Chapter 2
Mechanisms of Statistical Learning in Infancy

Scott P. Johnson

Abstract Statistical learning is the process of identifying patterns of probabilistic co-occurrence among stimulus features, essential to our ability to perceive the world as predictable and stable. Research on auditory statistical learning has revealed that infants use statistical properties of linguistic input to discover structure that may facilitate language acquisition. More broadly, statistical learning operates across sensory modalities and across species. Research on infants' visual statistical learning has revealed that statistical learning develops over time, yet the mechanisms (including developmental mechanisms) underlying infant performance remain unclear. This chapter examines competing models of statistical learning and how learning might be constrained by limits in infants' attention, perception, and memory.

The means by which humans acquire and represent knowledge is fundamental to cognitive science, and a central question asked by developmental psychologists concerns how infants learn so much in so little time without explicit instruction. For example, the rapidity and apparent ease with which infants and young children understand and produce speech, recognize faces, interpret others' mental states, detect violations of physical laws governing object properties, and discriminate different numbers of items have led some theorists to suggest that innate cognitive mechanisms—independent of learning and experience—provide the infant with some knowledge in each of these domains (Chomsky, 1965; Johnson & Morton, 1991; Leslie, 1997; Spelke, 1990; Wynn, 1992). Yet such views may neglect the role of environmental structure in guiding development, and studies of infant *statistical learning* (SL), the focus of this chapter, can help shed light on this issue. Statistical learning (SL) is a set of processes for learning stimulus features, objects, and events from *distributional* information over space and time: simple associations, probabilistic correspondence, frequencies, spatial positions, and order in sequence. SL contributes to segmentation of continuous information (such as speech) and the

---

S. P. Johnson (✉)
University of California, Los Angeles, Los Angeles, CA, USA
e-mail: spjohnson@g.ucla.edu

J. B. Childers (ed.), *Language and Concept Acquisition from Infancy Through Childhood*, https://doi.org/10.1007/978-3-030-35594-4_2

formation of representations of units in time and space, thus helping to shape cognitive development (Frost, Armstrong, Siegelman, & Christiansen, 2015; Siegelman & Frost, 2015), and it is an important part of language acquisition (see Chap. 4).

In this chapter, I discuss research efforts to discover the nature of SL in infancy, the kinds of statistical structure that infants are able to learn, the impact of different testing methods on infant learning, implications of infant SL for cognitive development and developmental disabilities, and, finally, mechanisms underlying statistical learning in infancy. As I will try to make clear, the importance of statistical learning for understanding cognitive development, language acquisition in particular, has become increasingly evident in the 20+ years since publication of the first paper describing SL in infants (Saffran, Aslin, & Newport, 1996). Yet much remains unknown about the foundational processes and mechanisms of SL in infancy.

Statistical Learning in Infancy

Research on detection of structure in complex input sequences has a considerable history (e.g., Gibson & Gibson, 1955; Reber, 1967). It has long been known that adult learners can detect patterns in the absence of explicit (articulable) knowledge (Reber, 1989), raising questions of learnability of complex sequences by nonverbal populations. SL in infants was first reported by Saffran et al. (1996) with a head-turn procedure. Eight-month-old infants listened to a continuous stream of computer-generated speech for 2 minutes, followed by a test phase during which segments of the familiarized speech stream, now separated by brief pauses, alternated with segments whose order was scrambled or whose parts had co-occurred relatively infrequently in the training set. One study, for example, familiarized infants with the pseudo-words *tupiro*, *golabu*, *padoti*, and *bidaku* in random order and with no pauses or immediate repetitions (e.g., *tupirogolabupadotibidakugolabutupirobidakupadotitupiro…*). The test phase involved two of the four original "words" (e.g., *tupiro*, *golabu*) and two "nonwords" (e.g., *dapiku*, *tilado*) formed from a random assembly of syllables; words and nonwords were separated by a 500 ms gap. Infants in a second experiment heard words alternating with "part-words" formed from the last syllable of a word combined with the first two syllables of a different word (e.g., *bupado*, *kugola*). Discrimination of words from nonwords and part-words was evaluated during the test phase by recording look durations toward a flashing light that accompanied repeated presentation of test stimuli, on the right or left side of a testing chamber, on the assumption that interest in the sound sequences could be operationalized as attention in the direction of the light. Infants in both experiments exhibited increased interest in the novel items (nonwords and part-words).

How were infants able to parse the speech stream into coherent words, recognize them when heard in isolation, and discriminate them from the part- and nonwords? One possibility is that infants learned words from differences in *transitional probabilities* (TPs) between adjacent syllables, because there were no other cues to segmentation, such as pauses and prosody, that typically mark word and phrase boundaries in real-world speech (Fougeron & Keating, 1997; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). TP is a statistical measure that describes the predictability of adjacent items in an array or sequence (Miller & Selfridge, 1950; the TP of successive element XY is defined as probability of Y|X = frequency of XY/frequency of X). In the Saffran et al. (1996) study, TPs within words such as *tupiro* were always 1.0, meaning *tu* perfectly predicted *pi*; in turn, *pi* perfectly predicted *ro* (see Fig. 2.1). TPs between words, however, were lower, averaging 0.33. This is because *ro* (in tupiro) was sometimes followed by *go* (in *golabu*), sometimes by *pa* (in *padoti*), and sometimes by *bi* (in *bidaku*). Thus nonwords and part-words heard during the test phase such as *dapiku* and *bupado* had lower TPs between syllables than words such as *tupiro* and *padoti*. The Saffran et al. results imply that infants detected the TP differences in the test stimuli and preferred to listen to the low-TP stimulus owing to its violation of word boundaries.

But there is an alternative explanation: Words in the familiarization stimulus were heard 3× more often than nonwords, and part-words were never heard, and so it is possible that infants preferred nonwords and part-words simply because they were unfamiliar, not due to lower TPs. To address this possibility, Aslin, Saffran, and Newport (1998) tested 8-month-olds with a "frequency-balanced" design in which the word and part-word heard at test were presented the same number of times during familiarization. TP differences, however, were the same as those in the Saffran et al. (1996) study. Infants showed increased interest in the part-word at test relative to the word, replicating the Saffran et al. results and providing evidence that segmentation and learning were based on TPs, not simple frequencies of syllables or words. TP differences between syllables, therefore, seem to facilitate the learning of sequence structure by signaling boundaries and units in an otherwise uninterrupted stream of items.

