The developing infant creates a curriculum for statistical learning

Linda B. Smith[*], **Swapnaa Jayaraman**, **Elizabeth Clerkin**, and **Chen Yu**

Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405, USA

Abstract

New efforts are using head-cameras and eye-trackers worn by infants to capture everyday visual environments from the infant learner's point of view. From this vantage point, the training sets for statistical learning develop as the infant's sensory-motor abilities develop, yielding a series of ordered data sets for visual learning that differ in content and structure between time points but are highly selective at each time point. These changing environments may constitute a developmentally ordered curriculum that optimizes learning across many domains. Future advances in computational models are needed to connect the developmentally changing content and statistics of infant experience to the internal machinery that does the learning.

What are the data for learning?

The world presents learners with many statistical regularities. Considerable evidence indicates that humans are adept at discovering those regularities across many different domains including language, vision, and social behavior [1]. Accordingly, there is much interest in "statistical learning" and how learners discover regularities from complex data sets like those encountered in the world. Much of this interest is directed to the statistical learning abilities of human infants. In the first two years of life, infants make strong starts in language, in visual object recognition, in using and understanding tools, in social behaviors, and more. By the benchmarks of speed, amount, diversity, robustness, and generalization, human infants are powerful learners. Any theory of statistical learning worth its salt needs to address the growth in knowledge that occurs during infancy.

All statistical learning depends on both the internal machinery that does the learning and the regularities in the data on which that machinery operates. The usual assumption is that the

[*]Correspondence: smith4@indiana.edu (L.B. Smith).

learning environment is rich but noisy with data for many different tasks all mixed together. Thus, the main theoretical debates concern the nature of the learning machinery and the constraints that enable learners to sort through and learn from messy data [2, 3]. Contemporary research takes two principle forms: (a) laboratory experiments with human participants that test predictions about hypothesized learning mechanisms [4,5] and (b) computational models showing that the hypothesized machinery can use key statistical regularities [6,7] to accomplish certain learning tasks. However, the relevance of these efforts to understanding the prowess of infants' everyday learning is limited because the data sets used in most experimental studies and in computational modeling differ from the data on which everyday learning depends (Figure 1).

What are the data relevant to infant statistical learning? They are the data that make contact with the infant sensors. These sensors are attached to the body and thus the sensory data for learning are *determined* by the disposition of the infant's body in space [Box 1]. Advances in wearable sensors now provide researchers with the means to capture the learning environment from the perspective of the learner [8–11]. Accordingly, and in contrast to the usual focus on learning mechanisms, this review focuses on the data for statistical learning from the infant's perspective.

Box 1

Ego-centric Vision

Real world visual experience is tied to the body as it acts in the world. As a consequence, the learner's view of the nearby environment is highly selective. The colored objects in Figure I are in the depicted infant's view and would be captured by the head camera. Many things in the room and spatially near the infant are not in view. Unless the infant turns her head and looks, the dog, the train, her mother's face, the window, and the crib are not in view. The perceiver's location, posture, and ability or motivation to change their posture systematically bias first-person visual information. The emerging area of research called ego-centric vision studies fundamental questions in vision from the perspective of freely moving individuals, and growing evidence suggests that many of the properties are fundamentally different from those found in laboratory studies that restrict body movements [17, 69–72]. For example, freely moving perceivers use their whole body to select visual information with eyes and head aligned. Gaze is predominantly directed to the center of a head-centered frame of reference, making head cameras – with or without eye-trackers – a useful method for capturing the ego-centric view [71–73]. A second example concerns the study of "natural statistics" of vision. Considerable progress in understanding adult vision has been made by studying the visual statistics of "natural scenes", and the sensitivities of the mature visual system appear biased to detect the statistically prevalent features [74]. However, the "natural scenes" used to determine these natural statistics are not the ego-centric perspective of perceivers as they move about the world; rather, they are photographs taken by adults and thus biased by the already mature visual system and the mature body that stands still and holds the camera to frame a picture. There is as no evidence as to whether the natural statistics of adult egocentric scenes align or differ from those of the adult-photographed world, but it seems

increasingly likely that those statistics – from the point of view of infants and children -- will change with development.
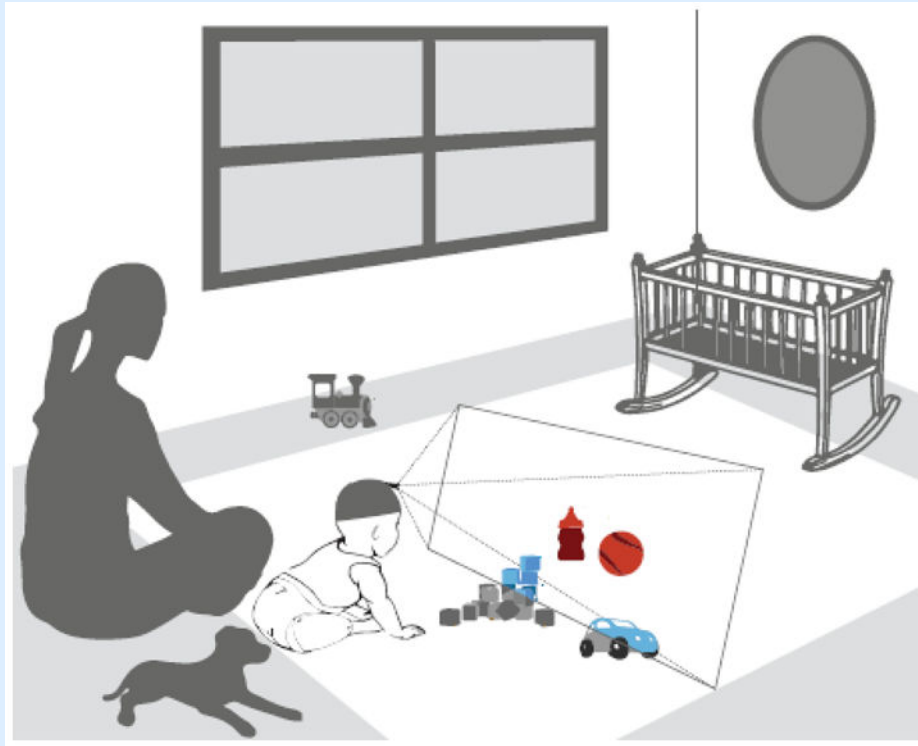
Fig I.
Egocentric vision is highly selective because it depends on the momentary location and posture of the perceiver.

