: masks of the object pixels illustrating mask orientation (MO) coded in degrees

images generated by infants during free-flowing toy play with parents, rather than play alone (cf. James et al., 2014; Pereira, James, Jones, & Smith, 2010), for two reasons. First, this is the natural context of object manipulation for infants at this age. Second, this context is likely to elicit attention to objects that do not derive solely from the infant's own manipulations, enabling us to compare image variability when the infant holds and manipulates objects with image variability during other moments in which the infant visually fixated those same objects. We measured infants' vocabulary at 15 months and 6 months after the play session, during the period of expansive growth in object name vocabulary (Goldfield & Reznick, 1990). The rationale for the experimental and analytic approach is this: If individual infant's propensity to manually generate many different visual images of individual objects during free play creates the real-time visual information for learning about object shape and thus object names—and if this toy-play session in the laboratory adequately assesses individual differences in this propensity, then infants who generate images with more variable MOs at 15 months of age should have larger object name vocabularies at 21 months. Further, if the key ingredient for visual learning is variability created by the infant's manipulation behavior, then the MO variability created during this play task by factors *other than* the child's manual behavior should *not* be predictive.

2 | METHODS

2.1 | Participants

Participants were part of a larger ongoing project to understand the real-time processes supporting early object recognition and word learning during toy play. Twenty-two infants (11 females) contributed toy play data when they were 15 months old (range 14.9 to 15.9 months) and parents reported on infant vocabulary when the infant was both 15 months old and 21 months old (range 20.6 to 21.8 months). Twenty-two infants were determined to be an appropriate sample size based on prior research using temporally dense sensory-motor measures (e.g., Kretch & Adolph, 2015; Smith, Yu, & Pereira, 2011; Yu & Smith, 2012; Yu, Suanda, & Smith, 2018). More specifically, each infant contributed on average 6,714 frames (*SD* = 1762) of gaze data directed to an object—each of which was coded for MO, percentage of object pixels, and manual manipulation—and two vocabulary measures. In brief, each infant contributed on average 20,142 data points. Three additional infants participated at 15 months of age but were excluded from the final sample due to incorrect positioning of the head camera (n = 1) or productive vocabulary below the 5th percentile (n = 2; Fenson et al., 1994). Families were recruited from a working and middle-class population of a Midwestern college town. Infants were exposed to only English at home. Participants were treated in accordance with University IRB #0906000439, and all families gave their informed consent prior to their inclusion in the study.

2.2 | Stimuli

Six unique novel objects (on average, about 288 cm^3; Figure 2a) were custom made from clay, wood, and plastic. The objects were organized in two sets of three. Within each set, one object was painted blue, one red, and one green.

2.3 | Experimental setup

Parents and infants sat across from each other at a small table (61 cm × 91 cm × 64 cm). The table, walls, and floor were white and participants wore white smocks leaving the toys, hands, and faces as the only nonwhite objects in the images (this supported computer object recognition, see below). Infants wore a head-mounted eye tracker (Positive Science, LLC) that included an infrared eye camera—mounted on the head and pointed to the right eye of the participant—that recorded eye images, and a scene camera (Figure 2b) that captured 90° of the infant's first-person visual field (less than their full visual field, but sufficient to capture their direction of gaze during the majority of the study). The eye-tracking system recorded both the egocentric-view video and gaze direction (x and y) in that view, with a sampling rate of 30 Hz.

2.4 | Procedure

Both participants put on white smocks. The child was then seated and distracted with a push-button pop-up toy while a second experimenter (from behind) placed the eye-tracking gear low on the infant's forehead. One experimenter then directed the child to push a button on a pop-up toy while the second experimenter adjusted the camera such that the button being pushed by the child was near to the center of the head camera image. To collect calibration points

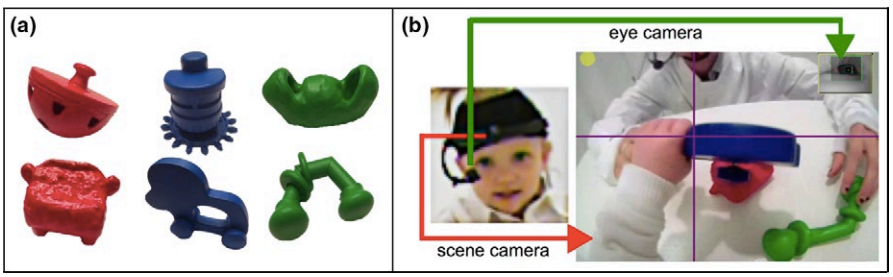

FIGURE 2 Experimental stimuli and eye-tacking paradigm. (a) Two sets (by row) of novel objects played with during the study. (b) Example infant scene camera frame overlaid with crosshairs indicating infant point of gaze

for eye tracking, the first experimenter then directed the infant's attention toward the beam of a laser pointer that the experimenter directed to various locations on the tabletop to ensure a sufficient number of calibration points, which were used after the session to calibrate eye-gaze within the head camera images. Parents were told that the goal of the study was simply to observe how their child played with toys and that they should try to interact as naturally as possible. The experimenters then left the room and the play session began. Each of the two toy sets was played with twice, in alternation, for about 1.5 min at a time, resulting in approximately 6 min of free-play data from each dyad. At both the 15-month session as well as when the infant was 21 months of age, the parent filled out the productive vocabulary sections of a MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al., 1994), a standardized checklist of an English-learning infant's early productive vocabulary.

