Visual objects as they are encountered by young language learners

Hanako Yoshida & Caitlin M. Fausey

How do infants and toddlers learn to talk about objects? We know that concrete visual objects like CUP, BALL, and SPOON, are among the first-named objects (Caselli et al., 1995; Gentner & Boroditksy, 2001). We know a lot about how what young learners *hear* shapes what they say (Goodman, Dale, & Li, 2008; Snow & Ferguson, 1977; Weisleder & Fernald, 2013). Research on the early stages of language learning has focused on language input and how infants find words within a speech stream (Jusczyk & Aslin, 1995; Saffran, Aslin, & Newport, 1996). However, we know very little about how what children *see* matters for their language achievements. In this chapter, we first review key points regarding visual learning that are relevant to early language learning, including how children segment and attend to visual objects. We then review evidence about children's egocentric views of objects and their relevance for language learning. We conclude the chapter by discussing two new directions — the role of language in creating visual experiences and atypical language development—for language learning research grounded in these recent discoveries about visual objects as they are encountered by young language learners.

Finding visual objects

For infants to learn about objects and their properties, to form memories of those objects, and ultimately to learn the names of objects and their categories, infants must first find the objects in potentially complex and cluttered scenes in which objects overlap and may be partially occluded. Developmental research in vision science has provided programmatic and elegant experiments showing that the ability to segment a partially occluded object and represent it as a unified entity starts developing early (Valenza & Bulf, 2011), but progresses gradually during the period from birth to 6 months (Johnson & Aslin, 1996). This process is highly dependent on object motion that is independent of background (Johnson & Aslin, 1995; Kellman, Gleitman, & Spelke, 1987; Slater, Morison, Somers, Mattock, Brown, & Taylor, 1990) and is related to infants' experiences and practice in tracking objects (Busking, Botha, & Post, 2010; Johnson, Davidow, Hall-Haro, & Frank, 2008; Valenza & Bulf, 2011). Experiments studying early segmentation have typically presented infants with single planar views of an object with lateral or looming motions, both of which support segmentation (Busking et al., 2010; Johnson et al., 2008; Valenza & Bulf, 2011; c.f., Graf, 2006; Hummel & Biederman, 1992). This is particularly interesting, because recent observations of how parents show objects to young infants suggest that parents show objects so that the child sees a lot of the objects' flat surfaces but rarely see any depth or angles. Parents show objects with shaking, looming, and lateral movements or planar rotations (Matatyaho & Gogate, 2008; Thelen & Smith, 1998). These are observed in parent-infant interaction, specifically in the moments when parents hold and move objects during word teaching contexts and/or play. Critical visual experiences such as the feature-rich single viewing of objects could be in part supported by these social interactions that may both capture attention and segregate the object from the background.

Attending to visual objects

To learn about an object, infants must also sustain focused attention on the object. Infants successfully learn about an object only when they catch relevant information about it, and

this requires rapid and well-controlled attention. One major milestone in attending to objects is when infants can follow moving objects and/or the hand pointing toward objects. Such attentional organization emerges in infants as young as three months, helps early learners identify word meanings, and serves potential communicative functions (Ruff & Rothbart, 2001).

Most research examining infant visual attention has measured looking behavior such as habituation or preferential looking tasks (Kellman & Banks, 1998; Reynolds, Courage, & Richards, 2013). Looking duration has also been used to index information processing and/or intelligence (Colombo, 1993; Colombo & Mitchell, 1990; Tamis-LeMonda & Bornstein, 1989) as well as cognitive functioning (Rose, Feldman, & Jankowski, 2012). In these efforts, visual attention is used primarily as an index of discrimination or recognition. For example, one of the most commonly used measures of recognition in infant research presents an infant with a stimulus for a set length of time, followed by comparison trials in which two visual stimuli are presented simultaneously to the left and right of midline. The sensitivity to novelty is the proportion of infant looking to the novel stimulus out of the total looking time to both stimuli, and this score serves as an index of learning. If the infant recognizes the familiar stimulus, then she/he would be expected to look longer toward the novel stimulus and demonstrate recognition.

The visual attention skills studied in this literature, aimed at discovering properties of early visual attention and the impact of attention on visual recognition memory (Colombo, 2001; Reynolds et al., 2013), are also highly relevant for tasks designed to measure how young children understand words. The most recent documentations of young infants' understanding of object-word associations have used experimental paradigms that measure infant looking behavior—a task known as "looking while listening" (Bergelson & Swingley, 2012; Fernald, Zangl, Portillo, & Marchman, 2008). In the typical setup, infants are shown two discrete images, one of which is labeled in a spoken sentence ("Do you see the apple?") to examine whether or not infants look toward the named item (correct item), indicative of their knowledge about the relation between the referent and its name. Findings from this literature have demonstrated that orienting attention to referents when hearing words is a critical skill for learning words and becomes more efficient with time and experience (e.g., Bergelson & Swingley, 2012; Weisleder & Fernald, 2013). One outstanding question is: How do everyday learning contexts and viewing experiences shape this changing attentional trajectory? What factors help to organize looking behavior and support linking heard names to seen objects?