The data for learning changes with development

The infant's personal view of the world changes as the infant's sensory-motor abilities change. Each new sensory-motor achievement – rolling over, reaching, crawling, walking, manipulating objects – opens and closes gates, selecting from the external environment different data sets to enter the internal system for learning (Figure 1). Studies using head cameras and head-mounted eye-trackers consistently show the tight dependence between the infants own developing abilities and visual experience. For example, newborns have limited acuity [12] and can do very little with their body. Much of what they see depends on what caregivers put in front of and close to the infant's face. What caregivers often put in front their infant is their own face [13–15]. In contrast, an older crawling baby can see much farther and can move to a distant object for a closer view. When moving, the crawler creates new patterns of dynamic visual information or optic flow [16]. However, when crawling, the infant sees only the floor and has to stop crawling and sit up to see social partners or the goal object [17, 18]. When older infants manually play with objects they create a data set of many different views of a single object, experiences that have been related to advances in object recognition and object name learning [19, 20]. Taken together, these studies, which focus on

specific behaviors and contexts show that infants changing abilities select and create data for learning that changes with development.

These age-related changes in abilities and contexts for learning create large-scale changes in the contents of everyday experience. To capture the natural statistics of experience at this larger scale, researches have embedded head cameras in hats and sent them home with infants to be worn by them during their daily lives [11–15, 21]. Analyses of these home-collected head-camera images reveal marked changes in infant visual experiences as they progress from 1 month to 24 months of age. For example, although people are commonly present in the head camera images of very young and older infants, the body parts in view change systematically with the age of the head-camera wearer [14]. The images of people captured by infants under 3 months of age showed predominantly close and frontal views of faces. However, by the time infants reach their first birthday, faces were rarely in the captured images [13]. Infants under 3 months of age received an average of 15 minutes of face views out of every hour of head-camera recording. In contrast, one-year old infants received only 6 minutes of faces views for every hour of head-camera recording. The finding that other people's faces are rarely in the toddler's egocentric view was a surprising result when first reported [22] but one that has now been documented by many different researchers in a variety of contexts using both head cameras and head-mounted eye-trackers [e.g., 17, 18, 23, 24]. After the first birthday, when people's faces continue to be rare, people's hands become pervasively present in the infant's first-person view [11]. In 80% of the cases when another person's hands are in the older infant's view, those hands are in in contact with and acting on an object. The pervasiveness of hands performing instrumental acts aligns with these older infants' increasingly sophisticated manual skills [25] and their increasingly sophisticated understanding of the goal-directed manual acts [26].

The learner's view of the world depends on learner's posture, location, and behavior. The infants body and sensory-motor abilities change systematically and markedly during the first two years of post-natal life. These changes create statistical data for learning that are partitioned into distinct sets: first faces and then hands (and their instrumental acts on objects).

Timing matters

A large literature shows that very young infants are learning a lot about faces. By the end of the first 3 months, infants' visual systems are biased to the specific properties of the specific faces in their environments as they preferentially discriminate their caregivers' faces and recognize and discriminate faces that are similar to their caregivers' in race and gender [27, 28]. Moreover, studies of infants who do not experience an early visual world dense with faces indicate that these early experiences may be critical to the development of mature face processing [29–31]. One line of relevant evidence comes from individuals with congenital cataracts that were removed by the time they were 4 to 6 months of age [30,31]. These individuals lack the early period of visual experience dense with close frontal views of faces. Many visual abilities, including some relevant to visual face recognition, show no long-term deficits. However, individuals with congenital cataracts removed as early as 4 months show permanent deficits in configural face processing. Configural face processing – the sensitivity

to second-order relations that adults can use effectively only with upright faces and only when low spatial frequencies are present – is a late developing property of the human visual system, one that only emerges in childhood and is not fully mature until adolescence. Thus, the deficit in early visual experience caused by congenital cataracts is characterized as a sleeper effect, a late emerging consequence of a much earlier sensory deprivation [31]. Deficits in face processing may not be the only sleeper effects resulting from limited early experience with close frontal views of faces. Adults who were born with cataracts removed before they were 4 months of age also show deficits in the dynamic processing of sight-sound synchronies [32]. This deficit may also derive from the lack of the early evolutionarily-expected experiences of close faces and the dynamically coupled audio-visual experiences caregivers generate when they put their faces in front of those their young infants [33].

