Title
Developmental See-Saws: Ordered visual input in the first two years of life

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Jayaraman, Swapnaa
Fausey, Caitlin
Smith, Linda

Publication Date
2013

Peer reviewed

Developmental See-Saws: Ordered visual input in the first two years of life

Swapnaa Jayaraman, Caitlin M. Fausey, & Linda B. Smith
({swapnaa, cfausey, smith4} @indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 East Tenth Street, Bloomington, IN 47405 USA

Abstract

The first two years of life are characterized by considerable change in all domains – perception, cognition, action, and social interactions. Here, we consider the statistical structure of visual input during these two years. Infants spanning the ages from 1 to 24 months wore body-mounted video cameras for 6 hours at home as they engaged in their daily activities. Our data strongly suggest that the statistical structure of the learning environment is dynamic and ordered. The available visual statistics are not stationary, but rather they are gated by young children's developmental level. We find a rolling wave of "See-Saw" patterns over developmental time in two classes of important social stimuli: First faces, then hands; and within hands, first other-then-self-then-touching-then-holding. These ordered environments may help learning systems "start small," find the optimal path to the optimal solution, and determine the architecture of the system that does the learning.

Keywords: natural statistics; first-person camera; faces; hands

Introduction

Growing evidence across many domains indicates that human learners, including infants, are highly sensitive to the statistical regularities in the learning environment (Saffran, Aslin & Newport, 1996) and that in many domains the regularities in the learning environment contain sufficient information to yield deep conceptual representations (Chater, Tenenbaum & Yuille, 2006) of the kind that appear responsible for syntax (Griffiths, Steyvers & Tenenbaum, 2007), semantics (Griffiths, et al., 2007), categories (Madole & Oakes, 1999), and human visual object recognition (Logothetis & Sheinberg, 1996; Quinn, Eimas & Tarr, 2001). As methods for understanding structure in large data sets advance, it seems likely that we will discover rich insights in even broader domains about how the statistical structure of learning environments shapes human learning and knowledge. The research presented in this paper makes two contributions to this endeavor: by extending the study of the statistical structure of the learning environment to social stimuli – faces and hands; and by showing that the statistical structure of the learning environment is not stationary. Rumelhart and McClelland's (1986) model of the learning of the past tense was once famously criticized (Pinker & Prince, 1988) as a cheat because the network was presented with learning examples in an ordered way rather than as batch statistics. However, the statistics of the learning environment *can* change substantially with development itself. The present findings provide one clear and dramatic example of this reality. We return to the more general issue of what constitutes a developmentally appropriate conceptualization of statistical learning in the General Discussion.

Faces and Hands as Important Social Stimuli

The first two years of life are characterized by considerable change in all domains – perception, cognition, action, and social interactions. Findings in all of these domains indicate the important role of other people, in scaffolding and supporting developmental process (Tomasello, 1988). Research into the social behaviors of mature partners that support infant learning have centered on two body regions – face and hands. Research into the adult actions that infants attend to as guides to learning and understanding the world also focus on two body regions – faces and hands. And, indeed, a very large literature suggests that infants are highly sensitive to what are very small movements in these body regions – a shift in eye gaze (Butterworth & Jarrett, 1991), a mouth opening (Moll & Tomasello, 2012), a point (Leung & Rheingold,1981), and a grasp (Woodward, 1998). It seems likely that faces and hands are everywhere in early infant experience. From the statistical learning perspective, this would mean that face and hand experiences present a very large data set for mining the structure and meaning of social gestures.

However, contemporary research is also consistent with the idea that this statistical learning might be modulated by internal (and innate, Meltzoff & Moore, 1977; Slater, 1999) biases that privilege faces early in development. Research on infant face perception begins from the perspective that faces are a special class of stimuli. A large set of findings using diverse tasks indicate that very young infants are differentially sensitive to face-like visual stimuli relative to other stimulus categories (Goren, Sarty & Wu, 1975; Johnson, et al., 1991) and can discriminate familiar faces from unfamiliar ones (Field et al, 1984) shortly after birth. Moreover, the earliest social interactions consist of face-to-face play (Stern, 1971) and these have been characterized as "proto-conversations" that teach critical components of turn-taking and seem a likely context for learning about the facial cues that modulate social interactions and infant learning. Other evidence suggests that this early sensitivity to faces plays a critical role in tuning face perception processes: By 6 months infants recognize faces that are similar to those that have dominated their visual experiences (same race) better than faces that are dissimilar (different race) (Kelly et al., 2007). Early visual deprivation appears to disrupt indices of face expertise such as configural processing of faces (Maurer, Le Grand & Mondloch, 2002), and identification of faces with altered orientations and expressions (Geldart et al, 2002).