Developmentally-gated contexts for learning about objects

Early visual experiences and the development of attention take place in a dynamically changing context that frequently involves other social beings. Social partners' referential cues — gaze, pointing, touching, object handling, object showing, tapping — strongly influence communication and language learning throughout a wide age range of typically and atypically developing children (Iverson et al., 1999; Landry & Chapieski, 1989; Leekam, Hunnisett, & Moore, 1998; Matatyaho & Gogate, 2008; Yu & Smith, 2016). Parents and caregivers also modify their behavior, based on the skills and interests of their infant, in ways that matter for access to visual objects (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). Specific to object naming, observations of how parents' object showing to very young infants suggest that the ways that parents hold and dynamically move objects near the baby's face is a potent force for early looking behavior, at least in the multimodal context of naming objects (Gogate, Bolzani, & Betancourt, 2006). In addition, parents change the way they play with

their child flexibly to adapt the child's needs— based on the child's age (Brand, Baldwin, & Ashburn, 2002), on developmental achievements like object knowledge (Dimitrova & Moro, 2013), communication skills (Iverson et al., 1999; Doussard-Roosevelt, Joe, Bazhenova, & Porges, 2003; Lemanek, Stone, & Fishel, 1993), and language development (Kasari & Sigman, 1997; Landa, Holman, Garrett-Mayer, 2007; Wray & Norbury, 2018). Parental adaptations influence young children's early visual experiences and attention, and as infant responses –looking, smiling, and vocalizing— become more complex, parents' responses become more complex and coordinated with their infant's responses (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998).

These dynamic social exchanges and experiences with object playing can also change dramatically as the infant's physical and motor experiences change. For example, early reaching movements appear to be related to increases in attention to faces and objects in face to face play (Libertus & Needham, 2011), in addition to object play per se during early interactions (Striano & Stahl, 2005). Also, the transition from crawling to walking typically develops over a very broad age range, 9 to 14 months, and with this transition there are many concurrent social changes: crawlers cannot easily carry objects, are less likely to share and show objects, and parents are less likely to respond to their bids for attention with an object; walkers can easily carry objects to parents, show objects more, make more bids for attention — even from a distance — and parents are more responsive to these bids (Campos & Bertenthal, 1990; Karasik, Tamis-LeMonda, & Adolph, 2011). These changing joint interactions with objects create unique and visual opportunities for the young learner; the objects in the visual input over the course of these interactions are the data available to young learners as they are learning to connect objects and their names.

The value to developmental researchers of direct access to what visual information is available to the child, and how it changes over time, is considerable in a wide range of fields, including the study of perceptual development, motor control, social development, and language learning. There is an increasing interest in understanding the origins of object name learning by measuring the dynamic first person view in natural contexts and activities. Researchers have now solved a number of technical challenges that had previously impeded progress, including recording devices that are tolerated by young children both inside and outside the lab (see Smith, Yu, Yoshida, & Fausey, 2015, for an overview).

Capturing egocentric views of objects

Recent innovations in a lightweight wearable camera system have made it possible to capture the first person views of young infants and children. There are now a number of researchers conducting studies in which the child's own view is recorded and analyzed, and different set ups have been developed to address different research questions. A typical set up for the lab setting includes multiple room cameras, which can include a wall-mounted camera for a side view of the scene and a ceiling camera for a bird's-eye view.

The cameras are often selected to be high-resolution digital cameras able to capture fast motion between frames. For the dynamic first person views, researchers use either a mini head camera or eye-tracking headgear. These are placed on the infant's forehead with a cap, headband, and/or glasses frame. The head-camera/eye-tracking headgear is small and lightweight – typically weighing between 48g to 83g (for head-camera ~20-30g). Head cameras use a single camera recording the visual field from the infant's perspective (e.g., ~75° diagonal, ~70° horizontal, ~50° vertical), eye-tracking headgear also uses another camera facing the infant's right eye to record the eye's movements. The infrared LED facing