More broadly, SL operates across sensory modalities and across species. In human adults, SL participates in fundamental perceptual and cognitive functions including visual search, object perception, motor planning, and event prediction (Baker, Olson, & Behrmann, 2004; Fiser & Aslin, 2002a; Hunt & Aslin, 2001; Turk-Browne, Scholl, Johnson, & Chun, 2010). Animal species learn statistically

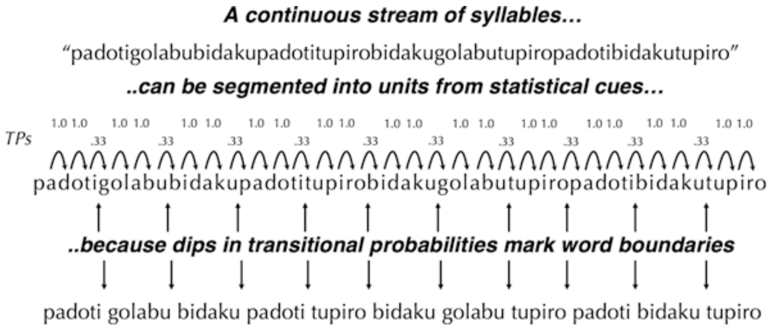

Fig. 2.1 Schematic description of how transitional probabilities between syllables mark word boundaries

structured speech streams (e.g., Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005), and human infants parse streams of musical tones based on statistical probabilities (Saffran, Johnson, Aslin, & Newport, 1999).

Experiments in my lab (Kirkham, Slemmer, & Johnson, 2002) provided the first demonstration of infants' SL in visual sequences with an experiment in which infants were habituated to a stream of looming colored shapes organized in pairs defined by TPs. (Habituation is defined as a decrement in looking across trials according to a predetermined criterion, e.g., a decline of 50% or more during four successive trials relative to the first four trials.) TPs within pairs were 1.0, and TPs across pairs were 0.33 (see Fig. 2.2). Each shape had a unique color and loomed from about 4 to 24 cm across in 1 s, with no pauses to mark pairs. Following habituation, infants viewed two test sequences with the same shapes: a "structured" sequence, defined by the same TPs as those in the habituation stimulus, and a pseudorandom sequence (no shape repetitions). Infants at 2, 5, and 8 months looked reliably longer at the random sequence, interpreted by Kirkham et al. as showing sensitivity to the statistical properties of the input—the TPs defining shape pairs in the habituation sequence—and noted when these statistics were violated. Infants at all three ages showed a reliable novelty preference for the random pattern, with no significant age differences aside from longer looking in general by the youngest group.

My colleagues and I then asked if visual SL may be available at birth with a replication of the Kirkham et al. (2002) methods, modified to suit newborns' limited color vision with the use of monochromatic stimuli (Bulf, Johnson, & Valenza, 2011). Newborn infants provided no evidence of discriminating random from struc-

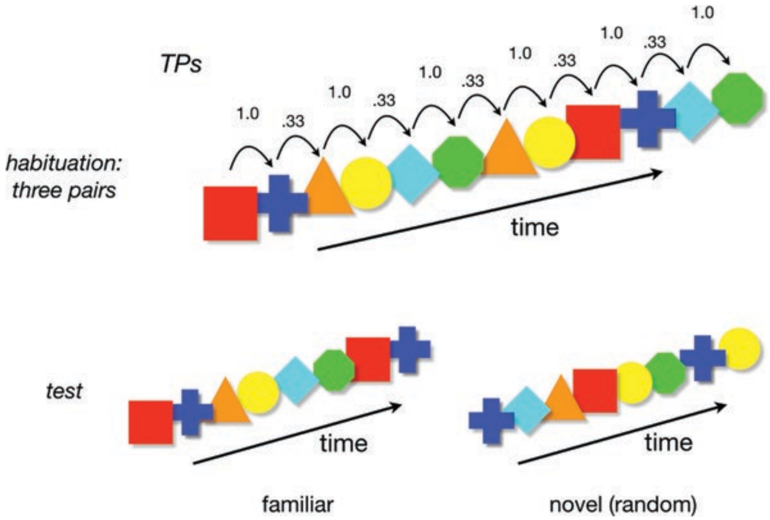

Fig. 2.2 Schematic depiction of habituation and test displays testing for visual statistical learning in infants (Kirkham et al., 2002)

tured six-item sequences. We reasoned that three pairs of shapes (a *high-demand* condition; e.g., ABCDEFCDABEFABCDABEF…) might overwhelm newborns' ability to track probabilities. This hypothesis was addressed with a two-pair, *low-demand* condition (e.g., ABCDCDABCDABABCDAB…). We observed a novelty preference for the random sequence in the low-demand condition, as did the older infants observed by Kirkham et al. who were tested with a high-demand condition. Thus the Bulf et al. study provides evidence that visual SL may be constrained by newborns' limited cognitive resources, perhaps preventing identification of relevant visual information necessary to detect statistical structure.

The Bulf et al. (2011) and Kirkham et al. studies (2002) documented young infants' ability to detect statistical information in sequences of discrete, looming shapes. By 8 months, infants detect probabilistic patterns in spatiotemporal visual sequences in which shapes appeared in locations defined by TPs of 1.0 or 0.33 (Kirkham, Slemmer, Richardson, & Johnson, 2007), and by 9 months, infants encode the underlying spatial statistical structure of multiple-element scenes in which shapes were arranged in groups defined by conditional probabilities among individual items (Fiser & Aslin, 2002b). These results led to claims of a domain-general SL device that is available early and operates across modalities, across time and space, and across species, suggesting that SL might be a predisposed, general associative mechanism (Kirkham et al., 2002). This hypothesis is supported by reports of SL and discrimination of visual and linguistic sequences in newborns (Bulf et al., 2011; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), constituting evidence for sensitivity to statistical information at birth in at least two modalities (vision and audition). SL is now a central feature of recent theories of human perception, cognition, and development (e.g., Aslin & Newport, 2012, 2014; Hasson, 2016; Krogh, Vlach, & Johnson, 2013; Thiessen, 2016; Turk-Browne, 2012).

