Infant Statistical Learning

Jenny R. Saffran[1] and **Natasha Z. Kirkham**[2]

[1]Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin 53706

[2]Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, United Kingdom

Abstract

Perception involves making sense of a dynamic, multimodal environment. In the absence of mechanisms capable of exploiting the statistical patterns in the natural world, infants would face an insurmountable computational problem. Infant statistical learning mechanisms facilitate the detection of structure. These abilities allow the infant to compute across elements in their environmental input, extracting patterns for further processing and subsequent learning. In this selective review, we summarize findings that show that statistical learning is both a broad and flexible mechanism (supporting learning from different modalities across many different content areas) and input specific (shifting computations depending on the type of input and goal of learning). We suggest that statistical learning not only provides a framework for studying language development and object knowledge in constrained laboratory settings, but also allows researchers to tackle real-world problems, such as multilingualism, the role of ever-changing learning environments, and differential developmental trajectories.

Keywords

statistical learning; infancy; cognitive development; language development; sequence learning; perceptual development; multisensory

1. Introduction

How do learners discern the structure organizing their environments? This question has been at the center of intellectual debates since the founding of the field of psychology, providing impetus for the theories of Ivan Pavlov and B.F. Skinner. In the domain of linguistics, a similar question—how learners discern the structure of natural languages—led to the two dominant perspectives of the twentieth century: the structural linguistics of Leonard Bloomfield and Zellig Harris and the generative linguistics of Harris's most famous student, Noam Chomsky.

All theories agree that learners must have some way to ascertain which patterns are relevant to acquire and store and which are not. But what factors determine which patterns merit

learning? This is where theoretical accounts diverge. Do the data themselves tell learners what matters and why? Or do learners receive guidance—via innate predispositions or knowledge—illuminating what to learn? As the structures to be learned become more abstract and less transparently mirrored in the input, the answers to these questions become less obvious. Similarly, as the number of possible patterns explodes combinatorially in complex input, it becomes less clear which patterns are tracked and why. As Gibson (1966) stressed, we need to understand the nature of the input before we can understand the nature of processing.

Statistical learning mechanisms have become prominent in cognitive and developmental science because they provide ways to test specific hypotheses about what is learned from any given set of input, and how. The term statistical learning originated in the machine learning literature and made contact with cognitive science through its application to problems in natural language processing and computer vision. In particular, connectionist models and other computational analyses of linguistic corpora demonstrated that, for suitably equipped learners, myriad statistical patterns are available in language input that could help learners to break the code of their native language. When creating models of human vision, it has become obvious that cortical-cell behavior is related to the statistics inherent in the natural environment (Field 1987).

1.1. Initial Evidence for Statistical Learning in Human Infants

The analyses described above made it clear that statistical patterns lurk in the natural world, including both the linguistic and visual environments. What remained unknown was whether human learners could take advantage of these patterns. In particular, the primary targets of interest for theories of unsupervised learning are infants, who have the most to learn and the least prior knowledge about how to allocate their efforts. Are infants statistical learners?

Several lines of research, beginning in the 1980s, have suggested that the answer is yes. For example, the developmental decline of sensitivity to non-native speech contrasts during the first year suggests that infants are sensitive to the distribution of individual speech sounds in their native language (e.g., Kuhl et al. 1992, Werker & Tees 1984). In the visual domain, researchers in infant cognition have found that infants are sensitive to spatial relationships among repetitive events. For example, young infants can learn simple (two-location), predictable spatial sequences in the visual expectation paradigm, which uses anticipatory eye movements as the index of learning (Haith 1993). By 10 months of age, infants can use correlational structure to discover simple visual categories (Canfield & Haith 1991, Younger 1985, Younger & Cohen 1986). Although these studies were not designed to assess statistical learning mechanisms per se, they provide clear evidence that infants are sensitive to statistical regularities.

1.2. Infant Statistical Language Learning: Initial Evidence

One particular learning problem has emerged as an important test case for claims about infant statistical learning: word segmentation. Speech, even speech addressed to infants, is essentially continuous (except at utterance boundaries). Thus, in order to segment speech into words, infants must have some way to break the speech stream into word-like units.

This problem captured researchers' interest for several related reasons. First, it is a very difficult problem to solve without knowing in advance what the words are, as evidenced by decades of research devoted to speech-to-text technology. Second, despite this difficulty, infants discover word forms in fluent speech sometime in the middle of the first year of postnatal life (e.g., Jusczyk & Aslin 1995). Finally, this is a problem that requires learning. Although there are certainly innate constraints that could be helpful (e.g., Seidl & Johnson 2006, Shukla et al. 2007), infants cannot know a priori which specific sounds are going to be words in their native languages.

The first infant study on word segmentation was published by Goodsitt et al. (1993). In this study, 7-month-olds heard utterances containing a target syllable preceded by two context syllables. The infants were sensitive to the statistical structure of the syllables that served as context for the target syllable. When the context syllables always occurred in the same order, infants were better able to detect the subsequent target syllable, supporting the hypothesis that infants can cluster syllables based on statistical patterns.

Subsequent studies by Saffran et al. (1996) assessed 8-month-olds' ability to track statistical patterns in continuous speech. The only cues available to chunk the speech into word-like units were the statistical regularities with which syllables co-occurred. After two minutes of exposure to the speech stream, infants could discriminate words from sequences of syllables spanning a word boundary (see Pelucchi et al. 2009b for related evidence using natural language stimuli). Importantly, these learning outcomes involved no instruction or explicit feedback, suggesting that statistical learning could be a mandatory response to structured input.

1.3. Infant Statistical Learning in Other Domains: Initial Evidence

A key question raised by these early studies concerns domain specificity. Are statistical learning abilities tailored specifically for a particular domain, like language? Or do they operate across multiple domains (e.g., music, vision, movement)? The first study to address this issue used a musical tone analog of the Saffran et al. (1996) task (Saffran et al. 1999). The results suggested that infants can successfully track nonlinguistic auditory statistics. Although these findings cannot tell us whether the same learning mechanisms subserve learning in both linguistic and nonlinguistic inputs, they are consistent with the view that statistical learning mechanisms are not tailored specifically for language.

Successive studies have expanded these investigations to the visual modality. Fiser & Aslin (2001) demonstrated that adult statistical learning of shape conjunctions (i.e., scenes of arbitrary complex shapes presented simultaneously on a grid) was not only spontaneous but also rapid. Participants learned first-order and higher-order statistics from the spatial arrangement of the shapes in the scene without being instructed to do so. In other words, not only did they learn the immediate relationships between the shapes, they also detected broader probabilistic regularities. Subsequent studies investigated similar capacities in infants. Kirkham et al. (2002) presented 2-, 5-, and 8-month-olds with a visual analog of the original Saffran et al. (1996) paradigm. During test trials, each age group showed heightened looking time to a randomly ordered presentation of the same shapes, suggesting a sensitivity to statistics in the original temporal sequence. Subsequent studies revealed that infants are

sensitive to many different statistical regularities in the visual domain across both temporal and spatial input, enabling them to extract patterns for further processing (Bulf et al. 2011, Fiser & Aslin 2002, Kirkham et al. 2007, Tummeltshammer & Kirkham 2013, Tummeltshammer et al. 2017, Wu et al. 2011).

1.4. Infant Statistical Language Learning

With these data in hand, one might ask whether statistical learning has any bearing on language learning. That is, if infants are able to track statistical regularities across myriad types of input, the original demonstrations of statistical language learning may have been unintentionally misleading in suggesting that statistical learning mechanisms subserve language development. As a case in point, consider the Saffran et al. (1996) study. In describing the results, the authors suggested that "our results raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language" (Saffran et al. 1996, p. 1928). Note, however, that the results of this study simply showed that infants could discriminate between high- and low-probability syllable sequences. Although this ability would certainly be useful for word segmentation, the study did not provide evidence for word segmentation per se.

To address this issue, Graf Estes et al. (2007) investigated whether the output of statistical tracking in fluent speech is actually word-like. They exposed 17-month-old infants to a stream of nonsense words, with only statistical cues to indicate word boundaries. Following exposure, the sound sequences were mapped to novel objects. Infants only acquired the words when the labels were statistically defined words in the fluent speech (for results using natural language stimuli, see Hay et al. 2011). When the labels spanned word boundaries in the fluent speech stream, infants failed to map them to novel objects. These results are consistent with the hypothesis that statistical learning mechanisms are harnessed in domain-relevant ways.In the case of the sequential statistics that characterize continuous speech, infants can exploit these regularities in the service of discovering candidate words in fluent speech (for related evidence in younger infants, see Erickson et al. 2014, Saffran 2001b, Shukla et al. 2011).