2.5 | Dependent measures

2.5.1 | Infant gaze

Three regions-of-interest (ROIs) were defined, one for each toy, in each play trial. These ROIs were coded by highly trained coders who code this variable for many different experiments and who were naïve with respect to the research questions. Each ROI was coded manually, frame by frame, by watching the infant-view video with crosshairs indicating gaze direction (Figure 2b) and annotating when the crosshairs indicating infant gaze overlapped any portion of an object and which object (Slone et al., 2018). A second coder independently coded a randomly selected 10% of the frames in the corpus, with the inter-coder reliability ranging from 82% to 95% (Cohen's kappa = 0.81).

2.5.2 | Object image variability

For each frame of the infant-view video in which an object ROI was coded, a mask of the gazed object was defined via a machine vision program (see Yu & Smith, 2012, Appendix A). Each mask was fitted with an ellipse that had the same normalized second central moments as the mask, and the orientation of that ellipse (MO) was specified in terms of the angle between a horizontal axis and the major axis of the ellipse (Figure 1b); MO ranged from −90° to 90°. To quantify the variation of the object images each infant observed during play, we calculated Shannon entropy (H) of each infant's MO histogram across frames in which the infant manipulated the gazed object (H_{manip}) and also across frames in which the infant did not manipulate the gazed object ($H_{no\ manip}$) (see the section on "Quantifying Variability" below). The results reported here are based on binning MOs as follows: the interval of −90° to +90° was divided into 12 equal bins of 15°, centered at −82.5° to +82.5° in increments of 15°, as shown in Figures 3 and 4; however, variability metrics were highly consistent across the different bin sizes we explored (9 bins of 20°, 12 bins of 15°, 18 bins of 10°).

2.5.3 | Object manipulation

Manual object manipulation was defined as any hand contact with an object, and may or may not have included physical movement of the object. Manipulation was coded by a set of highly trained coders who code this variable for many different experiments and who were naïve with respect to the research questions. Manipulation (who and which object) was coded manually, frame by frame, from the scene camera images and from two high-resolution (recording rate 30 Hz) third-person-view cameras. We developed a custom-coding program that allowed coders to access these three views simultaneously to determine which object was manually contacted frame by frame. In practice, coders most often rely on the view of the scene camera, but in case of uncertainty, they would consult with the other two views to make a decision. A second coder independently coded a randomly selected 10% of the frames in the corpus, with the inter-coder reliability ranging from 91% to 100% (Cohen's kappa = 0.94).

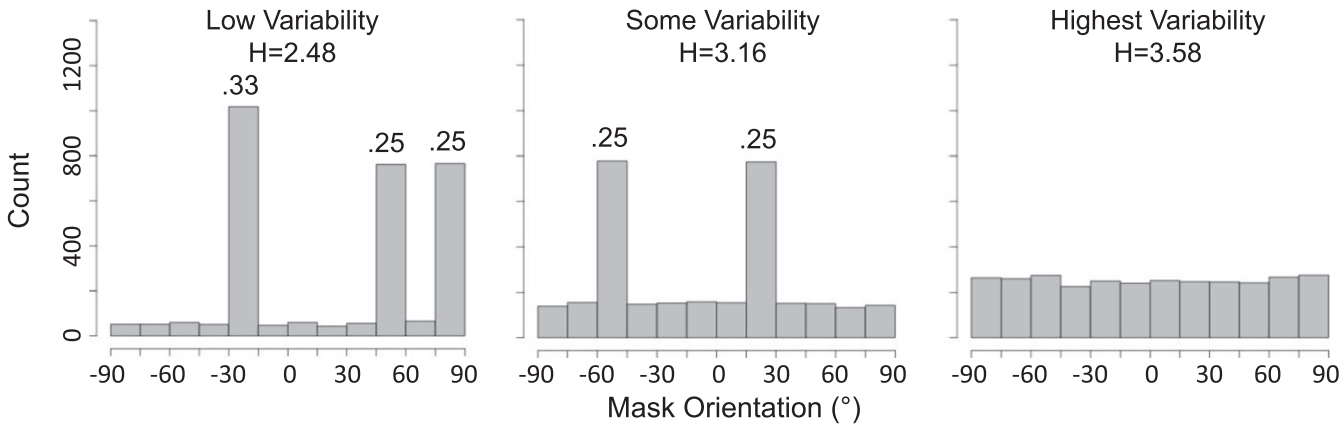

FIGURE 3 Simulated mask orientation (MO) histograms and associated Shannon entropy (H) metrics. Histograms depict 3,040 frames (the mean number of frames that an infant manipulated an object) with (left) 83% or (middle) 50% of the data falling into a small number of bins, and the rest of the data or (right) all of the data distributed randomly across the remaining bins; a number directly above a bin indicates the proportion of data that fell into that bin