Kinds of Statistical Structure Infants Are Able to Learn

As noted previously, early studies of SL were aimed largely at questions of (a) whether infants might use SL to segment continuous speech into discrete units (Aslin et al., 1998; Saffran et al., 1996) and (b) the possibility that infants' SL might be a domain-general learning mechanism (Kirkham et al., 2002; Saffran et al., 1999). These studies examined SL with methods involving a learning (familiarization or habituation) phase with streams of unsegmented auditory or visual sequences, followed by a test phase probing for recognition of clusters of items that were either high or low in TPs. Results were taken to indicate that SL in infancy was domain-general and innate: that is, SL was proposed to operate across multiple kinds of sensory inputs and available from birth (Kirkham et al., 2002). Yet only the Aslin et al. study was designed to rule out other kinds of statistical information, such as frequency, in favor of TPs. The Kirkham et al. study did not test for infants' TP learning or segmentation: Instead, infants were tested for simple discriminations of

TP-structured sequences vs. pseudorandom sequences. This kind of discrimination was later discovered to occur even without a learning phase: Five-month-olds were tested with two six-shaped visual sequences, seen in alternation, that either were ordered randomly or followed the statistical structure described previously. Interestingly, the infants looked longer at random vs. structured sequences of visual shapes, even without prior familiarization or habituation (Addyman & Mareschal, 2013), thus demonstrating a spontaneous preference for random sequences that does not require prior experience or learning. This implies that infants in the Kirkham et al. and Bulf et al. (2011) experiments did not necessarily learn anything during the experiment, calling into question the likelihood that SL operates from birth and undergoes little developmental change in infancy.

The Addyman and Mareschal (2013) results also imply that young infants can discriminate sequences solely from *ordinal* information—the orderings of items. Ordinal information, like TP information, is a kind of statistic, but recognition of ordinal violations may be less demanding than recognition of TP violations, especially when infants are also required to segment an input stream into units. Consistent with this possibility, infants as young as 3 months were reported to identify violations of serial order in audiovisual sequences (Lewkowicz, 2008); in contrast, 4.5- and 5-month-olds, but not younger infants, segmented visual sequences from TP differences (Marcovitch & Lewkowicz, 2009; Slone & Johnson, 2015). These studies highlight an important distinction between *discrimination* of different sequences based on statistical information and *learning* statistical information to segment sequences of items into clusters or units. The studies also highlight the distinction between different statistics that might be identified and/or learned. Furthermore, the Addyman and Mareschal results are important in demonstrating that infants' preferences for items in sequence might stem from differences in complexity (cf. Kidd, Piantadosi, & Aslin, 2012, 2014; Tummeltshammer & Kirkham, 2013).

Other kinds of inputs have been examined in infant SL tasks. For example, by 11 months, infants can learn probabilistic sequences of items appearing in predictable spatial locations, and 8-month-olds can learn spatiotemporal sequences when item location combines with color and shape cues (Kirkham et al., 2007; cf. Sobel & Kirkham, 2006; Tummeltshammer & Kirkham, 2013); 5-month-olds tested under identical conditions did not appear sensitive to spatial information for the sequence. Infants at 8 months also were reported to learn predictable co-occurrences of items in visual arrays, akin to TPs between items in sequence (Fiser & Aslin, 2002b), and at 9 months, infants' SL of object features in visual arrays was facilitated by a social cue: a woman seen to be looking in the location where a coherent configuration was displayed (Wu, Gopnik, Richardson, & Kirkham, 2011). Also, other cues to segmentation that are present in real-world speech, such as prosody (Thiessen & Saffran, 2003) and word length (Lew-Williams, Pelucchi, & Saffran, 2011; Lew-Williams & Saffran, 2012), interact with, and constrain, SL of speech sounds.

Finally, there have been claims that infant SL has an important role in development of abstract "rule learning," a kind of pattern learning involving identification of simple reduplicative patterns and generalization of the pattern to new items (e.g., Gerken, 2006; Marcus, Vijayan, Rao, & Vishton, 1999), and an important foundation for analogical reasoning (see Chap. 5). Infants' learning and generalization of simple abstract rules in sequential patterns were first investigated by Marcus et al. (1999), who exposed 7-month-old infants to strings consisting of computer-generated speech. In their first experiment, strings followed either an "ABA" pattern (e.g., *gah tee gah*, *nee lah nee*) or an "ABB" pattern (e.g., *gah tee tee*, *nee lah lah*). A and B items were separated by 250 ms of silence, strings by 1 s of silence. The speech stream was accompanied by a flashing light, mounted centrally in the testing chamber. After 2 minutes of continuous repetitions of one of these two familiarization patterns, the infants received trials of the same (familiar) pattern instantiated by different phonemes (e.g., *woh fei woh*, *dee koh dee*) and the second (novel) pattern on alternating trial, from a speaker located either to the left or right of the infant. Each kind of test stimulus was also accompanied by a flashing light, located either left or right, and learning was operationalized in terms of differences in looking time toward the light when the word or part-word was heard. The infants exhibited a reliable preference for the novel pattern, a result that extended to a test of ABB vs. AAB. The balance of phonetic features across familiarization and test stimuli ruled out the possibility that performance was based on learning sequences of low-level cues (such as voiced vs. unvoiced consonants). Importantly, the positive outcome of the ABB/AAB comparison obviated an account based on learning a simple reduplication pattern (i.e., adjacent repetition) without respect to its place in sequence (i.e., initial/final edge position).

The Marcus et al. (1999) task bears superficial similarities to the Saffran et al. (1996) task: Infants listened to a structured speech stream for 2 minutes, and they were tested for recognition of the underlying pattern using a head-turn method to generate preferences for a flashing light on one vs. the other side of a testing chamber. Yet there is a vital difference in what is tested in these two paradigms. In SL tasks such as the Saffran et al. study, infants are asked to segment a speech stream into units that are bounded by dips in TPs: that is, the words heard at test (now segmented) had higher internal TPs than nonwords or part-words. In abstract rule-learning tasks such as the Marcus et al. study, in contrast, infants are not required to segment the input (it is already segmented into units) nor are they required to recognize correspondences among items, learned during familiarization, to the same items at test. This is because no items from familiarization were heard at test. Instead, infants were required to learn an *abstract* pattern that, as noted previously, was *independent of surface features* (such as vowels and consonants).

