Visual constancies amidst changes in handled objects for 5- to 24-month-old infants

Joseph M. Burling

Department of Psychology, University of California, Los Angeles

Hanako Yoshida

Department of Psychology, University of Houston

Author Note

Word Count: **3999**

Correspondence:
Hanako Yoshida
Department of Psychology
126 Heyne Building, University of Houston
Houston, TX 77204-5022 USA
Email: yoshida@uh.edu

Acknowledgments

This research was supported in part by the University of Houston (The Laurie T. Callicutt Scholarship, Provost's Undergraduate Research Scholarship, Summer Undergraduate Research Fellowship), and a National Institutes of Health grant (R01 HD058620) awarded to the last author, Hanako Yoshida. We especially want to extend our gratitude to the parents and children in our community who supported the research and participated in the study, the Cognitive Development Lab for collection of the data and coding data, Dr. Linda B. Smith for valuable feedback on a previous version of the manuscript and Dr. Merrill Hiscock for numerous advices on the revision.

Abstract

Manual skills slowly develop throughout infancy and have been shown to create clear views of objects that provide better support for visually sustained attention, recognition, memory, and learning. One possibility is that these clear views coincide with the development of manual skill. Another possibility is that social scaffolding supports clear viewing experiences similar to the ones created through active object exploration generated by toddlers. The current study used a head-mounted eye tracker to record 5- to 24-month-olds' object views during repeated mother-infant play sessions. Results show an early beginning of scaffolding in which parents generate views similar to those of older infants, resulting in increased fixations to objects. The finding implicates parents as early scaffolders of object attention and learning.

Keywords: scaffolding, visual attention, eye tracking

Visual constancies amidst changes in handled objects for 5- to 24-month-old infants

Introduction

Scaffolding refers to the ideal level of support that a child needs to learn a new skill (Vygotsky, 1978). A large body of work in the problem-solving domain has documented the use of parental strategies with preschool and school-aged children (Berk & Winsler, 1995; Wood, Bruner, & Ross, 1976), and with younger children in the social, cognitive, and language-development domains (Bates, 1979; Carpenter et al., 1998; Tomasello, 1995; Striano and Rochat, 2000; Striano & Stahl, 2005). Interpersonal interactions promote children's learning via scaffolding, which extends well into education and broader social contexts (Duncan & Tarulli, 2003; Scrimsher & Tudge, 2003; Winsler, 2003). To date, little is known about how scaffolding may impact cognitive foundations, such as early perceptual and attentional abilities, and how an *ideal level* of support is maintained with development. In the present study, we show early indicators of scaffolding in infancy. We show constancies in infant visual exploration and object attention induced by object handling between parent and infant from 5 to 24 months of age amid changes in infant maturation.

Two lines of research motivate the present study. The first outlines parental object showing and holding behavior specific to infants as early as 3 months (Libertus & Needham, 2011; Clark & Estigarribia, 2011; Zukow, 1990; Yoshida & Burling, 2013). Active, parental object presentation is shown to be *well organized*, scaffolding the development of attention, memory, and learning (Dent-Read & Zukow, 1997; Matatyaho & Gogate, 2008; Nomikou & Rohlfing, 2011; Zukow, 1990). Caregivers' holding and showing actions provide "structure" in a

dynamic environment, thereby establishing multimodal synchrony among objects, gestures, and labels.

The second line focuses on the perspective of the infant. Active object exploration organizes visual experiences in toddlers (Pereira, James, Jones, & Smith, 2010), and is thought to be central to learning that involves visual discrimination, recognition, and memory for objects (e.g., Johnson, 2010). However, an overview of recent *infant-point-of-view* studies measuring infants' visual field suggests that this assumption is only partially correct, at least for children older than 12 months. Although there are frequent ambiguous moments when visual information is a jumble of candidates competing for attention, these moments are punctuated by instances of clear viewing—a single object dominating the child's field-of-view (Pereira, Smith, & Yu, 2014; Yoshida & Smith, 2008; Yu & Smith, 2012, 2013). These moments of visual clarity occur principally when the child is able to hold the object close to their face and eyes, resulting in partial occlusion of other objects in the scene (Yu & Smith, 2012, see also de Barbaro, Johnson, Forster, & Deák, 2016; James, Jones, Smith, & Swain, 2014; Soska, Adolph, & Johnson, 2010 for influence of self-regulate body movements on visual experiences).

The question is whether in-depth visual exploration happens *before* infants become skilled at reaching, handling, and object manipulation. We ask if parental scaffolding—by holding and showing objects—works in a manner that mimics older children's self-generated viewing experiences. In particular, we address if parental holding behavior isolates objects in the infant's visual field so that they are retinotopically large enough to partially occlude competing information.

One may argue that parent-guided visual experiences are not functionally similar to the self-generated experiences seen later in development. Young infants visually explore objects

indirectly through a separate system (parent) that has no first-hand knowledge of the infant's immediate visual input. Furthermore, parents must adapt their behaviors in response to the rapid changes taking place in infancy, e.g., providing static views for young infants in supine positions and more dynamic views for children who are crawling and talking. Given these challenges to parent effectiveness, infants might be forced to get by on minimal parental support until they reach physical maturity.

