Statistical-Sequential Learning in Development

Jennifer B. Misyak, Michael H. Goldstein and *Morten H. Christiansen*

Department of Psychology, Cornell University

Please address correspondence to:

Jennifer B. Misyak

Department of Psychology

Uris Hall

Cornell University

Ithaca, NY 14853

e-mail: jbm36@cornell.edu

Keywords: statistical-sequential learning, artificial grammar learning, serial reaction-time, visual expectation paradigm, word-segmentation studies, fixed sequences, context-dependent sequences, adjacent dependencies, nonadjacent dependencies, developmental change

1. Introduction

The child's mind is embedded in a flowing, stimuli-rich world of perceived regularities. Children learn to engage their surroundings skillfully, in a manner reflecting knowledge of astoundingly complex structural patterns. From infants to adults, studying the unsupervised, ostensibly unconscious nature underlying many of these early-acquired processes has continually fueled research in the fields of "implicit learning" and "statistical learning" since their onset. The convention has been to trace the empirical genesis of implicit learning research from the early work of Reber (1967) in the 60's, and to follow modern statistical learning developments since the seminal work of Saffran and colleagues in the 90's (Saffran, Aslin, and Newport 1996).

This chapter, though, is not about tracing the vibrant histories of the implicit and statistical learning fields, but rather about interrelating the accrued developmental findings from both literatures—while in the service of promoting a synergistic fusion between their two approaches. For despite their disparately-pursued lines of work to date, establishing this common discourse may be easier to accomplish than otherwise presupposed. Witness, for instance, the close overlap among operational definitions provided by their paradigm progenitors.[1] Implicit learning has thus been related as "the process by which knowledge about complex stimulus domains is acquired largely independent of conscious awareness of either the process or the products of acquisition" (Reber and Allen 2000: 227). Reber's initial account (1967) also included characteristics such as the "efficient responding" of the organism and the development of "a strong sensitivity to the lawfulness [. . .] [existing in a] stimulus array;" in other words, participants may "become sensitive to the statistical nature of their environment without using explicit or verbalizable strategies" when the stimuli they receive is "patterned" or "ordered." He emphasized how this was "closely akin to Gibson and Gibson's (1955) perceptual learning and is

[. . .] a rudimentary inductive process which is intrinsic in such phenomena as language learning and pattern perception" (863).

Regarding statistical learning in its modern form, Saffran and collaborators introduced this as a "powerful mechanism for the computation of statistical properties of the language input" (Saffran, Aslin, et al. 1996: 1926), emphasizing the rapidity and adeptness by which infant learners incidentally extracted relevant regularities. This was otherwise termed "the process of statistical learning, or the detection of patterns of sounds, words, and classes of words in the service of discovering underlying structure" (Saffran 2002: 172) and could be relevant for learning complex (i.e., hierarchical) linguistic forms (Saffran 2001). As subsequent findings suggested that such learning was widely applicable to a variety of nonlinguistic domains, this empirical definition became more general, e.g., "the ability to discover units via their statistical coherence" (Saffran 2003: 111), and construed more broadly by researchers, e.g., as pertaining to "the discovery of statistical structure in the environment" (Gómez 2006: 90).

Close terminological correspondences such as these invite parallels—or demand a more rigorous partitioning and defining of phenomena (see Perrig 2001 for analogous claims in conjunction with implicit memory and procedural knowledge). It is our contention, though, that the resemblances in this case point to a common learning mechanism(s)[2]—albeit historically seen from different theoretical orientations and empirical traditions—and hereafter referred to as statistical-sequential learning. That is, much of the phenomena revealed through statistical learning and implicit learning approaches concerns the learning of sequential material and largely taps into the same probabilistically-sensitive, associative-based mechanism(s) recognized as belonging to "statistical learning" proper.[3]

One potential discrepancy, however, between the descriptions above relate to Reber's notion of the process as proceeding via unconscious *rule abstraction*, in which the participant tacitly apprehends “a valid, if partial, representation of the actual underlying rules of the [finite-state] language” (Reber and Allen 1978: 191). Conversely, statistical learning is construed as a process driven by statistical properties of the input, which results in participants’ probabilistic knowledge of the constraints governing stimuli formation. However, the sensitivity of participants to the “lawfulness” per se in sequentially arrayed material of implicit learning experiments has been sharply contested and has given rise to statistically-based explanations (without symbolic or abstract rules) of the computational principles entailed by successful performance (e.g., Cleeremans 1993). Although still an active matter of debate, if one is convinced by converging evidence that such learning is indeed driven by sensitivity to the *statistical* properties of the stimuli, then this natural affinity between the two fields should be readily apparent. They both entail incidental learning of sequential patterns (in spatiotemporal, temporal, or visually-arrayed distributions) that are defined by statistical relations over units perceived by the learner and that are processed in a manner respecting intrinsic regularities or probabilistic constraints of the input.