1.5. Infant Statistical Learning: Now What?

Over the past two decades, there has been an explosion of research in the area of infant statistical learning. The original Saffran et al. (1996) infant statistical learning study has been cited over 4,000 times (Google Scholar, accessed 2017, https://scholar.google.com/scholar?hl=en&q=saffran+aslin+newport&btnG=&as_sdt=1%2C50&as_sdtp=) and has been applied to myriad learning problems, ages, species, disorders, and implementations. Although scholars disagree about just how useful these mechanisms may be for solving specific problems (e.g., Johnson & Tyler 2010, Lidz & Gagliardi 2015), there appears to be consensus that infants are sensitive to statistical regularities in their environments.

In the remainder of this review, we ask: Now what? There is abundant evidence that infants are sensitive to statistical regularities and that this sensitivity reflects a robust form of incidental learning. The question we hope to address is what this sensitivity to statistical

structure does for infants. To do so, we deconstruct statistical learning into the elements across which computations can occur and the statistics computed over those elements. We then turn to real-world problems where statistical learning approaches may provide novel explanations while raising new questions for future work. Finally, we take a step back and ask why we are statistical learners. The goal is to provide a selective review of the literature, organized in such a way as to motivate future research in this dynamic area.

2. Statistics of What? The Primitives Over Which Statistics are Computed

One of the main arguments leveled by Chomsky against classic learning theory accounts of language acquisition is known as the argument from the poverty of the stimulus (e.g., Chomsky 1965). The crux of this argument lies in the availability of the right kinds of data in the input given the linguistic target to be acquired. Children receive restricted input both quantitatively (in terms of the number of utterances they are exposed to) and qualitatively (in terms of how well the data point to the structures to be acquired). More than 50 years later, debates over the innateness of specific linguistic devices still turn on arguments based on poverty of the stimulus (for a current discussion, see Han et al. 2016, Piantadosi & Kidd 2016). Developmental arguments that hinge on the input extend far beyond the problem of language acquisition. Indeed, since William James first described it as a "blooming, buzzing confusion" (James 1890, p. 488), the infant's complex and noisy multisensory environment has been viewed as an obstacle to learning, obscuring signals and making information less accessible.

From a statistical learning perspective, the stimulus remains problematic. We still ask whether the data support the types of inferences and abstractions that characterize mature knowledge systems. But the quantitative issues are quite different. The question is not whether there is sufficient data in the input. The problem, instead, is that there is too much data. There are vastly many statistics that could be computed over any set of input. This is the case, in part, because of the number of potential computations themselves—a topic which we address below. But the problem of the richness of the stimulus also resides in the nature of the input itself. There are so many potential elements to track. How do infants determine which primitives—the elements over which computations occur—to learn about?

Consider a problem like word segmentation. How do learners know which information to prioritize in their computations? Learners might track the probabilities of co-occurrence of features, phonemes, or syllables, all of which would be reasonable primitives over which to perform computations (e.g., Newport & Aslin 2004). But what about a cue like pitch contour? Pitch is integral to lexical structure in tonal languages like Mandarin or Hmong, and tones are discoverable via statistical information in adult speech (e.g., Gauthier et al. 2007). But pitch contours, however linguistically relevant, are likely irrelevant to word boundary detection, even in a tonal language. Indeed, even speakers of tonal languages find pitch contours difficult to use for word segmentation (Wang & Saffran 2014).

Similar issues arise when considering the primitives over which visual statistical learning operates. Is each visual feature dimension (e.g., color, shape, orientation) independent? Or are features bound together to create higher-order multidimensional units? And does this

sensitivity to either single dimensions or feature chunks change across development? For example, both adults and infants track the statistics of human action sequences (Baldwin et al. 2008, Monroy et al. 2017, Stahl et al. 2014). One can imagine that, in this situation, the details of each visual element would not only be less important than the gestalt of the action being performed, but would also be a distraction from the task at hand.

Furthermore, consider the combination of auditory and visual stimuli. Some visual cues could be helpful insofar as they are correlated with auditory information (e.g., mouth movements) and vice versa (e.g., noting an intensification in sound as an object gets closer). But should learners track them? What about other visual cues, like eye blinks? These are not plausibly useful as cues to linguistic structure. But what is to keep language learners from tracking their statistics as well?

To some degree, this argument is absurd. Obviously, learners do not track the correlations between eye blinks and word boundaries. But why not? This is the problem of the richness of the stimulus. Learners are presumably constrained to consider some elements in their computations and not others. The interesting questions surround the determination of which types of units are tracked and why.

2.1. The Primitives that Enter Into Infants' Computations

The question of primitives matters in any consideration of statistical learning because changing the units that are tracked can change the outcome of learning. This issue was explicitly addressed in a series of developmental studies of statistical learning in tone sequences. As described above, Saffran et al. (1999) demonstrated that similar learning outcomes occurred for continuous sequences of musical tones as for sequences of syllables. This result, however, raised the interesting question of primitives. Consider a tone sequence like AC#E, created to be analogous to a syllable sequence like "golabu." One can compute transitional probabilities between the individual tones (absolute pitches: AC#E), as one would between syllables. But tone sequences contain another primitive that is not present in syllable sequences: musical intervals (relative pitches: ascending major third followed by ascending minor third). These two types of information were confounded in the original study by Saffran et al. (1999), making it unclear which primitives infants tracked.

Subsequent studies revealed interesting developmental differences in the prioritization of musical primitives. Whereas 8-month-olds appear to be biased to track absolute pitches given continuous streams of tones, adults are biased to track relative pitches (Saffran & Griepentrog 2001). Both groups of participants in these studies heard the same sequence of tones in the input, but the groups appear to have learned different things because they tracked different primitives. These preferences for particular primitives can be shifted by altering the input such that absolute pitches are no longer informative, leading infants to track relative pitches (Saffran et al. 2005), or by making the input more musical, leading adults to track absolute pitches (Saffran 2003).

In the case of visual statistics, the issue of what constitutes a visual primitive rears its much-debated head (e.g., Edelman et al. 2002, Marr 1982). Is the learner attending to single feature dimensions individually in a multi-element scene or chunking these elements

together and tracking across objects? In the early laboratory studies (e.g., Fiser & Aslin 2002, Kirkham et al. 2002), visual stimuli were created to be simple two-dimensional, unimodal elements, with the primitives being basic shapes or colors. Either the stimuli removed color from the equation (so that tracking occurred across monochrome individual shapes) or the shapes and colors were perfectly correlated. To determine which features infants were tracking, Kirkham et al. (2007) exposed 8- and 11-month-olds to a spatiotemporal sequence of identical shapes (i.e., the location of the shapes comprised the statistics). Only 11-month-olds showed evidence of learning; 8-month-olds required the shapes to be uniquely colored to pick up on the sequence. In other words, the younger infants needed more cues to the sequence to demonstrate learning. This finding suggested, for the first time, a developmental trajectory in sensitivity to specific visual statistics.

Using a different paradigm assessing infants' ability to track objects made up of multiple features, Kirkham and colleagues (2012) replicated this developmental trajectory; it was not until 10 months of age that infants could reliably unbind the features to track the informative ones. Further addressing the issue of binding across features, Turk-Browne et al. (2008) familiarized adults to a sequence of multifeatured objects and then tested them on objects either without their unique colors or without their shapes. Adults bound features together during learning, depressing their test performance when either of the features was removed. Although in comparable studies adults could easily track the statistics of monochrome shapes, the features presented during the learning phase are clearly important. Turk-Browne et al. (2008) interpreted their results to suggest that visual statistical learning not only depends on what has been encoded, but could actually provide the cues as to what is an object.

Subsequent studies manipulated the stimuli to look at clusters of visual elements (e.g., objects). In a series of studies looking at expectations about object integrity based on feature co-occurrence, Wu et al. (2011) showed 9-month-olds a temporal sequence of colorful multipart objects, within which some parts co-occurred more often than others. The results revealed that infants were sensitive to the differential statistics of the parts within the objects, suggesting that the infants were computing relations not only between the objects, but also within them. Statistical tracking occurs across a variety of different primitives depending on the learning objective.

This pattern of results suggests several general points that should be considered in the study of statistical learning. First, primitives matter; some types of units may be prioritized over others by dint of both perceptual biases (e.g., infant tracking of absolute pitch) and experience (e.g., adult tracking of relative pitch). In addition, primitives matter when considering the learning goal (e.g., predictions about upcoming shapes versus expectations about how objects should behave). Second, the structure of the input matters; the prioritization of units can be shifted when supported by the input. When the sequence of tone words is continually transposed, as in the study by Saffran et al. (2005), the statistics of absolute pitches lose their value—they fail to predict structure. Under those circumstances, infants appear to increase the weight of relative pitches.

The third general point pertains to domain specificity. The specificity of the primitives is one way to construe domain specificity. In this view, whether the computations themselves are general is distinct from considerations of the input representations. It seems clear that different domains of knowledge place distinct demands on perception. Music and language are both auditory, but they make use of different perceptual primitives for the most part (with the exception of some aspects of prosody). Shapes, objects, and action sequences are all part of the visual environment, but the learner must track across increasingly broadly defined primitives. In other words, the learner must chunk multiple individual features (*e.g.*, color, shape) together to fully represent a rich, dynamic, and complex sensory environment (*e.g.*, a sequence of actions). Many of the distinctions between different domains arise from the use of different inputs. The computations themselves may be quite similar, just computed over different types of elements (*e.g.*, Saffran 2008).