Alternatively, we hypothesize that parental scaffolding throughout infancy can be as effective as the experiences created by and older child, e.g., showing objects that are large and isolated in their visual field. This does not mean that the parent is the only one who shapes the infant's initial visual experience. The dynamic and active role of the infant in shaping his or her own experiences (Piaget, 1954) has been shown across domains. The infant's vocalizing, smiling, and reaching affect their own experiences and the way the parent plays with the infant (Libertus & Needham, 2011; Bertenthal & Camps, 1990; Karasik, Tamis-LeMonda, & Adolph, 2011; Brand, Baldwin, & Ashburn, 2002; O'Neill, Bard, Linnell, & Fluck, 2005).

With the parental scaffolding hypothesis in mind, we explored how caregivers provide—and infants actively exploit—opportunities for in-depth visual exploration. We observed the developmental trajectory of parental scaffolding that led to constancies in infant visual experiences, before and after their ability to actively explore. Timing is critical during changes in infant object manipulation and, in the framework of parental scaffolding, timing is a key component in providing an *ideal level* of support (Vygotsky, 1978).

To test our hypothesis, we observed semi-natural object play within parent-infant dyads. We used a longitudinal approach to analyze the consistency of infant viewing experiences by observing the interactions between parents and their infants of 5 to 24 months. We used a head-

mounted eye-tracking device to measure the degree of object isolation from first-person-views FPV, Figure 1A). Our focus was directed at the moment objects were held and manipulated, and we compared the moment with information from infant FPV. Recording at least two sessions of the same infant at different ages allowed us to observe transitions in the viewing size as a function of the individual and their actions.

Method

Longitudinal data collection

Parents and their infants from middle-class families from the greater Houston Area were recruited to participate in the study. Parents read and provided informed consent regarding their participation in multiple sessions at least 3 months apart. A total of 10 dyads (7 female) completed at least two sessions (maximum 5 sessions, with an average of 2.7 visits per parent-infant dyad) and were used for model fitting (see Figure 2). Eight different dyads that completed only one session were used for model testing purposes (the markers along the bottom of Figure 2). The age range of infants was 4.7 to 23.6 months (M = 12.1 mo, SD = 5.1 mo). The ethnic backgrounds represented were: Caucasian, including Hispanic or Latino (55%); African American (25%); Asian (8%); and Others (12%). Since our goal was to understand trajectories within a developmental period—as opposed to changes within a specific sample of individuals—we report the group/dyad-level results and also compare growth models to a separate sample of infants of the same age range and demographics. Two additional dyads were omitted from the analysis due to incomplete data, leaving a total of 18 dyads. Both the parent and infant were later provided with a small gift in exchange for their participation.

Stimuli and materials

Testing environment. Participants were observed in a laboratory setting that consisted of a table, two chairs, eight toy objects in a container, and the video capturing devices (see Figure 1). The parent and infant sat next to each other at a 75cm × 50cm table, which was used as a natural surface for jointly interacting with the objects. A wall-mounted camera at a distance of 2.5m captured interactions from a third-person view (TPV: Figure 1 left), and the parent's voice was also recorded.

Eye tracking headgear was placed on the infant before the play session started. An experimenter placed a fitted cap lined with Velcro on the infant while another experimenter fixed the eye tracking headgear to the cap. The eye tracking headgear, which weighed 51 g, consisted of two small cameras and an infrared LED. One of the cameras faced the infant's right eye and recorded the eye movements using corneal reflection (Positive Science, http://www.positivescience.com; Franchak, Kretch, Soska, & Adolph, 2011). A second camera recorded the visual field from the infants' perspective (FPV: 54.4° horizontal by 42.2° vertical). For coding purposes, videos were later joined and synchronized (time locked at 30 frames-per-second) to show the third-person-view (TPV) and first-person-view (FPV) with eye tracking coordinates superimposed over the image (pink circle indicating the fixation as shown in Figure 1A, right).

Toy objects used. Eight infant-friendly naturalistic toy objects (Figure 1B) were stored in a container placed near the parent's chair. To stimulate and structure their play, parents were asked to demonstrate early-learned words (*open*, *bunny*, *car*, *bottle*, *cookie*, *eat*, *drink*, *put*) by using the appropriate objects as they played with their infant. These words and objects were

selected based on the earliest learned words from the McArthur Child Development Inventory (Fenson et al., 2000, see Table 1 for the respective age of acquisition of the selected words).

Procedure

Play sessions. Parents were instructed to interact with their infant as naturally as possible using the objects provided, and to demonstrate the meaning of the early-learned words. They were informed that they would hear an audio prompt played from a speaker every 40 seconds, which cued one of the eight words, and that they should use the prompts to pace their interactions. The intervals (40s × 8 trials of unique objects) allowed for a total playtime of 5min 20s. Guiding the parent's play in this manner allowed for an equal number of samples across dyads and sessions, and it ensured that any variability in number of frames available for coding was due to how coders marked the onset (first verbal prompt) and offset ("thank you for participating…") timestamps, and not from age-related differences in the behaviors of interest. All videos were near 9600 frames in total with a range of ± 60 frames. The wall-mounted cameras and head-mounted eye tracking system were adjusted after the instructions were given. Once experimenters left the room, parents and infants ignored the recording devices. Infants actively looked at objects and parent's face, and they smiled; parents naturally engaged in the play.

