Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Statistical learning: A core mechanism in a developmental hierarchy

Richard N. Aslin and Jenny R. Saffran

Over the past 30 years, studies of learning in humans and other animal species have revealed powerful mechanisms that extract structured information from exposure to stimuli without the need for immediate reward or feedback. These implicit learning mechanisms are referred to as statistical learning because they are sensitive to the distributional information present in the ambient environment. Here we review the historical roots of statistical learning research as it evolved from the psychological and linguistic literature, later incorporating computational and cognitive neuroscience methods with the goal of fleshing out underlying mechanisms. We emphasize that statistical learning must be constrained by evolutionary and/or learning biases to function effectively and efficiently in a given context. We also discuss the component processes involved in statistical learning and how they may be implemented in the brain. Finally, we highlight the utility of studying statistical learning in early development. Immature neural systems lend themselves to an examination of their underlying components, which are more difficult to tease apart once they are consolidated as parallel interactive circuits in the mature brain.

Addresses
Yale University and University of Wisconsin-Madison, USA

Corresponding author: Aslin, Richard N. (richard.aslin@yale.edu)

Current Opinion in Neurobiology 2025, **95:**103124

This review comes from a themed issue on **Systems Neuroscience 2025**

Edited by **Robert Froemke**, **Máté Lengyel**, **József Fiser** and **Livia de Hoz**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.conb.2025.103124

Introduction

The brains of every species — from bees [1] to birds [2] to primates [3] — must acquire information from their environment and store that knowledge to efficiently and effectively guide their future behavior. The process by which information is acquired varies substantially across different contexts. At one extreme is a social context in which a teacher instructs a novice learner by providing explicit feedback to convey both positive and negative instances of information. At the other extreme there is no teacher, no feedback, and no immediate consequences. Regardless of the context, the naïve learner is confronted with the challenge of selecting the relevant information and filtering out the irrelevant, via some combination of evolutionary biases and constraints on learning.

The canonical paradigm for studying learning is to bring the animal under experimental control — that is, placing the animal in a context in which the experimenter knows precisely which independent variables are affecting the animal's behavior. A powerful tool for manipulating the animal's behavior is reward, whether positive or negative, because the causal role of a given independent variable can be confirmed quickly and more definitively than by observing the effects of multiple variables that are often partially confounded. Thus, learning in a reward context — *reinforcement learning* — has been the dominant paradigm in experimental psychology and neuroscience for the past century [4].

However, there is no question that a substantial amount of learning occurs in less well controlled contexts in which reward is either absent or irrelevant. For example, foraging bumblebees engage in a preliminary survey of their surroundings so that once a rich source of pollen is discovered they can efficiently return to the hive and convey that information to their fellow foragers [5]. There is no immediate reward provided by this preliminary aerial survey. Female hatchling Zebra finches exposed to the songs of their fathers (the mothers do not sing) never produce that song, and yet these naïve learners acquire knowledge of this song template so that months later, once mature, they can affiliate and mate with conspecific males [6]. Juvenile chimpanzees who observe an older chimpanzee use a stick to capture termites from an underground nest are able to mimic that tool-use to forage for termites on their own, despite no immediate reward during observational learning [7]. Thus, across different contexts in the natural ecology of all species, *unsupervised learning* is a powerful mechanism for acquiring knowledge and deploying that knowledge to solve particular tasks. The key distinction between supervised and unsupervised learning is that the former relies on feedback from the environment — via reward and/or labeled input — whereas the latter extracts structure from the environment without either type of feedback [8].

A suitable domain for investigating the role of unsupervised learning in humans is the acquisition of language — a complex system of vocal sounds (or manual signs) that conveys information from one person to another. Information in language is hierarchically organized such that sound elements (phonemes) are combined to form words (and morphemes that are parts of words) that convey meanings, both by themselves and in combination with other words which are organized according to a set of sequencing rules called grammar or syntax [9]. Due to the inherent constraints on studying infants and young children, much of what we know about language learning is based on naturalistic observations of parental language input and child language production [10,11]. Classic studies have confirmed that reinforcement learning is not the primary means by which children acquire their native language — parents rarely provide negative feedback when their child makes a grammatical error [12], although contingent social responsiveness to infant vocalizations can shape their speech productions to match the properties of parental speech [13]. Moreover, children acquire the phonemic categories of their native language by 12 months of age by merely listening to the speech sounds in their ambient environment [14]. Thus, it has long been assumed that humans must acquire language, especially in the early phases of language learning, by mere exposure — in other words, by unsupervised learning. However, it was unclear what kinds of unsupervised language learning tasks infants could perform, and the nature and complexity of the kinds of information that infants could track.