2.2. Experience as a Determinant of Primitives

The primitives that enter into statistical learning computations are affected by experience in a particular domain (e.g., Krogh et al. 2013). Some of the evidence to support this claim comes from comparisons between infants and adults, as in the studies on absolute versus relative pitch described in the previous section. Another example comes from a study by Thiessen (2010) examining the role of correlated cues in statistical learning. When adults were given a sequence of syllables paired with shapes, they were better able to learn the syllable statistics than when the shapes were not present. Infants, however, were equally good at tracking the syllable statistics whether the correlated shapes were present or not. Thiessen (2010) hypothesized that this pattern of results can be explained by differences in learners' prior experiences. Adults expect syllable strings to be paired with visual referents based on a lifetime of exposure to language, whereas 8-month-old infants do not yet have this expectation. To test this hypothesis, Thiessen (2010) tested adults on a tone sequence analog of the syllable task, reasoning that adults should not expect tone strings to be paired with shapes. Indeed, the presence of the referents did not improve learning, supporting the view that prior exposure shapes expectations in statistical learning.

A related example comes from the body of work on infant rule learning. This task, pioneered by Marcus and colleagues (1999), involved abstraction away from the specific sequences to which infants are exposed. Infants are better at this task in some domains than others (e.g., Marcus et al. 2007). Experience seems to mediate these effects. Learning is facilitated by the use of familiar rather than unfamiliar stimuli, such as animals rather than abstract shapes or upright faces rather than inverted faces (e.g., Bulf et al. 2015, Saffran et al. 2007). Experience can also inhibit learning. For example, younger infants are actually better than older infants at abstracting across tone sequences because older infants' knowledge of musical structure may inhibit some types of generalizations (Dawson & Gerken 2009). Even within-experiment manipulations can affect whether infants generalize in these tasks. Simply giving 7-month-old infants exposure to social agents who appear to be using tones communicatively leads infants to generalize beyond the tone sequences they have heard, something they do not do in the absence of this experience (Ferguson & Lew-Williams 2016).

As suggested by this last result, experience with the input can affect downstream learning. When learners are first exposed to a particular set of stimuli, some of the primitives may be opaque. An infant listening to a stream of speech initially has access only to statistics at the level of sounds (phonetic features, phonemes, syllables, etc.). Until she learns some of the words, statistics at the word level are invisible (e.g., Saffran & Wilson 2003, Sahni et al. 2010).

Experience with the input affects the primitives over which statistics are computed. This is because statistical learning is dynamic; the output of one learning experience can serve as the input to a new learning experience (Saffran 2008). Saffran & Wilson (2003) investigated this phenomenon by exposing 12-month-old infants to a fluent speech stream in which the words were organized according to a simple grammar. Infants were then tested on grammatical versus ungrammatical sentences, with both types of test items equated for the sequential probabilities of the syllables. Infants successfully discriminated between the test items, suggesting that they were able to solve the task at the level of word patterns rather than just the level of syllable patterns. Infants began the task by tracking syllables but ended up also tracking words.

The structure of the input drives statistical learning in other ways, as well. The input can call attention to some dimensions of the stimuli, highlighting them downstream in learning. For example, infants who are primed with a list of two-syllable nonsense words separated by pauses are subsequently better at using statistical regularities to detect new two-syllable words than to detect three-syllable words in fluent speech, and vice versa (Lew-Williams & Saffran 2012). We find similar results with other types of phonological patterns: Exposure to items that follow one particular pattern facilitates detection of similar items in fluent speech with only statistical cues to word boundaries (Saffran & Thiessen 2003, Thiessen & Saffran 2007). Infants can also use specific experiences—such as exposure to adjacent regularities— to help bootstrap the acquisition of more complex nonadjacent regularities across a single experiment (Lany & Gómez 2008).

The structure of the input helps learners to determine which types of generalizations to draw from the available data. Gerken (2006) adapted the infant rule learning task described above such that the input supported both a broad generalization (ABA versus ABB) and a narrow generalization (AAdi versus AdiA). Infants generalized in the way that was the most consistent with the structure of the input. In a follow-up study, Gerken (2010) made a small change to the input by adding a few counterexamples to the narrow generalization at the end of exposure. Just three counterexamples were enough to shift infants toward the broader generalization, suggesting that infants' learning outcomes are updated on something close to a trial-by-trial basis.

3. Which Statistics Do Learners Track?

These considerations of the primitives over which learning occurs lead us to the next major issue: the computations themselves. Which computations are occurring over these primitives? Do computations change across development and/or across primitives? Learners must be constrained to some degree in deciding which statistics to track (for discussions of

the computational constraints required for optimal visual statistical learning in adults, see Fiser & Aslin 2005, Fiser et al. 2007). One way in which the learner can be constrained is by the eventual goal of learning. If the learner is trying to predict an upcoming event based on previous events, then the statistics will look different than if she is trying to bind across different modalities to form a coherent representation of a scene or display. However, in most infancy paradigms, there is no specified goal or outcome; the infant is placed in front of a display showing a series of events, and their looking times and/or eye movements are measured. So what are the computations that are performed automatically? And do these computations change when a goal is specified?

3.1. Frequency, Transitional Probabilities, and Dependencies

The original work by Saffran et al. (1996) presented the elements of computation as transitional probabilities (i.e., one event in the stream is dependent upon others). These probabilities were higher within words (1.0) than between words (0.33). Additional studies across different domains followed suit. The transitional probabilities between syllables, shapes, objects, audiovisual events, and faces were similar to the original language studies, with within-event probabilities at 1.0 and between-event probabilities significantly lower (e.g., Bulf et al. 2011; Kirkham et al. 2002, 2007; Saffran et al. 1999; Wu et al. 2011).

Frequency counting as an alternative possible computation was ruled out quickly with frequency-controlled studies in both the auditory and the visual domain (Aslin et al. 1998, 2001). By at least 8 months of age, infants appeared to be tracking transitional probabilities regardless of frequency of appearance. Subsequent studies examined this issue in more detail. Marcovitch & Lewkowicz (2009) extended the work of Kirkham et al. (2002) by presenting infants with sequences of shape pairs, defined independently by transitional probabilities and by frequency. Although 2-month-olds failed to show a sensitivity to either computation, 5- and 8-month-olds could track both frequency information and transitional probabilities.

Any given set of inputs contains myriad levels of statistical regularities. What information do infants use to determine which level(s) to track? Research on nonadjacent dependency learning has pointed to some of the key variables that influence this process. For example, given three-word strings, infants tend to learn the adjacent probabilities between those words. However, when the variability of the middle item is increased, infants shift to learn the nonadjacent pairs spanning the middle word (Gómez 2002). Adults learning similar structures are able to track both adjacent and nonadjacent relationships in the same sets of inputs (e.g., Romberg & Saffran 2013a). Interestingly, adults are more aware of the nonadjacent relationships than the adjacent relationships, suggesting that, at least for adults, explicit representations may influence some aspects of statistical learning.

3.2. Complexity and Maximizing Information Gathering

In the real world, input can be measured as more or less complex (i.e., information can have higher or lower levels of redundancy). Complexity has direct implications for which statistics will be attended to. Addyman & Mareschal (2013) ran a modified version of Kirkham et al.'s (2002) experiment, omitting the habituation phase and using looks away as

the dependent measure. This dependent measure allowed for a subtler assessment of infant attention. The results suggested that, in temporally organized visual sequences, 5-month-olds are more sensitive to local repetitions than global statistics. Infants tended to look away during more repetitive portions of the sequence (e.g., during a patterned sequence versus a random sequence). In other words, when complexity was low, infants allocated less attention to the sequence.

These results have implications not only for discussions of which statistics are being computed, but also for thinking about how attention is deployed within these paradigms. Kidd and colleagues (2012, 2014) provided additional evidence suggesting differential attentional deployment as a function of complexity. In their experiments, infants observed visual and auditory episodes of varying complexity based on the predictability (or likelihood) of an upcoming event. In line with their predictions, infants were more likely to look away during episodes of either very low or very high complexity, preferring to allocate attention to events of intermediate complexity.

Learning itself is affected by complexity. In an eye-tracking study with 8-month-olds, three levels of predictability were embedded within one spatiotemporal sequence (Tummeltshammer & Kirkham 2013). Infants showed faster saccade (eye movement) latencies, more anticipation, and increased accuracy to items that were highly predictable relative to items that were either deterministic or unpredictable. In this case, learners may be maximizing information gathering by using likelihoods to constrain search (e.g., Dougherty et al. 2010, Gweon et al. 2010, Téglás et al. 2011, Yu et al. 2007). Because deterministic relations are unambiguous, they offer little information to the infant and, perhaps, little incentive to test possible outcomes with anticipatory looking. Low-probability relations have the most alternatives (and are perhaps most engaging for the infant), but the relevant hypotheses take longer to generate and test. Finally, high-probability relations offer the incentive to gain information but only a few alternatives to confirm or reject, making them a good target for an information-seeking infant with limited resources.

