Trends in Cognitive Sciences

CellPress REVIEWS

Review

The Developing Infant Creates a Curriculum for Statistical Learning

Linda B. Smith,[1,*] Swapnaa Jayaraman,[1] Elizabeth Clerkin,[1] and Chen Yu[1]

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

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 viewpoint 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 not only challenges current assumptions of statistical learning but can also inform computational models of statistical 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 datasets such as those encountered in the world. Much of this interest is directed to the statistical learning abilities of human infants. In the first 2 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 learning environment is rich but noisy, with data for many different tasks being 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 principal forms: (i) laboratory experiments with human participants that test predictions about hypothesized learning mechanisms [4,5], and (ii) computational models showing that the hypothesized machinery can use key statistical regularities [6,7] to accomplish some learning tasks. However, the relevance of these efforts to understanding the prowess of everyday learning by infants is limited because the datasets 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

[1]Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA

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

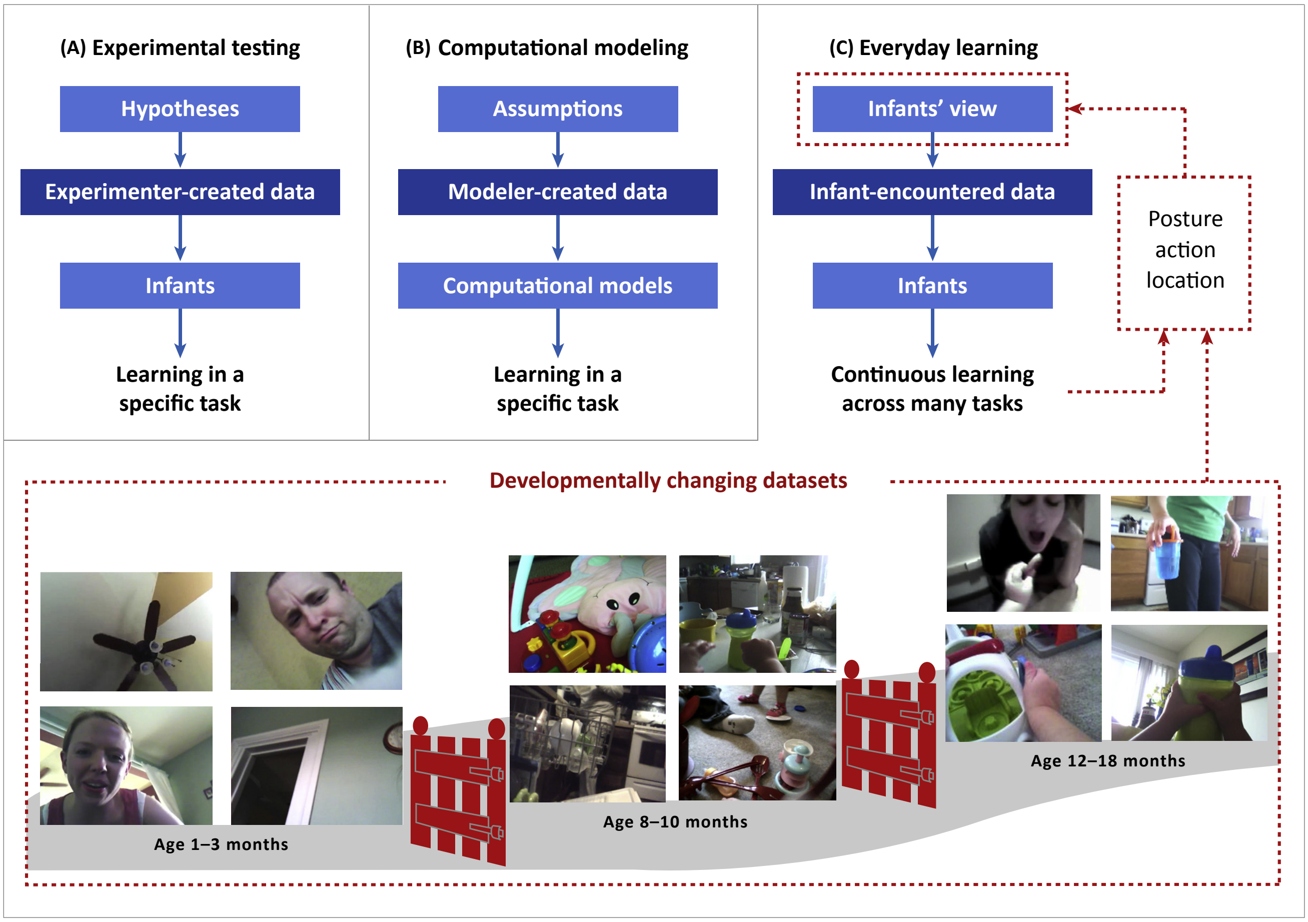


Figure 1. The Data Available for Everyday Learning Change with Development. Statistical learning depends on both the data experienced by the learner and the cognitive machinery that does the learning. Contemporary research is focused on the nature of the learning machinery, and uses two approaches. (A) In human laboratory experiments, researchers test hypotheses about internal cognitive processes by creating experimental training sets and testing the abilities of human learners to learn from them. (B) In human computational studies, researchers build models that instantiate their ideas about the learning machinery, and use the training sets of models created by the experimenter to test the ability of the model to find the regularities in the input data. Everyday learning by infants (C) differs substantially from these approaches in the nature of the data used for learning. Infants encounter data for learning from a specific vantage point that depends on the location, body posture, and behavior of the infant. Because infant locations, postures, and behaviors change systematically with development, the datasets for learning change systematically with development. The developing abilities of infants – sitting up, crawling, walking – open and close gates for visual experiences with different contents and statistical structures.

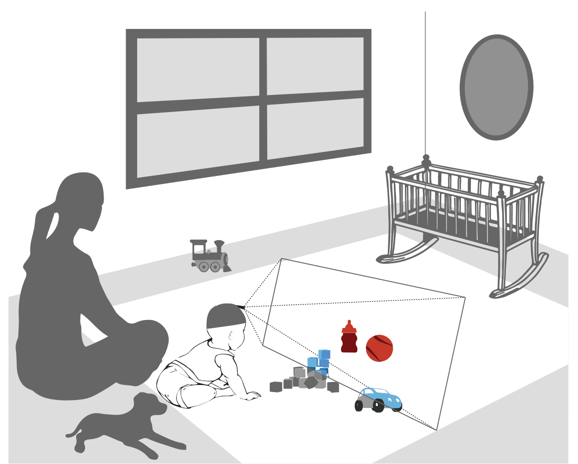
learning mechanisms, this review focuses on the data for statistical learning from the perspective of the infant.