The findings about the developmentally changing density of faces in infant's first-person views should constrain theories about the internal processes that underlie human face processing. In very early infancy, those processes work on substantial but not massive amount of data. By the end of the first 3 months, using the estimates of 15 minutes of face time per hour and 12 waking hours a day, an infant would have experienced 270 hours of predominately close frontal views of faces [13]. The consequences of missing these 270 hours of experience are permanent deficits that apparently cannot be counteracted by a lifetime of seeing faces. Thus, the first three months of postnatal life may be characterized as a sensitive period for face processing, a period of development in which specific experiences have an out-sized effect on long-term outcome. Sensitive periods may reflect fundamental changes in neural plasticity [34]. Developing environments – the gates to sensory input that open and close as the infant develops – may also create sensitive periods. Infants who have had their cataracts removed at 4 months may not "catch up" in configural face processing because they do not encounter the same structured data set: dense close frontal views of faces. Because the ego-centric view depends on the infant's own sensory-motor, cognitive, and emotional level of development, the gate on dense close experiences of faces may have been closed by the infant's own more mature behavior and interests.

Sampling is selective

An individual learner samples the information in the world from a localized perspective. Thus, of all the faces in the world that an infant might see, she is most likely to see the faces of family members because they are often in the same location as the young infant. An individual learner also samples the information in the world through the lens of their own actions in that world. Thus, of all the cups in a toddler's house, the toddler who can only drink without spilling from a sippy cup, is mostly likely to see that type of cup. An extensive literature [21, 35–37] shows that everyday learning environments are characterized by frequency distributions in which a very few types (the mother's face, the sippy cups) are very frequent, but most types (all the different faces encountered at a grocery store, all the cups in the cupboard) are encountered quite rarely.

Analyses of images from head cameras worn in the home show these characteristic extremely-skewed frequency distributions in which a very few instances are highly frequent

and most other instances are encountered quite rarely. Thus, the visual world of very young infants is not just dense with faces but with the faces of a very few people: just 2 to 3 individuals account for over 80% of all the images in which faces appear [13]. Although older infants' head camera images have relative few with faces in them, the faces of just three individuals account for 80% of the faces that appear [13]. Nonetheless, from this highly selective and non-uniform sampling of faces, infants become able to learn to recognize and discriminate faces in general. Because these skewed distributions are common in natural environments and because infants readily learn in these everyday environments, we can be confident that infants are equipped with learning machinery that learns from these kinds of data. How does this work?

The frequency distribution of everyday objects captured in infant head-camera images provides a case example for thinking about the question. The analyzed images for this case example were collected by 8- and 10- month old infants as they went about their daily activities [21]. Infants this age are beginning to sit steadily, to crawl, and to play with objects, but their manual skills are still quite limited as compared to older infants. Analyses of in-home head camera images for these infants suggest that neither faces nor hands acting on objects are statistically dominant; instead, their head camera images contain mixture of the various body parts of nearby people [11, 14]. Analyses of mealtimes scenes for these infants show them to be highly cluttered [21], each scene containing many different objects (Figure 2). This clutter poses an interesting theoretical problem. In laboratory studies, infants this age look to named objects upon hearing those names [38, 39]. They must have learned these names by linking heard names to seen objects. But given highly cluttered scenes, how could they know the object being named?

The right-skewed frequency distribution of objects in the images provides a straightforward solution to this problem. Of the many objects in each of these cluttered scenes, a very small number were pervasively present across the corpus of 8- to 10-month-olds' mealtime scenes. The frequency distribution [Figure 2] was extremely skewed such that a very small number of object categories (spoons, bottles, sippy cups, bowls, yogurt) were repeatedly present whereas most object categories (jugs, salad tongs, catsup bottles) were present in only a very few scenes. The pervasive repetition of instances from a select set of categories could help infants find and attend to that select group of categories in the clutter of many other things, and in this way provide a foundation for linking those objects to their names [21]. There are several interrelated hypotheses as to how the extremely skewered frequency distributions may facilitate statistical learning (Figure 3).

Consistency

Learners could just ignore the rarer items and in so doing create a small and consistent training set [21]. However, the items that comprise this small high-frequency set are also likely to change with development. For example, cheerios but not grapes are pervasive first finger foods for 8- to 10-month-old infants who are in danger of choking; in contrast, grapes and chicken nuggets are plentiful for two-year-old children. If the high frequency objects in the infant's view change systematically with development, then learners would be presented with a series of small lessons of consistent and repeated training items.

Bootstrapping

Low frequency items may be learned along with high frequency items because they benefit being encountered in the context of high frequency items [36, 40, 41]. For example, rarer object categories in scenes with high frequency object categories may be segmented more robustly even in a cluttered scene. The resolution of ambiguities – such as the partial occlusion of a known object by an unknown one -- may be more rapidly resolved because of the constraints provided by the better-known higher frequency object.

Desirable difficulties

A general problem in statistical learning is overfitting a solution to the specific items used in training. For example, if all the cups a toddler sees are green, then the toddler could use color to identify cups. However, the many different low frequency items – that are green but not cups – would help prevent this overfitting, enabling the learner to find the right features for representing the high frequency items. Properties of training data that would seem to make learning more difficult --- clutter or distractors -- but that actually make that learning more robust, have been called "desirable difficulties" [42,43].

Infants create a curriculum for learning

Although infants begin learning object names prior to their first birthday, by all indications this learning is fragile, errorful, and progresses slowly [44–46]. However, the rate and nature of learning changes noticeably as infants reach their second birthday [46]. Laboratory studies show that 2-year-olds can rapidly infer the extension of a whole category from a single instance of that category. For example, if a 2-year-old child encounters their very first tractor – say, a green John Deere working in a field – while hearing its name, the child is likely from that point forward to recognize all variety of tractors as tractors – red Massey-Fergusons, rusty antique tractors, ride-on mowers – but not backhoes or trucks [48 – 50]. What changes? The child most certainly has changed in cognitive and language skills. The visual learning environment has also changed in ways that may support the rapid learning and generalization of an object name.