Systematic attention to hand actions as social indicators – as pointers to objects to which infants should attend -- has been shown in 4 month olds (Rohlfing, Longo & Bertenthal, 2012) but as far as we know this is the youngest demonstration of an understanding of a hand action. Most of the evidence indicating infant attention to and understanding of the meaning of hand actions – both in the context of language learning (Bates et al, 1989) and in the context of understanding the causal structure of events (Baldwin, 1991; Woodward, 1998) – focuses on older infants. For example, 10 and 12 months olds have been shown to use hand actions to predict causal sequences (Sommerville & Woodward, 2005), 11 month olds have been shown to use the structure of a hand action to predict where an event will occur (Canon & Woodward, 2012), in a large number of experiments 9 to 14 month olds have been shown to use points and other hand gestures to determine the intended referent of a heard word (Rader & Zukow-Goldring, 2012), and 18 month olds may even understand hand gestures that mimic actions as pointers to objects and events more readily than words (Namy & Waxman, 1998), as may 2-4 year olds (Hahn & Gershkoff-Stowe, 2010). Several recent studies on 12 to 18 month old infants' attention in naturalistic contexts indicate that these older infants differentially – and perhaps systematically – look to the hand actions of mature partners when engaged in joint play with the parent (de Barbaro, Chiba & Deák, 2011; Franchak, et al., 2010; Yoshida & Smith, 2008), a result that suggests that these older infants know that hand actions contain important social information.

These findings indicate that infants may know about faces as sources of social information before they know about hands and suggest the following hypothesis: Although faces and hands are equally ubiquitous in the learning environments of infants, learning about these two classes of social cues is gated by infants' early differential sensitivity to faces.

Ordered input

Traditional approaches to statistical learning have concentrated on non-incremental learning tasks, tasks in which the entire training set is fixed at the start of learning and then is either presented in its entirety or randomly sampled. From this perspective, if learning needs to be constrained in some way or directed to some portion of the input, it must be accomplished by internal constraints on the learning system (Markman & Hutchinson, 1984; Pinker, 1989), such as an innate sensitivity or interest in faces. But infants do not encounter the world as a single set of fixed statistics; they encounter it one learning instance at a time. Because of this, they may encounter and experience selected regularities in just one small region of the total batch-statistics learning environment (Smith & Gasser, 2005). Given the dramatic changes in the skills of human infants over the first two years of life, this seems highly likely. Given the ordered nature of these changes – first rolling over, then reaching, then sitting stably, then crawling, and then walking -- these selected statistics will also be ordered. West and King (1987) proposed the concept of ontogenetic niche: the idea that developmental level orders experiences in ways that constrains and canalizes developmental process. For example, in humans (and most mammal and some bird species) the young require constant caretaking and this constant caretaking limits as well as structures the input, and thus the regularities that can be learned one at a time. Humans' changing sensory motor abilities seem likely to constrain and expand visual experiences in different ways at different times. Human infants spend their first 6 months where others place them – on the floor, in infant seats, in a crib, in arms – and see what is in those places and what their mature caretakers care to show them. By 12 months, infants are much more masters of their own visual environments – placing themselves in different locations and actively selecting what they will show themselves (Adolph et al., 2012).

These considerations raise an alternative hypothesis about faces and hands: Although faces and hands are equally ubiquitous in human environments, they are not equally ubiquitous in the visual environments of infants of different ages; instead, experiences of faces and hands are ordered, with dense experiences of faces characterizing the early ontogenetic niche and dense experiences of hands characterizing the later ontogenetic niche.