The Data for Learning Change with Development

The infant's personal view of the world changes as sensorimotor abilities change. Each new sensorimotor achievement – rolling over, reaching, crawling, walking, manipulating objects – opens and closes gates, selecting from the external environment different datasets 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 developing abilities of the infant and their 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

Box 1. Egocentric 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 items 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 location and posture of the perceiver, and their ability or motivation to change their posture systematically, bias first-person visual information. The emerging area of research called 'egocentric vision' studies fundamental questions in vision from the perspective of freely moving individuals, and growing evidence suggests that the properties of the visual system appear profoundly different when studied from this perspective than when studied in the context of restricted body movements [17,67–70]. 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 egocentric view [69–71]. 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 to be biased to detect the statistically prevalent features [72]. However, the 'natural scenes' used to determine these natural statistics are not the egocentric perspective of perceivers as they move about the world; instead, they are photographs taken by adults – and are 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 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.



Trends in Cognitive Sciences

Figure I. Egocentric Vision Is Highly Selective because It Depends on the Momentary Location and Posture of the Perceiver.

the face of the infant. What caregivers often put in front their infant is their own face [13–15]. By 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 must 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 dataset 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 change 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 per hour of head-camera recording. By contrast, infants aged 1 year received only 6 minutes of face views for every hour of head-camera recording. The finding that the faces of other people are rarely in the egocentric view of the toddler 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 the faces of other people continue to be rare, their hands become pervasively present in the first-person view of the infant [11]. In 80% of the cases when the hands of another person are in the view of the older infant, those hands are in contact with and are acting on an object. The pervasiveness of hands performing instrumental acts aligns with the increasingly sophisticated manual skills of these older infants [25] and their increasingly sophisticated understanding of goal-directed manual acts [26].