The early slow learning about a few objects and their names before first birthday may be Lesson 1. The visual pervasiveness of a select set of objects in a cluttered visual field may build strong visual memories for these few things enabling infants them to remember seen things and their heard names [44, 51]. Lesson 2 may use very different training data. After their first birthday, toddlers' egocentric views differ from those of 8- to 10- month olds as well as from adults. Toddlers visual experiences of objects are shaped by what they can manually do with those objects [52–54]: Hand actions on objects create visual scenes in which the acted upon objects are visually foregrounded, often centered in the field of view. When the toddler acts on the object, the object is visually large (because it is close, given the very short arms of toddlers) in the field of view, and by being close and large in the visual field, the held object often obscures the clutter in the background. These scenes with a clear focal object are often long lasting (c. 4 secs) and coincide with the stilling of head movements by the infant [52]. They also invite joint attention with a mature social partner and the naming of those objects by the social partner [52, 55]. These less cluttered scenes

and the obvious interest of the infant in the held object to the mature partner may be major factors in the rate of object name learning after the first birthday. The training data are better for 1½-year-old infants than for 10-month-old infants: they contain more frequent naming by mature social partners and less visual ambiguity as to the intended referent of the heard name.

Other evidence shows that visual object recognition and object name learning co-develop [19, 20, 50]. Infants first-person views of their own object manipulations also provides a very rich data set for learning about 3-dimensional object shape. The character and frequency of toddler object manipulations have been shown to predict object memory, object recognition, shape processing, object name learning, and vocabulary size. In brief, toddlers' active manual engagement with objects may create and elicit a learning environment specialized for learning about visual objects and their names.

What is a developmental approach to statistical learning?

This review focused on the data for learning from the infant's perspective and in everyday contexts as a corrective to the over-emphasis on the internal learning mechanisms without regard to the natural statistics of the learning environment. However, the data for learning and the machinery that does that learning cannot be studied separately. The meaning of the regularities in infant everyday learning environments can only be determined through the learning mechanisms that operate on the data. Many influential approaches to statistical learning are inherently non-developmental, assuming a computational problem, a data set, and a learning machinery that is more or less constant. Such models cannot exploit the developmentally changing and ordered regularities observed in infant visual experience. One classic developmental approach to statistical learning that has been implemented in computational models is the "starting small" hypothesis [56]: The idea is that infant's learning machinery is limited by weak memory and attentional processes and that these weaknesses yield better statistical learning and generalization because only learning only operates on the most pervasive statistical regularities. This hypothesis has support [57, 58] and may be part of the explanation of why infants in some domains (e.g., language) appear better learners than adults. However, the premise of the "starting small" hypothesis is that the external data for learning is the same for younger and older learners what differs is their internal processes. Thus, these models also do not address the developmental changes in the data delivered by the sensory system to that learning machinery.

There are contemporary computational approaches to learning, most of which are not concerned with infants, that do consider learning from ordered data sets and thus may provide some guidance as to the kinds of models that could exploit the changing content and structure in infant perspective scenes. Machine learning approaches such as curriculum learning and iterative teaching explicitly seek to optimize learning by ordering the training material in time [59–61]. Other approaches model statistical learning and inference in data that emerges in time [62, 63]. Active learning approaches focus [64] on the learner's role in selecting the training sequence and how learners could, through their own behavior, select the information that is optimal, given the current state of knowledge, for moving learning forward. Current approaches within this larger framework include training attention in deep

learning networks such that the data selected for learning changes with learning [65] and using curiosity to shift attention to new learning problems as learning progresses [67–68].

What is the future of statistical learning? Ego-centric vision is just one approach to studying everyday learning environments from the perspective of individual learners [Box 2]. Future theories of statistical learning will need to handle these natural developmental statistics by connecting the internal machinery with statistics selected by developing learners.

Box 2

Advances in the study of infant environments

The study of infant learning environments is expanding at a rapid pace and in many directions. Researchers are using wearable sensors on shoes, motion trackers, and head-mounted eye trackers, to study the self-generated activities of learning to walk [10, 75]. Infant walkers – both skilled and novice – take several thousand steps a day, fall 17 times *an hour*, and rarely walk in a straight line. They walk forward, backward, sideways, and typically with no obvious goal for their movement. The training activities for learning to walk bear little similarity to traditional assessments of walking skill or to hypotheses about how infants become skilled walkers. Other researchers recorded parent talk to children in structured play contexts and in unstructured settings [76]. Parents talked differently during unstructured and structured activities; during unstructured activities, there was more silence, less talk overall, and much less diversity in the words used. Yet structured activities with children have thus far been the main contexts for studying language learning environments. These and other findings [77] about contextual variations in the statistics of parent talk offer a cautionary note that the field may know less – or have wrong assumptions – about the statistics for language learning. Fueled by the considerable evidence showing that the amount of parent talk in the early years determines vocabulary development and *future* school achievement [78,79], exciting new efforts are underway to capture language learning environments at scale using day long recording systems [80].