Some researchers have realized this connection. For instance, following in similar vein to the perspective informing the collection by Ellis (1994), Saffran, Newport, Aslin, Tunick, and Barrueco (1997) recognized that the literatures on "incidental learning" (which includes implicit learning and frequency estimation research) and natural language acquisition (which includes statistical learning, by this view) “would each be well served by a consideration of the theoretical and empirical concerns of the other" since these mutually suggest "pertinent" mechanisms for respective researchers (104). Perruchet and Pacton (2006: 237) delineated the growing convergence between results in the implicit learning and statistical learning fields, concluding that they appear to be "one phenomenon" that explores "the same domain-general incidental learning processes." They also note the increasing cross-referential synonymy of "implicit" and "statistical" learning terms, mentioning the example of Conway and Christiansen's (2006) coinage of "implicit statistical learning" as emblematic of their potentially future confluence.

However, beyond such claims and cross-references, there has been little (if any) attempt towards truly integrating and synthesizing findings across these wide literatures. Stronger efforts for communication between the literatures should be encouraged, as it would simultaneously widen and deepen the knowledge base for researchers in both fields. This chapter is a modest step in that direction; namely, it aims to underscore and support the theorized affinity of the statistical and implicit learning fields by providing a synthesis among findings to date. Its scope is confined to a human developmental context for two reasons: to fill in pre-adulthood timeline gaps from the canonical statistical learning literature alone, as well as to complement studies from the implicit learning literature that yield some equivocal findings during infancy; and to direct attention to the largely unasked but important question of developmental change.

In addressing developmental change, this chapter may be admittedly considered unorthodox. Developmental invariance is one of the central tenets, or corollaries, falling out of the theory on unconscious cognition posited by Reber, in which implicit learning is viewed as recruiting upon phylogenetically conserved and evolutionarily stable processes of high, basic adaptive value since antiquity ("the primacy of the implicit"; Reber, 1993). By Reber's framework, implicit learning has been expected to exhibit age independence, neurobiological robustness, little intraindividual variation, and remarkable cross-species commonality. These assumptions have in turn deterred many implicit learning researchers from directly seeking developmental trends. And the assumption of age-independence seems to even have seen its way borrowed into the canonical statistical learning literature; see especially Saffran et al. (1997, but note the conflicting evidence for developmental differences later found in Saffran 2001).

Such largely unconscious processes may indeed have basic and evolutionarily old roots, as well as recruit upon mechanisms shared across species. Fittingly, researchers from both implicit and statistical learning fields have specifically invoked parallels to principles from the classical Rescorla-Wagner model (1972) of animal learning along several lines, e.g., regarding the detection of predictive co-variation of stimulus events (Reber and Allen 2000), the similar subjection to perceptual constraints (Creel, Newport, and Aslin 2004), the use of prediction-based estimation from conditional statistics, or contingent probabilities (Aslin, Saffran, and Newport 1998; Hunt and Aslin 2001; Swingley 2005), and in relation to attention-based accounts of statistical learning (Pacton and Perruchet 2008). Despite some cross-species commonalities[4], though, earlier claims of implicit learning's (and by inference, statistical learning's) neurobiological robustness across impaired populations, and reputed lack of substantive differences across individuals, are being eroded by converging, recent evidence suggesting that systematic variations do in fact exist. Even Reber and Allen (2000) more recently have conceded the existence of some individual differences, referring back for support to findings from Reber, Walkenfeld, and Hernstadt (1991); they conversely argue now for which theoretical framework should be used to interpret these differences. Amid such shifting ground, this chapter's ancillary aim is to reappraise the remaining, fundamental postulate of developmental invariance, with ensuing implications for our understanding of the nature of statistical-sequential learning and the factors by which it may be influenced.