Correspondence between eye images FPV images was achieved using a manual calibration procedure utilizing a 60cm × 40cm board and displaying nine spatially distributed stickers. We calibrated gaze direction at the beginning and end of each session by pointing to each sticker using an LED glove to attract the infant's attention to that point. A minimum calibration correlation of 0.9 between FPV and eye position was obtained. When the calibration was completed, the experimenter started the recorder for all video capturing devices and left the

play area after a final inspection. The entire session took approximately ten minutes. Experimenters restarted the session from the beginning if the infant became fussy, or if he or she removed or shifted the eye tracking device. There were four instances of such a restart, all of which occurred during the first or second trial.

Video processing and annotation. Mapping eye position onto the FPV image coordinates was completed offline using the Yarbus software (Franchak et al., 2011). The exported video with superimposed eye tracking information had a resolution of 640 × 480 pixels. These eye position and FPV data were synchronized and imported into the Datavyu coding software (datavyu.org), which allowed gaze behaviors to be manually annotated.

For annotations, we specifically considered the viewing size of the focal object (as estimated from the eye tracker) along with the frequency in which the parent or infant manipulated, held, or touched each object within a given trial. To achieve this, we randomly selected one frame every 1.5 seconds (± 120 msec), for a total of 248 frames per session. Four research assistants were trained to annotate each sampled frame with the following information: 1) the viewing size of the focal object—defined as the total pixel count for objects within the selected frame (see Figure 3A), 2) the labels of all objects in contact with any individual's hands (if any), and 3) the person holding (parent or infant) the focal object (if any). Reliability was measured by duplicating 25% of the frames and checking inter-rater agreement for objects, holding person, and viewing size.

Analytic approach

We analyzed data from annotations for the infant's perspective (FPV images) and their surrounding context (TPV images). We counted the frequency of object manipulation moments (using both FPV and TPV) and estimated object viewing size from FPV. We then analyzed

changes in holding frequency and size according to the infant's age and the person holding the object. Our focus was on how the parent changes the way he or she handles and displays objects at specific points of infant development.

To estimate changes within and between dyads, we used Bayesian hierarchical generalized linear models. The *holding model* used object holding frequency as the dependent measure, and the *size model* used object size. Both models were fit using a negative binomial likelihood and log link function, which is appropriate for predicting changes in count data such as the number of annotated frames or number of pixels. The person holding (infant vs. parent), infant age (in months), and interaction between person and age were used as predictors in the models. Weakly informative student-*t* priors ($\nu = 4, \mu = 0, \sigma = 10$) were used to estimate the regression coefficients. A hierarchical model was selected to estimate varying slopes and intercepts among dyads, and the coefficients were estimated with a half-cauchy prior to form a common population covariance matrix. This accounts for repeated measures and within-subjects effects, and also permits generalization to new dyads.

The coefficients for the fitted models were estimated from a subset of the longitudinal data that comprised only dyads that completed at least two sessions (10 total, see Figure 2). Posterior distributions of parameters were obtained from each model, and Bayesian credible intervals (CIs) were used throughout for interpreting significant results. The sample size is reflected in the posterior distribution as uncertainty/variance in the parameters, with smaller samples increasing posterior width, making it less likely to detect differences among posteriors. The hierarchical approach is in essence a weighted average between the dependent variable and every annotation as well as the average performance across groups/dyads. Higher variability of the parameters is also reflected in the variability among dyads, with sample size being a factor in

variance estimation. We used a model comparisons approach to assess the main effects and interactions. Variants of the holding and size models were compared using their negative log-likelihood and computing the fit index PSIS-LOO-CV (pareto-smoothed leave-one-out cross-validation, a Bayesian version of AIC, see Vehtari, Gelman, & Gabry, 2016). We tested the main effect of person holding and its interaction with infant age by analyzing the change in LOO while penalizing models with more parameters. The single-session dataset was used to estimate the generalizability of the fitted models towards new dyads.

Results

Infant fixations over time

We first analyzed object fixation behaviors to confirm that infants were attending to objects throughout the observation period. Counting the number of annotations where we found infants viewing any of the 8 objects (the same annotations used to collect viewing size data), we observed that fixations to objects increased linearly over time from 5 months to 24 months. The trend seen in Figure 4 establishes that infants are already fixating target objects frequently before they actively manipulate objects.

Changes in the frequency of object holding over time

We annotated videos to count the frequency with which *anyone* was holding or manipulating at least one of the toy objects out of the total number of sampled video frames. Both FPV and TPV sources were examined. The trend shown in Figure 5 depicts changes in holding frequency. During this developmental window, the average object holding frequency remains stable over time (gray dashed line), indicating consistent object viewing across this age range. We also looked at holding frequency when the parent was the only person holding during