Nevertheless, there have been proposals for a common mechanism supporting infant SL and abstract rule learning (see Chap. 5 for additional discussion), perhaps because (a) language experience facilitates both SL (e.g., Saffran & Wilson, 2003) and abstract rule learning (Marcus, Fernandes, & Johnson, 2007), (b) simple connectionist models can explain both sets of results (e.g., Christiansen & Curtin, 1999), (c) simple reduplications may comprise a "perceptual primitive" as a basis for pattern extraction (e.g., Gerken, Dawson, Chatila, & Tenenbaum, 2015; Gómez & Gerken, 2000), or (d) abstract categories can arise from purely statistical input (Aslin & Newport, 2012, 2014; see Reeder, Newport, & Aslin, 2013 for evidence from adults). However, to my knowledge, there are no reports of any direct demonstrations in infants that SL and abstract rule learning stem from a single learning mechanism. Indeed, experiments in my lab designed to test SL and abstract rule learning from identical four-item audiovisual sequences found that 11-month-olds could learn about specific items and their positions in sequence—that is, statistical information, in this case order of items in a series. In contrast, the infants did not appear to learn a simple reduplication—that is, an abstract rule that was independent of surface features (Schonberg, Marcus, & Johnson, 2018; see Fig. 2.3).

In summary, studies of SL in infancy have tended to focus on infants' learning of TPs in segmentation tasks. Other kinds of statistical information are also available (ordinal information, frequency, repetition, linguistic cues), but their roles in segmenting and learning, and their interactions with TPs between stimulus features as contributions to learning, are not well understood at present.

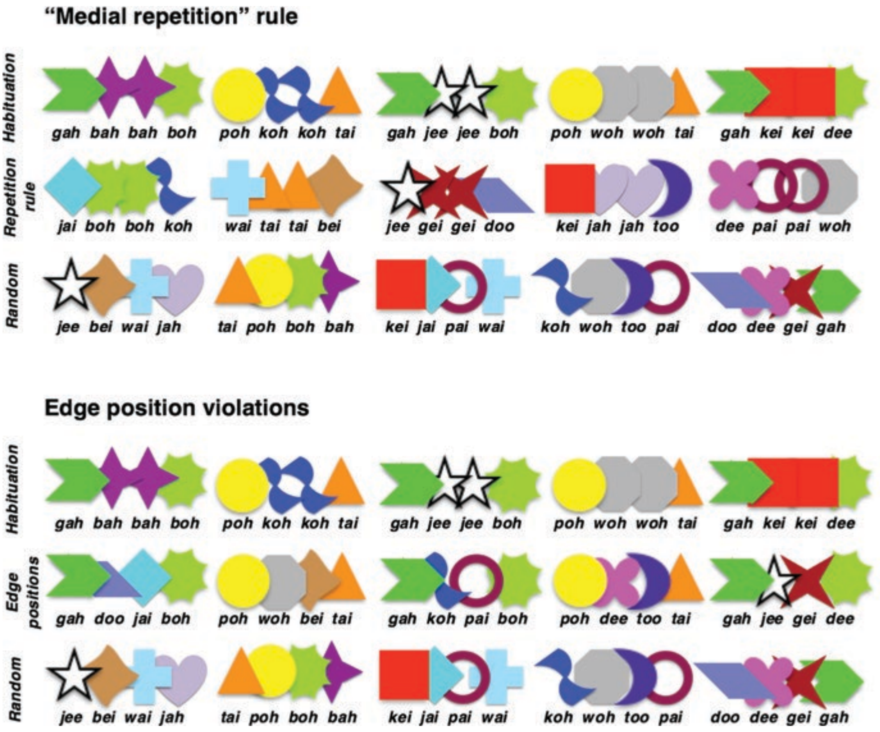

Fig. 2.3 Schematic depiction of stimuli used to test for infants' abstract rule learning, a "medial repetition rule" (top panel), and statistical learning, the specific positions of items in their ordinal positions (bottom panel). Each condition used identical habituation stimuli but tested for learning of either an abstract pattern or one based on items in sequence. Eleven-month-olds appeared to learn edge position violations, but not the abstract repetition rule. (Adapted from Schonberg et al. (2018))

Testing Methods

The majority of published infant SL studies have employed a learning phase (familiarization or habituation) followed by a test phase in which infants are observed for evidence of segmentation of continuous input, undifferentiated except by virtue of TP differences among adjacent items, and recognition of parsed units vs. foil stimuli consisting of a reordering of individual items (see Saffran & Kirkham, 2018, for review). Effects of variations in testing methods, such as the use of different stimuli in the same paradigm, are not well understood (see Chap. 4 for further discussion), but there is some evidence that their investigation can be fruitful (Kirkham et al., 2007). For example, Lewkowicz (2004) examined infants' detection of violations of serial order of items in sequence and found that ordinal information was more readily identified in sequences of linearly moving objects than looming objects presented in a single location (as in the Kirkham et al., 2002, method).

Eye-tracking and brain-based methods have provided complementary and, in some cases, unique insights into infants' SL. Eye-tracking methods involve records of infants' point of gaze as they view displays on a monitor (Gredebäck, Johnson, & von Hofsten, 2010). SL studies have examined eye movement (saccadic) latencies to items in sequence, the prediction being that spatial locations of more predictable items, by virtue of high TPs between items, will be fixated more quickly. As noted previously, evidence in support of this prediction was provided by the Kirkham et al. (2007) experiment in which infants were found to look toward locations in which a predictable item appeared vs. one of the other five locations on the display. More recently, Tummeltshammer and Kirkham (2013) examined 8-month-olds' saccadic latencies when viewing six-location visual arrays with sequences of spatiotemporal events. Arrays resembled a house or storefront with windows in which shapes appeared one at a time in a probabilistic sequence comprising three shape pairs. Each shape appeared in a particular window, disappeared, and subsequently reappeared in a different window according to its assigned probability. Items appeared in sequence with TPs of 1.0, 0.75, or 0.5, and one group of infants viewed arrays with additional competing visual distracters. Items with higher TPs were attended more often and with fewer errors (i.e., predictive looks) overall, and this effect interacted with the presence of distracters: With no distraction, latencies were fastest to high-probability (0.75) TP events, but with distracters, latencies were fastest to "deterministic" events with TPs of 1.0. These findings suggest that infants' SL guides predictive behavior and that predictions are influenced by distributional properties of the entire scene, even events (distracters) unrelated to the predictable items.