The remainder of this chapter is organized into five main divisions. The next section considers various paradigm implementations used in the developmental statistical and implicit learning literatures. Subsequently in Sec. 3, attention is directed towards areas of overlap between implicit and statistical learning research, the convergence of which delimits the statistical-sequential learning phenomena discussed throughout this chapter. In Sec. 4, we highlight various aspects of infants' and children's statistical-sequential learning as they relate to the processing of sequential relations and the tracking of probabilistic dependencies. With this punctuated empirical review in place, major developmental trends are then identified and further elaborated upon with regard to potential underlying factors (Sec. 5). The conclusion then ties together prospects and future directions for one way of bridging implicit and statistical learning literatures within a developmental context.

2. Common ground amid paradigmatic diversity

Before commencing our empirical overview/synthesis, a few words on methodology are in order. Understanding the basics and logic of four prominent paradigms will stand the reader in good stead through the remainder of the larger discussion that follows. Our exposition of these paradigms will proceed in chronological order of their introduction.

In the first paradigm—artificial grammar learning (AGL; Reber 1967) —participants are typically instructed to memorize or observe exemplars presented during a training phase. Often, these exemplars are visual letter-strings (e.g., PTTVPS) generated from an artificial finite-state grammar, but they can in principle be composed of any distinctive set of stimulus tokens varying along a perceptual dimension (e.g., auditory nonwords, musical tones, shapes) that are arranged in sequence according to the grammar (see Figure 1). Importantly, participants are not apprised of the existence of any underlying regularities until the test phase, when they are informed that stimulus strings follow a set of rules specifying the particular orderings among constituents. Without being told however the precise nature of these rules, participants are then asked to classify additional strings as either "grammatical" or "ungrammatical," relying upon intuitions or impressions of familiarity to guide their judgments. Participants typically achieve classification levels of above-chance performance on the task, even when test items comprise grammatical exemplars that were never directly encountered in training (i.e., requiring generalizations of the grammar to new strings) and despite being unable to provide verbal reports of actual patterns or rules. (Participants usually claim that they were merely "guessing.") These performances have been construed as evidence for participants' incidental encoding of the regularities of the grammar during training and their manifestation of this knowledge through meeting the task demands during test; and as alluded to earlier, are well-suited to statistical-based accounts (though not always without dispute) of the computational processes that mediate successful performance. Thus, participants may evince knowledge for complex, statistical relationships, even in the absence of reported awareness for any underlying structure and without direct intentions to discover such regularities.

Although artificial grammar learning remains a fruitful paradigm within both implicit learning and statistical learning work, few studies have been conducted with children, especially as standard grammaticality judgments require metacognitive skills not present very early in development. There are a few informative exceptions, though, for work with older children ages nine to eleven (Don, Schellenberg, Reber, DiGirolamo, and Wang 2003; Gebauer and Mackintosh 2007; van den Bos 2007), and a couple studies reported in the standard statistical learning literature, with children of six to seven years (Saffran 2001, 2002). Furthermore, similar methodological principles can be seen as informing the design and interpretation of other additional experiments in very young children. Accordingly, classic statistical learning studies with children and infants have involved familiarizing participants to carefully manipulated, frequency-balanced subsets of stimuli strings, sequences, or streams from a grammatical ‘corpus’ or miniature language, and then probing for sensitivity to the statistical relations by measuring more naturalistic behavioral responses to statistically consistent and inconsistent test items (more on this below).

A paradigm that has been successfully extended to both adults and children alike is the serial reaction-time paradigm (SRT; Nissen and Bullemer 1987; informative studies with children participants include Bremner, Mareschal, Destrebecqz, and Cleeremans 2007; Meulemans, Van der Linden, and Perruchet 1998; Thomas and Nelson 2001; and Thomas et al. 2004). In a prototypical task instantiation, participants are asked to respond as quickly and accurately as possible to trials of presented “targets” (e.g., illuminated lights) occurring at discrete locations on a computer screen, with each location mapping onto a particular response key or button. Unbeknownst to participants, target appearances follow a repeating or probabilistic sequence of locations. After many trials, participants become increasingly adept in anticipating and responding swiftly to the targets. When there is a disruption however to the predictive sequence, either through “noisy,” interspersed sequence-breaks or a continuous block of trials consisting of randomly-generated target locations, accuracy performance decreases and response latencies increase; when target locations conform again to the training sequence, participants’ reaction time (RT) performances dramatically “recover” (e.g., Schvaneveldt and Gómez 1998; Thomas and Nelson 2001). Because of the indirectness of the instructions and the task demands for speeded responses that discourage explicit reflection/strategizing, SRT work has yielded convincing demonstrations for participants' sensitivity to violations of sequential structure and incidental learning for sequence-embedded patterns.