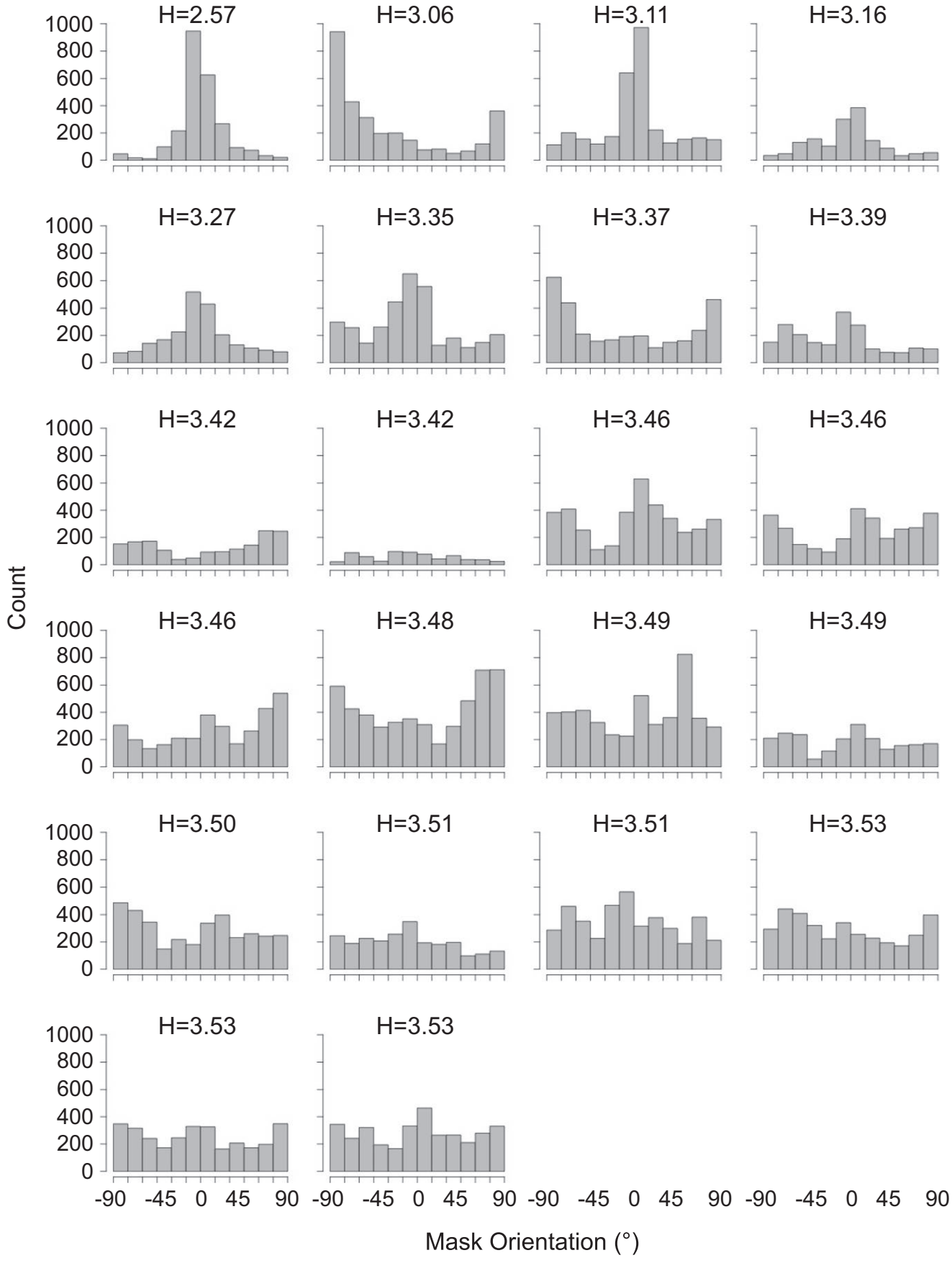

FIGURE 4 Mask orientation (MO) histograms and associated Shannon entropy (H) metrics for frames containing infant manipulation of the gazed object, for all 22 subjects

2.5.4 | Infant vocabulary

Parents completed the productive vocabulary sections of two MCDIs, standardized checklists of early-learned words that are predominantly composed of object names (Fenson et al., 1994). When infants were 15 months of age parents completed the vocabulary section of the Words and Gestures (Infant form) MCDI, and when infants were 21 months of age parents completed the vocabulary section of the Words and Sentences (Toddler form) MCDI. The Toddler MCDI contains a checklist of 680 words (all 396 of the words on the Infant MCDI, as well as 284 later-learned words), including 312 nouns, 232 of which are considered count nouns (Samuelson & Smith, 1999)—nouns that typically label solid, countable objects. While total vocabulary size estimates an infant's overall productive vocabulary ability, count noun vocabulary size specifically estimates an infant's word-object mapping ability. Infants were given scores for total vocabulary size and count noun vocabulary size at both 15 and 21 months of age based on the number of words in the corresponding sections of the MCDI that their parent reported the infant produced. Difference scores for both total and count noun vocabulary sizes (21 month vocabulary minus 15 month vocabulary) were calculated to estimate vocabulary growth.

3 | RESULTS

The data reported in this paper have been archived with Databrary and can be accessed via the following link: https://nyu.databrary.org/volume/705. Table 1 provides descriptive statistics concerning the amount of time infants spent looking at and manipulating objects during the play session at 15 months. Variability in the object images observed during play was assessed separately for frames containing (46% of analyzed frames) and not containing (54% of analyzed frames) infant manipulation of the gazed object.