The learner’s view of the world depends on posture, location, and behavior. The body and sensorimotor abilities of an infant change systematically and markedly during the first 2 years of postnatal 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, the visual systems of infants are biased to the specific properties of the specific faces in their environments: infants preferentially discriminate the faces of their caregivers, and recognize and discriminate faces that are similar to those of 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 crucial 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 4–6 months of age [30,31]. These individuals lack the early period of visual experience that is 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 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 audiovisual experiences that 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 first-person views should constrain theories about the internal processes that underlie human face processing. In very early infancy, these processes work on a 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 3 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 dataset: dense close frontal views of faces. Because the egocentric view depends on the sensorimotor, cognitive, and emotional level of development of the infant, 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, despite the variety of cups in the 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 relatively rarely.

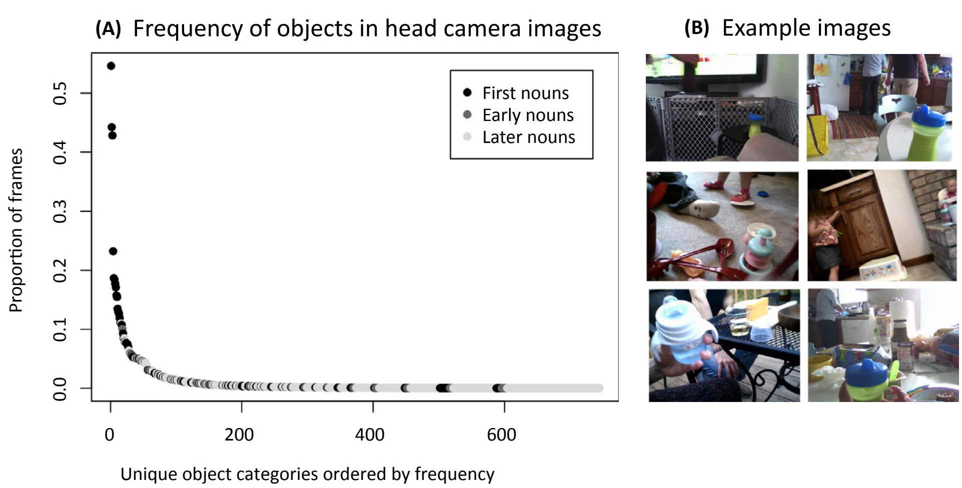
Analyses of images from head cameras worn in the home demonstrate characteristic and extremely skewed frequency distributions in which very small number of instances are highly frequent whereas most other instances are encountered very rarely. Thus, the visual world of very young infants is not only dense with faces but these are the faces of a very few people: only 2–3 individuals account for >80% of all the images in which faces appear [13]. Although head-camera images from older infants have relatively few with faces in them, the faces of only three individuals also account for 80% of the faces that appear [13]. 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 types 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 infants aged 8 or 10 months 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 mealtime 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 of 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. Nevertheless, given highly cluttered scenes, how could they know the object being named?

The right-skewed frequency distribution of objects in the images provides a straight-forward solution to this problem. mealtime scenes of infants aged 8–10 months. The (Figure 2) was extremely skewed. Of the many objects in each of these cluttered scenes, a very small number were pervasively present in the corpus of mealtime scenes (Figure 2). That is, a very small number of object categories (spoon, bottles, sippy cups, bowls, yogurt) were repeatedly present, whereas most object categories (jugs, salad tongs, ketchup bottles) were present in only a very few scenes. The pervasive repetition of instances from a select set of categories could help infants to 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 skewed frequency distributions may facilitate statistical learning (Figure 3).

