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.

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...
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
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, researchers have embedded head cameras in hats and sent them home with

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.

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

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 straightforward 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

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 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.
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 their developing behaviors in those surroundings.

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


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.