the infant's right eye tracks the pupil and corneal reflection in addition to the camera recording the visual field from the infant's perspective (Franchak, Kretch, Soska, & Adolph, 2011). After placing the camera(s) on children's heads, both the head camera and the eye tracking methods take care to calibrate the views. For example, a manual calibration procedure uses a board and displays some spatially distributed stickers, and gaze direction is calibrated during the beginning and end of a task session by the experimenter pointing to each sticker to attract the infant's attention to that point in the image space. The same procedure can be used for all ages of participants, and any portion of the process can be automatized. Videos along with audio data are often joined and synchronized (either on-line or off-line, and time locked at the appropriate rate) to show the multiple views including the first person view with eye tracking coordinates superimposed over the image (pink circle indicating the fixation as shown in Figure 7.1, top-right) for later annotation. A number of laboratories now have extensive experience using these kinds of systems and procedures, and report high success rates of placing and calibrating the eye-tracker with multiple age groups of participants.

<FIGURE 7.1 HERE>

The images in Figure 7.1 illustrate the views in front of infants and toddlers recorded by these systems. They are different from what a camera positioned on the ceiling or a tripod captures. In these views, objects are often brought up close to the child, largely viewed by the child (Smith et al., 2015; Yu & Smith, 2012), and differ from third-person perspective scenes (Aslin, 2009; Yoshida & Smith, 2008; Yurovsky, Smith, & Yu, 2013). Also, visual experience is intimately tied to body and movement. Every eye-gaze direction shift, every head turn, every hand action, every step taken changes the information available to the visual system. Many studies have demonstrated that early visual experiences are indeed influenced by the size and morphology of the infant body, and what they can do with that body (James, Swain, Jones, & Smith, 2013; Kretch, Franchak, & Adolph, 2014; Pereira, James, Jones, & Smith, 2010). These views are also different from views that reach adult eyes, with adult heads on adult bodies and adult motor repertoires (Smith, Yu, & Pereira, 2011), and thus they are not easily predicted by adult intuitions (Franchak, et al., 2011; Yurovsky et al., 2013). Across studies using different cameras, in different contexts, with different ages, it is now clear that the young learner's egocentric view differs dramatically from other views. The contents of the egocentric view are essential for researchers to understand what visual object information is available to young learners.

An exciting development in recent years has been to capture egocentric views not only in the lab, but also in everyday contexts at home (Fausey, Jayaraman, & Smith, 2016; Jayaraman et al., 2015; Smith et al., 2018; see also Bergelson & Aslin, 2017). This method prioritizes the everyday scenes that infants encounter, and yields datasets of egocentric views that are not distorted by the presence of an experimenter or pre-designed tasks. By using a lightweight camera with sufficient battery life and video storage, and that parents can easily position on their child, we can discover the content of everyday scenes and how this may change over development (see Figure 7.2).

<FIGURE 7.2 HERE>

Discoveries with egocentric object views

Over the last decade, we have learned a great deal about how visual input matters for early word learning by outfitting infants and toddlers with egocentric cameras during object play and naturalistic activities. One lesson is that the sensory input from the toddler's point of view is selective (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Smith et al., 2011; Yoshida & Smith, 2008) and object names associated with these selective views are especially transparent and likely to be learned (Pereira, Smith, & Yu, 2014; Yu & Smith, 2012; Yurovsky et al., 2013). Figure 7.3 illustrates the dramatic difference in the visual information available to a toddler holding an object compared to the room view.

<FIGURE 7.3 HERE>

One major gate to these egocentric views is the toddler's own action. Toddlers have short arms, so when they hold an object it is close to their eyes and dominates over potential competitor objects (Smith et al., 2011). Further, toddlers move their heads less when holding an object (Smith et al., 2011) and when reaching for an object (Yoshida & Smith, 2008) compared to other moments and so they create a relatively stable view. We now know that this uncluttered and stable view facilitates object name learning: toddlers are more likely to learn an object's name if they hear it while they are holding the object that is large-and-stable in view compared to other moments (Pereira et al., 2014; Yu & Smith, 2012). Another gate to the egocentric views emerges through social context where these views are made available by parents to infants who are still learning to reach efficiently and manipulate objects (Yoshida & Burling, 2013). In a semi-longitudinal study, infant's object fixation was tracked as a function of object manipulation of child and/or parent with 5- to 24-month-olds. The results suggest a robust sustained attention before they actively manipulate objects throughout the developmental transition between parent-generated and self-generated exploration of objects.