Consistency

Learners could simply 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 infants aged 8–10 months who are in danger of choking; by contrast, grapes and chicken nuggets are plentiful for two-year-old children. If the high-frequency objects in the view



Trends in Cognitive Sciences

Figure 2. The Right-Skewed Frequency Distribution of Objects in the First-Person Mealtime Views of Infants Aged 8–10 Months. (A) The frequency of common objects in head-camera images show an extremely skewed distribution. The most frequent objects (cup, spoon, bowl) are very frequent, but most objects in the images are rare. This creates a select set of highly frequent visual categories. The common names of these high-frequency objects are normatively the first nouns that children acquire (black dots). The many more but rarer objects in these scenes have names that are normatively acquired later in infancy (early words) or in childhood (later nouns). (B) Example images show the clutter characteristic of the first-person view of infants aged 8–10 months, and illustrate the many different and but repeated experiences of a high-frequency object (cup) that infants encounter in their everyday experiences.

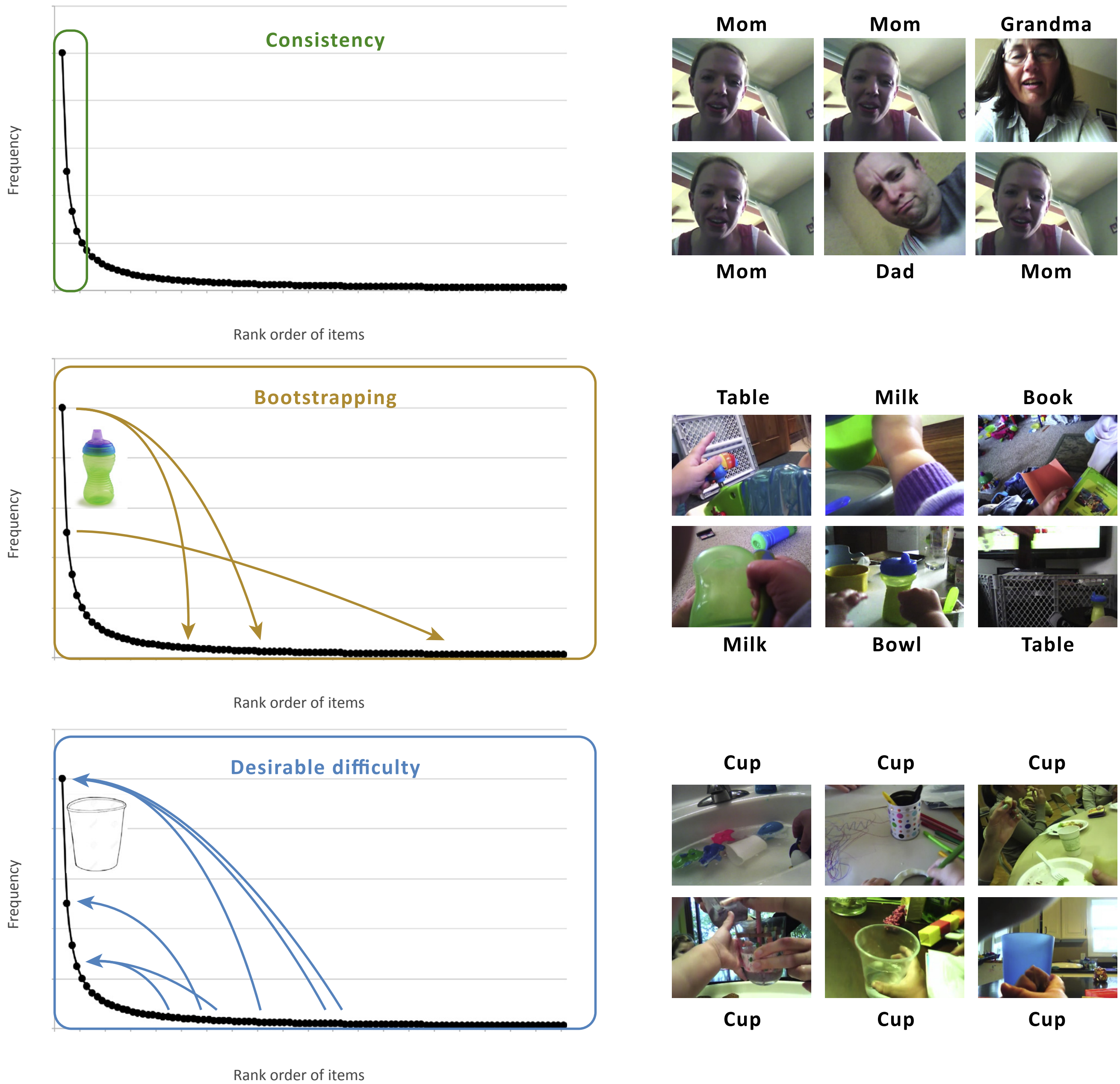


Figure 3. Three Hypotheses about How Heavy-Tailed Distributions May Support Learning. Theoretical frequency distributions are plotted such that the frequency of an object in visual scenes is plotted as function of the rank order of frequency of all objects. How might learners make use of the data with this distributional structure? One possibility, the 'consistency hypothesis', is that they only learn 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 objects, and that learning about the high-frequency objects supports learning about the 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 is in a context with high-frequency visual objects. The third hypothesis, 'desirable difficulties', is that learners may learn most robustly about high-frequency objects, but that learning is helped by encountering the high-frequency object (e.g., the cup) in many different scenes with many other different objects.

of the infant 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 together with high-frequency items because they benefit from 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 to 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 before 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 infants aged 2 years 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 – for example, 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 [47–49]. 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 items, enabling infants them to remember seen things and their heard names [44,50]. Lesson 2 may use very different training data. After their first birthday, the egocentric views of toddlers differ from those of infants aged 8–10 months as well as from adults. Toddlers visual experiences of objects are shaped by what they can manually do with those objects [51–53]: 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. 