Brain-based methods have been used to examine cortical loci of SL with functional MRI under various testing conditions in adults (e.g., Karuza et al., 2013; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Turk-Browne, Scholl, Chun, & Johnson, 2009) and children (McNealy, Mazziotta, & Dapretto, 2011). Electrophysiological methods, in particular event-related potentials (ERPs), have yielded evidence concerning the time course of "online" learning in adults from changes in the timing and strength of electrical cortical potentials (viz., ERP components) recorded at the scalp (e.g., Abla, Katahira, & Okanoya, 2008; Abla & Okanoya, 2009). ERPs have been used as an index of differences in visual SL between children with autism spectrum disorder (ASD) and typically developing children and have revealed impairments in some children with ASD (Jeste et al., 2015). ERP methods are more feasible for use with young populations relative to fMRI, and they have been used to examine SL in infants. For example, Teinonen et al. (2009) observed ERP differences to statistically structured vs. unstructured speech sequences in sleeping neonates, and Marin et al. (2019) observed ERP differences during a visual SL task between 3-month-old infants at elevated risk for ASD (due to high genetic load) and low-risk infants. The Jeste et al. and Marin et al. studies are discussed in more detail in the next section.

In sum, eye-tracking and brain-based methods, in particular electrophysiological methods, require specialized designs and equipment but can provide particularly sensitive measures of SL. This can be especially important for infant studies. Infants' control of eye movements is well-established even at birth (Gredebäck et al., 2010), and clever research designs can exploit infants' tendency to explore novel scenes and learn contingencies among events, including probabilistic events. ERPs, likewise, can be used in infants at all ages (de Haan, 2007) and can reveal cortical activity in response to probabilistic events that more overt behaviors cannot necessarily reveal.

Implications of Infant SL for Cognitive Development and Developmental Disabilities

There is extensive evidence that SL is related to and perhaps facilitates language acquisition (see Romberg & Saffran, 2010, for review). In 8-month-olds, for example, nonsense words acquired via SL are treated as "candidate" words when embedded in new linguistic contexts (Saffran, 2001); moreover, SL provides candidate words that can become associated with novel objects at 17 months (Graf Estes, Evans, Alibali, & Saffran, 2007) and with novel object categories at 22 months (Lany & Saffran, 2010). In addition, 8.5-month-olds' performance on a visual SL task was correlated with the infants' vocabulary size, assessed by parental report (Shafto, Conway, Field, & Houston, 2011). Six-month-old infants' oculomotor responses to events in a visual pattern-learning task predicted vocabulary size 16 months later (Ellis, Gonzalez, & Deák, 2014), and 6- to 8-year-olds' visual SL performance predicted their comprehension of native-language syntax (Kidd & Arciuli, 2016). However, measures of cognitive development more broadly (i.e., independent of language), such as performance on the Bayley Scales (Bayley, 2005), general IQ, and gesture comprehension, were not related to SL performance (Ellis et al., 2014; Kidd & Arciuli, 2016; Shafto et al., 2011).

Evidence for how SL might affect developmental disabilities is consistent with these findings: SL is related to language acquisition and performance but may have somewhat less impact on cognitive function. For example, the possibility that SL is impaired in ASD has received mixed support. Some studies report impaired SL (e.g., Jeste et al., 2015; Scott-Van Zeeland et al., 2010), but others report little or no impairment (Mayo & Eigsti, 2012) or even *enhanced* SL in ASD (Roser, Aslin, McKenzie, Zahra, & Fiser, 2015). ASD, however, is a heterogeneous disorder that remains poorly understood at the level of individual differences (Jeste et al., 2015), and notably, these studies of SL in ASD used varying methods and tested different populations (e.g., children with unknown symptom severity vs. high-functioning adults), making direct comparisons of results difficult. In infants with Williams syndrome, a developmental disorder characterized by strong language skills but impaired intellectual capacity, SL seems to be intact (Cashon, Ha, Graf Estes, Saffran, & Mervis, 2016). A recent meta-analysis found strong and consistent evidence for reduced SL in individuals with specific language impairment but mixed evidence for reduced SL in individuals with ASD (Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016).

Recently, Jeste et al. (2015) investigated ERP correlates of SL in children with ASD vs. typically developing controls. ERP was recorded as children watched streams of looming shapes, similar to methods described previously with infants (Kirkham et al., 2002), and after a learning phase, they introduced a violation of the expected sequence by showing an unexpected shape. This study revealed two important findings. First, the ASD group showed attenuated evidence of SL in two ERP components: a reduced “N1” component, which was theorized to signify early visual recognition, akin to the N100 in adults (Coull, 1998) and a reduced P300 component, which represents attention to salient information and probabilities of a target stimulus (Picton, 1992). Second, analyses of individual differences in the ASD group revealed a positive correlation between N1 amplitude difference and nonverbal IQ and a positive correlation between P300 amplitude difference and adaptive social function. This study demonstrates, therefore, that ASD is highly variable among individuals, and variability in learning capacity may help explain deficits in social, and perhaps cognitive, function.

In infants, my colleagues and I recently recorded ERPs in 3-month-old infants at elevated or low risk for ASD, due to the presence (or not) of one or more close family members having received a diagnosis of ASD (Marin et al., 2019). We asked whether visual SL at 3 months, recorded as described previously for the Jeste et al. (2015) experiment, might predict cognitive function and ASD symptoms at 18 months. Interestingly, higher-risk infants demonstrated increased neural responses (late slow wave and N700 components) to the probabilistic event, whereas low-risk infants demonstrated increased neural responses to the deterministic (expected) event. Moreover, individual differences in these ERP components at 3 months predicted visual reception ability and ASD symptoms at 18 months of age. The reasons for these differences so early in infancy are not yet clear, but the potential predictive value for emerging ASD symptoms from such observations may be an important finding.