Rationale for the present approach

The findings reported here are part of a larger program of research examining the statistical structure of natural visual environments as it relates to social cues and language learning. We build on the approach of a growing number of researchers using ego-centric cameras (Fathi, Hodgkins & Rehg, 2012; Kanade, 2009) to capture first person visual environments. Studies of infants' first person perspectives (mostly small laboratory studies, Aslin, 2009; Franchak et al., 2011; Smith, Yu, & Pereira, 2011) have shown that these first person environments are characterized by properties of early visual experience that are not evident from third-person observer perspectives (Yoshida & Smith, 2008) and have also documented the impact of infant body movements on infant visual experience (Kretch et al., 2012). Intriguingly, all the head-camera studies conducted with *toddlers* to date have noted that faces are rarely in the head camera images whereas hands – the child's and social partner's – are often in view (Franchak et al, 2011; Frank, 2012; Smith et al, 2011; Yoshida & Smith, 2008). These studies, however, did not broadly sample the natural or representative experiences of participants. The present study was designed to do just this. Infants spanning the ages from 1 to 24 months wore body-mounted video cameras for 6 hours at home as they engaged in their daily activities. The first questions we asked of these data, the results we report here, are these: How prevalent are faces and hands in the visual environment? Do the frequencies of faces and hands change systematically with development?

Method: Capturing early visual environments

Participants

23 infants and toddlers provided up to 6 hours of video each. This visual corpus consists of four subsets grouped by age. All videos within an age range are treated as a set.

Table 1: Infant and toddler visual corpus

Age (months)	n	Hours of video	Frames coded
1-3	7	19.26	13,865
7-9	5	21.88	15,754
18	6	22.64	16,303
24	5	14.59	10,505
Total	23	78.37	56,427

Materials and Procedure

A small, lightweight camera was used to record the visual environments of infants and toddlers (Looxcie 2, Looxcie, Inc.). The diagonal FOV is 62 degrees with a 2":infinity depth of focus. The camera was secured to a wearable hat or harness. Parents were given a camera, hat and/or harness, and instructions about camera operations. Parents recorded up to 6 hours of video when their child was awake.

Video pre-processing

Recorded videos were screened for private content and blank screens (e.g., camera was left turned on while not on child). Remaining videos were converted to images sampled at one frame for every five seconds of video. This first-of-a-kind corpus has approximately 78 hours of video and 56,000 frames of the natural visual environments of infants and toddlers in the first two years of life.

Video coding: A reliable crowd-sourced approach

Frames were presented to coders on Amazon's Mechanical Turk (mturk.com) and analyses consider only those frames for which at least 75% of coders agreed (across all coding passes, 93.7% reliable judgments). Coding proceeded in six separate passes through the data. For each pass, coders viewed an instructions page with example images.

Faces and Hands

The first broad passes coded for the presence of Faces and Hands. The infant (1-3, 7-9 months) and toddler (18, 24 months) data were coded with slightly different protocols. For infants, each coder saw up to eight frames and answered several questions about each frame. The two relevant questions were: (1) Do you see a human face or face part? and (2) Do you see other body parts or skin? If yes, which do you see? (a) bare hands/fingers, (b) bare feet/toes, (c) other body parts (neck, shoulder, knee, etc.), (d) body parts covered in clothes, (e) two or more of the above. In these analyses, only responses that indicated the presence of bare hands were further analyzed. Four unique coders judged each frame. For toddlers, each coder saw up to 100 frames and answered the same yes-or-no question for all frames. In separate passes, coders answered either (1) Do you see a human face in this picture? or (2) Do you see a human hand in this picture? Five unique coders judged each frame.

Free, touching and holding hands

The next coding passes focused specifically on hands. First, we identified whether hands in the visual input belonged to the child or to someone else. Then, we identified whether the child's own hand was free, touching something, or holding a small object. In four distinct passes, coders answered one of these questions: Does any hand you see belong to the child wearing the camera?, Is the child's own hand touching something?, Is the child's own hand holding onto something?, Is the child's own hand holding something that can be carried?