the session, and when the manipulations were from the infant alone (and excluding those instances in which both the parent and the infant were simultaneously interacting with an object). Holding instances for both persons consisted of times when the infant and parent both had an object in hand, but the objects were different (15.3% of sampled frames), and times when both the infant and parent were touching the same object (12% of annotated data). The trends for parent holding and infant holding, as shown in Figure 5, indicate dramatically different trajectories. We found no main effect for the person holding (Δ LOO = 4.3, *CI* = -1.9, 15.5), but the frequency of object holding depends on both person holding and infant age, as indicated by the significant interaction (Δ LOO = 7.1, *CI* = 29.6, 58.1). Parents are more active with their infants by handling objects more frequently during the early months, after which there is a gradual decline in parental showing frequency and a sharp increase in infant holding behavior. This clear shift in object engagement starts to occur after 12 months, and shows how, before this period, infants' own attempts at object manipulations are relatively infrequent compared to the parents', with some infant frequencies near zero. These low frequencies are expected at an early age, given the limited mobility of young infants who are still learning to reach efficiently and manipulate objects (E. J. Gibson, 1988). However, a steep, positive trend takes place over time until the infant, sometime between 15 and 24 months, is the one more likely to actively engage in objects, as compared with their caregiver. We also found the marginalized coefficients (predicting holding frequency to generalize to the set of dyads that completed a single session. The R^2 value for the holding model fitted on the longitudinal data (training) was 0.49 and the estimated coefficients for the test data (single session dyads) showed similar explained variance (R^2 = 0.45) compared with a model variant without the interaction or person holding main effect

(see Table 2). These results indicate a robust trend for the developmental transition between parent-generated and self-generated exploration of objects.

Changes in object size over time

Given that the parent is initially responsible for reaching for objects and bringing them within range for infant fixations, what do these objects look like when attended from the infant's perspective, and how does this view change over the time course observed in this study? To address these questions, we analyzed how the visual size of the object changes over time, with an emphasis on size changes during the earliest months (when the parent is most likely to display objects) versus later periods of development (when infants start generating their own object viewing experiences). We analyzed the size of the focal object as a function of the person holding *infant only* or *parent only*), the infant's age (in months), and the interactions between person and age. We also treated toy objects as a random effect and estimated random intercepts for each object to reflect size differences among objects. The object size model showed a main effect for person holding, indicating that the viewing size changes as a function of whether the parent or infant is holding the item, as opposed to an alternative model without this variable (Δ LOO = 20.1, *CI* = 51.3, 130.1). These results are similar to previous findings of enlarged views of objects when the child is holding the object (Pereira et al., 2014). Viewing size is dependent not only on the person doing the holding, but most critically, on the age of the infant during the play session. For example, Figure 6 shows a reduction in size for object views as the infants get older. Changes in viewing size over time are similar for both parents and infants. However, as indicated in the previous analysis, self-generated views of objects during early infancy were infrequent. A decrease in parent involvement as the infant gets older and learns to manipulate objects more efficiently may lead to the increase in infant-generated visual experiences, but the

older infant is less likely to view objects as being as large as they appeared to be in earlier months. After including the coefficients estimated from the longitudinal data in the single-session dyads, the model with the interaction term was similar in performance to the model without the interaction ($R^2 = 0.175$, Table 2). These analyses reveal a similar pattern: an infant's perception of object size depends on a number of factors, but parents show consistent holding behavior while their infants undergo rapid changes in development (as seen before 12 months) and together provides optimal viewing experiences while these changes unfold.

Discussion

Our results show that young infants have extensive object viewing experiences *before* proficiency in reaching and object manipulation. We aimed to strike a balance between a naturalistic and lab-controlled environment for measuring repeated learning instances. This accommodation invariably leads to certain limitations, and in this instance, the geometric and longitudinal aspects affecting interpretability and generalization (see Smith et al., 2014 for a discussion on other limitations regarding head-mounted systems). We note that, because of the focus on moment-to-moment gaze behaviors in a lab setting, the child might have experienced less cluttered views from one seating arrangement than in some other situations. Additionally, we acknowledge the incomplete longitudinal sample for each time interval, as well as the use of the small sample sizes of 10 (longitudinal dyads) and 8 (single sessions dyads) as the basis for our analyses. Given these limitations, the reported developmental transition point from parental scaffolding to self-handling of objects cannot be guaranteed to generalize to more naturalistic environments or a more general population of infants of the same age range (Clerkin, Hart, Rehg, Yu, & Smith, 2017). Nonetheless, the statistical methods employed to try to mitigate these

effects show clear results that point to the importance of parental scaffolding for establishing visual constancies during infants' exploration and attention to objects.

The collaborative handling-viewing feedback loop

To make complex information more accessible, parents provide additional, "support" behaviors specific to infants. These include the types of gestures used for attentional navigation (Booth, McGregor, & Rohlfing, 2008; Namy & Nolan, 2004), and for highlighting and synchronizing speech during word teaching (Fernald, 1992; Gogate, Bahrick, & Watson, 2000; Matatyaho & Gogate, 2008; Namy, Acredolo, & Goodwyn, 2000; Zukow-Goldring, 1997; Zukow and Rader, 2001). A large literature documenting these supporting behaviors reveals the importance of parental responsiveness for enhancing infant learning opportunities (see also, Chang, de Barbaro, & Deák, 2016; Iverson, Capirci, Longobardi, & Cristina Caselli, 1999). Conboy et al. (2015) documented that language learning was more successful when instruction accompanied interpersonal interaction. Similarly, in the case of parental object handling and infant viewing, coordinated actions between the two separate systems (parent and infant) are necessary if the visual experiences are to be effective for early learning. This can be challenging for the parent because the actions must be coordinated in real time and with precise timing. Social coordination is vital in infancy. A large body of literature shows that the quality (e.g., precision) of turn-taking is associated with learning, successful joint coordination, affect, and perceived quality of the interaction (Jaffe et al., 2001; Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). Little is known, however, about the development of turn-taking dynamics when a child is engaged in parent-child play, which is when much of early learning takes place. To a collaborative feedback loop that optimizes viewing, the parent and infant must lock themselves in a feedback loop of object showing and viewing so that both participants process sent and