Statistical learning in human infants

In the early 1990s, our group (Jenny Saffran, Elissa Newport, Richard Aslin) embarked on a test of unsupervised learning with 8-month-old infants [15]. Around this age, infants must begin to solve a key problem in language learning: determining where one word ends and the next begins when confronted with fluent multiword speech. This is not a trivial task because there are no reliable pauses or other acoustic markers at word boundaries [16]. Moreover, simple low-level solutions to this task of word segmentation are unlikely. For example, single-word utterances by parents to infants are rare [17], but even if they were common, it would nevertheless be unclear to the child whether a single multisyllable word spoken in isolation was one word or multiple words. Similarly, while some languages provide prosodic cues to word boundaries [18] (e.g., stress patterns in multi-syllabic words), these cues are language-specific and thus must themselves be learned.

Of course, nothing in science proceeds de novo, but rather builds on earlier work in both the same general discipline and in related disciplines. We benefitted from earlier work in linguistics and computer science, as well as three seminal behavioral studies with adults and infants, described below. Linguists analyzing an unfamiliar language can chart the patterning of sounds across a corpus of that language and make inferences about its structure, even without access to native-speaking informants [19]. These connections between the distributions of sounds/words and underlying structure were applied by structural linguists in the mid-20th century to successfully unravel the hierarchical structure of unfamiliar languages [20].

In computer science, statistical approaches to clustering and pattern learning have been used for decades [21], but these computational approaches, like that of structural linguists, were applied to large datasets in a "batch" manner, in which the algorithm was not under real-time pressure to make inferences about the underlying information structure. This situation contrasts with the task facing the infant, who hears fluent speech at the daunting rate of 3—4 syllables per second. Given the known limitations on sensory encoding and working memory in young infants [22], it seemed implausible that the rapid sequential structure contained in fluent speech could be extracted by infants.

However, behavioral evidence from adults suggested that structural properties of sound sequences could be learned by mere exposure. Hayes and Clark [23] created sequences of tones whose pitch varied in seemingly random ways but whose actual ordering was highly constrained within subunits in the sequence. Adults successfully distinguished familiar sequences from novel sequences formed from the same tones after a few minutes of passive listening. And we knew from prior work that infants can recognize a target word when embedded in a multiword utterance [24] and that detection of the target word is facilitated by the statistical structure of that utterance [25].

With the Hayes and Clark study as a model, we replaced their tones with speech syllables. The syllables were concatenated into "words" (*golabu*, *padoti*, etc.), and the words were randomized into a continuous stream of syllables. After just several minutes of passive listening, adults and infants judged familiar syllable sequences as different from the same syllables in novel orders. We hypothesized that learners successfully discriminated familiar from novel syllable sequences by extracting the transitional probabilities of syllable-pairs, and not the relative frequency of those syllable-pairs. Subsequent studies [26] confirmed this hypothesis, although the relative frequency of syllable-pairs can also be used, illustrating the multiple levels of statistical information present even in simple corpora. Given a 1996 computational linguistics monograph by Eugene Charniak [27] entitled *Statistical Language Learning*, we borrowed that term to describe our behavioral results.

Extensions, interpretations, and challenges for statistical learning

Our initial infant statistical learning results [15] have been cited over 7700 times in the past 30 years and spawned a huge variety of extensions and clarifications. The temporal ordering of elements was extended from speech syllables to visual objects [28], and visual objects were clustered spatially to define scenes [29], thereby extending the importance of transitional probabilities in time to conditional probabilities in space. Other studies revisited the tones of Hayes and Clark to explore statistical learning in music [30] and extended the use of visual objects to letters [31] to explore the relevance of statistical learning to reading. These extensions have raised the possibility that statistical learning is domain-general, although the encoding of information is clearly domain-specific [32]. These extensions have also raised the possibility that statistical learning proceeds in the absence of overt attention (i.e., conscious awareness), although overt attention clearly facilities the speed of learning [33]. In summary, there is no question that statistical learning is an example of unsupervised learning that occurs by mere exposure (with or without instruction or explicit attention), and that it is ubiquitous across domains, sensory modalities, and species [34].