TABLE 1 Means, standard deviations (*SD*), and ranges of object looking and manipulation measured at 15 months, and vocabulary measured at 15 and 21 months. Numbers in parentheses in measures 2–4 refer to aforementioned measure numbers

Measure	Mean	*SD*	Range
Behavioral measure			
1. Total time (mins) infant looked at objects in 6 min play	3.7	1.0	1.1–5.0
2. Proportion of (1) that infant manipulated the gazed object	0.46	0.11	0.27–0.68
3. Proportion of (1) that infant did *not* manipulate the gazed object	0.54	0.11	0.32–0.73
4. Proportion of (3) that the *parent* manipulated the gazed object	0.50	0.13	0.24–0.78
Vocabulary measure			
5. Total vocabulary size at 15 months	35	29	4–113
6. Count noun vocabulary size at 15 months	14	16	0–67
7. Total vocabulary size at 21 months	234	127	46–419
8. Count noun vocabulary size at 21 months	97	50	19–173
9. Total vocabulary growth	199	108	42–371
10. Count noun vocabulary growth	83	41	19–166

3.1 | Quantifying variability

The maximum variability of image properties as measured by the orientation of the most elongated axis of the image or MO would be indicated by a uniform distribution in which all of the possible orientations occurred equally often. The minimal variation possible would be a distribution in which only one MO occurred. Beginning here, Figure 3 illustrates three different simulated distributions that fall between these extremes. The Low Variability histogram illustrates a simulation in which 83.3% of the data fell into only three of the 12 bins, with the rest of the data distributed randomly across the remaining nine bins. The Some Variability histogram illustrates a simulation in which 50% of the data fell into two of the 12 bins, with the rest of the data distributed randomly across the remaining 10 bins. The Highest Variability histogram illustrates a simulation in which 100% of the data was distributed randomly across the 12 bins. Figure 4 provides histograms of each infant's real MO distribution across frames in which the infant manipulated the gazed object (see Figure S1 for a similar figure containing infants' MO distributions across frames in which the infant did *not* manipulate the gazed object).

One well-used and mathematically well-understood measure of the variability or informativeness in distributions such as those in Figures 3 and 4 is Shannon entropy (H) (Shannon, 1948). H measures the information value of any data point in the distribution. Each image in a more uniform distribution of MO provides more unique information about the object than does each image from a lower variability distribution. Generally, more uniform (i.e., flatter) distributions have higher H values and more peaked distributions have lower H values. (An alternative approach to measuring the variability would be to use the standard deviation of variance of the measured MOs for each infant. This measure is not appropriate in the present case because the MO distribution need not have a central tendency.)

Comparing the histograms and H values of the simulated (Figure 3) and real (Figure 4) MO histograms suggests a relatively high degree of variability in the object images observed by individual infants. As is evident from Figure 4, there was also considerable variability across infants in distributions of observed visual images. The average variation for infants' MO distributions for frames containing object manipulation (hereafter H_{manip}) fell between that of the *Some* and *Highest Variability* distributions ($M = 3.37$, $SD = 0.23$), as did the average variation for frames without infant object manipulation (hereafter $H_{no\ manip}$) ($M = 3.40$, $SD = 0.13$). The average variability did not differ when infants did or did not create that variability through their own manipulations: $t(21) = 1.16$, $p = 0.26$; this empirical fact strengthens the test of the hypothesis that it is the infant's propensity for object manipulations generating different object images that is predictive of vocabulary learning.

TABLE 2 Coefficient estimates (B, β), adjusted R^2, and Bayesian information criterion (BIC) from five regression models predicting count noun vocabulary growth from visual object entropy (H), object looking and manipulation times, and visual object size measures

Model	Predictor	*B* [95% CI]	*β*	Adjusted R^2	BIC
Child manipulation					
1	Child manipulation only			0.17*	228.4
	H_{manip}	83 [13, 153]	0.46*		
2	Child no manipulation only			−0.02	232.9
	$H_{no\ manip}$	57 [−81, 196]	0.18		
Parent manipulation					
3	Parent manipulation only			−0.08	236.3
	$H_{parent\ manip}$	41 [−103, 185]	0.13		
	Time (mins) parent manipulated gazed objects	7 [−53, 38]	−0.07		
Child manipulation versus other factors					
4	Other characteristics only			−0.10	238.5
	Total time (mins) infant looked at objects	−8 [−35, 19]	−0.18		
	Time (mins) infant manipulated gazed objects	−4 [−49, 40]	−0.06		
	Mask size during infant manipulation (% scene)	−4 [−28, 19]	−0.09		
5	Other characteristics versus child manipulation			0.20*	233.2
	H_{manip}	104 [32 ,176]	0.58*		
	Total time (mins) infant looked at objects	−11 [−237, 215]	−0.27		
	Time (mins) infant manipulated gazed objects	−6 [−44, 32]	−0.09		
	Mask size during infant manipulation (% scene)	−12 [−33, 8]	−0.25		

*$p < 0.05$.

3.2 | Predicting word learning

If it is the variability in images generated by children's *own actions* that supports object name learning, then H_{manip} but not $H_{no\ manip}$ should predict later vocabulary growth. In Models 1–2 (Table 2), we regressed count noun vocabulary growth between 15 months and 21 months on H_{manip} and on $H_{no\ manip}$, respectively.[1] H_{manip} was a significant positive predictor of count noun vocabulary growth (Model 1), whereas $H_{no\ manip}$ was not (Model 2). Infants who observed more self-generated object image variability exhibited greater count noun vocabulary growth over the next 6 months.

In Models 3–5 (Table 2), we considered alternative hypotheses to incorporate other plausible factors and thus further support the main findings shown in Models 1–2. Model 3 tested the hypothesis that parent manipulation matters. That is, although MO entropy across frames without infant object manipulation did not relate to infant vocabulary, parents were in manual contact with the gazed object in approximately half of these frames (Table 1) and it is possible that the amount of parent manipulation or the variation generated by parent object manipulation matters for vocabulary growth. Neither of these factors, however, was a significant predictor of count noun vocabulary growth (Model 3).