received information simultaneously (Figure 7). Our results suggest that before the *internal* feedback loop can be stable in infants, the *collaborative* feedback loop must be established with a social partner. Thus, parental participation plays a major role in initiating the coupling and helping to facilitate the early stages of visual processing and perception in infancy. However, the *collaborative* feedback loop can be shaped by infant's active participation and development. It is possible that infant's increased reaching skills may motivate the parent to hold/place objects further, influencing the infant's object viewing experiences. The current findings provide key insights into mechanisms of early scaffolding, and thus further the understanding of social coordination and development, and how infants' own development—such as object manipulation and sensitivity to social rhythms—in return, influences parents' behaviors.

References

Bates E. 1979. Intentions, conventions, and symbols. In Bates E, Benigni L, Bretherton I, Camaioni L, Volterra V (eds). *The Emergence of Symbols*. Academic Press: New York; 33–42.

Bertenthal, B. I., & Campos, J. J. (1990). A systems approach to the organizing effects of self-produced locomotion during infancy. *Advances in infancy research*.

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington: National Association for the Education of Young Children.

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-Pragmatics and Attention: Contributions to Gesturally Guided Word Learning in Toddlers. *Language Learning and Development*, *4*(3), 179–202. https://doi.org/10.1080/15475440802143091

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, *44*(3), 867–874. https://doi.org/10.1037/0012-1649.44.3.867

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*, 5(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i-174. https://doi.org/10.2307/1166214

Chang, L., de Barbaro, K., & Deák, G. O. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, *41*(5-8), 342–361. https://doi.org/10.1080/87565641.2016.1274313

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*(1), 1–23. https://doi.org/10.1075/gest.11.1.01cla

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Conboy, B. T., Brooks, R., Meltzoff, A. N., & Kuhl, P. K. (2015). Social interaction in infants learning of second-language phonetics: An exploration of brain-behavior relations. *Developmental Neuropsychology*, *40*(4), 216–229.

https://doi.org/10.1080/87565641.2015.1014487

Dent-Read, C., & Zukow, P. G. (Eds.). (1997). *Evolving explanations of development: ecological approaches to organism-environment systems* (1st ed). Washington, DC: American Psychological Association.

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–Infant–Object interactions. *Child Development*, *87*(2), 494–512. https://doi.org/10.1111/cdev.12464

Duncan, R. M., & Tarulli, D. (2003). Play as the Leading Activity of the Preschool Period: Insights from Vygotsky, Leont'ev, and Bakhtin. *Early Education and Development*, *14*(3), 271–292. https://doi.org/10.1207/s15566935eed1403_2

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*(1), 95–116.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal Vocal Communication: Comparative and Developmental Approaches*. Cambridge University Press.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*(1), 1–42. https://doi.org/10.1146/annurev.ps.39.020188.000245

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A Study of Multimodal Motherese: The Role of Temporal Synchrony between Verbal Labels and Gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of Dialogue in Infancy: Coordinated Timing in Development. *Monographs of the Society for Research in Child Development*, *66*(2), i–149. Retrieved from http://www.jstor.org/stable/3181589

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: Official Journal of the Cognitive Development Society*, *15*(3), 393–401. https://doi.org/10.1080/15248372.2012.749481

Johnson, S. P. (2010). How infants learn about the visual world. *Cognitive Science*, *34*(7), 1158–1184. https://doi.org/10.1111/j.1551-6709.2010.01127.x

Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental Science*, *14*(6), 1355–1364. https://doi.org/10.1111/j.1467-7687.2011.01084.x

Matatyaho, D. J., & Gogate, L. J. (2008). Type of Maternal Object Motion During Synchronous Naming Predicts Preverbal Infants' Learning of Word–Object Relations. *Infancy*, *13*(2), 172–184. https://doi.org/10.1080/15250000701795655

Namy, L. L., & Nolan, S. A. (2004). Characterizing changes in parent labelling and gesturing

and their relation to early communicative development. *Journal of Child Language*, *31*(4), 821–835. https://doi.org/10.1017/S0305000904006543

Namy, L. L., Acredolo, L., & Goodwyn, S. (2000). Verbal Labels and Gestural Routines in Parental Communication with Young Children. *Journal of Nonverbal Behavior*, *24*(2), 63–79. https://doi.org/10.1023/A:1006601812056

Nomikou, I., Rohlfing, K. J., & Szufnarowska, J. (2013). Educating attention: Recruiting, maintaining, and framing eye contact in early natural mother–infant interactions. *Interaction Studies*, *14*(2), 240–267. https://doi.org/10.1075/is.14.2.05nom

O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, 8(4), 352–359. https://doi.org/10.1111/j.1467-7687.2005.00423.x

Piaget, J. (1954). The construction of reality in the child. New York, NY, US: Basic Books. https://doi.org/10.1037/11168-000

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*(11), 22. https://doi.org/10.1167/10.11.22