Conclusion

Learning, and development, is a personal journey. The learner's personal vantage point determines the data for learning for that individual. Emerging evidence from studies of infant ego-centric vision suggest that the structure of the infant's changing personal view is key to how and why their learning – across so many different tasks – is so robust. From the infant's perspective, different learning problems are segregated in time. There is extensive training on a small set of instances and rarer encounters with many others. The infant's own changing sensory-motor abilities open and close environments for learning. The developing infant's increasing autonomy puts the infant in control of generating the data sets for learning. The developmental study of the changing statistical structure of learning environments may also yield a deeper understanding of individual differences in early cognitive development. The source of differences – and interventions to support healthy development in all children -- may emerge in part from the developmental structure of the

data for learning, data that is determined by the learner's immediate surroundings and the learner's developing behaviors in those surroundings.

Acknowledgments

References

1. Aslin RN. Statistical learning: a powerful mechanism that operates by mere exposure. Wiley Interdisciplinary Reviews: Cognitive Science. 2017; 8(1–2):1–7. DOI: 10.1002/wcs.1373

2. Griffiths TL, et al. Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences. 2010; 14(8):357–364. [PubMed: 20576465]

3. McClelland JL, et al. Letting structure emerge: Connectionist and dynamical systems approaches to cognition. Trends in Cognitive Sciences. 2010; 14(8):348–356. [PubMed: 20598626]

4. Saffran JR, Kirkham NZ. Infant Statistical Learning. Annual Review of Psychology. 2017; 69(1)

5. Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: The paradox of statistical learning. Trends in cognitive sciences. 2015; 19(3):117–125. [PubMed: 25631249]

6. Gopnik A, Bonawitz E. Bayesian models of child development. Wiley interdisciplinary reviews: cognitive science. 2015; 6(2):75–86. [PubMed: 26263064]

7. Blythe RA, Smith AD, Smith K. Word learning under infinite uncertainty. Cognition. 2016; 151:18–27. [PubMed: 26927884]

8. Allé MC, et al. 2017Wearable cameras are useful tools to investigate and remediate autobiographical memory impairment: a systematic PRISMA review. Neuropsychology review. :1–19. [PubMed: 27878426]

9. Yu C, Smith LB. The social origins of sustained attention in one-year-old human infants. Current Biology. 2016; 26(9):1235–1240. [PubMed: 27133869]

10. Lee DK, et al. The cost of simplifying complex developmental phenomena: a new perspective on learning to walk. Developmental Science. 2017

11. Fausey CM, Jayaraman S, Smith LB. From faces to hands: Changing visual input in the first two years. Cognition. 2016; 152:101–107. [PubMed: 27043744]

12. Braddick O, Atkinson J. Development of human visual function. Vision research. 2011; 51(13):1588–1609. [PubMed: 21356229]

13. Jayaraman S, Fausey CM, Smith LB. The faces in infant-perspective scenes change over the first year of life. PloS one. 2015; 10(5):e0123780. [PubMed: 26016988]

14. Jayaraman S, Fausey CM, Smith LB. Why are faces denser in the visual experiences of younger than older infants? Developmental psychology. 2017; 53(1):38. [PubMed: 28026190]

15. Sugden NA, Moulson MC. Hey baby, what's "up"? One-and 3-month-olds experience faces primarily upright but non-upright faces offer the best views. The Quarterly Journal of Experimental Psychology. 2017; 70(5):959–969. [PubMed: 26912287]

16. Ueno M, et al. Crawling Experience Relates to Postural and Emotional Reactions to Optic Flow in a Virtual Moving Room. Journal of Motor Learning and Development. 2017:1–24.

17. Kretch KS, Franchak JM, Adolph KE. Crawling and walking infants see the world differently. Child development. 2014; 85(4):1503–1518. [PubMed: 24341362]

18. Franchak JM, Kretch KS, Adolph KE. See and be seen: Infant – caregiver social looking during locomotor free play. Developmental science. 2017; doi: 10.1111/desc.12626

19. James KH, et al. Young children's selfD generated object views and object recognition. Journal of Cognition and Development. 2014a; 15(3):393–401. [PubMed: 25368545]

20. James KH, et al. Some views are better than others: Evidence for a visual bias in object views self-generated by toddlers. Developmental Science. 2014b; 17(3):338–351. [PubMed: 24410976]

21. Clerkin EM, et al. Real-world visual statistics and infants' first-learned object names. Phil Trans R Soc B. 2017; 372(1711):20160055. [PubMed: 27872373]

22. Yoshida H, Smith LB. What's in view for toddlers? Using a head camera to study visual experience. Infancy. 2008; 13(3):229–248. [PubMed: 20585411]

23. Deak GO, et al. Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. Developmental science. 2014; 17(2):270–281. [PubMed: 24387193]

24. Yu C, Smith LB. Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloS one. 2013; 8(11):e79659. [PubMed: 24236151]

25. Lockman JJ, Kahrs BA. New Insights Into the Development of Human Tool Use. Current directions in psychological science. 2017; 26(4):330–334. [PubMed: 28943723]

26. Gerson SA, Woodward AL. Learning from their own actions: The unique effect of producing actions on infants' action understanding. Child development. 2014; 85(1):264–277. [PubMed: 23647241]

27. Lee K, Quinn PC, Pascalis O. Face race processing and racial bias in early development: A perceptual-social linkage. Current Directions in Psychological Science. 2017; 26(3):256–262. [PubMed: 28751824]

28. Rennels JL, et al. Caregiving experience and its relation to perceptual narrowing of face gender. Developmental Psychology. 2017; 53(8):1437. [PubMed: 28594188]

29. Moulson MC, et al. The Effects of Early Experience on Face Recognition: An Event-Related Potential Study of Institutionalized Children in Romania. Child development. 2009; 80(4):1039–1056. [PubMed: 19630892]

30. Maurer D. Critical periods re-examined: Evidence from children treated for dense cataracts. Cognitive Development. 2017; 42:27–36.