Models 4 and 5 tested the hypothesis that characteristics of the play session other than infants' MO variability—the total time that infants looked at objects (not manipulation per se), or the total time that they looked at manipulated objects—predicted infants' vocabulary growth. Because our specific prediction is that the relevant image variability is about shape, we also included an additional image measure in Model 4, the median visual size of the mask (as a percentage of the pixels in the entire scene) during infant object manipulation. Mask size varies principally with the distance of the object from the viewer. None of these factors, however, predicted vocabulary growth (Model 4), whereas H_{manip} was a significant predictor of vocabulary growth even with these other factors in the model (Model 5), and significantly increased model fit compared to a model with these other factors alone (Model 4), $\chi^2 = 8.34$, $p = 0.010$. Infants who observed more self-generated object image variability exhibited greater count noun vocabulary growth, even controlling for the amounts of time infants looked at and manipulated objects and for objects' visual sizes. Moreover, comparison of Bayesian information criterion (BIC) values for each of the models suggests that a model including H_{manip} alone (Model 1) provided the best model fit (lowest BIC) (Raftery, 1995). In total, these results provide compelling support for the idea that self-generated variability in the shape properties of the visual image supports the processes important to early word learning.

3.3 | Image variability of individual objects

To ensure sufficient data for measuring individual differences in self-generated variability, the entropy measures in the main analyses were calculated across all objects without taking into account the potentially different ways that *individual* objects may have been positioned and manipulated. Infant's propensity to generate variable images when handling objects in this task is assumed to be a measure of their likelihood in general. Under this assumption, the present measure across all objects seems likely to provide a reasonable index of this propensity. However, the hypothesized mechanism through which these self-generated variable images are presumed to work is by providing the perceiver with multiple views of a single object, the training data for recognizing an object under different viewing conditions. It is possible, in the present study, that different objects' affordances (Figure 2a) could lead an individual object's MO distribution to look very different from the MO distribution of all objects combined, which would limit the generalizability of the present measure. Therefore, we calculated entropy of MOs for each object, per subject, separately for frames in which the infant did and did not manipulate the gazed object. On average, infants contributed enough data to calculate entropy of MOs from frames containing infant object manipulation for 4.4 of the 6 objects (total corpus across all 22 infants = 97 objects), and from frames not containing infant object manipulation for 4.7 of the 6 objects (total corpus across all 22 infants = 104 objects). For each infant we calculated (a) mean per-object MO entropy, and (b) the proportion of objects with MO entropy above the corpus median, both for frames containing and not containing infant object manipulation. Infants' mean per-object MO entropy as well as the proportion of objects with MO entropy above the median were highly correlated with the MO entropy of all objects combined that were the basis for the main analyses, both for frames in which the infant manipulated the gazed object ($rs > 0.64$, $p < 0.01$) and for frames in which the infant did not manipulate the gazed object ($rs > 0.50$, $p < 0.05$). These findings suggest that the degree of observed variability in MOs was fairly consistent across objects within individual participants, such that infants who experienced greater self-generated per-object variability (i.e., higher mean per-object entropy as well as infants who observed a greater proportion of objects with MO entropy above the median) also tended to experience greater self-generated variability in the overall MO distribution (across all objects combined).

4 | DISCUSSION

Objects names are a major component of early vocabularies (Fenson et al., 1994) and learning object names depends on *visually* recognizing objects in the world. However, the visual information and visual problems young learners must solve have rarely been studied by early language researchers. Although category learning—the variety of things that are dogs or cars, for example—has been studied at a conceptual level (Bloom & Markson, 1998; Gelman & Meyer, 2011), the fundamental visual question of moment-to-moment variations in visual appearance presented by objects has not. Rather, many laboratory studies of infant word learning leap over the step of building an object representation that works over multiple viewing contexts, instead presenting a single view of an object to be linked with a word (e.g., Houston-Price, Plunkett, & Harris, 2005; Smith & Yu, 2008).

Yet early learning of object names is in part a visual problem, as young learners must find and recognize individual objects given the highly variable nature of the proximal stimulus, the 2-dimensional image received by the eye. The present results make clear that the visual information presented by a single object to the viewer's sensors is highly variable. They also show that this variability is not a negative factor for learning, but a positive one: the amount of self-generated variability in an infant's own view positively predicts later vocabulary growth. Why should this be the case? The answer, we propose, is that an essential component of the data for visual learning about objects is this self-generated variability. Previous research has shown that infant object manipulation skills are developmentally linked to changes in visual object recognition (Soska et al., 2010), that experimental conditions that foster different object manipulation experiences change shape perception and object categorization (Smith, 2005; see also James et al., 2014), and that poor object manipulation skills predict later deficits in language learning (Zuccarini et al., 2017). Other studies have shown that developmental changes in visual object recognition are correlated with noun vocabulary size (Jones & Smith, 2005; Smith, 2003) and predict later changes in object name learning (Yee, Jones, & Smith, 2012). This is the first

demonstration that self-generated variability in the proximal 2-dimensional image is related to vocabulary development, a potential mechanistic pathway from manipulation to object name learning.