Further, toddlers' self-generated visual experiences may be central to development of visual object recognition. Pereira et al. (2010) reported developmental changes in how toddlers hold objects during visual exploration, across the same developmental period as the normative vocabulary burst. In this study, 12- to 36-month-old children first participated in visual and manual exploration of held objects and then later completed an object recognition test with sparse geometric versions of those objects. Pereira et al. (2010) reported that older children (but not younger children) showed a preference for planar in which the major axis of the object is parallel to the line of sight (James, Humphrey, & Goodale, 2001) and an increasing sensitivity to the geometric structure of the objects, which has been linked to the number of nouns in their productive vocabularies (Pereira & Smith, 2009; Smith, 2003). Further correlational evidence suggests a link between infants' object exploration and object recognition by showing that 5- to 8-month-old infants' history in sitting, holding and visually exploring objects predicts their ability to recognize an object from a previously unseen view (Soska, Adolph, & Johnson, 2010). These studies together point to an emerging consensus that developmental processes involving visual object recognition and the way young children hold and manual objects are tightly linked (Bertenthal & Clifton, 1998; Needham, 2000; Rochat, 1989) and may play a critical role in early object name learning (Smith, 2013).

A second lesson from recent analyses of young children's egocentric views is that the set of views available to toddlers may be an especially useful object recognition 'training dataset'. Variable views of single objects, together with highly non-uniform distribution of views across objects, may be computationally advantageous to early learning. Several recent studies demonstrate the value of the natural visual experiences that toddlers create for themselves when holding objects. Yu, Bambach, Zhang, and Crandall (2017) showed that state of the art

machine learning algorithms (Simonyan & Zisserman, 2014) learned to recognize objects in new contexts more successfully when trained on toddler views than adult views. This generalization success could be attributed to the variability in views (Simonyan & Zisserman, 2014). Slone, Smith, and Yu (2017) pursued this hypothesis and reported that greater variability in self-generated views at 15 months of age predicted the number of object names that toddlers knew six months later (see Montag, Jones, & Smith, 2015; Perry, Samuelson, Malloy, & Schiffer, 2010 for the potential processes). Clerkin et al. (2017) also recently reported that everyday scenes of one activity (mealtime) are highly cluttered, but within this clutter there is a small set of repeating objects. These most pervasive objects are also among the earliest named objects according to CDI norms. We now know that specific properties of the egocentric views available to young children are quite different from the typical the "visual diet" fed to computer vision algorithms, which presents exciting opportunities for collaboration and future insights to both developmental and computer science (see Smith & Slone, 2017, for further discussion).

Finally, a third lesson revealed by capturing and analyzing young children's egocentric views is that these views change over time. In the social domain, young infants see many more faces than hands while toddlers see many more hands than faces (Fausey et al., 2016). The properties of scenes with early named objects also change over time, with 8-10 month-olds encountering views relatively cluttered with objects and toddlers encountering more selective object views (Clerkin et al., 2017; Smith et al., 2018). These discoveries highlight the need to take changing input into account in our theories of developmental change.

Implications and future directions

We return to a foundational question for theories of language learning: What supports linking heard names to seen objects? How do recent discoveries based on egocentric views of objects guide next steps for studying the role of visual attention, memory, and learning in language development? We suggest that researchers are now in an excellent position to link the statistical structure of encountered objects to the available linguistic input, and to examine how the visual and linguistic streams may vary across contexts and learners.

Evidence from adults and children suggests that speech and eye movements are strongly coupled (Borovsky, Elman, & Fernald, 2012; Griffin & Bock, 2000; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). For example, Griffin and Bock (2000) demonstrated that speakers have a strong tendency to look toward objects referred to by speech and that words begin roughly one second after speakers gaze at their referents. Developmental studies in controlled lab settings document that heard words guide attention throughout development in a variety of task contexts (e.g., preferential looking, word comprehension, visual search, and sentence processing, among others). Recent studies have shown the power of words to direct visual attention in infants as young as 6-months-old (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012). Other studies have shown that three-year-old children are faster to find objects in cluttered scenes if they are cued with the object name than if they are cued with the object picture (Vales & Smith, 2015), suggesting that labels play a role in children's visual working memory representations and visual target identification. What is the origin of this linguistically-mediated attention and object recognition, and how robust is it?

We do not yet know the origins or developmental pathways that support linguistically-cued attention, but the multimodal structure of the encountered input is likely relevant. The idea that statistical regularities are a strong (so-called top down) force on attention is a critical

component of all influential models of attention (Desimone & Duncan, 1995; Egeth & Yantis, 1997; Fecteau & Munoz, 2006; Hayhoe & Ballard, 2005; Treisman, 2009; Wolfe, 2007), but we are only beginning to understand these regularities in infants' experiences. What are the words that co-occur with the egocentric views of objects, especially in everyday contexts? Perhaps parents' object handling helps to organize co-occurring sounds, supporting phonological segmentation and word-referent mappings for specific objects. Or, developmentally ordered input like massive repetition of visual objects for months before encountering the objects' names may support rapid learning of multiple words during the typical toddler 'word burst' period. The power of a word to direct attention may depend on its multimodal history, and vary across the early years as children build their vocabularies. Determining the extent to which visual and linguistic streams offer concurrent and/or ordered support for linking objects and their names will constrain theories about the origins of linguistically-cued attention. The head camera approach will help researchers discover the microstructure of moment-to-moment instances in which a word meets its visual scene, creating a history of language-vision co-occurrences that may support linguistically-cued attention.