Mechanisms Underlying Statistical Learning in Infancy

As noted in prior sections, SL is a powerful means by which infants learn about a structured environment, and studies of SL can be particularly informative about learning in children with developmental disabilities. Yet the specific processes underlying SL remain unclear. Recently, research in my lab (Slone & Johnson, 2018) investigated two types of models underlying statistical learning: “statistical” (or “transition-finding”) and “chunking” (or “clustering”) models that have been proposed to account for SL in adults (Thiessen, Kronstein, & Hufnagle, 2013).

The goal of both statistical and chunking models is to account for sensitivity to sequential structure and the units that are learned, but they differ in the proposed representations stored in memory. Statistical or TP-learning models can be instantiated in computational models known as simple recurrent networks (e.g., Elman, 1990) that compute and represent statistical relations between items, such as TPs, in memory. Representing TPs informs the model of the likelihood of two items occurring together and allows the model to predict individual items based on previous items in a sequence. In the syllable stream used by Saffran et al. (1996), for example (Fig. 2.1), the model would learn that the probability of *pi* after *tu* and the probability of *ro* after *pi* are high, because items *tu*, *pi*, and *ro* always appear in order (in the word *tupiro*). The probability of *pa* after *ro*, in contrast, will be lower because *padoti* follows *tupiro* only 1/3 of the time in the familiarization sequence. In this way, statistical models can distinguish statistically coherent units of information contained within a sequence (e.g., *tupiro*) from less coherent units like part- words (e.g., *ropado*). Importantly, statistical models do not explicitly represent statistically coherent units (e.g., words); rather, they represent statistical relations between items (e.g., syllables) as TPs.

Chunking models, in contrast, represent statistically coherent units of information in memory. One such computational model, the “truncated recursive autoassociative chunk extractor” (TRACX), forms groupings simply by joining items that tend to co-occur (French, Addyman, & Mareschal, 2011; Mareschal & French, 2017). Groupings, or chunks, become single units that can be stored in memory. Representations of units whose component items co-occur regularly are progressively strengthened in memory, whereas representations of units whose component items do not co-occur regularly are forgotten. In the Saffran et al. (1996) sequence, for example, the model could initially capture the sequence *tupiropadoti* in three separate chunks: *tupi*, *ropa*, and *doti*. Over time, chunks *tupi* and *doti* will be reinforced in memory because their component items always co-occur. In contrast, chunk *ropa* will only be weakly represented because its component items co-occur less frequently. Moreover, once the sequences *tupi* and *piro* become represented as single chunks, it becomes possible for *tupiro* to be captured as an even larger chunk (i.e., as the aggregate of *tupi* and *piro*). Thus, with sufficient exposure, the model will form strong representations of statistically coherent units of information (e.g., *tupiro*) and distinguish them from weakly represented part-words (e.g., *ropado*). Statistical relations among items are not retained in memory over time—only the chunks.

Several studies have suggested that adults’ SL is best accounted for by chunking models (Fiser & Aslin, 2002a; Giroux & Rey, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008; Perruchet & Poulin-Charronnat, 2012), but others have provided evidence that statistical TP-learning models may often provide a better fit for adult performance in SL tasks (Endress & Langus, 2017; Endress & Mehler, 2009). Moreover, it remains unknown which type of model best accounts for infants’ SL performance.

We addressed the question of whether statistical or chunking was the best account of infant sequence learning in three experiments with 8-month-olds (Slone & Johnson, 2018). In the first experiment, infants were familiarized with five-item sequences for 5 minutes. Sequences were constructed such that certain items were shared across units (see Fig. 2.4a). Following habituation, infants were tested for recognition of a familiar triplet (tantamount to a word in the Saffran et al., 1996,

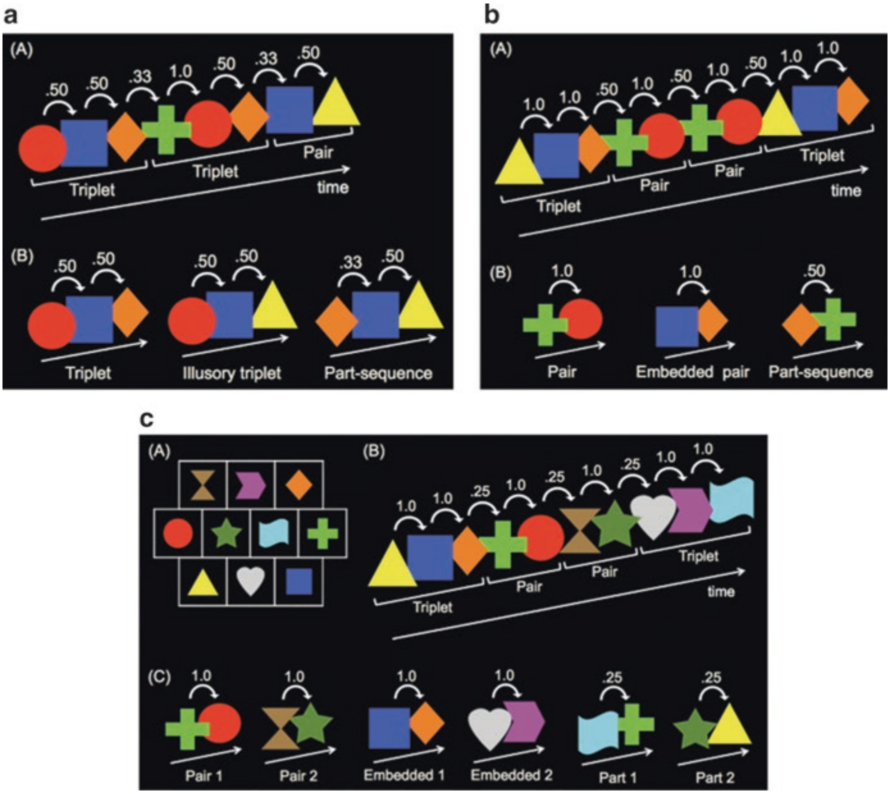

Fig. 2.4 Schematic depiction of familiarization and test sequences in experiments testing statistical vs. chunking models (see text for details). Numbers above adjacent shapes represent TPs during familiarization. Familiarization sequences are seen at the top in each panel and test sequences at the bottom. Brackets below shapes indicate the unit structure of the familiarization sequences. (**a**) Illusory triplet, (**b**) embedded pair, (**c**) embedded pair with increased exposure. (Adapted from Slone and Johnson (2018))

study), a part-sequence (triplet), and an "illusory" triplet, composed of two pairs of items that had high TPs but had not been seen together. We reasoned that if infants had learned a chunk (the triplet) during familiarization, they would recognize the triplet when seen in isolation at test, but not the illusory triplet or the part-sequence. If infants recognized the illusory triplet, however, this would support statistical models, because the TPs of the familiar and illusory triplets were identical. The first prediction was supported, in line with chunking models.