The primary measure of visual object variability used here—entropy of a MO distribution—is measured directly from object image pixels and thus unencumbered by assumptions about the nature of varying object views or the specific forms of visual information that may be critical to object recognition. We see this as a strength. The young visual learner begins with the image. For example, prior research has shown that toddlers are biased to generate planar object views and the planar bias is related to better visual object recognition (James et al., 2014). But planar views are defined by the relation of the viewer to the distal object (an elongated axis perpendicular or parallel to the line of sight) and not by the image information. The visual system cannot know that a view is planar in these terms without knowing the overall shape of the object in the world; "planar" is not an image property but a higher-level descriptor of a whole object property that must, somehow, be recovered from the image and variations in the image as the relation between the object and the viewer change in time. By hypothesis, manual play with objects is critical to the development of visual object recognition because it generates this variation from which the visual system can discover the higher-level properties of distal objects and their views. Nevertheless, variability in the proximal image was not redundant with amount of object manipulation, but rather was created through object manipulation combined with other factors including gaze, head movements relative to the hands, and occlusion of objects by hands and other objects. The present results show that variability in the proximal image of an object is considerable in a single session and that the differing propensities of individual children to self-generate that variability predicts later vocabulary growth beyond sheer amount of object manipulation. Future work elucidating the visual statistics of the views infants experience—and what lessons they may teach the visual system about 3-dimensional objects—constitutes a crucial missing link in current understanding. The present study provides a first step to filling that gap in current knowledge.

Why does the image variability created by the child matter but not the image variability of the visually attended object at other moments? The developing infant and their behavioral tendencies are likely the dominant factor in creating the infant's own visual experiences—because after all, the infant's eyes are connected to their own body and are coordinated with their hand actions in goal directed action. Thus, it could be that what matters for predicting later vocabulary development is the day-in and day-out experience of different object views, however those views are generated. The infant's own activity in this experiment—not the parents' activity with the objects—may simply be the best predictor of that day-in day-out variability in everyday life that is the real-world training ground for the visual system. However, it is also possible that self-generated object manipulation is essential. Previous research demonstrates that placing hands near an object can enhance visual perception, processing, and memory of objects (Brockmole, Davoli, Abrams, & Witt, 2013), which in turn likely strengthen object representations. Prediction learning (Apps & Tsakiris, 2014; Lowe, 1999) from one's own actions to changes in visual information may also be an important part of the process. Another possibility concerns the mechanism that enables learners to link one momentary image of an object to another different image of that same object. Holding an object may provide key information to the young perceiver that a set of varying views is all of the same object.

This study is correlational in nature, predicting from behavior at one time point to a later outcome. Thus, it is possible that more "advanced" babies come to more informatively manipulate objects to create more variability in objects' images and then also through some independent pathway build larger vocabularies. This typology, however, leaves unanswered the considerable literature showing that giving infants and children the chance to manipulate objects leads to better visual object memory and visual discrimination (Bushnell & Baxt, 1999; Needham, 2000; Ruff, 1982; Wilcox et al., 2007). The importance of a deeper understanding of the visual information—and the role of infant manipulations in generating that information—extends both to fundamental questions about visual object recognition and to understanding early object name learning. There are many empirical indicators that these are causally related developments including developmental changes in visual object recognition just prior to and during the period of explosive acquisition of object names (James et al., 2014; Smith, 2003, 2009). There are also increasing indicators that children who are slow to grow their vocabularies show delays or disruptions in visual processes (Behrmann, Thomas, & Humphreys, 2006; Collisson, Grela, Spaulding, Rueckl, & Magnuson, 2015; Jones & Smith, 2005). Finally, there is considerable evidence that early disruptions of object manipulation are diagnostic markers of later learning problems (Provost, Lopez, & Heimerl, 2007). The specific contribution of the present study is that it provides a direct link from object manipulation to the variability of the visual information at the level of 2-dimensional images to later object name learning. This is a key step to understanding how that image variability supports visual learning and object name learning.

Most achievements in human development are multicausal, with many contributing factors. Vocabulary growth is predicted by a host of factors in addition to object image variability, including factors related to the infants' behavior and development—habituation (Tamis-LeMonda & Bornstein, 1989), sustained attention (Yu et al., 2018), walking (Walle & Campos, 2014)—and factors related to parent behavior—language input (Cartmill et al., 2013), parental responsiveness (Tamis-LeMonda, Bornstein, Baumwell, & Melstein Damast, 1996)—among others. A particular focus of past word-learning research has been the role of language input and the information available for word learning in the immediate context of heard words. However, the present study and recent research by Clerkin and colleagues motivates the importance of studying infants' visual experiences of objects in its own right. Clerkin, Hart, Rehg, Yu, and Smith (2017) outfitted infants with head-mounted cameras while in their own homes and demonstrated, perhaps unsurprisingly, that the most frequent objects in infants' visual fields were those whose noun labels infants typically learn earliest in life (i.e., before

16 months of age). More surprising is their finding that infants very rarely heard the labels for these nouns while they were in view (only 3% of episodes when a visual object was present included the label for that object) (Clerkin & Smith, in preparation), suggesting that the visual presence of objects may contribute to object name learning in general and independently of object naming, setting up the early learning system for when it later hears object labels. The take-away is not that language input does not matter, but rather that the everyday visual information infants observe about objects may be related to progress in early word learning even apart from simultaneous object naming.