The challenge of too many statistics

Despite the growing body of literature focused on statistical learning, understanding its role in the natural environment faces several significant challenges. First, there is an infinite number of statistics that could characterize information in any domain, and yet naïve learners manage to find, at minimum, the subset of statistics required to solve a particular task. How do they do that efficiently without conducting an exhaustive search of all possible statistics? Evidence accumulated over the past 40 years supports a multi-layered representational system in which the input to statistical learning at any given moment is influenced by the current knowledge state of the learner [35]. For example, consider the hierarchical structure of language. A naïve learner might start out representing the input as a sequence of syllables, whereas a learner who has already acquired words would have access to a new level of statistical regularities – sequences of words. As learners iterate through the input, there would be an ongoing process of encoding the elements being perceived, detecting extant statistical regularities, and making decisions about how to use those statistics to solve a particular task. Priors, in the form of previous learning and other aspects of the learning context, constrain what regularities must be considered each step of the way. Figure 1 illustrates these hierarchical properties in a general framework that begins with sensory input, forms perceptual primitives constrained by top-down feedback, extracts low-level statistics constrained by limited cognitive resources, and formulates hierarchical structures that enable generalization beyond the specific inputs received by the learner.

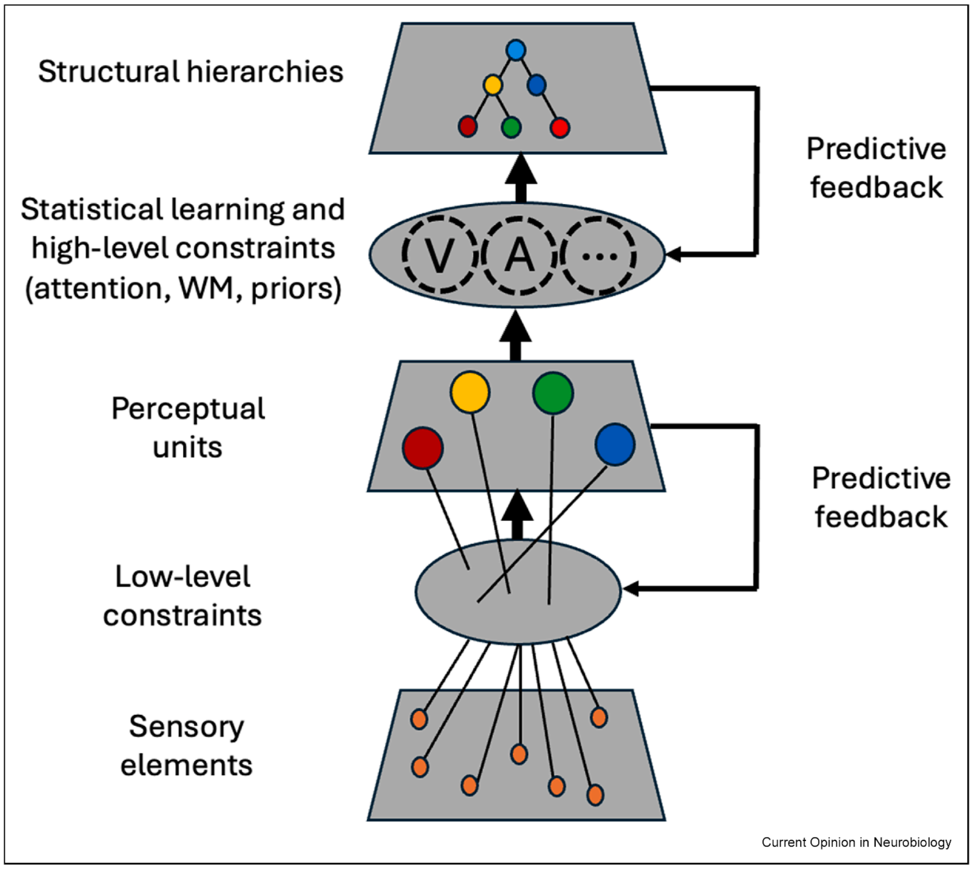
The challenge of generalization

Early considerations of statistical learning focused on the extraction of information contained in the environment to which the learner was directly exposed (i.e., learn and remember the input). For example, Saffran, Aslin, and Newport [15] examined learning and memory for specific exemplars (trisyllabic "words"). But many other studies, beginning with Marcus et al. [36], have shown that infants can also generalize beyond the exemplars to which they were exposed. Moreover, infants who fail to generalize a pattern in a "difficult to learn" context can be primed by exposure to a related pattern in an "easy to learn" context [37,38]. Importantly, the scope of generalization is licensed by the distributional properties of the learning corpus [39]. For example, if very few exemplars of a pattern are present in the input, generalization tends to be restricted, whereas when many different exemplars are present in the input, broad generalization is more likely. These same principles that modulate the scope of generalization in infants also hold for learning much more complex grammatical structures in adults [40] and children [41] and may help to explain how learners hone in on the right statistics (computations and units over which to perform those computations) for a given learning problem. Thus, it is perhaps not surprising that, depending on the context and the complexity of the task, many different neural systems are activated during statistical learning.

The challenge of linking statistical learning to the brain

The neural mechanisms that support statistical learning appear to be widespread in the brain rather than concentrated in a particular focal substrate, such as the superior colliculus/tectum for sensory integration [42]. For example, neural correlates of statistical learning in humans have been reported in inferior frontal gyrus [43], striatum [44], anterior cingulate cortex [45], hippocampus in both adults [46] and infants [47], as well as inferotemporal cortex in monkeys [48], with changes in connectivity to many other areas of the brain strengthened as statistical learning improves [49]. If statistical learning does not require a particular region, hub, or core neural circuit, then perhaps it is not a computational module but rather a ubiquitous neurophysiological process such as LTP [50]. Moreover, redundant neural mechanisms that support statistical learning provide a