Clearly, the power of a word to direct attention is important in many contexts – from learning words, to finding relevant parts of a scene, to following instructions in a classroom. However, as for robustness, we know relatively little about potential individual differences in the developmental trajectories of linguistically-cued attention. Can a single utterance effectively direct attention in all children, or could multiple repetitions of the same word be required for some children? Individual differences in the suite of early visual, language, social, and attentional experiences are particularly relevant for understanding the robustness of linguistically-cued attention in typical and atypical development.

A growing number of studies indicate that individual differences in the ability of 6- to 18-month olds to establish joint attention are strongly predictive of language ability at 24- to 36-months (Markus, Mundy, Morales, Delgado, & Yale, 2000; Morales, Mundy, & Rojas, 1998). Mundy and colleagues have been particularly interested in the early diagnosis of children with autism spectrum disorders (Delgado, Mundy, Crowson, Markus, Yale, & Schwartz, 2002). The task they used is similar to those that are used to measure the extent of the perceptual field. In this task, a target that is socially indicated by a parent's gaze direction or point is located within or outside of the visual field, and children's gaze shift to that target is recorded from a room camera facing the child. Typically developing children (12 to 18 months) readily follow eye gaze (or point) to a target outside of the visual field. Mundy and colleagues have used performance in this task to predict children at risk for difficulties in learning language and children with autism show specifically marked difficulties in the use of social cues in this task. Interestingly, however, a recent study in which parents and their young children with autism wore head-cameras during a social interaction documented parental scaffolding that supported the child's sustained attention to referential cues and joint attention (Yoshida & Kushalnagar, 2009). In this study, children with autism experienced joint attention moments at a similar rate as their typically developing peers. They also experienced more joint attention moments than their typically developing peers immediately after parents gestured. Parents appear to alter their behavior according to the developmental level of their child, including increased scaffolding for their children with autism, with measurable consequences for the child's visual experiences in social interactions. Head-cameras are particularly well-suited to address questions about visual patterns of cause and consequence, feedback loops, and coordination in dyadic play (see also Yu & Smith, 2016),

with clear public health relevance and the potential to guide evidence-based supportive parenting interventions.

<FIGURE 7.4 HERE>

Autism research has also reported links between atypical sensorimotor development and atypical patterns of object exploration (de Campos, Savelsbergh, & Rocha, 2012), visual processing (Behrmann, Thomas, & Humphreys, 2006) and attention (Takarae, Luna, & Sweeney, 2012), indicating potential cognitive factors constraining visual experiences. The co-occurrence of words and referents with these atypical visual experiences, and the predictability structure of the input, may also be atypical —leading to different and challenging trajectories of both word learning and linguistically-cued attention. Recent studies using head-cameras to capture interactions between parents and children, however, highlight a wide range of co-occurrence and predictability structures relating the linguistic and visual streams within experiences of *typically developing children* (Castellanos, Pisoni, Yu, Chen, & Houston, accepted; Yoshida & Kushalnagar, 2009). In one study, typically developing three to five year old deaf children's (of deaf parents) egocentric viewing was recorded in a social interaction with objects (Yoshida & Kushalnagar, 2009). The preliminary results indicate that during interactions between a child and a parent whose linguistic input is visually encoded dominantly (e.g., American Sign Language), the parent's hands and the child's own hands dominate the child's visual experiences (see Figure 7.4). Compared to hearing children, single object play is also especially apparent. Hence, egocentric views may reveal not only potentially atypical encounters with visual objects, but also how these encounters and the coordination of visual, language, and social input, support multiple pathways of language development. Egocentric views reveal properties of the visual environment that are available to young learners and therefore available to shape their attention and learning (see Jayaraman, Fausey, & Smith, 2017, Figure 7.1). Insights into the structure of this input, and its variation across typical and atypical development, will advance our understanding of how learning emerges and changes, and guide future interventions designed to support strong beginnings of language development.

Conclusion

Children learn to talk about objects. Recent efforts using head cameras have captured the visual objects that young learners actually encounter. We have learned that these egocentric scenes are often selective with respect to the objects in view, change over developmental time, and have properties that are computationally advantageous to early learning. We look forward to future discoveries about the linguistic and social cues that co-occur with egocentric views in order to support early language learning in many kinds of learners in many kinds of contexts.