The present study examined how and why visual object information might matter for object name learning. We argue that visual variability likely contributes to more generalizable object representations, such that objects are better recognized and therefore better mapped to their word labels when those labels are heard. This possibility is supported by neural network modeling work by Bambach et al. (2016) and Bambach, Crandall, Smith, & Yu (2018). These researchers demonstrated that first-person view images of objects from naturalistic toy play can be used to train CNN-based (convolutional neural network-based) object models that generalize to recognizing new object instances unseen in model training. Specifically, the researchers found that toddlers' object views were more visually diverse than those of their parents, and that the networks trained on a set of diverse toddler-generated object views consistently recognized objects better than the networks trained on a set of parent-generated object views (Bambach et al., 2016) and the networks trained on a set of less diverse toddler-generated object views (Bambach et al., 2018). This modeling work supports the possibility that diverse views of objects may be critical to the development of generalizable object representations. The present work provides valuable insights into the type of view diversity that may be most important for developing generalizable object representations: self-generated shape-related variability in object images.

4.1 | Next steps

The present study measured self-generated variability in object images during a 6-min toy-play session in a laboratory. The fact that this variability predicted vocabulary growth over the next six months suggests that the way infants manipulate objects to explore and create variable views of objects in their everyday lives may be a tendency stable enough to show up in a short laboratory session, and opens exciting new directions for more precise questions about why and how the proximal visual stimulus relates to learning about objects. Are infants actively creating the object views they want to see or are these views bi-products of the types of object manipulations infants engage in? Is the relation between object image variability and vocabulary growth permissive, with visual variability allowing for better visual object representations in general? Or does visual variability lead directly to better representations of specific objects, facilitating word learning when those objects' names are heard? To answer these questions, we can create and compare experimental conditions in which infants either actively create visual variability or are passively exposed to the same amount of variability. Moreover, we can directly measure infants' knowledge and representations of visual objects after toy play, as well as their learning of the labels of those objects, and directly link visual variability created during object play with various kinds of learning outcomes for those objects. More generally, future studies need to rely on both experimental and observational/correlational approaches to examine the pathway from self-generated object manipulation, to visual object variability and to vocabulary growth. An important goal of language acquisition research is to continue to elucidate how the many factors shown to predict word learning work together in the moment and in developmental time.

ACKNOWLEDGMENTS

This research was funded by NIH R01HD28675, R01HD074601, and T32HD007475-22, NSF BCS1523982 and BCS1730146, and by Indiana University through the Emerging Area Research Initiative—Learning: Brains, Machines, and Children. The authors appreciate the participation of the infant and parent volunteers and the assistance of Computational Cognition and Learning Laboratory members, especially Sven Bambach, in data collection and processing.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ENDNOTE

[1]Entropy measures were investigated in separate regression models to avoid multicollinearity among these predictors (H_{manip} and $H_{no\ manip}$: $r = 0.81$; H_{manip} and $H_{parent\ manip}$: $r = 0.79$). All analyses reported in this paper were conducted with count noun vocabulary growth as the dependent measure, however, count noun vocabulary growth and total vocabulary growth were highly correlated ($r = 0.98$) such that the same patterns obtain when total vocabulary growth is modeled (see Table S1).

ORCID

Lauren K. Slone https://orcid.org/0000-0003-0068-5866

REFERENCES

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2016). Active viewing in toddlers facilitates visual object learning: An egocentric vision approach. In D. Grodner, D. Mirman, A. Papafragou & J. Trueswell (Eds.)

Proceedings of the 38th annual conference of the cognitive science society (pp. 1631–1636). Austin, TX: Cognitive Science Society.

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences*, *10*, 258–264. https://doi.org/10.1016/j.tics.2006.05.001

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*, *2*, 67–73. https://doi.org/10.1016/S1364-6613(98)01121-8

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2018). Toddler-Inspired Visual Object Learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett (Eds.), *Paper presented at Neural Information Processing Systems (NIPS)* (pp. 1209–1218).

Brockmole, J. R., Davoli, C. C., Abrams, R. A., & Witt, J. K. (2013). The world within reach: Effects of hand posture and tool use on visual cognition. *Current Directions in Psychological Science*, *22*, 38–44. https://doi.org/10.1177/0963721412465065

Bushnell, E. W., & Baxt, C. (1999). Children's haptic and cross-modal recognition with familiar and unfamiliar objects. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1867.

Cadieu, C. F., & Olshausen, B. A. (2008, December). Learning transformational invariants from natural movies. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Paper presented at Neural Information Processing Systems (NIPS)* (pp. 209-216). Retrieved from https://papers.nips.cc/paper/3378-learning-transformational-invariants-from-naturalmovies.pdf

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

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. https://doi.org/10.1098/rstb.2016.0055

Clerkin, E.M., & Smith, L.B. (in preparation). Word learning in context: Visual objects and their names in infants' daily lives.