Figure 1



Schematic illustration of the hierarchy of levels of analysis operating on environmental input and how predictive feedback at each level enables a description of the underlying information structure relevant to solving a particular task via unsupervised learning. The dashed circles at the level of statistical learning indicate that it operates within domains (e.g., vision, audition) but that interactions between domains are possible given the influence of predictive feedback on learned priors (WM = working memory).

hedge against loss of function from brain damage as well as flexible developmental pathways to maturity.

How might statistical learning be instantiated at the neural level, taking into account context-specificity, task-complexity, and generalization? First, the specific neural mechanisms that support statistical learning undoubtedly involve populations of neurons [51]. The firing rates of these population responses have a mean and variance that are proportional to the identity and reliability, respectively, of a given stimulus [52]. When exposed to a complex set of stimuli, population responses are elicited in parallel by the vast number of hierarchically related sources of information. Given intrinsic noise and effects of prior learning (e.g., priming that entails context memory), some of those population responses will “stick out” above the background.

Second, the robust responses that exceed background noise must be evaluated for their relevance in solving any task, regardless of how those responses are extracted by a learning mechanism. For example, the relative frequency of stimuli lying along a single dimension could form a unimodal or a multimodal distribution, which in turn provides evidence for a single or multiple categories (i.e., clusters) along that dimension. A decision-making process must evaluate the characteristics (i.e., shape) of the entire frequency distribution to extract the number of categories, and there is evidence in both infants [53] and adults [54] of this category-extraction even in the absence of reward. In sum, it should not be surprising that statistical learning is based on principles of probabilistic neural responses across a wide swathe of the brain’s computational specializations — sensory encoding, decision-making, and memory [55].

Future directions

What, then, is a reasonable path forward to unravel the role played by different subcomponents of the brain in statistical learning? Our view is that a focus on *development* will be a particularly fruitful avenue for making progress on addressing this question. Mature learners have acquired enormously complex learning systems that are so efficient that it's difficult to tease apart their underlying mechanisms at the behavioral level. And even at the neural level, the complex interplay among various subcomponents makes it very difficult to assign causal roles to these subcomponents, especially if limited to non-invasive recordings such as fMRI and EEG/MEG [56]. In contrast, immature learners not only have vastly less experience, but they also have far less efficient neural responses at all levels of the information hierarchy. As a result, it is easier to observe the effects of new experiences in younger organisms than older organisms. Moreover, when new experiences occur, their neural responses are typically slower and therefore easier to decompose into their component parts, making their emergence with age more accessible than later in development when they are all functioning in parallel. Clearly, there are ethical limitations on conducting invasive neuroscience on human infants, and therefore the systematic and coordinated combination of non-invasive studies of neural development in human infants with invasive studies of developing non-human animals are likely to be a productive approach to understanding the neural mechanisms of statistical learning.