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 (ca 4 s) and coincide with the stilling of head movements by the infant [51]. They also invite joint attention with a mature social partner and the naming of those objects by the social partner [51,54]. These less-cluttered scenes and the obvious interest of the mature partner in the object held by the infant may be major factors in the rate of object name learning after the first birthday. The training data are better for infants aged 1½ years than for those aged 10 months: 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 codevelop [19,20,49]. First-person views of infants of their own object manipulations also provide a very rich dataset for learning about 3D 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 has focused on the data that underpin learning from the perspective of the infant and in everyday contexts as a corrective to the overemphasis 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 dataset, and a learning machinery that are 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 [55]: the idea is that the learning machinery of an infant 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 [56,57], and may be part of the explanation of why infants in some domains (e.g., language) appear to be better learners than adults. However, the premise of the starting small hypothesis is that the external data for learning are the same for younger and older learners, but 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 many contemporary computational approaches to learning; although most of these are not concerned with infants, several do consider learning from ordered datasets, and thus may provide some guidance as to the types of model 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 [58–60]. Other approaches model statistical learning and inferences in data that emerge in time [61,62]. Active-learning approaches focus [63] on the role of the learner in selecting the training sequence and on 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 change with learning [64], and the use of curiosity to shift attention to new learning problems as learning progresses [65,66,80].

What is the future of statistical learning? Egocentric vision is only 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 [10,73]. Infant walkers – both skilled and novice – take several thousand steps a day, fall 17 times each hour, and rarely walk in a straight line. They walk forward, backward, and sideways, 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 [74]. 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. However, structured activities with children have thus far been the main contexts for studying language-learning environments. These and other findings [75] 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 [76,77], exciting new efforts are underway to capture language-learning environments at scale using day-long recording systems [78,79].

Concluding Remarks

Development is a personal journey. The personal vantage point of the learner determines the data for learning for that individual. Emerging evidence from studies of infant egocentric vision suggest that the structure of the changing personal view of an infant is key to how and why infant learning – across many different tasks – is so robust. From the perspective of the infant, different learning problems are segregated in time. There is extensive training on a small set of instances and rarer encounters with many others. The changing sensorimotor abilities of the infant open and close environments for learning. The increasing autonomy of the developing infant puts them in control of generating the datasets 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 are determined by the immediate surroundings of the learner and the their developing behaviors in those surroundings.

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 real-time activities of the learner? How do they interact with potentially changing or different learning mechanisms?