Collisson, B. A., Grela, B., Spaulding, T., Rueckl, J. G., & Magnuson, J. S. (2015). Individual differences in the shape bias in preschool children with specific language impairment and typical language development: Theoretical and clinical implications. *Developmental Science*, *18*, 373–388. https://doi.org/10.1111/desc.12219

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, i–185. https://doi.org/10.2307/1166093

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, *3*, 194–200. https://doi.org/10.1162/neco.1991.3.2.194

Gelman, S. A., & Meyer, M. (2011). Child categorization. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 95–105. https://doi.org/10.1002/wcs.96

Gleitman, L. R., & Trueswell, J. C. (2018). Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science*, https://doi.org/10.1111/tops.12352

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*, 171–183. https://doi.org/10.1017/S0305000900013167

Houston-Price, C., Plunkett, K., & Harris, P. (2005). 'Word-learning wizardry' at 1; 6. *Journal of Child Language*, *32*, 175–189. https://doi.org/10.1017/S0305000900400661

James, K. H., Jones, S. S., Smith, L. B., & Swain, S. N. (2014). Young children's self-generated object views and object recognition. *Journal of Cognition and Development*, *15*, 393–401. https://doi.org/10.1080/15248372.2012.749481

Jones, S. S., & Smith, L. B. (2005). Object name learning and object perception: A deficit in late talkers. *Journal of Child Language*, *32*, 223–240. https://doi.org/10.1017/S0305000090400664

Kretch, K. S., & Adolph, K. E. (2015). Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science*, *18*, 736–750. https://doi.org/10.1111/desc.12251

LeBarton, E. S., & Iverson, J. M. (2016). Associations between gross motor and communicative development in at-risk infants. *Infant Behavior and Development*, *44*, 59–67. https://doi.org/10.1016/j.infbeh.2016.05.003

Lockman, J. J. (2000). A perception–action perspective on tool use development. *Child Development*, *71*, 137–144. https://doi.org/10.1111/1467-8624.00127

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, *2*, 1150–1157. https://doi.org/10.1109/ICCV.1999.790410

Nagai, Y., Asada, M., & Hosoda, K. (2006). Learning for joint attention helped by functional development. *Advanced Robotics*, *20*, 1165–1181. https://doi.org/10.1163/156855306778522497

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*, 131–156. https://doi.org/10.1207/S15327647JCD010201

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated object views. *Journal of Vision*, *10*, 22–22. https://doi.org/10.1167/10.11.22

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology*, *4*, e27. https://doi.org/10.1371/journal.pcbi.0040027

Provost, B., Lopez, B. R., & Heimerl, S. (2007). A comparison of motor delays in young children: Autism spectrum disorder, developmental delay, and developmental concerns. *Journal of Autism and Developmental Disorders*, *37*, 321–328. https://doi.org/10.1007/s10803-006-0170-6

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163. https://doi.org/10.2307/271063. https://www.jstor.org/stable/271063

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*, 43–53. https://doi.org/10.1111/j.1467-7687.2011.01096.x

Ruff, H. A. (1982). Role of manipulation in infants' responses to invariant properties of objects. *Developmental Psychology*, *18*, 682–691. https://doi.org/10.1037/0012-1649.18.5.682

Ruff, H. A. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology*, *20*, 9–20. https://doi.org/10.1037/0012-1649.18.5.682

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*, 1–33. https://doi.org/10.1016/S0010-0277(99)00034-7

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C., Franchak, J. M., Pearcy, D., … Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *Journal of Visualized Experiments*, *141*, e58496. https://doi.org/10.3791/58496

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science*, *14*, 244–250. https://doi.org/10.1111/1467-9280.03439

Smith, L. B. (2005). Action alters shape categories. *Cognitive Science*, *29*, 665–679. https://doi.org/10.1207/s15516709cog0000_13

Smith, L. B. (2009). From fragments to geometric shape: Changes in visual object recognition between 18 and 24 months. *Current*

Directions in Psychological Science, *18*, 290–294. https://doi.org/10.1111/j.1467-8721.2009.01654.x

Smith, L. B., Suanda, S., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Science*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates 3D object completion. *Developmental Psychology*, *46*, 129–138. https://doi.org/10.1037/a0014618

Tamis-LeMonda, C. S., & Bornstein, M. H. (1989). Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. *Child Development*, *60*, 738–751. https://doi.org/10.2307/1130739

Tamis-LeMonda, C. S., Bornstein, M. H., Baumwell, L., & Melstein Damast, A. (1996). Responsive parenting in the second year: Specific influences on children's language and play. *Early Development and Parenting: an International Journal of Research and Practice*, *5*, 173–183. https://doi.org/10.1002/(SICI)1099-0917(199612)5:4<173:AID-EDP131>3.0.CO;2-V

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology*, *50*(2), 336. https://doi.org/10.1037/a0033238

Wilcox, T., Woods, R., Chapa, C., & McCurry, S. (2007). Multisensory exploration and object individuation in infancy. *Developmental Psychology*, *43*, 479–495. https://doi.org/10.1037/0012-1649.43.2.479

Yee, M. N., Jones, S. S., & Smith, L. B. (2012). Changes in visual object recognition precede the shape bias in early noun learning. *Frontiers in Psychology*, *3*, 533. https://doi.org/10.3389/fpsyg.2012.00533

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., Smith, L. B., Shen, H., Pereira, A. F., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *1*, 141–151. https://doi.org/10.1109/TAMD.2009.2031513

Yu, C., & Suanda, H. S., & Smith, L. B. (2018). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*, e12735. https://doi.org/10.1111/desc.12735

Zuccarini, M., Guarini, A., Savini, S., Iverson, J. M., Aureli, T., Alessandroni, R., … Sansavini, A. (2017). Object exploration in extremely preterm infants between 6 and 9 months and relation to cognitive and language development at 24 months. *Research in Developmental Disabilities*, *68*, 140–152. https://doi.org/10.1016/j.ridd.2017.06.002

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Slone LK, Smith LB, Yu C. Self-generated variability in object images predicts vocabulary growth. *Dev Sci*. 2019;e12816. https://doi.org/10.1111/desc.12816