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Scrimsher, S., & Tudge, J. (2003). The Teaching/Learning Relationship in the First Years of School: Some Revolutionary Implications of Vygotskya's Theory. *Early Education and Development*, *14*(3), 293–312. https://doi.org/10.1207/s15566935eed1403_3

Smith, L., Yu, C., Yoshida, H., & Fausey, C. M. (2014). Contributions of head-mounted

cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in Development: Motor Skill Acquisition Facilitates 3D Object Completion. *Developmental Psychology*, *46*(1), 129–138. https://doi.org/10.1037/a0014618

Striano, T., & Rochat, P. (2000). Emergence of Selective Social Referencing in Infancy. *Infancy*, *1*(2), 253–264. https://doi.org/10.1207/S15327078IN0102_7

Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science*, *8*(4), 333–343. https://doi.org/10.1111/j.1467-7687.2005.00421.x

Tomasello M. (1995). Joint attention as social cognition. In Moore, C., & Dunham, P. J. (Eds). 103-130. *Joint attention: its origins and role in development*.

Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. https://doi.org/10.1007/s11222-016-9696-4

Vygotsky, L. S., & Cole, M. (1981). *Mind in society: the development of higher psychological processes* (Nachdr.). Cambridge, Mass.: Harvard Univ. Press.

Winsler, A. (2003). INTRODUCTION TO SPECIAL ISSUE: Vygotskian Perspectives in Early Childhood Education: Translating Ideas into Classroom Practice. *Early Education and Development*, *14*(3), 253–270. https://doi.org/10.1207/s15566935eed1403_1

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yoshida, H., & Burling, J. M. (2013). Dynamic shift in isolating referents: From social to self-generated input. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1–2). https://doi.org/10.1109/DevLrn.2013.6652570

Yoshida, H., & Smith, L. B. (2008). What's in View for Toddlers? Using a Head Camera to Study Visual Experience. *Infancy*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Zukow, P. G. (1990). Socio-perceptual bases for the emergence of language: an alternative to innatist approaches. *Developmental Psychobiology*, *23*(7), 705–726. https://doi.org/10.1002/dev.420230711

Table 1. The age (in months) at which each word reaches the 50% reporting level, according to the McArthur Child Development Inventory (MCDI).

Word	Age of comprehension	Age of production
Open	14	22
Bunny	14	19
Car	11	25
Bottle	8	16
Cookie	12	16
Eat	12	19
Drink	12	20
Put	16	25

Table 2: Proportion of explained variance for the holding and size models. The train section shows R^2 and its 95% CI for each model type and variant. The test section shows the R^2 value from applying the corresponding train model to the single-session dataset.

Models	R^2
Longitudinal data (train)	
Hold Freq. = Age, Person, Age × Person	0.486 *(0.380, 0.655)*
Hold Freq. = Age, Person	0.031 *(0.000, 0.053)*
Hold Freq. = Age	0.008 *(0.000, 0.017)*
Obj. Size = Age, Person, Age × Person	0.208 *(0.185, 0.230)*
Obj. Size = Age, Person	0.208 *(0.186, 0.230)*
Obj. Size = Age	0.195 *(0.173, 0.216)*
Single-session data (test)	
Hold Freq. = Age, Person, Age × Person	0.453 *(0.276, 0.688)*
Hold Freq. = Age, Person	0.066 *(0.000, 0.124)*
Hold Freq. = Age	0.032 *(0.000, 0.066)*
Obj. Size = Age, Person, Age × Person	0.174 *(0.133, 0.222)*
Obj. Size = Age, Person	0.175 *(0.134, 0.225)*
Obj. Size = Age	0.157 *(0.118, 0.206)*

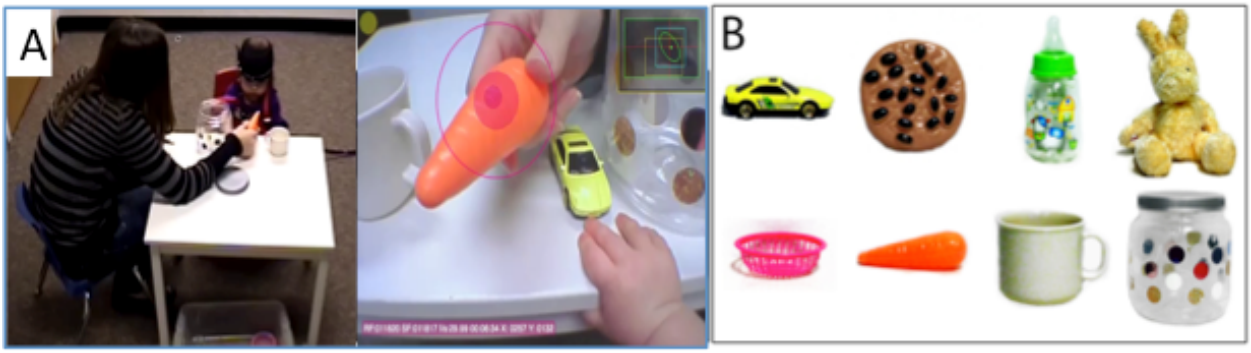

Figure 1. *A*. Testing environment. Videos were joined and synchronized to show the third-person-view (TPV) and first-person-view (FPV) with infant eye tracking coordinates superimposed over the image. *B*. A complete set of object images used throughout sessions.