Does the order of developmentally segregated datasets – 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, or in letter recognition?

If sensitive periods are formed in part by the closing of sensorimotor gates on key experiences, can a sensitive period for learning be reopened by reopening those sensorimotor 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 sensorimotor development.

Acknowledgments

L.B.S. was supported in part by National Science Foundation grant BCS-1523982; E.M.C. was supported by the National Institute of Child Health and Human Development (NICHD; T32HD007475-22), L.B.S., S.W., and C.Y. were supported by Indiana University through the Emerging Area of Research Initiative – Learning: Brains, Machines, and Children.

References

1. Aslin, R.N. (2017) Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdiscip. Rev. Cogn. Sci.* 8, 1–7
2. Griffiths, T.L. *et al.* (2010) Probabilistic models of cognition: exploring representations and inductive biases. *Trends Cogn. Sci.* 14, 357–364
3. McClelland, J.L. *et al.* (2010) Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356
4. Saffran, J.R. and Kirkham, N.Z. (2018) Infant statistical learning. *Annu. Rev. Psychol.* 69, 181–203
5. Frost, R. *et al.* (2015) Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* 19, 117–125
6. Gopnik, A. and Bonawitz, E. (2015) Bayesian models of child development. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 75–86
7. Blythe, R.A. *et al.* (2016) Word learning under infinite uncertainty. *Cognition* 151, 18–27
8. Allé, M.C. *et al.* (2017) Wearable cameras are useful tools to investigate and remediate autobiographical memory impairment: a systematic PRISMA review. *Neuropsychol. Rev.* 27, 81–99
9. Yu, C. and Smith, L.B. (2016) The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* 26, 1235–1240
10. Lee, D.K. *et al.* (2017) The cost of simplifying complex developmental phenomena: a new perspective on learning to walk. *Dev. Sci.* Published online October 22, 2017. http://dx.doi.org/10.1111/desc.12615
11. Fausey, C.M. *et al.* (2016) From faces to hands: changing visual input in the first two years. *Cognition* 152, 101–107
12. Braddick, O. and Atkinson, J. (2011) Development of human visual function. *Vision Res.* 51, 1588–1609
13. Jayaraman, S. *et al.* (2015) The faces in infant-perspective scenes change over the first year of life. *PLoS One* 10, e0123780
14. Jayaraman, S. *et al.* (2017) Why are faces denser in the visual experiences of younger than older infants? *Dev. Psychol.* 53, 38
15. Sugden, N.A. and Moulson, M.C. (2017) Hey baby, what's 'up'? One-and 3-month-olds experience faces primarily upright but non-upright faces offer the best views. *Q. J. Exp. Psychol.* 70, 959–969
16. Ueno, M. *et al.* (2017) Crawling experience relates to postural and emotional reactions to optic flow in a virtual moving room. *J. Motor Learn. Dev.* http://dx.doi.org/10.1123/jmld.2016-0063