References

Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science: Official Publication of the American Academy of Optometry, 86*(6), 561–565. doi:10.1097/opx.0b013e3181a76e96

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends in cognitive sciences*, *10*(6), 258-264. doi:10.1016/j.tics.2006.05.001

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds.

Proceedings of the National Academy of Sciences, *114*(49), 12916-12921. doi:10.1073/pnas.1712966114

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA,109,*3253-3258. doi:10.1073/pnas.1113380109

Bertenthal, B. I., & Clifton, R. K. (1998). Perception and action. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 51–102). New York, NY: Wiley.

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

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*, 417-436. doi:10.1016/j.jecp.2012.01.005

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

Busking, S., Botha, C. P., & Post, F. H. (2010). Dynamic multi-view exploration of shape spaces. *Computer Graphics Forum*, 29(3), 973-982. doi:10.1111/j.1467-8659.2009.01668.x

Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*(2), 159-199. doi: 10.1016/0885-2014(95)90008-X

Campos J. J., & Bertenthal B. I. (1990). Locomotion and psychological development in infancy. In F. Morrison, K. Lord, & D. Keating (Eds.), *Applied developmental psychology: Psychological development in infancy* (pp. 229-258). New York, NY: Academic Press.

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. doi:10.2307/1166214

Castellanos, I., Pisoni, D. B., Yu, C., Chen, C., & Houston, D. M. (accepted). Embodied cognition in prelingually deaf children with cochlear implants: Preliminary findings. In H. Knoors, & M. Marschark (Eds.), Educating Deaf Learners: New Perspectives. New York, NY: Oxford University Press.

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*, 372: 20160055. doi: 10.1098/rstb.2016.0055

Colombo, J. (1993). Infant cognition: Predicting later intellectual functioning. Newbury Park, CA: Sage. doi:10.4135/9781483326481

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology, 52*(1), 337–367. doi:10.1146/annurev.psych.52.1.337

Colombo, J. & Mitchell, D. W. (1990). Individual and developmental differences in infant visual attention: Fixation time and information processing. In Colombo, J. & Fagen J. W. (Eds.), *Individual differences in infancy* (pp. 193-227). Hillsdale, NJ: Erlbaum

de Campos, A. C., Savelsbergh, G.J., & Rocha, N. A. (2012). What do we know about the atypical development of exploratory actions during infancy? *Research in Developmental Disabilities*, *33*(6), 2228-2235. doi:10.1016/j.ridd.2012.06.016

Delgado, C., Mundy, P., Crowson, M., Markus, J., Yale, M., & Schwartz, H. (2002). Responding to joint attention and language development: A comparison of target locations. *Journal of Speech, Language and Hearing Research, 45*, 1715–1719. doi:10.1044/1092-4388(2002/057)

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222. doi:10.1146/annurev.neuro.18.1.193

Dimitrova, N. & Moro, C. (2013). Common ground on object use associates with caregivers' gestures, *Infant Behavior and Development, 36*(4), 618-26. doi:10.1016/j.infbeh.2013.06.006

Doussard-Roosevelt, J.A., Joe, C.M., Bazhenova, O.V., & Porges, S.W. (2003). Mother-child interaction in autistic and nonautistic children: Characteristics of maternal approach behaviours and child social responses. *Development and Psychopathology, 15*, 277-295. doi:10.1017/s0954579403000154

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48, 269-297. doi:10.1146/annurev.psych.48.1.269

Fausey, C. M., Jayaraman, S. & Smith, L. B. (2016) From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101-107. doi:10.1016/j.cognition.2016.03.005

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, 382-390. doi:10.1016/j.tics.2006.06.011

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds. & Trans.), *Developmental Psycholinguistics: On-line Methods in Children's Language Processing* (pp. 97-135). Amsterdam: John Benjamins. doi:10.1075/lald.44.06fer

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. doi:10.1111/j.1467-8624.2011.01670.x

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256).Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511620669

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259-288. doi:10.1207/s15327078in0903_1

Goodman, J. C., Dale, P.S., & Li, P. (2008) Does frequency count? Parental input and the acquisition of vocabulary. *Child Language, 35*, 515–531. doi:10.1017/s0305000907008641

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin, 132(6)*, 920-945. doi:10.1037/0033-2909.132.6.920

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274-279. doi:10.1111/1467-9280.00255

Hayhoe, M. M., & Ballard, D. H. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, 188–194. doi:10.1016/j.tics.2005.02.009