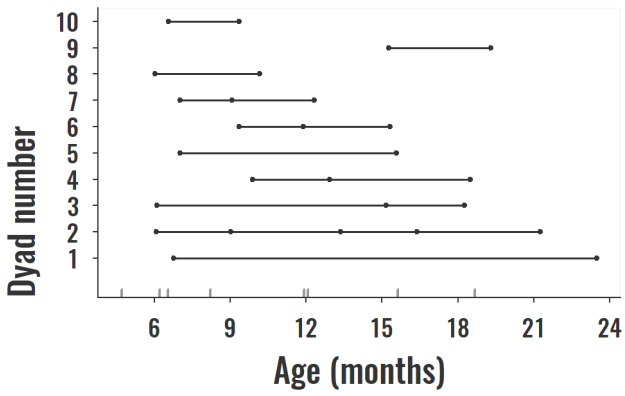

Figure 2. The sample of dyads and infant age at each session. Each row is a unique dyad (the same parent and infant) along different measured time points (indicated by the dots). Markers along the bottom axis show single-session dyads.

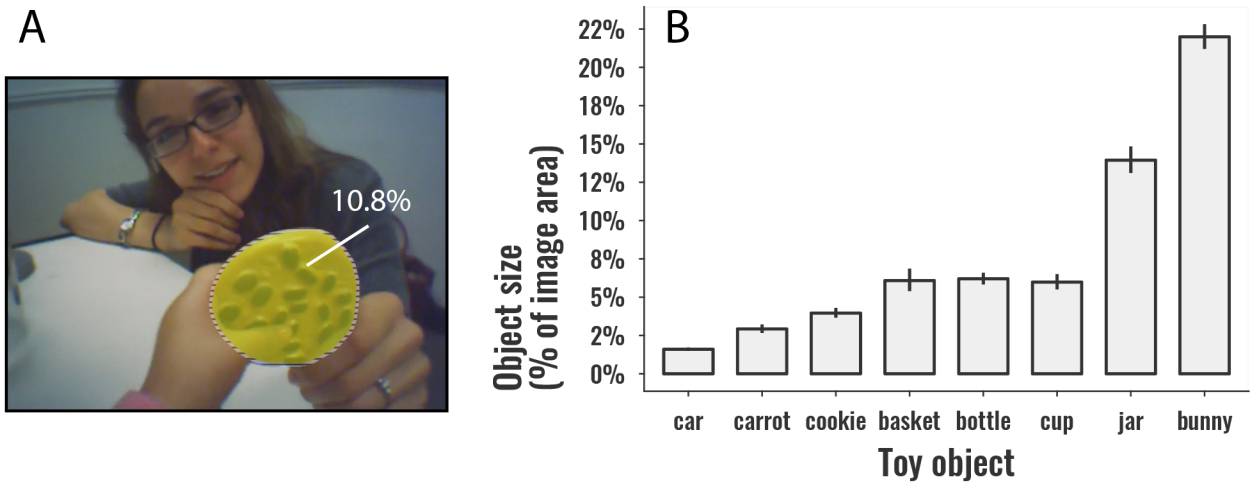

Figure 3. *A*. Illustrating the size of *cookie* for a single captured frame (i.e., the percent area of an object from a FPV image after manually detecting its boundary). *B*. Mean size for each toy object, averaged across all annotated frames.

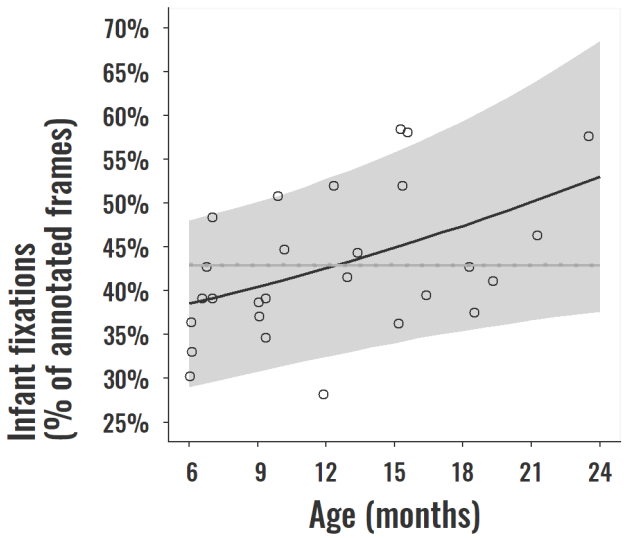

Figure 4. Number of infant fixations to toy objects as a function of age. The dotted gray line is from the model excluding the age effect.