3.3. Issues of Input Specificity

In the visual domain, learners must track statistics not just temporally but also spatially. This differs from auditory input, in which the information to be learned is primarily arrayed in time, not space. In the original visual statistical learning paradigm with infants, Kirkham et al. (2002) presented each group of infants with a temporal sequence of shapes, looming one at a time from the middle of a screen. Results showed that infants were as capable in the visual domain as in the auditory domain, providing a clear analog to the Saffran et al. (1996) study. However, an important aspect of the ability to perceive the visual environment as coherent and intelligible is understanding objects' spatial locations and what their present locations might predict about future events. Acquisition of this type of knowledge is essential for motion perception and for the production of action sequences; one has to learn not only which actions are appropriate, but also where and when they should be performed. For example, if, while looking out the window of your house, you see your child walking up the path to the front door, you can reasonably predict that you will see her next in the doorway of your house. You can use this information to guide appropriate anticipatory

behavior, such as moving to a location that provides a view of the door to greet your child as she comes inside. In other words, each visual event is temporally related both to the previous event and to the future event and occurs within a spatial context.

Indeed, by 8 months of age, infants can learn temporally ordered statistics that involve informative spatial relations (Kirkham et al. 2007, Sobel & Kirkham 2006, Tummeltshammer & Kirkham 2013) and predictable co-occurrences in multi-element scenes (Fiser & Aslin 2002). As mentioned above, 8-month-olds' success in Kirkham et al.'s (2007) spatiotemporal paradigm occurred only when the elements were easily differentiable (e.g., differently colored shapes, each bound to an individual location). This suggests an interesting developmental trajectory in the effect of stimulus type on tracking statistics and highlights the importance of stimuli in processing the input.

Visual streams containing both backward and forward conditional probabilities provide an interesting opportunity to evaluate input specificity. Whereas some statistics, such as frequency, do not contain any information about order or direction, conditional probabilities can differ when computed with respect to the forward direction (i.e., X followed by Y) or the backward direction (i.e., Y preceded by X). Research in the auditory domain has demonstrated that both infants and adults are sensitive to statistical regularities defined in the backward as well as the forward direction (Jones & Pashler 2007, Pelucchi et al. 2009a, Perruchet & Desaulty 2008). However, language is inherently temporal, which suggests a need to be receptive to temporal order. Sensitivity to backward and forward statistics could be modality specific rather than domain general. Indeed, when 8-month-olds were familiarized to either temporal or spatial visual displays, they did not encode the visual regularities in the same way across both temporal and spatial dimensions (Tummeltshammer et al. 2017). Infants computed the predictive direction only in the temporal condition, with chunking occurring in the spatial condition. These data are consistent with the view that the computations performed by learners are susceptible to the specifics of the input.

Modality constraints observed in some studies of statistical learning can be construed as perceptual biases that affect domain-general computational principles (Frost et al. 2015). Studies with adults suggest substantial modality effects (e.g., Conway & Christiansen 2005, Emberson et al. 2011, Saffran 2002). For example, Saffran (2001b) developed an artificial grammar learning task in which the presence of a statistical cue to syntactic phrase structure (predictive dependencies between elements of phrases) was manipulated across conditions. Adults, children, and infants were better able to learn the grammar when predictive dependencies were present (Saffran 2001a, Saffran et al. 2008). The same pattern of results was obtained when adults were trained on auditory non-linguistic sequences (computer alert sounds) and on spatial arrays of visual images (Saffran 2002). However, when presented with sequences of visual images—organized like auditory information in time, but presented visually—the benefits afforded by the statistical regularity were not observed (Saffran 2002). These results are consistent with the modality effects on learning described above. Visual information, unlike auditory information, is typically less transient, with patterns organized in space rather than time. These differences appear to impact the outcome of statistical learning.

4. Real-World Problems

Researchers considering what statistical learning can do for learners have approached a number of important and interesting problems through this lens. These approaches have both suggested novel answers and raised new questions for researchers.

4.1. Multilingualism

Since the earliest discussions of the possible role of statistical learning in language development, questions about bilingual learners have come to the fore. If statistical regularities play a key role in such language learning processes as phonemic learning, word segmentation, and word learning, what happens in bilingual environments? Can learners track multiple sets of statistics simultaneously? If so, what cues do they use to help them determine which bits of input go with which language? Strikingly, infants in bilingual environments acquire language at roughly the same pace as their monolingual peers, despite having twice as much to learn—and half the amount of input (e.g., Byers-Heinlein & Fennell 2014, Costa & Sebastián-Gallés 2014, Hoff et al. 2012).

The first study to examine the problem of bilingual statistical learning placed adults in a simulated bilingual environment created by interleaving two artificial languages (Weiss et al. 2009). By design, the languages contained overlapping syllable inventories. In order to recover the correct underlying statistics from each language, learners needed to keep the two languages separate. As long as an indexical cue—speaker voice—was available to highlight the presence of two languages, learners successfully tracked the two systems independently.

In another study, Antovich & Graf Estes (2017) tested 14-month-old infants in a simulated bilingual exposure task in which two artificial language streams were interleaved. Again, an indexical cue was present to indicate to learners that multiple streams were present. In contrast to the results of Weiss et al. (2009), monolingual infants failed to demonstrate learning of dual interleaved speech streams. However, bilingual infants were able to track both sets of regularities. Also in contrast to the results of Weiss et al. (2009), the two languages did not overlap in their syllable inventories. It is thus unclear whether the bilingual infants treated the input as being drawn from two languages or whether they acquired one larger set of words. Regardless, these findings suggest that infants who have had more experience dealing with complex and highly variable sets of input—i.e., bilingual infants—are better able to cope with this rich set of experimental input than monolingual infants.

In these studies, indexical information—a change in speaker voice—helped to mark the presence of two distinct speech streams. In monolingual language input, however, infants must learn to collapse over speaker identity. That is, the pitch of a word does not change its meaning (at least in nontonal languages). This observation raises an interesting question for infant statistical language learning research: Do infants collapse statistics across speakers within a single language? A recent study by Graf Estes & Lew-Williams (2015) suggests that the answer hinges on variability. When infants were exposed to an artificial language spoken by eight different female voices, infants successfully tracked the sequential statistics in the input. However, when just two voices were present, infants failed to demonstrate

learning, presumably because they did not collapse the statistics across the two voices. Taken with the previously discussed studies about simulated bilingual acquisition, these results raise important questions about the role of variability in statistical learning. The distribution of exemplars in memory has been argued to be highly sensitive to variability, helping to explain patterns of results across myriad statistical learning tasks (Thiessen & Pavlik 2013).

4.2. Individual Differences

Another area where statistical learning approaches have been gaining traction is the study of individual differences (e.g., Siegelman & Frost 2015, Siegelman et al. 2017). This issue is of interest both in terms of individual differences in learning in themselves and insofar as individual differences in learning help to explain variability in key outcomes, such as native language learning.

Much of the research in this area has focused on adults, to facilitate correlations between statistical learning results and measures of cognitive or academic achievement. For example, English-speaking adults who performed better at a visual statistical learning task showed higher levels of performance in the acquisition of the Hebrew writing system, which is highly patterned (Frost et al. 2013). Experience with Mandarin in the college classroom improved adults' performance on an auditory statistical language learning task (Potter et al. 2016). Skill at auditory statistical learning, but not visual statistical learning, appears to be related to musical skill (Vasuki et al. 2016).

Few studies have addressed individual differences in statistical learning in infancy. This is due at least in part to methodological constraints. Tasks like the head-turn preference procedure—used in many infant auditory learning studies—are not amenable to individual difference studies. They provide a single score for each infant—a difference score for looking on novel versus familiar trials. There is no evidence that the size of that difference is meaningful—that is, that an infant with a larger novelty preference learned more than an infant with a smaller novelty preference. Issues of direction of preference also complicate attempts to use preferential looking procedures to study individual differences. Unless there is a habituation component to the task, it is often not possible to make strong a priori predictions about the expected direction of preference.

Visual statistical learning tasks hold promise for studies of individual differences in infancy because they permit the collection of continuous measures that are clearly interpretable. For example, Shafto et al. (2012) used a reaction time measure in a visual anticipation task to assess statistical learning in 8.5-month-old infants. The results were correlated with the infants' vocabularies, as assessed by parental report. Indeed, infants' processing speed in sequential learning tasks predicts vocabulary size months later (Ellis et al. 2014). Studies with child learners suggest a similar pattern of results: 6- to 8-year-olds' visual statistical learning skills predict their level of performance on measures of native language syntax comprehension (Kidd & Arciuli 2016). Research investigating the relationship between visual attention in infancy (from a visual pattern prediction task) and later childhood behavior and temperament showed that mean fixation duration infancy was positively

associated with effortful control and negatively associated with surgency, hyperactivity, and inattention in childhood (Papageorgiou et al. 2014).