Another key aspect of the path forward is to move beyond studying isolated learning problems (e.g., syllable triplets) and embrace the complexity of the input facing real-world learners. Infants do not perform a single learning "task" at one time, like those we give them in the lab. Instead, their goal is likely something quite different: to make sense of the input [57], to represent and encode the input [58], and to make predictions based on the input [59]. Moreover, learning is dynamic and may change during development [60]. As infants learn something new — for example, finding words amidst a stream of syllables — that knowledge changes how they represent future input [61]. Much as contemporary large language models learn from every encounter, so do infants [62]. By embracing the complexity of the challenges facing naïve learners, we may find ourselves with novel answers to long-standing questions about the nature of human learning.

CRediT roles

Conceptualization; Funding acquisition; Investigation; Resources; Writing — review & editing: RNA and JRS

Data curation; Formal analysis; Methodology; Project administration; Software; Supervision; Validation: Not applicable

Visualization; Roles/Writing — original draft: RNA

Declaration of competing interest

I and my co-author have nothing to declare.

Acknowledgements

The authors were supported, in part, by the following grants while this article was being drafted and finalized for publication. RNA: NIH HD-037082 (Elissa Newport and Richard Aslin, MPIs); NSF BNS-2323360 (Richard Aslin and Sara Sanchez-Alonso, MPIs); NIH HD-103931 (David Lewkowicz and Richard Aslin, MPIs). JRS: NIH HD-105313 (Jenny Saffran, PI).

Data availability

No data were used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Chittka L: *The mind of a bee*. Princeton University Press; 2022.
2. Emery N: *Bird brain: an exploration of Avian intelligence*. Princeton University Press; 2017.
3. Tomasello M, Call J: *Primate cognition*. Oxford University Press; 1997.
4. * Gershman SJ, Daw ND: **Reinforcement learning and episodic memory in humans and animals: an integrative framework**. *Annu Rev Psychol* 2017, **68**:101–128.
This publication provides a comprehensive review of the main alternative to statistical learning – reinforcement learning.
5. Osborne JL, Smith A, Clark SJ, Reynolds DR, Barron MC, Lim KS, Reynolds AM: **The ontogeny of bumblebee flight trajectories: from naïve explorers to experienced foragers**. *PLoS One* 2013, **8**:e78681.
6. Chen Y, Clark O, Woolley SC: **Courtship song preferences in female zebra finches are shaped by developmental auditory experience**. *Proc Royal Society B: Biol Sci* 2017, **284**:20170054.
7. Lonsdorf EV, Sanz CM: **Behavioral and cognitive perspectives on the evolution of tool use from wild chimpanzees**. *Curr Opinion Behav Sci* 2022, **46**:101144.
8. Dayan P, Daw ND: **Decision theory, reinforcement learning, and the brain**. *Cognit Affect Behav Neurosci* 2008, **8**:429–453.
9. Baker MC: *The atoms of language: the mind's hidden rules of grammar*. Basic Books; 2008.
10. ** Brown R: *A first language: the early stages*. Harvard University Press; 1973.
This book is a classic in the field of child language development and provides an overview of the methods and initial findings in this field.
11. Gleason JB, Ratner NB: *The development of language*. 10th ed. Plural Publishing; 2024.
12. Hirsh-Pasek K, Treiman R, Schneiderman M: **Brown & Hanlon revisited: mothers' sensitivity to ungrammatical forms**. *J Child Lang* 1984, **11**:81–88.
13. Goldstein MH, Schwade JA: **Social feedback to infants' babbling facilitates rapid phonological learning**. *Psychol Sci* 2008, **19**:515–523.
14. * Kuhl PK: **Early language acquisition: cracking the speech code**. *Nat Rev Neurol* 2004, **5**:831–843.

This publication provides a classic review of the research on infant speech perception and early language development.

15. ** Saffran JR, Aslin RN, Newport EL: **Statistical learning by 8-month-old infants**. *Science* 1996, **274**:1926–1928.
This publication represents the initial finding in the field of infant statistical learning, documenting the robust and rapid learning of the structure of syllable sequences.