Results: Ordered visual input

Body parts in the visual environment

How prevalent are faces and hands in the visual environments of infants and toddlers? The relative frequency of these two key body parts depends on the developmental stage of the child (Figure 1). The visual environments of the infants had more faces than hands (1-3 months: .29 Faces, .01 Hands; 7-9 months: .15 Faces, .06 Hands). For toddlers, hands were more prevalent than faces (18 months: .11 Faces, .28 Hands; 24 months: .07 Faces, .32 Hands); $\chi^2(3, N = 17962) = 6936.84$, $p < .001$.

Faces and hands appear to trade-off, suggesting ordered visual input: Faces first, then hands. The developmental trend is not just increased variability of body parts in the visual input: The total proportion of faces and hands together is more stable across the first two years of life (.30, .21, .34, .34, for each age range respectively). The key finding is a "See-Saw" pattern: What is first available to infants (here, Faces; "See") fades to developmental history ("Saw") as infants creates new tasks for themselves with advancing motor, language and social skills.

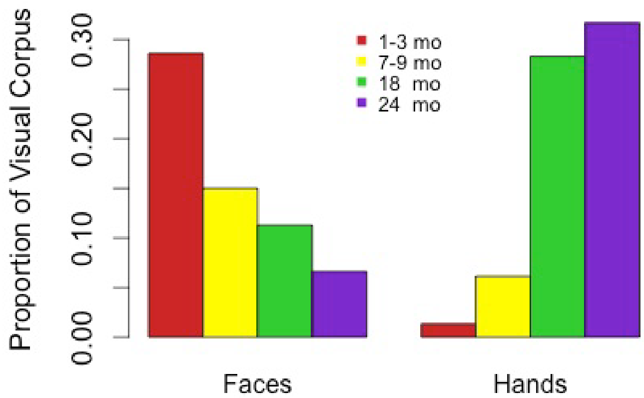

Figure 1: Faces and hands in early visual environments

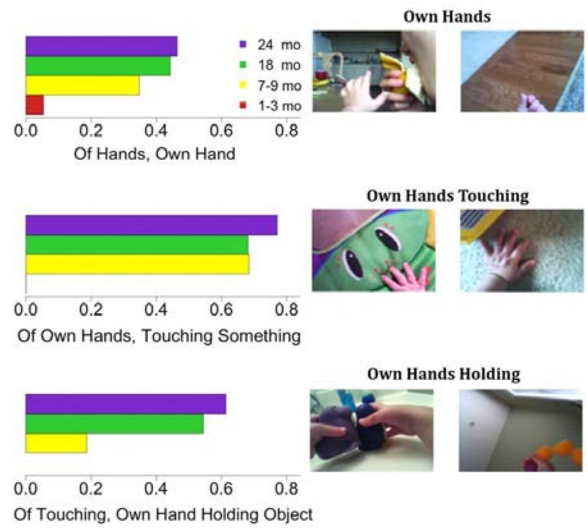

Figure 2: Ordered visual input of Hands

Hands in the visual environment

What kinds of hands are potentially in view for infants and toddlers? The answer to this question changes over the first two years of life. Here, we focus on three kinds of developmentally relevant input: hand identity, contact between hands and the world, and holding small objects. For each, we find patterns of ordered visual input: What is visually available early is replaced by something else later.

Hands Identity: Other-to-Own Whose hands are available in the visual input to infants and toddlers? Over the first two years of life, the child's own hands are increasingly available (Figure 2a). Early, the kinds of hands in the visual environment are overwhelmingly other people's hands, but by toddlerhood the child's own hands are nearly half the available hand visual input (1-3 months: .05 Own; 7-9 months: .35 Own; 18 months: .44 Own; 24 months: .46 Own), $\chi 2(3, N = 9096) = 152.64$, $p < .001$. The pattern is: First someone else's hands, then your own hands.

Hands Contact: Free-to-Touch When your visual environment includes your own hand, what else is potentially in view? Over the first two years of life, infants increasingly make manual contact with the world (Figure 2b). Early, hands are free -- flailing and reaching. The visual environment of 1-3 month-old infants does not include their own hands contacting the world. But, by 24 months, over two-thirds of the views of their own hands also include something they touch (1-3 months: 0 Touching; 7-9 months: .68 Touching; 18 months: .68 Touching; 24 months: .77 Touching), $\chi^2(3, N = 3936) = 61.76$, $p < .001$. The pattern is: First hands free, then hands touching the world.