31. Maurer D, Mondloch CJ, Lewis TL. Sleeper effects. Developmental Science. 2007; 10(1):40–47. [PubMed: 17181698]

32. Chen YC, et al. Early binocular input is critical for development of audiovisual but not visuotactile simultaneity perception. Current Biology. 2017; 27(4):583–589. [PubMed: 28190731]

33. Bremner AJ. Multisensory development: calibrating a coherent sensory milieu in early life. Current Biology. 2017; 27(8):R305–R307. [PubMed: 28441564]

34. Oakes LM. Plasticity may change inputs as well as processes, structures, and responses. Cognitive Development. 2017; 42:4–14. [PubMed: 28993717]

35. Montag JL, Jones MN, Smith LB. Quantity and diversity: Simulating early word learning environments. Cognitive Science. 2018

36. Salakhutdinov, R., Torralba, A., Tenenbaum, J. Learning to share visual appearance for multiclass object detection. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on; IEEE; 2011. p. 1481-1488.

37. Piantadosi ST. Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic Bulletin & Review. 2014; 21(5):1112–1130. [PubMed: 24664880]

38. Bergelson E, Swingley D. Early word comprehension in infants: Replication and extension. Language Learning and Development. 2015; 11(4):369–380. [PubMed: 26664329]

39. Bergelson E, Aslin R. Semantic Specificity in One-Year-Olds' Word Comprehension. Language Learning and Development. 2017; 13(4):481–501. [PubMed: 29200981]

40. Kurumada C, Meylan SC, Frank MC. Zipfian frequency distributions facilitate word segmentation in context. Cognition. 2013; 127(3):439–453. [PubMed: 23558340]

41. Kachergis G, Yu C, Shiffrin RM. A bootstrapping model of frequency and context effects in word learning. Cognitive Science. 2017; 41(3)

42. Bjork RA, Kroll JF. Desirable difficulties in vocabulary learning. The American Journal of Psychology. 2015; 128(2):241. [PubMed: 26255443]

43. Twomey KE, Ma L, Westermann G. All the right noises: Background variability helps early word learning. Cognitive Science. 2017

44. Kucker SC, Samuelson LK. The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. Infancy. 2012; 17(3):295–323. [PubMed: 22661907]

45. Vlach HA, Johnson SP. Memory constraints on infants' cross-situational statistical learning. Cognition. 2013; 127(3):375–382. [PubMed: 23545387]

46. Bobb SC, Huettig F, Mani N. Predicting visual information during sentence processing: Toddlers activate an object's shape before it is mentioned. Journal of experimental child psychology. 2016; 151:51–64. [PubMed: 26687440]

47. Frank MC, et al. Wordbank: An open repository for developmental vocabulary data. Journal of child language. 2017; 44(3):677–694. [PubMed: 27189114]

48. Landau B, Smith LB, Jones SS. The importance of shape in early lexical learning. Cognitive development. 1988; 3(3):299–321.

49. Smith LB. It's all connected: Pathways in visual object recognition and early noun learning. American Psychologist. 2013; 68(8):618–629. [PubMed: 24320634]

50. Yee M, Jones SS, Smith LB. Changes in visual object recognition precede the shape bias in early noun learning. Frontiers in psychology. 2012; 3:533.doi: 10.3389/fpsyg.2012.00533 [PubMed: 23227015]

51. Vlach HA. How we categorize objects is related to how we remember them: the shape bias as a memory bias. Journal of experimental child psychology. 2016; 152:12–30. [PubMed: 27454236]

52. Yu C, Smith LB. Embodied attention and word learning by toddlers. Cognition. 2012; 125(2):244–262. [PubMed: 22878116]

53. Pereira AF, Smith LB, Yu C. A bottom-up view of toddler word learning. Psychonomic bulletin & review. 2014; 21(1):178–185. [PubMed: 23813190]

54. Bambach, S., et al. Objects in the center: How the infant's body constrains infant scenes. Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2016 Joint IEEE International Conference on; IEEE; 2016. p. 132-137.

55. Yu C, Smith LB. Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloS one. 2013; 8(11):e79659. [PubMed: 24236151]

56. Elman, Jeffrey L. Learning and development in neural networks: The importance of starting small. Cognition. 1993; 48(1):71–99. [PubMed: 8403835]

57. Yurovsky D, et al. Probabilistic cue combination: Less is more. Developmental science. 2013; 16(2):149–158. [PubMed: 23432826]

58. Kam CLH, Newport EL. Getting it right by getting it wrong: When learners change languages. Cognitive psychology. 2009; 59(1):30–66. [PubMed: 19324332]

59. Bengio, Y., et al. Curriculum learning. Proceedings of the 26th annual international conference on machine learning; ACM; 2009. p. 41-48.

60. Gregor K, et al. DRAW: A recurrent neural network for image generation. 2015 arXiv preprint arXiv:1502.04623.

61. Krueger KA, Dayan P. Flexible shaping: How learning in small steps helps. Cognition. 2009; 110(3):380–394. [PubMed: 19121518]

62. Thiessen ED. What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. Phil Trans R Soc B. 2017; 372(1711):20160056. [PubMed: 27872374]

63. Sanborn AN, Griffiths TL, Navarro DJ. Rational approximations to rational models: alternative algorithms for category learning. Psychological review. 2010; 117(4):1144. [PubMed: 21038975]

64. Agrawal P, et al. Learning to poke by poking: Experiential learning of intuitive physics. Advances in Neural Information Processing Systems. 2016:5074–5082.