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*(3), 480–517. doi:10.1037/0033-295x.99.3.480

Iverson J. M., Capirci O., Longobardi E., Caselli M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14*, 57–75. doi:10.1016/s0885-2014(99)80018-5

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: where the action is. *Canadian Journal of Experimental Psychology*, *55*(2), 111-120. doi:10.1037/h0087358

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

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS One, 10*(5), e0123780. doi:10.1371/journal.pone.0123780

Jayaraman, S., Fausey, C.M., & Smith, L.B. (2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology, 53*(1), 38–49. doi:10.1037/dev0000230

Johnson, S. P., & Aslin, R. N. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology, 31*, 739–745. doi:10.1037/0012-1649.31.5.739

Johnson, S. P., & Aslin, R. N. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development, 11*(2), 161–180. doi:10.1016/s0885-2014(96)90001-5

Johnson, S. P., Davidow, J., Hall-Haro, C., & Frank, M. C. (2008). Development of perceptual completion originates in information acquisition. *Developmental psychology*, *44*(5), 1214. doi:10.1037/a0013215

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1–23. doi:10.1006/cogp.1995.1010

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development, 82*, 1199–1209. doi:10.1111/j.1467-8624.2011.01595.x

Kasari, C., & Sigman, M. (1997). Linking parental perceptions to interactions in young children with autism. *Journal of Autism and Developmental Disorders, 27*(1), 39-57. doi:10.1023/a:1025869105208

Kellman, P.J., & Banks, M.S. (1998). Infant visual perception. In D. Kuhn & R.S. Siegler (Eds.), Cognition, perception, and language: Vol. 2, Handbook of child psychology (pp. 103– 146). New York: Wiley

Kellman, P. J., Gleitman, H., & Spelke, E. S. (1987). Object and observer motion in the perception of objects by infants. *Journal of Experimental Psychology: Human Perception and Performances, 13*(4): 586-93. doi: 10.1037/0096-1523.13.4.586

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development, 85*(4), 1503–1518. doi: 10.1111/cdev.12206

Landa, R. J., Holman, K. C., Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Arch Gen Psychiatry, 64*(7), 853-64. doi:10.1001/archpsyc.64.7.853

Landry, S. H, & Chapieski, M. L. (1989). Joint attention and infant toy exploration: Effects of Down syndrome and prematurity. *Child Development, 60,* 103–118. doi:10.111/j.1467-8624.1989.tb02700.x

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: gaze-following in children with autism. *Journal of Child Psychology and Psychiatry, 39*(7), 951-62. doi:10.1017/s0021963098003035

Lemanek, K.L., Stone, W.L., & Fishel, P.T. (1993). Parent-child interactions in handicapped preschoolers – The relation between parent behaviors and compliance. *Journal of Clinical Child Psychology, 22*, 68-77. doi:10.1207/s15374424jccp2201_7

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

Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., & Yale, M. (2000). Individual differences in infant skill as predictors of child-caregiver joint attention and language. *Social Development, 9*, 302–315. doi:10.111/1467-9507.00127

Matatyaho, D. & Gogate, L. J. (2008). Type of maternal object motion in synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy, 13*(2), 172-184. https://doi.org/10.1080/15250000701795655

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological science*, *26*(9), 1489-1496. doi:10.1177/0956797615594361

Morales, M., Mundy, P., & Rojas, J. (1998). Gaze following and language development in six-month-olds. *Infant Behavior and Development, 21,* 373–377. doi:10.1016/s0163-6383(98)90014-5

Needham, A. (2012). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development, 1*, 131–156. doi:10.1207/s15327647jcd010201

Pereira, A., 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), 1–13. doi:10.1167/10.11.22

Pereira, A. & Smith, L. B. (2009). Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental Science, 12*, 67–80. doi:10.1111/j.1467-7687.2008.00747.x

Pereira, A., Smith, L. B. & Yu, C. (2014) A Bottom-up View of Toddler Word Learning. *Psychonomic Bulletin & Review, 21*(1), 178-185. doi:10.3758/s13423-013-0466-4

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010) Learn locally, think globally: Exemplar variability supports higher-generalization and word learning. *Psychological Science, 21*(12), 1894-902. doi:10.1177/0956797610389189

Reynolds, G. D., Courage, M. L., & Richards, J. E. (2013). The Development of Attention. Oxford Handbooks Online. doi:10.1093/oxfordhb/9780195376746.013.0063

Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology, 25*, 871–884. doi:10.1037/0012-1649.25.6.871

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of Infant Cognition for Executive Functions at Age 11. *Psychological Science, 23*(11), 1345-1355. doi:10.1177/0956797612444902

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: Themes and variations*. Oxford University Press. doi:10.1093/acprof:oso/9780195136326.001.0001

Saffran, J. R., Aslin, R. N. and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928. doi:10.1126/science.274.5294.1926

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. doi:10.1007/978-3-319-16865-4_35

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development, 13*,3349. https://doi.org/10.1016/0163-6383(90)90004-R.