16. Remez RE, Rubin PE, Pisoni DB, Carrell TD: **Speech perception without traditional speech cues**. *Science* 1981, **212**: 947–950.

17. Brent MR, Siskind JM: **The role of exposure to isolated words in early vocabulary development**. *Cognition* 2001, **81**: B33–B44.

18. Peperkamp S, Vendelin I, Dupoux E: **Perception of predictable stress: a cross-linguistic investigation**. *J Phonetics* 2010, **38**: 422–430.

19. Malmberg B: *Structural linguistics and human communication*, vol. 2. Springer Science Publishing; 2012.

20. Harris Z: *Methods of structural linguistics*. University of Chicago Press; 1951.

21. Bishop CM: *Pattern recognition and machine learning*. Springer Publishing; 2006.

22. Newman RS, Simpson VM: **Infants' short-term memory for consonant–vowel syllables**. *J Exp Child Psychol* 2023, **226**: 105567.

23. ** Hayes JR, Clark HH: **Experiments in the segmentation of an artificial speech analogue**. In *Cognition and the development of Language*. Edited by Hayes JR, Wiley and Sons Publishing; 1970:221–234.
This chapter provided the springboard for the research on statistical learning of word segmentation, which was modeled on this study's research on learning the structure of tone sequences.

24. Jusczyk PW, Aslin RN: **Infants' detection of the sound patterns of words in fluent speech**. *Cogn Psychol* 1995, **29**:1–23.

25. Goodsitt JL, Morgan JV, Kuhl PK: **Perceptual strategies in prelingual speech segmentation**. *J Child Lang* 1993, **20**: 229–252.

26. Aslin RN, Saffran JR, Newport EL: **Computation of conditional probability statistics by 8-month-old infants**. *Psychol Sci* 1998, **9**:321–324.

27. Charniak E: *Statistical Language learning*. MIT Press; 1996.

28. Kirkham NZ, Slemmer JA, Johnson SP: **Visual statistical learning in infancy: evidence for a domain general learning mechanism**. *Cognition* 2002, **83**:B35–B42.

29. Fiser J, Aslin RN: **Statistical learning of new visual feature combinations by infants**. *Proc Natl Acad Sci* 2002, **99**: 15822–15826.

30. Morgan E, Fogel A, Nair A, Patel AD: **Statistical learning and Gestalt-like principles predict melodic expectations**. *Cognition* 2019, **189**:23–34.

31. Chetail F: **What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing**. *Cognition* 2017, **163**:103–120.

32. ** Saffran JR, Kirkham NZ: **Infant statistical learning**. *Annu Rev Psychol* 2018, **69**:181–203.
This publication provides a comprehensive review of the many facets of infant statistical learning in different sensory modalities, domains, and contexts.

33. Toro JM, Sinnett S, Soto-Faraco S: **Speech segmentation by statistical learning depends on attention**. *Cognition* 2005, **97**: B25–B34.

34. Santolin C, Saffran JR: **Constraints on statistical learning across species**. *Trends Cognit Sci* 2018, **22**:52–63.

35. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND: **How to grow a mind: statistics, structure, and abstraction**. *Science* 2011, **331**:1279–1285.

36. * Marcus GF, Vijayan S, Bandi Rao S, Vishton PM: **Rule learning by seven-month-old infants**. *Science* 1999, **283**:77–80.
This publication provides the first evidence that infants can learn an abstract pattern from syllable sequences by generalizing beyond the specific stimuli presented during a learning phase.

37. Marcus GF, Fernandes KJ, Johnson SP: **Infant rule learning facilitated by speech**. *Psychol Sci* 2007, **18**:387–391.

38. Lany J, Gómez R: **Twelve-month-old infants benefit from prior experience in statistical learning**. *Psychol Sci* 2008, **19**: 1247–1252.

39. Gerken L: **Decisions, decisions: infant language learning when multiple generalizations are possible**. *Cognition* 2006, **98**:B67–B74.

40. Reeder PA, Newport EL, Aslin RN: **From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes**. *Cogn Psychol* 2013, **66**:30–54.