65. Mnih V, Heess N, Graves A. Recurrent models of visual attention. Advances in neural information processing systems. 2014:2204–2212.

66. Oudeyer PY, Smith LB. How Evolution may work through Curiosity-driven Developmental Process. Topics in Cognitive Science. 2016; 8:492–502. [PubMed: 26969919]

67. Kidd C, Hayden BY. The psychology and neuroscience of curiosity. Neuron. 2015; 88(3):449–460. [PubMed: 26539887]

68. Twomey KE, Westermann G. Curiosity-based learning in infants: a neurocomputational approach. Developmental science. 2017; doi: 10.1111/desc.12629

69. Schmitow C, Stenberg G. What aspects of others' behaviors do infants attend to in live situations? Infant Behavior and Development. 2015; 40:173–182. [PubMed: 26164419]

70. Solman GJ, Foulsham T, Kingstone A. Eye and head movements are complementary in visual selection. Royal Society Open Science. 2017; 4(1):160569. [PubMed: 28280554]

71. Fathi, A., Ren, X., Rehg, JM. Learning to recognize objects in egocentric activities. Computer Vision and Pattern Recognition (CVPR); 2011 IEEE Conference On; IEEE; 2011 Jun. p. 3281-3288.

72. Foulsham T, Walker E, Kingstone A. The where, what and when of gaze allocation in the lab and the natural environment. Vision research. 2011; 51(17):1920–1931. [PubMed: 21784095]

73. Smith LB, Yu C, Yoshida H, Fausey CM. Contributions of head-mounted cameras to studying the visual environments of infants and young children. Journal of Cognition and Development. 2015; 16(3):407–419. [PubMed: 26257584]

74. Simoncelli EP. Vision and the statistics of the visual environment. Current opinion in neurobiology. 2003; 13(2):144–149. [PubMed: 12744966]

75. Adolph KE, et al. How do you learn to walk? Thousands of steps and dozens of falls per day. Psychological science. 2012; 23(11):1387–1394. [PubMed: 23085640]

76. Tamis-LeMonda CS, et al. Power in methods: language to infants in structured and naturalistic contexts. Developmental science. 2017; 20(6):e12456.

77. Montag JL, Jones MN, Smith LB. The words children hear: Picture books and the statistics for language learning. Psychological science. 2015; 26(9):1489–1496. [PubMed: 26243292]

78. Newman RS, Rowe ML, Ratner NB. Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. Journal of child language. 2016; 43(5):1158–1173. [PubMed: 26300377]

79. Weisleder A, Fernald A. Talking to children matters: Early language experience strengthens processing and builds vocabulary. Psychological science. 2013; 24(11):2143–2152. [PubMed: 24022649]

80. VanDam, M., et al. Seminars in speech and language. Vol. 37. Thieme Medical Publishers; 2016. HomeBank: An online repository of daylong child-centered audio recordings; p. 128-142.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- The nature of the environment that supports learning is fundamental to understanding human cognition. Advances in wearable sensors are enabling research to study the everyday environments of infants at scale and with precision.

- Egocentric vision is an emerging field that uses head cameras and head-mounted eye trackers to study visual environments from the view of acting and moving perceivers.

- Studies of infant visual environments from the first-person view show that these environments change systematically with development, effectively creating a curriculum for learning.

- The structure of infant visual environments both challenges current assumptions of statistical learning and can also inform computational models of statistical learning.

Outstanding questions

- How can we experimentally test whether and how the structures found in first-person recordings of infant visual environments provide the required curriculum for infant learning?

- What are the real-time properties of data used for learning that change both over developmental time and over the learners' real time activities? How do they interact with potentially changing or different learning mechanisms?

- Does the order of developmentally segregated data sets – such as first faces and then objects matter to developmental outcomes? Do early face experiences support later visual development in other domains, in object perception, in letter recognition?

- If sensitive periods are formed in part by the closing of sensory-motor gates on critical experiences, can a sensitive period for learning be re-opened by re-opening those sensory-motor gates?