4.3. Developmental Disabilities

A related approach to understanding individual differences involves comparisons between groups of infants or children who are known to be following different developmental trajectories. These studies ask whether relative strengths and weaknesses in statistical learning can help to explain the patterns of deficits observed in infants and children with developmental disabilities. The first study to take this approach concerned adolescents with specific language impairment (SLI)—weakness in native language skill relative to other academic and cognitive skills (Tomblin et al. 2007). Participants with grammatical language impairment performed worse than their typically developing peers on a serial reaction time task requiring detection of visual patterns. Similar findings emerged from a study of grade school–aged children with SLI tracking statistical patterns inaword segmentation task (Evans et al. 2009). Compared to a nonverbal IQ–matched comparison group, the children did poorly on the statistical language learning task. Interestingly, the children with SLI performed even worse on a version of the word segmentation task using tone sequences rather than syllables, suggesting, in line with the Tomblin et al. (2007) findings, that the children's learning challenges are not limited to linguistic materials.

A recent meta-analysis of the extant literature confirmed this general pattern: Children with SLI perform worse on statistical learning tasks than children who are typical language learners (Obeid et al. 2016). These findings are interesting given the potential links between statistical learning and native language acquisition. Similar conclusions were drawn by a study comparing children with developmental dyslexia (DD) and children with typical development (Gabay et al. 2015). Children with DD performed more poorly on linguistic and tone-sequence statistical learning tasks than children in the comparison group. Moreover, performance on both the linguistic and nonlinguistic statistical learning tasks was correlated with reading measures. These data are consistent with the view that challenges in procedural learning underlie at least some aspects of DD (e.g., Lum et al. 2013).

It is not the case, though, that all developmental language learning deficits can be attributed to challenges in statistical learning. The same meta-analysis examined the extant literature on statistical learning in children and adolescents with autism spectrum disorders (ASD) (Obeid et al. 2016). Strikingly, the data across numerous studies suggest that autistic individuals do not show difficulties in statistical learning tasks. For example, Mayo & Eigsti (2012) tested autistic children with the same materials previously used by Evans et al. (2009) in their study of children with SLI. The autistic children showed the same pattern of performance as children with typical development. Thus, the tracking of sequential statistics does not, in itself, provide a way to differentiate between language disorders that may have quite different etiologies. That said, it is worth noting that the vast majority of studies investigating language learning in autistic children have sampled relatively high-functioning children (Obeid et al. 2016). Autistic children who show more language deficits may exhibit different patterns of functioning.

Using event-related electrophysiological methodology, Jeste et al. (2015) presented children with ASD with an oddball paradigm version of the Kirkham et al. (2002) test in which the sequence was infrequently interrupted by a deviant or unexpected stimulus. Results showed a positive association between visual statistical learning and both nonverbal IQ and social function in children with ASD. Children with high nonverbal IQ scores demonstrated a larger (more negative) response to the unexpected trials (i.e., the 10% of the trials during which a shape was followed by an unmatching shape), as quantified by the N1 component difference (see Vogel & Luck 2000 for a discussion of the N1 component). This was opposite to the response of the typically developing group, who showed a greater response to the expected trials, suggesting greater allocation of attention to unexpected events. This preliminary work suggests that there may be statistical learning processing differences between children with ASD and typically developing children.

It would be ideal to be able to test infants at risk for developmental disorders on these sorts of tasks. Doing so would allow researchers to disentangle the starting state—abilities to detect sequential regularities, for example—from the effects of experience in detecting sequential regularities. Diagnoses like SLI, DD, and ASD currently cannot be made in early infancy. However, other types of developmental disorders, such as genetic syndromes arising from deletions or point mutations (in which part of a chromosome or DNA sequence is lost during replication), are diagnosed in infancy and provide fascinating opportunities to examine early learning abilities. Williams syndrome (WS) is a genetic disorder that is associated with significant intellectual disabilities, although with relative sparing of language abilities. Cashon et al. (2016) tested a group of infants with WS on the Saffran et al. (1996) artificial language segmentation task. The data provide the first evidence that infants with developmental disorders can track sequential statistics. Understanding how these abilities are used—and how they are combined with detection of other types of regularities—may help us to understand the developmental trajectories characterizing children with different disorders and individual differences more generally (e.g., Thomas et al. 2009). For example, some evidence suggests that infants with WS are more reliant on prosodic cues than their typically developing peers, at least after the first year of postnatal life (Nazzi et al. 2003). More complex approaches to studying early learning, asking how infants integrate multiple sources of information rather than focusing on how infants track single cues, will be necessary to develop a deeper understanding of the emergence of these complex phenotypes.

4.4. Noise, Distraction, and Context

In the real world, the infant learner is faced by another set of problems: noise and distraction. The literature reviewed above has shown the infant to be a robust learner, sensitive to the statistics in the input across a wide variety of situations. However, in the real world, those statistics can be in a fierce competition for attention among other, equally enticing cues. Tummeltshammer & Kirkham (2013) asked whether attention to highly probabilistic events would shift with the addition of noise. In a paradigm looking at three different spatiotemporal probabilities, infants showed heightened learning of a highly predictable event (as opposed to deterministic or low-probability events). However, when noise was added to the paradigm (in the guise of a light going on and off in a separate area

of the screen), despite absolute attention to the events being the same as in the previous condition, infants now showed better learning of the deterministic events. This suggests that sensitivity to input statistics is context dependent. A series of studies by Tummeltshammer and colleagues (2014a,b; 2017) have expanded upon the issue of context, showing that visual statistical learning is mediated by stimulus salience, source reliability, and mode of presentation.

5. Why are We Statistical Learners?

As we have described throughout this review, there is ample evidence that human infants, along with other learners, are sensitive to the statistical structure of their environment. We have also tried to highlight the ways in which statistical learning abilities serve infants well, given the structure of the environment into which they are born. At least some of these abilities are observed in other species (e.g., Abe & Watanabe 2011, Santolin et al. 2016, Toro & Trobalón 2005). Almost 50 years ago, Rescorla (1968) demonstrated that conditioning is actually dependent on more complicated factors than just contiguity of the pairing. Conditioning is affected by the base rate of unconditioned stimulus occurrence, against which a conditioned stimulus/unconditioned stimulus (CS/US) contiguity takes place. In other words, if a tone always occurs just before a shock is administered to a rat, but shocks also occur in absence of a tone, the CS/US pairing is not learned. If, later, the same number of tones is presented but the shocks only occur after the tone, then the animal learns, even though in both cases the rate of tone–shock pairings is identical. Thus, it is not the absolute frequency of the pairings that is important, but the general probabilistic relationship between the variables. Conditioning takes place only when the US has predictive value. Rescorla's (1968) work refuted the traditional belief that the CS/US pairing frequency was the crucial aspect of classical conditioning and showed that animals are sensitive, instead, to the statistics of each individual situation.

5.1. Relationships Between Statistical Learning and the Environment

To some extent, we can construe infants' learning abilities as well tailored to their environments. This way of thinking makes sense when considering the visual world, which predates human evolution. The modality-specific constraints described above are a good example of tailoring: Learners appear to be better at tracking sequential statistics in auditory stimuli, which tend to be more temporally fleeting than stimuli in the visual environment. The latter are more temporally stable, with structures organized more in terms of space than of time. The degree to which our learning abilities reflect these differences may be due to modality constraints on learning and memory that are present prior to experience in these domains. Alternatively, infants' divergent early experiences with auditory and visual inputs, possibly dating to prenatal exposure, may have shaped modality-specific constraints on learning and memory. Regardless of the locus of the constraints, they are a good fit to the world.

Another way to think about the relationship between statistical learning and the world is to consider the idea that our learning abilities themselves may have played a role in shaping our environments. Consider the structure of natural languages. An enduring puzzle in the study

of linguistics concerns cross-linguistic similarities. Although languages appear to be very different on their surface, much of the underlying structure of human languages is remarkably similar. Whereas some of these similarities likely reflect historical relationships, others seem unlikely to be explainable in those terms. Indeed, these considerations played a major role in the original positing of a language acquisition device containing innate knowledge about possible human languages.

Statistical learning accounts are not sophisticated enough to be able to account for many detailed aspects of language acquisition, especially some of the more complex linguistic structures that do not appear to be transparently mirrored in the input (e.g., Han et al. 2016). However, both experimental tasks and computational models have suggested specific ways in which constraints on statistical learning might have influenced the structure of natural languages (e.g., Christiansen & Chater 2008, Saffran 2001b, Smith et al. 2017). The general idea is that language structures that are more learnable, particularly by infants and young children, should be more prevalent in the languages of the world than structures that are more difficult to learn. Moreover, if the constraints on learning precede the structures that they have shaped, the same constraints on learning should be evident in learning nonlinguistic structures (e.g., Saffran 2002).