In the second experiment, infants were familiarized with five-item sequences composed of one unique triplet and one unique pair (no shared items; see Fig. 2.4b). At test, infants viewed a familiar pair, a part-sequence (pair), and an "embedded pair," composed of items that were part of the triplet. We reasoned that infant looking at test would reveal whether they formed a triplet chunk that excluded the embedded pair, consistent with chunking models: recognition of the familiar pair but not the part-sequence or embedded pair. This prediction was also supported, again in line with chunking models.

Finally, in a third experiment, we asked if we might capture a point in time during familiarization when infants had learned TPs among adjacent items but not yet formed full chunks. We did this with a condition testing for recognition of embedded pairs, as in the previous experiment, but now employing twice the numbers of items and units: two unique triplets and two unique pairs, comprising 10 items in total (see Fig. 2.4c). Exposure time was kept the same, however, requiring infants to track more relations among items and thus perhaps impairing chunk formation. In support of this prediction and in contrast to the second experiment, infants in the third study appeared to recognize both familiar and embedded pairs, evidence that infants learned TPs between adjoining items, but exposure time had been insufficient for learning chunks of triplets. Taken together, these results inform the nature of infants' SL: As a first step in sequence learning, TPs between items are acquired, and then chunks are learned from the accumulation of TP-linked pairs. But whether TPs are immediately discarded may depend on the learning requirements in context (cf. Endress & Langus, 2017).

Conclusions and Broader Implications

Results of statistical learning studies can provide important constraints for theories of cognitive development, in particular computational models of associative learning in developmental disorders (Tovar, Westermann, & Torres, 2018), cross-situational/multimodal computational models of language acquisition (Monaghan, 2017), and Bayesian computational models of category learning (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Yet many models of infant cognition do not take account of possible effects of stimulus modality on learning or possible constraints in infant attention, memory, and learning capacity (e.g., Franz & Triesch, 2010; Rogers, Rakison, & McClelland, 2004; Tenenbaum et al., 2011).

In sum, much remains to be discovered with respect to infants' SL, despite important progress in our understanding of SL as a vital part of language acquisition and as a window into the nature of some developmental disabilities. For example, links between infants' SL and abstract rule learning remain unexplored but may involve comparison processes between items and relations (see Chap. 5). In addition, neural processes that give rise to statistical learning are becoming understood as interactions between the declarative and nondeclarative memory systems of the brain (Batterink, Paller, & Reber, 2019), but little is known about how these interactions develop early in life. Nor is the developmental time course of SL in individuals well understood (Siegelman & Frost, 2015). Finally, an important question concerns the role of SL in infants' learning of real-world events. For example, when children begin to learn relations between objects, these may become chunked into a unit of "causal action" and associated with a label (e.g., a verb). Evidence for this and other possible contributions of SL to cognitive development await future study.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*, 952–964. https://doi.org/10.1162/jocn.2008.20058

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research, 64*, 185–190. https://doi.org/10.1016/j.neures.2009.02.013

Addyman, C., & Mareschal, D. (2013). Local redundancy governs infants' spontaneous orienting to visual-temporal sequences. *Child Development, 84*, 1137–1144. https://doi.org/10.1111/cdev.1206

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*, 170–176. https://doi.org/10.1177/0963721412436806

Aslin, R. N., & Newport, E. L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language Learning, 64: Cognitive Neuroscience Supplement, 2*, 86–105. https://doi.org/10.1111/lang.12074

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324. https://doi.org/10.1111/1467-9280.00063

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15*, 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science, 11*, 482–503. https://doi.org/10.1111/tops/12420

Bayley, N. (2005). *Bayley scales of infant and toddler development-III*. San Antonio, TX: Pearson.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121*, 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

Cashon, C. H., Ha, O. R., Graf Estes, K., Saffran, J. R., & Mervis, C. B. (2016). Infants with Williams syndrome detect statistical regularities in continuous speech. *Cognition, 154*, 165–168. https://doi.org/10.1016/j.cognition.2016.05.009

Chomsky, N. A. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Curtin, S. (1999). Transfer of learning: Rule acquisition or statistical learning? *Trends in Cognitive Sciences, 3*, 289–290. https://doi.org/10.1016/S1364-6613(99)01356-X

Coull, J. T. (1998). Neural correlates of attention and arousal: Insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in Neurobiology, 55*, 343–361. https://doi.org/10.1016/S0301-0082(98)00011-2

De Haan, M. (2007). *Infant EEG and event-related potentials*. London, UK: Psychology Press.

Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: What is the association with later vocabulary? *Language Learning and Development, 10*, 36–50. https://doi.org/10.1080/15475441.2013.799988

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211. https://doi.org/10.1207/s15516709cog1402_1

Endress, A., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351–367. https://doi.org/10.1016/j.jml.2008.10.003

Endress, A. D., & Langus, A. (2017). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology, 92*, 37–64. https://doi.org/10.1016/j.cogpsych.2016.11.004

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467. https://doi.org/10.1037//0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences (USA), 99*, 15822–15826. https://doi.org/10.1073/pnas.232472899

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728–3740. https://doi.org/10.1121/1.418332

Franz, A., & Triesch, J. (2010). A unified computational model of the development of object unity, object permanence, and occluded object trajectory perception. *Infant Behavior and Development, 33*, 635–653. https://doi.org/10.1016/j.infbeh.2010.07.018

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review, 118*, 614–636. https://doi.org/10.1037/a0025255

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*, 117–125.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74. https://doi.org/10.1016/j.cognition.2005.03.003

Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science, 18*, 80–89. https://doi.org/10.1111/desc.12183

Gibson, E. J., & Gibson, J. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 32–41. https://doi.org/10.1037/h0048826