41. Culbertson J, Schuler K: **Artificial language learning in children**. *Ann Rev Linguistics* 2019, **5**:353–373.

42. Isa T, Marquez-Legorreta E, Grillner S, Scott EK: **The tectum/superior colliculus as the vertebrate solution for spatial sensory integration and action**. *Curr Biol* 2021, **31**: R471–R762.

43. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK: **Neural evidence of statistical learning: efficient detection of visual regularities without awareness**. *J Cognit Neurosci* 2009, **21**: 1934–1945.

44. Lim SJ, Fiez JA, Holt LL: **Role of the striatum in incidental learning of sound categories**. *Proc Natl Acad Sci* 2019, **116**: 4671–4680.

45. Hill MR, Boorman ED, Fried I: **Observational learning computations in neurons of the human anterior cingulate cortex**. *Nat Commun* 2016, **7**:12722.

46. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM: **Statistical learning of temporal community structure in the hippocampus**. *Hippocampus* 2016, **26**:3–8.

47. ** Ellis CT, Skalaban LJ, Yates TS, Bejjanki VR, Córdova NI, B N, Turk-Browne: **Evidence of hippocampal learning in human infants**. *Curr Biol* 2021, **31**:3358–3364.
This publication provides the first neural evidence that the infant hippocampus is engaged in a visual statistical learning task.

48. Meyer T, Olson CR: **Statistical learning of visual transitions in monkey inferotemporal cortex**. *Proc Natl Acad Sci* 2011, **108**: 19401–19406.

49. Karuza EA, Li P, Weiss DJ, Bulgarelli F, Zinszer BD, Aslin RN: **Sampling over nonuniform distributions: a neural efficiency account of the primacy effect in statistical learning**. *J Cognit Neurosci* 2016, **28**:1484–1500.

50. Titley HK, Brunel N, Hansel C: **Toward a neurocentric view of learning**. *Neuron* 2017, **95**:19–32.

51. ** Fiser J, Berkes P, Orbán G, Lengyel M: **Statistically optimal perception and learning: from behavior to neural representations**. *Trends Cognit Sci* 2010, **14**:119–130.
This publication provides a unifying framework to studies of perception and learning that bridges the behavioral and neural levels.

52. Meyniel F, Sigman M, Mainen ZF: **Confidence as Bayesian probability: from neural origins to behavior**. *Neuron* 2015, **88**:78–92.

53. * Maye J, Werker JF, Gerken L: **Infant sensitivity to distributional information can affect phonetic discrimination**. *Cognition* 2002, **82**:B101–B111.
This publication provides the first evidence that infants can utilize the statistical patterning from a distribution of speech stimuli to form speech categories.

54. Schatz T, Feldman NH, Goldwater S, Cao XN, Dupoux E: **Early phonetic learning without phonetic categories: insights from large-scale simulations on realistic input**. *Proc Natl Acad Sci* 2021, **118**: e2001844118.

55. Conway CM: **How does the brain learn environmental structure? Ten core principles for understanding the**

neurocognitive mechanisms of statistical learning. *Neurosci Biobehav Rev* 2020, **112**:279–299.

56. * Turk-Browne NB, Aslin RN: **Infant neuroscience: how to measure brain activity in the youngest minds**. *Trends Neurosci* 2024, **47**:338–354.
This publication provides a state-of-the-art review of current non-invasive methods for recording from the infant brain.

57. Csibra G, Gergely G: **Natural pedagogy**. *Trends Cognit Sci* 2009, **13**:148–153.

58. Rosenberg RD, Feigenson L: **Infants hierarchically organize memory representations**. *Dev Sci* 2013, **16**:610–621.

59. Schwarzer G, Jovanovic B: **Infants' predictive minds: the role of motor experience**. *Child Develop Perspect* 2024, **18**:123–128.

60. Forest TA, Schlichting ML, Duncan KD, Finn AS: **Changes in statistical learning across development**. *Nat Rev Psychol* 2023, **2**:205–219.

61. Saffran JR, Wilson DP: **From syllables to syntax: multilevel statistical learning by 12-month-old infants**. *Infancy* 2003, **4**: 273–284.

62. * Vong WK, Wang W, Orhan AE, Lake BM: **Grounded language acquisition through the eyes and ears of a single child**. *Science* 2024, **383**:504–511.
This publication begins to bridge the gap between Large Language Models that are trained on millions of input sentences to the much smaller corpora available to an infant in the natural language-learning environment.