Although there is not a great deal of data to support this theoretical perspective, the extant studies are promising. For example, infants are better able to track phonotactic patterns—the statistics of phoneme co-occurrence conditioned by position within syllables and words—when the observed patterns mirror the types of regularities present in natural languages (Saffran & Thiessen 2003). Similarly, infants are better able to acquire linguistic phrase structure when it contains distributional regularities—within-phrase predictive elements—that mirror structures found in natural languages (Saffran et al. 2008). Similar constraints on learning appear given non-linguistic input designed to simulate language structures (C. Santolin & J.R. Saffran, unpublished manuscript; Thiessen 2011).

5.2. Memory and Prediction

Statistical learning mechanisms are well suited to our environments, and our environments, in turn, may have been shaped by our learning mechanisms—at least for structures that are culturally transmitted. In this section, we turn to an even more speculative question: Why do we track statistics in the first place?

One possibility is that the detection of statistical patterns is a result of the structure of memory (e.g., Perruchet & Vintner 1998, Thiessen & Pavlik 2013). For example, the iMinerva model proposed by Thiessen & Pavlik (2013) simulates a range of statistical learning phenomena using principles of long-term memory: activation, decay, integration, and abstraction. Thus, sensitivity to statistical regularities is due to the properties of memory and forgetting. Memory-based approaches to statistical learning permit the integration of multiple learning tasks that appear quite different on the surface but that may be explainable under the umbrella of memory considerations (e.g., Thiessen 2017). This leads nicely to another major question currently debated in visual statistical learning: Is this type of learning chunking or statistical computation (see Perruchet & Pacton 2006)? One alternative interpretation for infants' seeming sensitivity to statistical distributions across visual input is

that this may be the outcome of a broader associative learning strategy (i.e., chunking; Miller 1956). This implies that co-occurring elements in a scene are extracted and stored as a structured chunk (Perruchet & Peereman 2004), allowing infants to recall regularities encountered in the environment without relying on sophisticated computational abilities [see the models PARSER (word segmentation; Perruchet & Vintner 1998) and TRACX (sequence segmentation and chunk extraction; French et al. 2011, Mareschal & French 2017)]. However, as noted by Perruchet and colleagues (Perruchet & Pacton 2006, Perruchet & Peereman 2004), statistical learning and chunking explanations may not be mutually exclusive, and, indeed, chunking may arise from an initial sensitivity to statistical regularities.

Another way to think about why we track statistics entails a shift in focus from learning statistics to using statistics (e.g., Hasson 2017). Statistical information sharpens predictions. To the extent that our brains and, by extension, our cognitive and linguistic systems are engaged in reduction of uncertainty, statistical information should be informative. Note that this way of framing the issues—around prediction—is neutral concerning the specific types of statistical information that are relevant; any type of information, from Bayesian priors to transitional probabilities to the weights in connectionist networks, could, in principle, help to tune predictions. By tuning predictions, learners have the opportunity to reduce errors to better anticipate outcomes.

Predictive contexts also provide the opportunity for learners to generate internal error signals, which may supplement bottom-up statistical information and facilitate learning. For example, consider the following simple predictive learning study by Romberg & Saffran (2013b). In one condition, infants saw a brief video repeatedly on the left side of the screen, and learned to saccade predictively to the left, anticipating the reward. After several such trials, the reward occurred on the right side of the screen. The question of interest was what infants would do on the next trial. Would they respond by anticipating on the left based on the overall statistical information (the reward was much more likely to occur on the left), or would they weight the error signal more highly and look to the right? Interestingly, the infants' behavior was unaffected by a single unexpected trial; they continued to predict the reward on the left. However, after a second unexpected trial, the infants updated their predictions and became less biased to the left side. Across the experiment, the evidence suggested that infants updated their predictions based on the evidence they saw, but not based on a single counterexample.

6. Conclusion

We began this review by placing theories of statistical learning firmly in the middle of the big question of how the structured environment is detected by infants. We have presented evidence to suggest that infants' sensitivity to statistical structure is not only broad, applied across modalities and domains, but also focused, attending to the specifics of the input and the varying goals of perception. We have suggested that statistical learning is part of a reciprocal determinism between the brain mechanisms and the environment, in which each helps shape the other, perhaps crucially related to the structure of human memory itself.

Normal perception is concerned with real-life events: dynamic multimodal scenes that involve language, objects, and action. For statistical learning to be of use, it must be a mechanism that is flexible enough to encompass all of these dimensions. For example, solely attending to the relationships between individual features in a sensory scene would eventually create a computational bottleneck that would strangle the system. Thus, the primitives of statistical computation matter, changing the outcome of learning. They are affected by experience and modified by perceptual biases. And what are the actual computations operating over these primitives? In the field of infant statistical learning, transitional probabilities, frequencies, redundancies, dependencies, and conditional probabilities (temporal and spatial) have all played a part in this discussion. Although the research continues, it is clear from the work to date that these computations depend on both the age of the infant (perhaps shifting from frequency counting or attention to local redundancies to transitional probabilities across the first year of life) and the specifics of the input. At present, numerous different implementations can account for the empirical findings. It is up to the field to generate experimental results that can tease them apart.

For statistical learning to be useful, it has to tell us something about real-world problems. And it does: It offers insights into issues specific to multilinguals and tackles problems of real-world chaos (e.g., noise and distraction). Recently, studies using statistical learning paradigms have begun to shine light on individual differences in perception and certain developmental disabilities.

In sum, statistical learning is a rich and robust learning mechanism allowing infants to find structure (and meaning) in the blooming, buzzing confusion. We recognize that we have only touched the surface of the field in this review. But we hope that we have raised issues and questions that will help to motivate the next generation of research on statistical learning.

Acknowledgments

Literature Cited

Abe K, Watanabe D. Songbirds possess the spontaneous ability to discriminate syntactic rules. Nat Neurosci. 2011; 14(8):1067–74. [PubMed: 21706017]

Addyman C, Mareschal D. Local redundancy governs infants' spontaneous orienting to visual-temporal sequences. Child Dev. 2013; 84(4):1137–44. [PubMed: 23432603]

Antovich DM, Graf Estes K. Learning across languages: Bilingual experience supports dual language statistical word segmentation. Dev Sci. 2017 In press.

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. Psychol Sci. 1998; 9(4):321–24.

Aslin, RN., Slemmer, JA., Kirkham, NZ., Johnson, SP. Statistical learning of visual shape sequences; Presented at Biennial Meet. Soc. Res. Child Dev; Apr. 19–22; Minneapolis, MN. 2001.

Baldwin D, Andersson A, Saffran J, Meyer M. Segmenting dynamic human action via statistical structure. Cognition. 2008; 106(3):1382–407. [PubMed: 18035346]

Bulf H, Brenna V, Valenza E, Johnson SP, Turati C. Many faces, one rule: the role of perceptual expertise in infants' sequential rule learning. Front Psychol. 2015; 6:1595. [PubMed: 26539142]

Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. Cognition. 2011; 121(1):127–32. [PubMed: 21745660]

Byers-Heinlein K, Fennell CT. Perceptual narrowing in the context of increased variation: insights from bilingual infants. Dev Psychobiol. 2014; 56(2):274–91. [PubMed: 24114364]

Canfield RL, Haith MM. Young infants' visual expectations for symmetric and asymmetric stimulus sequences. Dev Psychol. 1991; 27:198–208.

Cashon CH, Ha OR, Estes KG, Saffran JR, Mervis CB. Infants with Williams syndrome detect statistical regularities in continuous speech. Cognition. 2016; 154:165–68. [PubMed: 27299804]

Chomsky, N. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press; 1965.

Christiansen MH, Chater N. Language as shaped by the brain. Behav Brain Sci. 2008; 31(05):489–509. [PubMed: 18826669]

Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. J Exp Psychol Learn Mem Cogn. 2005; 31(1):24–39. [PubMed: 15641902]

Costa A, Sebastián-Gallés N. How does the bilingual experience sculpt the brain? Nat Rev Neurosci. 2014; 15(5):336–45. [PubMed: 24739788]

Dawson C, Gerken L. From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. Cognition. 2009; 111(3):378–82. [PubMed: 19338982]

Dougherty M, Thomas R, Lange N. Toward an integrative theory of hypothesis generation, probability judgment, and hypothesis testing. Psychol Learn Motiv. 2010; 52:299–342.

Edelman, S., Intrator, N., Jacobson, JS. Proc Int Workshop Biol Motiv Comp Vis 2nd Tübingen Ger. Berlin: Springer; 2002. Unsupervised learning of visual structure; p. 629-42.

Ellis EM, Gonzalez MR, Deák GO. Visual prediction in infancy: What is the association with later vocabulary? Lang Learn Dev. 2014; 10(1):36–50.

Emberson LL, Conway CM, Christiansen MH. Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. Q J Exp Psychol. 2011; 64(5):1021–40.

Erickson L, Thiessen ET, Graf Estes K. Statistically coherent labels facilitate categorization in 8-month-olds. J Mem Lang. 2014; 72:49–58.

Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. J Speech Lang Hear Res. 2009; 52(2):321–35. [PubMed: 19339700]

Ferguson B, Lew-Williams C. Communicative signals support abstract rule learning by 7-month-old infants. Sci Rep. 2016; 6:25434. [PubMed: 27150270]

Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A. 1987; 4(12):2379–94. [PubMed: 3430225]

Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychol Sci. 2001; 12(6):499–504. [PubMed: 11760138]

Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. PNAS. 2002; 99(24):15822–26. [PubMed: 12429858]

Fiser J, Aslin RN. Encoding multielement scenes: statistical learning of visual feature hierarchies. J Exp Psychol Gene. 2005; 134(4):521–37.

Fiser J, Scholl BJ, Aslin RN. Perceived object trajectories during occlusion constrain visual statistical learning. Psychon Bull Rev. 2007; 14(1):173–78. [PubMed: 17546749]

French RM, Addyman C, Mareschal D. TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. Psychol Rev. 2011; 118(4):614–36. [PubMed: 22003842]

Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. Trends Cogn Sci. 2015; 19(3):117–25. [PubMed: 25631249]

Frost R, Siegelman N, Narkiss A, Afek L. What predicts successful literacy acquisition in a second language? Psychol Sci. 2013; 24(7):1243–52. [PubMed: 23698615]

Gabay Y, Thiessen ED, Holt LL. Impaired statistical learning in developmental dyslexia. J Speech Lang Hear Res. 2015; 58(3):934–45. [PubMed: 25860795]

Gauthier B, Shi R, Xu Y. Learning phonetic categories by tracking movements. Cognition. 2007; 103(1):80–106. [PubMed: 16650399]

Gerken L. Decisions, decisions: infant language learning when multiple generalizations are possible. Cognition. 2006; 98(3):B67–74. [PubMed: 15992791]

Gerken L. Infants use rational decision criteria for choosing among models of their input. Cognition. 2010; 115(2):362–66. [PubMed: 20144828]

Gibson, JJ. The Perception of the Visual World. Boston: Houghton Mifflin; 1966.

Gómez RL. Variability and detection of invariant structure. Psychol Sci. 2002; 13(5):431–36. [PubMed: 12219809]

Goodsitt JV, Morgan JL, Kuhl PK. Perceptual strategies in prelingual speech segmentation. J Child Lang. 1993; 20(2):229–52. [PubMed: 8376468]

Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning Psychol Sci. 2007; 18(3):254–60. [PubMed: 17444923]

Graf Estes K, Lew-Williams C. Listening through voices: infant statistical word segmentation across multiple speakers. Dev Psychol. 2015; 51(11):1517–28. [PubMed: 26389607]

Gweon H, Tenenbaum JB, Schulz LE. Infants consider both the sample and the sampling process in inductive generalization. PNAS. 2010; 107(20):9066–71. [PubMed: 20435914]

Haith, MM. Future-oriented processes in infancy: the case of visual expectations. In: Granrud, CE., editor. Visual Perception and Cognition in Infancy. Hillsdale, NJ: Erlbaum; 1993. p. 235-64.

Han CH, Musolino J, Lidz J. Endogenous sources of variation in language acquisition. PNAS. 2016; 113(4):942–47. [PubMed: 26755580]

Hasson U. The neurobiology of uncertainty: implications for statistical learning. Phil Trans R Soc B. 2017; 372(1711):20160048. [PubMed: 27872367]

Hay JF, Pelucchi B, Estes KG, Saffran JR. Linking sounds to meanings: infant statistical learning in a natural language. Cogn Psychol. 2011; 63(2):93–106. [PubMed: 21762650]

Hoff E, Core C, Place S, Rumiche R, Senor M, Parra M. Dual language exposure and early bilingual development. J Child Lang. 2012; 39:1–27. [PubMed: 21418730]

James, W. The Principles of Psychology. Cambridge, MA: Harvard Univ. Press; 1890.

Jeste SS, Kirkham N, Senturk D, Hasenstab K, Sugar C, et al. Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. Dev Sci. 2015; 18(1):90–105. [PubMed: 24824992]

Johnson EK, Tyler MD. Testing the limits of statistical learning for word segmentation. Dev Sci. 2010; 13(2):339–45. [PubMed: 20136930]

Jones J, Pashler H. Is the mind inherently forward looking? Comparing prediction and retrodiction. Psychon Bull Rev. 2007; 14(2):295–300. [PubMed: 17694916]

Jusczyk PW, Aslin RN. Infants' detection of the sound patterns of words in fluent speech. Cogn Psychol. 1995; 29(1):1–23. [PubMed: 7641524]

Kidd E, Arciuli J. Individual differences in statistical learning predict children's comprehension of syntax. Child Dev. 2016; 87(1):184–93. [PubMed: 26510168]

Kidd C, Piantadosi ST, Aslin RN. The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. PLOS ONE. 2012; 7(5):e36399. [PubMed: 22649492]

Kidd C, Piantadosi ST, Aslin RN. The Goldilocks effect in infant auditory attention. Child Dev. 2014; 85(5):1795–804. [PubMed: 24990627]

Kirkham NZ, Richardson DC, Wu R, Johnson SP. The importance of "what": Infants use featural information to index events. J Exp Child Psychol. 2012; 113(3):430–39. [PubMed: 22867888]

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition. 2002; 83(2):B35–42. [PubMed: 11869728]

Kirkham NZ, Slemmer JA, Richardson DC, Johnson SP. Location, location, location: development of spatiotemporal sequence learning in infancy. Child Dev. 2007; 78(5):1559–71. [PubMed: 17883448]

Krogh L, Vlach HA, Johnson SP. Statistical learning across development: flexible yet constrained. Front Psychol. 2013; 3:598. [PubMed: 23430452]

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science. 1992; 255:606–8. [PubMed: 1736364]

Lany J, Gómez RL. Twelve-month-old infants benefit from prior experience in statistical learning. Psychol Sci. 2008; 19(12):1247–52. [PubMed: 19121132]

Lew-Williams C, Saffran JR. All words are not created equal: Expectations about word length guide infant statistical learning. Cognition. 2012; 122(2):241–46. [PubMed: 22088408]

Lidz J, Gagliardi A. How nature meets nurture: universal grammar and statistical learning. Annu Rev Linguistics. 2015; 1:333–53.

Lum JA, Ullman MT, Conti-Ramsden G. Procedural learning is impaired in dyslexia: evidence from a meta-analysis of serial reaction time studies. Res Dev Disabil. 2013; 34:3460–76. [PubMed: 23920029]

Marcovitch S, Lewkowicz DJ. Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. Dev Sci. 2009; 12(6):1020–25. [PubMed: 19840056]

Marcus GF, Fernandes KJ, Johnson SP. Infant rule learning facilitated by speech. Psychol Sci. 2007; 18(5):387–91. [PubMed: 17576276]

Marcus GF, Vijayan S, Rao SB, Vishton PM. Rule learning by seven-month-old infants. Science. 1999; 283(5398):77–80. [PubMed: 9872745]

Mareschal D, French RM. TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. Phil Trans R Soc B. 2017; 372(1711):20160057. [PubMed: 27872375]

Marr, D. Vision. San Francisco: WH Freeman & Co; 1982.

Mayo J, Eigsti IM. Brief report: a comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. J Autism Dev Disord. 2012; 42(11):2476–85. [PubMed: 22382606]

Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956; 63(2):81–97. [PubMed: 13310704]

Monroy CD, Gerson SA, Hunnius S. Toddlers' action prediction: statistical learning of continuous action sequences. J Exp Child Psychol. 2017; 157:14–28. [PubMed: 28103496]

Nazzi T, Paterson S, Karmiloff-Smith A. Early word segmentation by infants and toddlers with Williams syndrome. Infancy. 2003; 4(2):251–71.

Newport EL, Aslin RN. Learning at a distance: 1. Statistical learning of non-adjacent dependencies. Cogn Psychol. 2004; 48:127–62. [PubMed: 14732409]

Obeid R, Brooks PJ, Powers KL, Gillespie-Lynch K, Lum JA. Statistical learning in specific language impairment and autism spectrum disorder: a meta-analysis. Front Psychol. 2016; 7:1245. [PubMed: 27602006]

Papageorgiou KA, Smith TJ, Wu R, Johnson MH, Kirkham NZ, Ronald A. Individual differences in infant fixation duration relate to attention and behavioral control in childhood. Psychol Sci. 2014; 25(7):1371–79. [PubMed: 24815614]

Pelucchi B, Hay JF, Saffran JR. Learning in reverse: Eight-month-old infants track backward transitional probabilities. Cognition. 2009a; 113(2):244–47. [PubMed: 19717144]

Pelucchi B, Hay JF, Saffran JR. Statistical learning in a natural language by 8-month-old infants. Child Dev. 2009b; 80(3):674–85. [PubMed: 19489896]

Perruchet P, Desaulty S. A role for backward transitional probabilities in word segmentation? Mem Cogn. 2008; 36:1299–305.

Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. Trends Cogn Sci. 2006; 10:233–38. [PubMed: 16616590]

Perruchet P, Peereman R. The exploitation of distributional information in syllable processing. J Neurolinguist. 2004; 17:97–119.

Perruchet P, Vintner A. PARSER: a model for word segmentation. J Mem Lang. 1998; 39:246–63.

Piantadosi ST, Kidd C. Endogenous or exogenous? The data don't say. PNAS. 2016; 113(20):E2764. [PubMed: 27099292]

Potter C, Wang T, Saffran JR. Second language experience facilitates statistical learning of novel linguistic materials. Cogn Sci. 2016; 41(S4):913–27. [PubMed: 27988939]

Rescorla RA. Probability of shock in the presence and absence of CS in fear conditioning. J Comp Physiol Psychol. 1968; 66(1):1–5. [PubMed: 5672628]

Romberg AR, Saffran JR. All together now: concurrent learning of multiple structures in an artificial language. Cogn Sci. 2013a; 37(7):1290–320. [PubMed: 23772795]

Romberg AR, Saffran JR. Expectancy learning from probabilistic input by infants. Front Psychol. 2013b; 3:610. [PubMed: 23439947]

Saffran JR. The use of predictive dependencies in language learning. J Mem Lang. 2001a; 44:493–515.

Saffran JR. Words in a sea of sounds: the output of statistical learning. Cognition. 2001b; 81:149–69. [PubMed: 11376640]

Saffran JR. Constraints on statistical language learning. J Mem Lang. 2002; 47(1):172–96.

Saffran JR. Absolute pitch in infancy and adulthood: the role of tonal structure. Dev Sci. 2003; 6(1): 35–43.

Saffran, JR. What can statistical learning tell us about infant learning?. In: Needham, A., Woodward, A., editors. Learning and the Infant Mind. Oxford, UK: Oxford Univ. Press; 2008. p. 29-46.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–28. [PubMed: 8943209]

Saffran JR, Griepentrog GJ. Absolute pitch in infant auditory learning: evidence for developmental reorganization. Dev Psychol. 2001; 37(1):74–85. [PubMed: 11206435]

Saffran JR, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. Cognition. 2008; 107(2):479–500. [PubMed: 18082676]

Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999; 70(1):27–52. [PubMed: 10193055]

Saffran JR, Pollak SD, Seibel RL, Shkolnik A. Dog is a dog is a dog: Infant rule learning is not specific to language. Cognition. 2007; 105(3):669–80. [PubMed: 17188676]

Saffran JR, Reeck K, Niebuhr A, Wilson D. Changing the tune: The structure of the input affects infants' use of absolute and relative pitch. Dev Sci. 2005; 8(1):1–7. [PubMed: 15647061]

Saffran JR, Thiessen ED. Pattern induction by infant language learners. Dev Psychol. 2003; 39(3): 484–94. [PubMed: 12760517]

Saffran JR, Wilson DP. From syllables to syntax: multilevel statistical learning by 12-month-old infants. Infancy. 2003; 4(2):273–84.

Sahni SD, Seidenberg MS, Saffran JR. Connecting cues: Overlapping regularities support cue discovery in infancy. Child Dev. 2010; 81(3):727–36. [PubMed: 20573101]

Santolin C, Rosa-Salva O, Vallortigara G, Regolin L. Unsupervised statistical learning in newly hatched chicks. Curr Biol. 2016; 26(23):R1218–20. [PubMed: 27923125]

Seidl A, Johnson EK. Infant word segmentation revisited: Edge alignment facilitates target extraction. Dev Sci. 2006; 9(6):565–73. [PubMed: 17059453]

Shafto CL, Conway CM, Field SL, Houston DM. Visual sequence learning in infancy: domain-general and domain-specific associations with language. Infancy. 2012; 17(3):247–71. [PubMed: 22523477]

Shukla M, Nespor M, Mehler J. An interaction between prosody and statistics in the segmentation of fluent speech. Cogn Psychol. 2007; 54(1):1–32. [PubMed: 16782083]

Shukla M, White KS, Aslin RN. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. PNAS. 2011; 108(15):6038–43. [PubMed: 21444800]

Siegelman N, Bogaerts L, Christiansen MH, Frost R. Towards a theory of individual differences in statistical learning. Phil Trans R Soc B. 2017; 372(1711):20160059. [PubMed: 27872377]

Siegelman N, Frost R. Statistical learning as an individual ability: theoretical perspectives and empirical evidence. J Mem Lang. 2015; 81:105–20. [PubMed: 25821343]

Smith K, Perfors A, Fehér O, Samara A, Swoboda K, Wonnacott E. Language learning, language use and the evolution of linguistic variation. Phil Trans R Soc B. 2017; 372(1711):20160051. [PubMed: 27872370]

Sobel DM, Kirkham NZ. Blickets and babies: the development of causal reasoning in toddlers and infants. Dev Psychol. 2006; 42:1103–15. [PubMed: 17087545]

Stahl AE, Romberg AR, Roseberry S, Golinkoff RM, Hirsh-Pasek K. Infants segment continuous events using transitional probabilities. Child Dev. 2014; 85(5):1821–26. [PubMed: 24749627]

Téglás E, Vul E, Girotto V, Gonzalez M, Tenenbaum JB, Bonatti LL. Pure reasoning in 12-month-old infants as probabilistic inference. Science. 2011; 332(6033):1054–59. [PubMed: 21617069]

Thiessen ED. Effects of visual information on adults' and infants' auditory statistical learning. Cogn Sci. 2010; 34(6):1093–106. [PubMed: 21564244]

Thiessen ED. Domain general constraints on statistical learning. Child Dev. 2011; 82(2):462–70. [PubMed: 21410909]

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

Thiessen ED, Pavlik PI. iMinerva: a mathematical model of distributional statistical learning. Cogn Sci. 2013; 37(2):310–43. [PubMed: 23126517]

Thiessen ED, Saffran JR. Learning to learn: infants' acquisition of stress-based strategies for word segmentation. Lang Learn Dev. 2007; 3(1):73–100.

Thomas MS, Annaz D, Ansari D, Scerif G, Jarrold C, Karmiloff-Smith A. Using developmental trajectories to understand developmental disorders. J Speech Lang Hear Res. 2009; 52(2):336–58. [PubMed: 19252129]

Tomblin JB, Mainela-Arnold E, Zhang X. Procedural learning in adolescents with and without specific language impairment. Lang Learn Dev. 2007; 3(4):269–93.

Toro JM, Trobalón JB. Statistical computations over a speech stream in a rodent. Atten Percept Psychophys. 2005; 67(5):867–75.

Tummeltshammer K, Amso D, French RM, Kirkham NZ. Across space and time: Infants learn from backward and forward visual statistics. Dev Sci. 2017 In press.

Tummeltshammer KS, Kirkham NZ. Learning to look: Probabilistic variation and noise guide infants' eye movements. Dev Sci. 2013; 16(5):760–71. [PubMed: 24033580]

Tummeltshammer KS, Mareschal D, Kirkham NZ. Infants' selective attention to reliable visual cues in the presence of salient distractors. Child Dev. 2014a; 85(5):1981–94. [PubMed: 24646174]

Tummeltshammer KS, Wu R, Sobel DM, Kirkham NZ. Infants track the reliability of potential informants. Psychol Sci. 2014b; 25(9):1730–38. [PubMed: 25022277]

Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA. Multidimensional visual statistical learning. J Exp Psychol Learn Mem Cogn. 2008; 34(2):399–407. [PubMed: 18315414]

Vasuki PRM, Sharma M, Demuth K, Arciuli J. Musicians' edge: a comparison of auditory processing, cognitive abilities and statistical learning. Hear Res. 2016; 342:112–23. [PubMed: 27770623]

Vogel EK, Luck SJ. The visual N1 component as an index of a discrimination process. Psychophysiology. 2000; 37(2):190–203. [PubMed: 10731769]

Wang T, Saffran JR. Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. Front Psychol. 2014; 5:953. [PubMed: 25232344]

Weiss DJ, Gerfen C, Mitchel AD. Speech segmentation in a simulated bilingual environment: a challenge for statistical learning? Lang Learn Dev. 2009; 5(1):30–49. [PubMed: 24729760]

Werker J, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav Dev. 1984; 7:49–63.

Wu R, Gopnik A, Richardson DC, Kirkham NZ. Infants learn about objects from statistics and people. Dev Psychol. 2011; 47(5):1220–29. [PubMed: 21668098]

Younger BA. The segregation of items into categories by ten-month-old infants. Child Dev. 1985; 54:858–67.

Younger BA, Cohen LB. Developmental change in infants' perception of correlations among attributes. Child Dev. 1986; 57:803–15. [PubMed: 3720405]

Yu, C., Smith, LB., Klein, KA., Shiffrin, RM. Proc Annu Meet Cogn Sci Soc 29th Nashville TN. Austin, TX: Cogn. Sci. Soc; 2007. Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same?; p. 737-42.