Hands on Objects: Touch-to-Hold How often do early visual environments include one's own hand holding an object? Over the first two years of life, an increasing proportion of visual instances of touching the world are instances of holding objects (Figure 2c). Before 18 months, the visual environment includes few instances of hands together with objects (7-9 months: .19 of all touching instances). Toddlers' visual input, however, includes many of these instances (18 months: .54; 24 months: .61). That is, infants and toddlers find themselves in very different visual circumstances with respect to hands-on-objects, $\chi^2(2, N = 2814) = 141.61$, $p < .001$. The pattern is: First hands touching, then hands holding.

General Discussion

Our corpus of visual environments is unprecedented in scope: We are capturing visual regularities throughout the first two years of human life. Importantly, we capture environments *throughout* these two years, rather than zooming in to focus on one unique time, or zooming out to collapse across many different times. Everyday acting and thinking happens within nested timescales and complete theories of how environmental regularities matter for human cognition demand evidence from each scale: from realtime measures of in-the-moment attention through summaries of long-term experience. Our project provides critical insight into a scale currently missing from theories of statistical learning: *developmental* time.

Our data strongly suggest that input is dynamic and ordered. Visual regularities in developmental time may be a rolling wave of "See-Saw" patterns. Here, we see this across two classes of important social stimuli - faces, then hands. We also see this within hands, going from Other-to-Self-to-Touching-to-Holding. If our investigation into early visual statistics had zoomed into 3-month-old infants, we would have missed important regularities about hands; if we had zoomed out to batch statistics over the first two years of life, we would have concluded that the environments of infants and toddlers include roughly one-third body parts. Instead, we find a key pattern in environmental regularities: Developmental statistics are dynamic and ordered.

The available visual statistics are gated by young children's developmental level. A 3-month-old infant who is placed and carried finds herself in different visual environments than a walking, talking 24-month-old. Like other species, our data suggest that humans experience distinct ontogenetic niches as they progress toward adult-like motor, social, and language abilities. These visual niches may do a lot of important filtering for young learners: rather than sophisticated internal attentional control, "starting small" in structured input may be accomplished by other developmental constraints. Of course, the fact that developmental changes in many skills constrain the visual environment does not rule out the possibility of additional attentional gating. It may, however, reduce the challenges that attentional gating must resolve.

Does this temporal ordering of statistical regularities matter? It could be that outcomes at 2 years are best predicted by the total set of regularities and not by the order of those visual environments. Alternatively, some paths through the search space may be optimal, and mother-nature may optimize social learning by guiding the learner along optimal paths. More radically, the order of these experiences may not just enhance the optimal solution, but may determine the class of outcomes. Developmental process consists not just in the sampling of information but also in the change in the very internal structure of the learner. Considerable evidence from a psycho-biological perspective shows that the ordering and timing of sensory information play a critical role in brain development (Held & Hein, 1963; Lord, 2012; Turkewitz & Kenny, 2004). Reordering the usual sensory experiences within a developmental individual changes the architecture of the brain, not just what is known but what is knowable (Knudsen, 2006). A related idea, from cognitive theorists is the “starting small” hypothesis: limits that arise from the immaturity of the neural system constrain the input and, rather than holding back development, play a role in fostering development (Dominguez & Jacobs, 2003; Elman, 1993; Fox, Levitt & Nelson, 2010; Newport, 1990; Westermann, 2000). Between birth and 2 years, human infants travel through a set of highly distinct developmental environments determined first by their early immaturity and then by their growing emotional, motor, and cognitive competence. These ordered environments may help learning systems “start small,” find the optimal path to the optimal solution, and determine the architecture of the system that does the learning.

Acknowledgments

This work was supported by NICHHD 28675, NIH NRSA HD007475-18, and NIH R21HD068475. The authors wish to thank Char Wozniak and Ariel La for data collection.

References

Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., & Sotsky, R. B. (2012) How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science*.