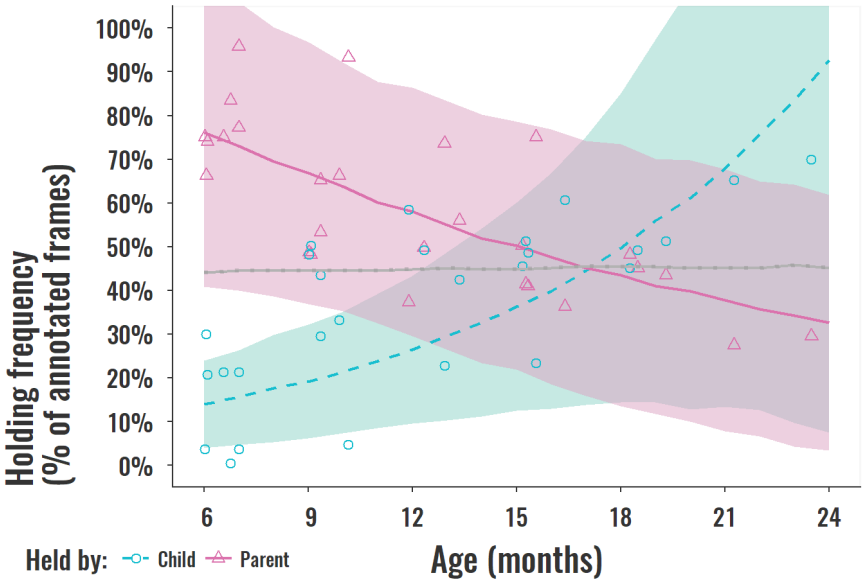

Figure 5. Changes in object holding frequency between parent and infant over time. The dotted gray line is the trend for the model excluding the indicator of person holding. Mean holding frequencies at each age are shown for parents (triangles) and infants (circles).

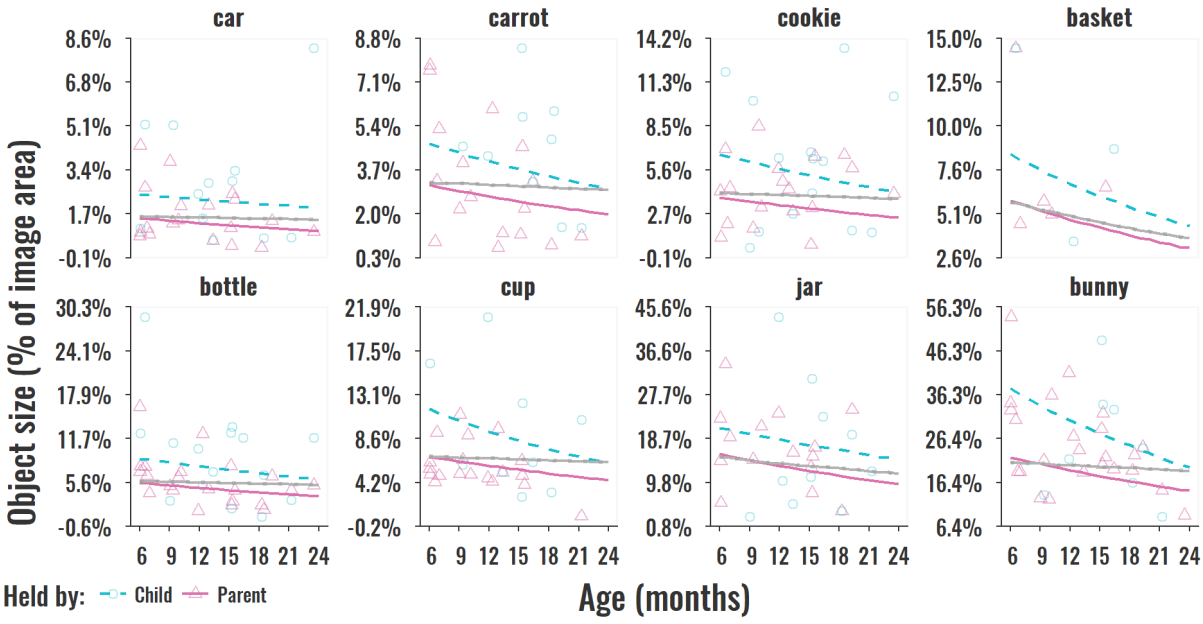

Figure 6. Decreases in object sizes over time. Separate lines are shown for the parent and infant. The gray line is the trend for the model excluding the person holding indicator.

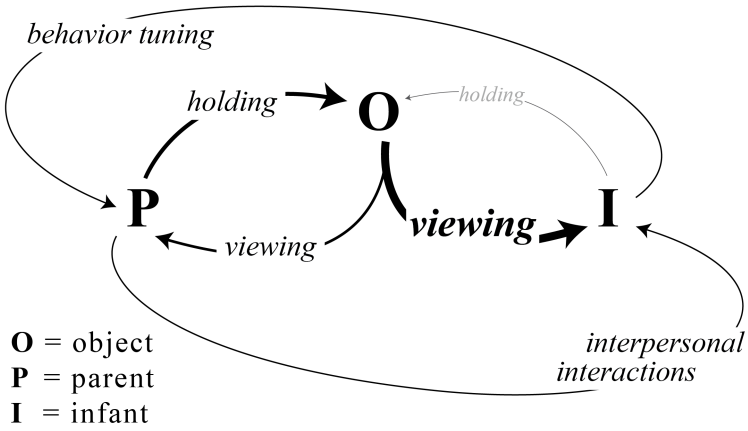

Figure 7. Collaborative feedback loops during early infancy. Within individuals (**P** and **I**), object holding is updated based on visual feedback (**O**). Collaboration emerges when parents provide frequent demonstrations of holding behavior (and thus object views) for the infant. Both parents and infants update their behavior over time given the social feedback from the partner.