Slone, L. K. Smith, L. B. & Yu, C. (2017). Self-generated variability in object images predicts later vocabulary size. Poster presented at the Cognitive Development Society Biennial Conference, Portland, OR.

Smith, L. B. (2013) It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist, 68*(8), 618-629. doi.10.1037/a0034185

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The Developing Infant Creates Curriculum for Statistical Learning. *Trends in cognitive sciences, 4*, 325-336. PMID: 29519675 doi:10.1016/j.tics.2018.02.004

Smith, L. B. & Slone, L. K. (2017) A Developmental Approach to Machine Learning? *Frontiers in Psychology, 8,* 2124. doi:10.3389/fpsyg.2017.02124

Smith, L. B., Yu, C., & Pereira, A. F. (2011) Not your mother's view: the dynamics of toddler visual experience. *Developmental Science, 14*(1), 9-17. doi:10.1111/j.1467-7687.2009.00947.x

Smith, LB., Yu, C., Yoshida, H., & Fausey, C. (2015) Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development, 16*(3):407-419. doi:10.1080/15248372.2014.933430

Snow, C., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press doi:10.2307/412603

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: motor skill acquisition facilitates three-dimensional object completion. *Developmental psychology*, *46*(1), 129. doi:10.1037/a0014618

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

Takarae, Y., Luna, B., & Sweeney, J. A. (2012). Development of Visual Sensorimotor Systems and Their Cognitive Mediation in Autism. In *Handbook of Growth and Growth Monitoring in Health and Disease* (pp. 1379-1393). Springer New York. doi:10.1007/978-1-4419-1795-9_83

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*(3), 738–751. doi:10.2307/1130739

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 16*, 268(5217), 1632-4. doi:10.1126/science.7777863

Thelen, E. & Smith, L.B. (1998). Dynamic systems theories. In Handbook *of Child Psychology: Theoretical Models of Human Development* vol. 1, R. M. Lerner, 5th Eds. (pp. 563-634), John Wiley: New York. doi:10.1002/9780470147658.chpsy0106

Tincoff, R. & Jusczyk, P.W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy, 17*(4), 432-444. doi:10.1111/j.1532-7078.2011.00084.x

Treisman, A. (2009). Attention: Theoretical and psychological perspectives. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 189–204). Cambridge, MA: MIT Press.

Valenza, E., & Bulf, H. (2011). Early development of object unity: Evidence for perceptual completion in newborns. *Developmental science*, *14*(4), 799-808. doi:10.1111/j.1467-7687.2010.01026.x

Vales, C., & Smith, L.B. (2015). Words, shapes, visual search and visual working memory in 3-year-old children. *Developmental Science, 18*, 65-79. doi:10.1111/desc.12179

Wray, C., & Norbury, C. F. (2018). Parents modify gesture according to task demands and child language needs. *First Language, 38*(4), 419-439. doi:10.1177/0142723718761729

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143-2152. doi:10.1177/0956797613488145

Wolfe, J. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195189193.003.0008

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. doi:10.1080/15250000802004437.

Yoshida, H., & Burling, J. M. (2013). Dynamic shift in isolating referents: From social to self-generated input. In Proceedings of the 3rd IEEE International Conference on Development and Learning and on Epigenetic Robotics. DOI: 10.1109/DevLrn.2013.6652570

Yoshida, H. & Kushalnagar, P. (2009 April) Attentional flexibility: Head-cameras reveal different strategies for hearing children, deaf children, and children with ASD. Talk presented at pre-conference for the biennial meeting of the Society for Research on Child Development, Denver, CO.

Yu, C., Bambach, S., Zhang, Z., & Crandall, D. J. (2017). Exploring inter-observer differences in first-person object views using deep learning models. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). doi:10.1109/iccvw.2017.326

Yu, C. & Smith, L. B. (2012) Embodied Attention and Word Learning by Toddlers. *Cognition, 125*, 244-262. doi:10.1016/j.cognition.2012.06.016

Yu, C. & Smith, L. B. (2016) The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology, 26*(9), 1235-1240. doi:10.1016/j.cub.2016.03.026

Yurovsky, D., Smith, L. B. & Yu, C. (2013) Statistical Word Learning at Scale: The Baby's View is Better. *Developmental Science, 16*, 959-966. doi:10.1111/desc.12036.