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.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child development*, *62*(5), 874-890.

Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology*, *25*(6), 1004.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British journal of developmental psychology*, *9*(1), 55-72.

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. Developmental Science.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7),287-291.

de Barbaro, K., Chiba, A., & Deák, G. O. (2011). Micro-analysis of infant looking in a naturalistic social setting: insights from biologically based models of attention. *Developmental Science*, *14*(5), 1150-1160.

Dominguez, M., & Jacobs, R.A. (2003). Developmental constraints aid the acquisition of binocular disparity sensitivities. *Neural Computation*, *15*(1), 161-182.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71-99.

Fathi, A., Hodgins, J. K., & Rehg, J. M. (2012, June). Social interactions: A first-person perspective. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1226-1233). IEEE.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development*, *7*(1), 19-25.

Fox, S.E., Levitt, P., & Neslon III, C.A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child development, 81*(1), 28-40.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010, March). Head-mounted eye-tracking of infants' natural interactions: a new method. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 21-27). ACM.

Frank, M. C. (2012). Measuring children's visual access to social information using face detection. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.*

Geldart, S., Mondloch, C. J., Maurer, D., De Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science*, 5(4), 490-501.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544-549.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological review*, *114*(2), 211.

Hahn, E. R., & Gershkoff-Stowe, L. (2010). Children and adults learn actions for objects more readily than labels. *Language Learning and Development*, *6*(4), 283-308.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology, 56*(5), 872.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1-19.

Kanade, T. (2009). First-person, inside-out vision. In *IEEE Workshop on Egocentric Vision, CVPR* (Vol. 1).

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084-1089.

Knudsen, E. I. (2006). Neural Derivation of Sound Source Location in the Barn Owl. *Annals of the New York Academy of Sciences, 510*(1), 33-38.

Kretch, K., Franchak, J., Brothers, J., & Adolph, K. (2012). What infants see depends on locomotor posture. *Journal of Vision*, *12*(9), 182-182.

Leung, E. H., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology*, *17*(2), 215

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, *19*(1), 577-621.

Lord, K. (2012). A Comparison of the Sensory Development of Wolves (Canis lupus lupus) and Dogs (Canis lupus familiaris). *Ethology*.

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*,*19*(2), 263-296.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, *16*, 1-27.

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255-260.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75-78.

Moll, H., & Tomasello, M. (2010). Infant cognition. Current *Biology, 20*(20), R872-R875.

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child development,69*, 295-308.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. The MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*,*28*(1), 73-193.

Quinn, P. C., Eimas, P. D., & Tarr, M. J. (2001). Perceptual categorization of cat and dog silhouettes by 3-to 4-month-old infants. *Journal of experimental child psychology*, *79*(1), 78-94.

Rader, N. D. V., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony.*Language Sciences*.

Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science*.

Rumelhart, D. E., McClelland, J. L., & CORPORATE PDP Research Group. (1986). Parallel distributed processing: explorations in the microstructure of cognition, vol. 2: psychological and biological models.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants.

Slater, A. M. (1999). *Perceptual development: Visual, auditory and speech perception in infancy*. Psychology Press.

Smith, L. B., Yu, C., & Pereira, A. F. (2010). Not your mother's view: The dynamics of toddler visual experience. *Developmental science, 14*(1), 9-17.

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life*, *11*(1-2), 13-29.

Sommerville, J. A., & Woodward, A. L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition*, *95*(1), 1-30.

Stern, D. N. (1971). A Micro-Analysis of Mother-Infant Interaction. Behavior Regulating Social Contact Between a Mother and her 3 1/2 Month-Old Twins. *Journal of the American Academy of Child Psychiatry*, *10*(3), 501-517.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences*, *10*(1), 69-88.

Turkewitz, G., & Kenny, P.A. (2004). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology, 15*(4), 357-368.

West, M. J., & King, A. P. (1987). Settling nature and nurture into an ontogenetic niche. *Developmental psychobiology*, *20*(5), 549-562.

Westermann, G. (2000). *Constructivist neural network models of cognitive development* (Doctoral dissertation, University of Edinburgh).

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34.

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.