- What role do disruptions in the real-world data for learning play in the cognitive developmental trajectories of children with developmental disorders? This will shed light on the cognitive developmental disorders that are characterized by atypical patterns of sensory-motor development.
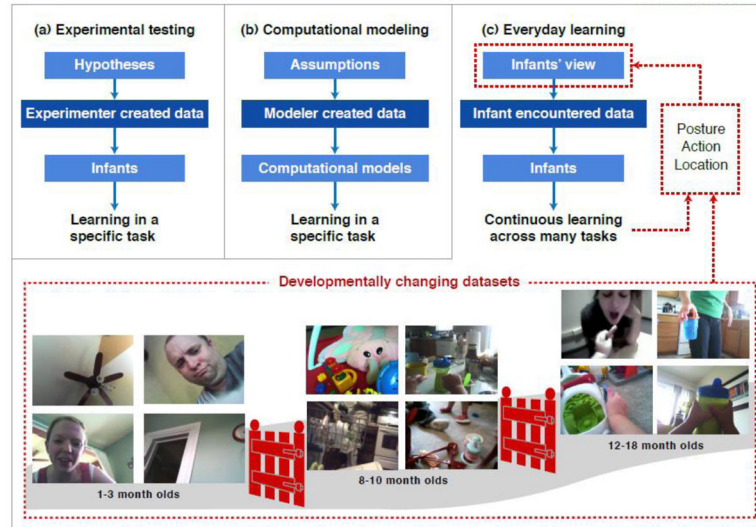
Fig. 1. The data available for everyday learning changes with development
Statistical learning depends on the data – the training trials experiences of the learner – and the learner's cognitive machinery that does the learning. Contemporary research is focused on the nature of the learning machinery and uses two approaches: (a) In human laboratory experiments, researchers test hypotheses about internal cognitive processes by creating experimental training sets and testing human learners' abilities to learn from them. (b) In human computational studies, researchers build models that instantiate their ideas about the learning machinery and feed those models training sets created by the experimenter to tests the model's ability to find the regularities in the input data. Every day learning by infants (c) differs substantially from these approaches in the nature of the data for learning. Infants encounter data for learning from a specific vantage point that depends on where they the location, the body's posture, and behavior of the infant. Because infant locations, postures, and behaviors change systematically with development, the data sets for learning change systematically with development. The infants developing abilities – sitting up, crawling, walking – open and close gates or visual experiences with different content and different statistical structure.
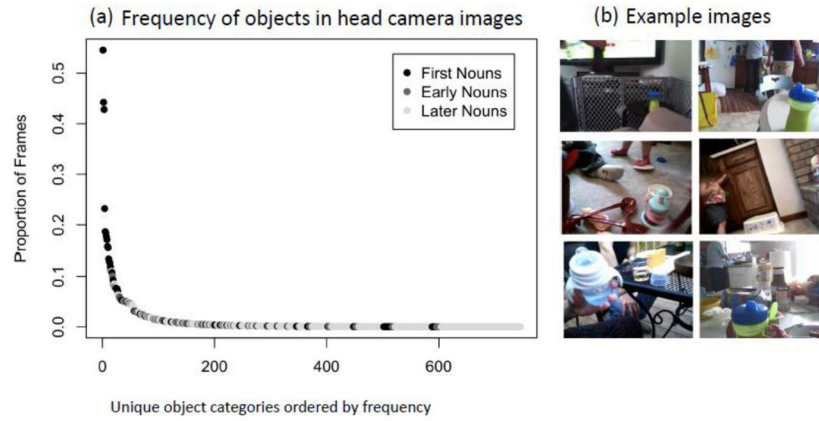
Fig. 2. The right-skewed frequency distribution of objects in 8- 10-month-old infants' first-person views of meal time

(a) The frequency of common objects in head camera images show and extremely skewed distribution. The most frequent objects (cup, spoon, bowl) are very frequent but most objects in the images rare. This creates a select set of highly frequent visual categories. These high frequency objects all have common names are normatively the first nouns that children acquire (black dots). The many more rarer objects in these scenes have names normatively acquired later in infancy (early words) or in childhood (later nouns). (b) Example images show the clutter characteristic of 8- to 10-month old infant's first person view and illustrates the many different and but repeated experiences of a high-frequency object (cup) that infants encounter in their everyday experiences.
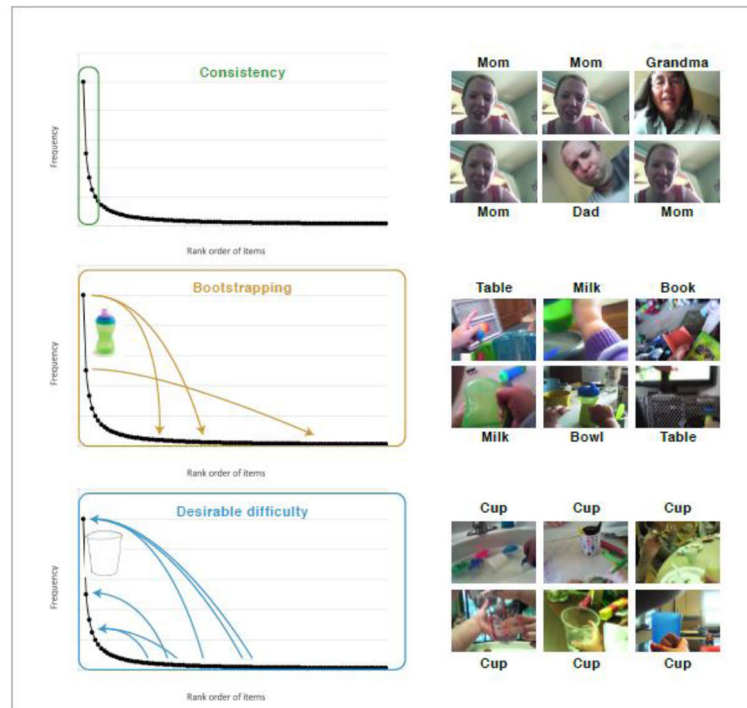
Fig. 3. Three hypotheses about how heavy-tailed distributions may support learning
Theoretical frequency distributions are plotted such that the frequency of an object in visual scenes is plotted as function of the rank order of frequency of all objects. How might learners make use of the data with this distributional structure? One possibility, the Consistency hypothesis, is that they learn only about the high frequency items, for example only about the high frequency faces and that learning is relatively unaffected by the properties of other more rarely encountered faces. A second possibility, the Bootstrapping hypothesis, is that learners learn about both the high frequency and low frequency and low frequency objects and that learning about the high frequency objects supports learning about the other rarer objects. For example, the learner may be better able to visually find a low frequency object in clutter (e.g., the book) if it in a context with high frequency visual objects. The third hypothesis, Desirable Difficulties, is that learners may learn most robustly about the high frequency objects but that learning is helped by encountering the high-frequency objects (the cup) in many different scenes with many other different objects.