17. Kretch, K.S. *et al.* (2014) Crawling and walking infants see the world differently. *Child Dev.* 85, 1503–1518
18. Franchak, J.M. *et al.* (2017) See and be seen: infant-caregiver social looking during locomotor free play. *Dev. Sci.* Published online October 26, 2017. http://dx.doi.org/10.1111/desc.12626
19. James, K.H. *et al.* (2014) Young children's self-generated object views and object recognition. *J. Cogn. Dev.* 15, 393–401
20. James, K.H. *et al.* (2014) Some views are better than others: evidence for a visual bias in object views self-generated by toddlers. *Dev. Sci.* 17, 338–351
21. Clerkin, E.M. *et al.* (2017) Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B* 372, 20160055
22. Yoshida, H. and Smith, L.B. (2008) What's in view for toddlers? Using a head camera to study visual experience. *Infancy* 13, 229–248
23. Deak, G.O. *et al.* (2014) Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. *Dev. Sci.* 17, 270–281
24. Yu, C. and Smith, L.B. (2013) Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One* 8, e79659
25. Lockman, J.J. and Kahrs, B.A. (2017) New insights into the development of human tool use. *Curr. Dir. Psychol. Sci.* 26, 330–334
26. Gerson, S.A. and Woodward, A.L. (2014) Learning from their own actions: the unique effect of producing actions on infants' action understanding. *Child Dev.* 85, 264–277
27. Lee, K. *et al.* (2017) Face race processing and racial bias in early development: a perceptual–social linkage. *Curr. Dir. Psychol. Sci.* 26, 256–262
28. Rennels, J.L. *et al.* (2017) Caregiving experience and its relation to perceptual narrowing of face gender. *Dev. Psychol.* 53, 1437
29. Moulson, M.C. *et al.* (2009) The effects of early experience on face recognition: an event-related potential study of institutionalized children in Romania. *Child Dev.* 80, 1039–1056
30. Maurer, D. (2017) Critical periods re-examined: evidence from children treated for dense cataracts. *Cogn. Dev.* 42, 27–36
31. Maurer, D. *et al.* (2007) Sleeper effects. *Dev. Sci.* 10, 40–47
32. Chen, Y.C. *et al.* (2017) Early binocular input is critical for development of audiovisual but not visuotactile simultaneity perception. *Curr. Biol.* 27, 583–589
33. Bremner, A.J. (2017) Multisensory development: calibrating a coherent sensory milieu in early life. *Curr. Biol.* 27, R305–R307
34. Oakes, L.M. (2017) Plasticity may change inputs as well as processes, structures, and responses. *Cogn. Dev.* 42, 4–14
35. Montag, J.L. *et al.* (2018) Quantity and diversity: simulating early word learning environments. *Cogn. Sci.* Published online February 7, 2018. http://dx.doi.org/10.1111/cogs.12592
36. Salakhutdinov, R. *et al.* (2011) Learning to share visual appearance for multiclass object detection. In *2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1481–1488, IEEE
37. Piantadosi, S.T. (2014) Zipf's word frequency law in natural language: a critical review and future directions. *Psychon. Bull. Rev.* 21, 1112–1130
38. Bergelson, E. and Swingley, D. (2015) Early word comprehension in infants: replication and extension. *Lang. Learn. Dev.* 11, 369–380
39. Bergelson, E. and Aslin, R. (2017) Semantic specificity in one-year-olds' word comprehension. *Lang. Learn. Dev.* 13, 481–501
40. Kurumada, C. *et al.* (2013) Zipfian frequency distributions facilitate word segmentation in context. *Cognition* 127, 439–453
41. Kachergis, G. *et al.* (2017) A bootstrapping model of frequency and context effects in word learning. *Cogn. Sci.* 41, 590–622
42. Bjork, R.A. and Kroll, J.F. (2015) Desirable difficulties in vocabulary learning. *Am. J. Psychol.* 128, 241
43. Twomey, K.E. *et al.* (2017) All the right noises: background variability helps early word learning. *Cogn. Sci.* Published online September 23, 2017. http://dx.doi.org/10.1111/cogs.12539
44. Kucker, S.C. and Samuelson, L.K. (2012) The first slow step: differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy* 17, 295–323
45. Vlach, H.A. and Johnson, S.P. (2013) Memory constraints on infants' cross-situational statistical learning. *Cognition* 127, 375–382
46. Bobb, S.C. *et al.* (2016) Predicting visual information during sentence processing: toddlers activate an object's shape before it is mentioned. *J. Exp. Child Psychol.* 151, 51–64
47. Landau, B. *et al.* (1988) The importance of shape in early lexical learning. *Cogn. Dev.* 3, 299–321
48. Smith, L.B. (2013) It's all connected: pathways in visual object recognition and early noun learning. *Am. Psychol.* 68, 618–629
49. Yee, M. *et al.* (2012) Changes in visual object recognition precede the shape bias in early noun learning. *Front. Psychol.* 3, 533
50. Vlach, H.A. (2016) How we categorize objects is related to how we remember them: the shape bias as a memory bias. *J. Exp. Child Psychol.* 152, 12–30
51. Yu, C. and Smith, L.B. (2012) Embodied attention and word learning by toddlers. *Cognition* 125, 244–262
52. Pereira, A.F. *et al.* (2014) A bottom-up view of toddler word learning. *Psychon. Bull. Rev.* 21, 178–185
53. Bambach, S. *et al.* (2016) Objects in the center: how the infant's body constrains infant scenes. In *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics*, pp. 132–137, IEEE
54. Yu, C. and Smith, L.B. (2013) Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PloS One* 8, e79659
55. Elman, J.L. (1993) Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99
56. Yurovsky, D. *et al.* (2013) Probabilistic cue combination: less is more. *Dev. Sci.* 16, 149–158
57. Kam, C.L.H. and Newport, E.L. (2009) Getting it right by getting it wrong: when learners change languages. *Cogn. Psychol.* 59, 30–66
58. Bengio, Y. *et al.* (2009) Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48, ACM
59. Gregor, K. *et al.* (2015) DRAW: a recurrent neural network for image generation. *arXiv* Published online February 16, 2015, http:\arxiv.org/abs/1502.04623
60. Krueger, K.A. and Dayan, P. (2009) Flexible shaping: how learning in small steps helps. *Cognition* 110, 380–394
61. Thiessen, E.D. (2017) What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Phil. Trans. R. Soc. B* 372, 20160056
62. Sanborn, A.N. *et al.* (2010) Rational approximations to rational models: alternative algorithms for category learning. *Psychol. Rev.* 117, 1144
63. Agrawal, P. *et al.* (2016) Learning to poke by poking: experiential learning of intuitive physics. In *Advances in Neural Information Processing Systems* (Lee, D.D., ed.), pp. 5074–5082, Neural Information Processing Systems Foundation
64. Mnih, V. *et al.* (2014) Recurrent models of visual attention. In *Advances in Neural Information Processing Systems* (Ghahramani, Z., ed.), pp. 2204–2212, Neural Information Processing Systems Foundation
65. Kidd, C. and Hayden, B.Y. (2015) The psychology and neuroscience of curiosity. *Neuron* 88, 449–460
66. Twomey, K.E. and Westermann, G. (2017) Curiosity-based learning in infants: a neurocomputational approach. *Dev. Sci.* Published online October 26, 2017. http://dx.doi.org/10.1111/desc.12629
67. Schmitow, C. and Stenberg, G. (2015) What aspects of others' behaviors do infants attend to in live situations? *Infant Behav. Dev.* 40, 173–182
68. Solman, G.J. *et al.* (2017) Eye and head movements are complementary in visual selection. *R. Soc. Open Sci.* 4, 160569