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33*, 260–272. https://doi.org/10.1111/j.1551-6709.2009.01012.x

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Gredebäck, G., Johnson, S. P., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*, 1–19. https://doi.org/10.1080/87565640903325758

Hasson, U. (2016). The neurobiology of uncertainty: Implications for statistical learning. *Philosophical Transactions of the Royal Society B, 372*, 20160048. https://doi.org/10.1098/rstb.2016.0048

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*, 658–680. https://doi.org/10.1037//0096-3445.130.4.658

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., et al. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science, 18*, 90–105. https://doi.org/10.1111/desc.12188

Johnson, M. J., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*, 46–54. https://doi.org/10.1016/j.bandl.2012.11.007

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7*, e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*, 1795–1804. https://doi.org/10.1111/cdev.12263

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development, 87*, 184–193. https://doi.org/10.1111/cdev.12461

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78*, 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology, 3*, 598. https://doi.org/10.3389/fpsyg.2012.00598

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science, 21*, 284–291. https://doi.org/10.1177/0956797609358570

Leslie, A. M. (1997). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Lewkowicz, D. J. (2004). Perception of serial order in infants. *Developmental Science, 7*, 175–184. https://doi.org/10.1111/j.1467-7687.2004.00336.x

Lewkowicz, D. J. (2008). Perception of dynamic and static audiovisual sequences. *Child Development, 79*, 1538–1554. https://doi.org/10.1111/j.1467-8624.2008.01204.x

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*, 1323–1329. https://doi.org/10.1111/j.1467-7687.2011.01079.x

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 122*, 241–246. https://doi.org/10.1016/j.cognition.2011.10.007

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning with a balanced chunk strength design. *Journal of Cognitive Neuroscience, 16*, 427–438. https://doi.org/10.1162/089892904322926764

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science, 12*, 1020–1025. https://doi.org/10.1111/j.1467-7687.2009.00838.x

Marcus, G., Fernandes, K., & Johnson, S. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*, 387–391. https://doi.org/10.1111/j.1467-9280.2007.01910.x

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month- old infants. *Science, 283*, 77–80. https://doi.org/10.1126/science.283.5398.77

Mareschal, D., & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B, 372*, 20160057. https://doi.org/10.1098/rstb.2016.0057

Marin, A., Hutman, T., Ponting, C., McDonald, N. M., Dickinson, A., Dapretto, M., et al. (2019). *Electrophysiological signatures of visual statistical learning in three-month old infants at familial and low risk for autism spectrum disorder*. Manuscript submitted for publication.

Mayo, J., & Eigsti, I. M. (2012). A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders, 42*, 2476–2485. https://doi.org/10.1007/s10803-012-1493-0

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2011). Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science, 14*, 1261–1282. https://doi.org/10.1111/j.1467-7687.2011.01075.x

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176–185. https://doi.org/10.2307/1418920

Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science, 9*, 21–34. https://doi.org/10.1111/tops.12239

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology, 7*, 1245. https://doi.org/10.3389/fpsyg.2016.01245

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences (USA), 105*, 2745–2750. https://doi.org/10.1073/pnas.0708424105

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language, 66*, 807–818. https://doi.org/10.1016/j.jml.2012.02.010

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*, 456–479.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*, 30–54. https://doi.org/10.1016/j.cogpsych.2012.09.001

Rogers, T. T., Rakison, D. H., & McClelland, J. L. (2004). U-shaped curves in development: A PDP approach. *Journal of Cognition and Development, 1*, 137–145. https://doi.org/10.1207/s15327647jcd0501_14

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 906–914. https://doi.org/10.1002/wcs.78

Roser, M. E., Aslin, R. N., McKenzie, R., Zahra, D., & Fiser, J. (2015). Enhanced visual statistical learning in adults with autism. *Neuropsychology, 29*, 163–172. https://doi.org/10.1037/neu0000137

Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition, 81*, 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology, 69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273–284. https://doi.org/10.1207/S15327078IN0402_07

Schonberg, C., Marcus, G. F., & Johnson, S. P. (2018). The roles of item repetition and position in infants' abstract rule learning. *Infant Behavior and Development, 53*, 64–80. https://doi.org/10.1016/j.infbeh.2018.08.003

Scott-Van Zeeland, A. A., McNealy, K., Wang, A. T., Sigman, M., Bookheimer, S. Y., & Dapretto, M. (2010). No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biological Psychiatry, 88*, 345–351. https://doi.org/10.1016/j.biopsych.2010.01.011

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2011). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy, 17*, 247–271. https://doi.org/10.1111/j.1532-7078.2011.00085.x

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2- and 5-month-olds' segmentation of continuous visual sequences. *Journal of Experimental Child Psychology, 133*, 47–56. https://doi.org/10.1016/j.jecp.2015.01.007

Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition, 178*, 92–102. https://doi.org/10.1016/j.cognition.2018.05.016

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*, 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56. https://doi.org/10.1016/0364-0213(90)90025-R

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*, 21. https://doi.org/10.1186/1471-2202-10-21

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*, 1279–1285. https://doi.org/10.1126/science.1192788

Thiessen, E. D. (2016). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B, 372*, 20160048. https://doi.org/10.1098/rstb.2016.0056

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*, 792–814. https://doi.org/10.1037/a0030801

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*, 867–875. https://doi.org/10.3758/BF03193539

Tovar, A. E., Westermann, G., & Torres, A. (2018). From altered synaptic plasticity to atypical learning: A computational model of Down syndrome. *Cognition, 171*, 15–24. https://doi.org/10.1016/j.cognition.2017.10.021

Tummeltshammer, K. S., & Kirkham, N. Z. (2013). Learning to look: Probabilistic variation and noise guide infants' eye movements. *Developmental Science, 16*, 760–771. https://doi.org/10.1111/desc.12064

Turk-Browne, N. B. (2012). Statistical learning and its consequences. In M. D. Dodd & J. H. Flowers (Eds.), *The influence of attention, learning, and motivation on visual search (Nebraska Symposium on Motivation)* (pp. 117–146). New York, NY: Springer.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience, 30*, 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 1707–1717. https://doi.org/10.1121/1.402450

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology, 47*, 1220–1229. https://doi.org/10.1037/a0024023

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750. https://doi.org/10.1038/358749a0