69. Fathi, A. *et al.* (2011) Learning to recognize objects in egocentric activities. In *2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3281–3288, IEEE
70. Foulsham, T. *et al.* (2011) The where, what and when of gaze allocation in the lab and the natural environment. *Vision Res.* 51, 1920–1931
71. Smith, L.B. *et al.* (2015) Contributions of head-mounted cameras to studying the visual environments of infants and young children. *J. Cogn. Dev.* 16, 407–419
72. Simoncelli, E.P. (2003) Vision and the statistics of the visual environment. *Curr. Opin. Neurobiol.* 13, 144–149
73. Adolph, K.E. *et al.* (2012) How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychol. Sci.* 23, 1387–1394
74. Tamis-LeMonda, C.S. *et al.* (2017) Power in methods: language to infants in structured and naturalistic contexts. *Dev. Sci.* 20, e12456
75. Montag, J.L. *et al.* (2015) The words children hear: picture books and the statistics for language learning. *Psychol. Sci.* 26, 1489–1496
76. Newman, R.S. *et al.* (2016) Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *J. Child Lang.* 43, 1158–1173
77. Weisleder, A. and Fernald, A. (2013) Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24, 2143–2152
78. VanDam, M. *et al.* (2016) HomeBank: an online repository of daylong child-centered audio recordings. *Semin. Speech Lang.* 37, 128–142
79. Frank, M.C. *et al.* (2017) Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* 44, 677–694
80. Oudeyer, P.Y. and Smith, L.B. (2016) How evolution may work through curiosity-driven developmental process. *Top. Cogn. Sci.* 8, 492–502