For the youngest of subjects, however, other methods are necessary to assess incidental sequence knowledge. In the implicit learning literature, the visual expectation paradigm (VExP; Haith, Hazan, and Goodman 1988; Haith, Wentworth, and Canfield 1993) has been used with infants as young as two months to investigate their formation of expectations for upcoming visual events comprising predictable sequential patterns. A video monitor displays pictures for brief durations, separated by interstimulus intervals (without intervening visual input), and projected in distinct locations relative to the center of the infant's visual field (i.e., left versus right, up-down, left-center-right or within a triadic-pivot arrangement); the location and/or timing of visual events furthermore accord with either a predictive or randomly ordered series. There are generally three dependent variables of main interest: reaction times (RTs) for eye saccades to correctly anticipated upcoming stimulus locations in the predictive sequences (compared against RTs for non-predictive sequences); the frequency of accurate anticipatory (i.e., non-reactive, as opposed to elicited) saccades to target locations comprising the predictive series; and any facilitation effect on RTs (i.e., shorter latency to shift fixation to a predictive location upon the onset of a visual event). Importantly, infants do shift their visual fixations to the location where a future picture will appear prior to the timing of that picture's actual onset. Thus, it has been possible for researchers to obtain a behavioral index for infants' expectations of visual event sequences through measuring anticipatory RTs (assessed against an appropriate RT baseline for when events unfold in a relatively unpredictable manner). Such work has indicated that infants at a very early age rapidly form expectations based on detecting basic spatiotemporal regularities governing the predictive sequences.

Finally, the early infant statistical learning studies (e.g., Aslin et al. 1998; Gómez and Gerken 1999; Saffran, Aslin et al. 1996) have employed variants or adaptations of existing infant habituation-dishabituation and preference methods to investigate statistical learning. They have used syllables, nonwords, tones, or shapes for the elements instantiating the statistical relations of their artificial training grammars or sequences. Thus, for example, the word-segmentation study of Aslin et al. used a familiarization method to expose eight-month-old infants to a three-minute continuous speech stream (e.g., *pabikugolatudaropitibudo*...) consisting of four trisyllabic nonce words concatenated together in random order without immediate word repetitions and with each word occurring with a controlled frequency across the stream. While there were no acoustic cues (pauses, prosodic contours, etc.) marking artificial word boundaries, the edges of the nonce words could be successfully segmented or "extracted out" from the continuous sound sequence by utilizing statistical information governing sequence-element transitions: namely, that word-internal, successive syllable transitions ($P = 1.00$) contain higher conditional probabilities than pairwise syllables straddling word boundaries (for this study, $P = .50$). To assess whether infants were in fact sensitive to such cues for segmenting the stream, testing involved twelve trial presentations of two types of test sequences: repetitions of either single words (e.g., *pabiku*), or repetitions of part-words (e.g., *tudaro*). Infants demonstrated that they could discriminate between the two types on the basis of the relevant distributional statistics by orienting reliably longer to the direction of the loudspeaker on trials in which it emanated the words of the artificial language. Capitalizing on the attentional and natural orienting responses of infants, differential looking performance at test thus provided a measure of successful discrimination, and hence statistical sensitivity, in preverbal infants. This work exemplifies a prototypical design structure of many infant statistical learning studies to date, and these in turn have contributed valuably towards our understanding of statistical learning mechanisms.

3. Phenomenological boundaries

Across various experimental designs employed in the implicit and statistical learning literatures, and from infancy to early adolescence and beyond, individuals display an exquisite sensitivity to statistical aspects of their environments. The range of statistical information permitted on such accounts is theoretically broad in principle, as the underlying perspective endorsed here is that the statistical computations realized by the learner are carried out by mechanisms capable of extracting and integrating statistical information from multiple sources to home in on the most reliable regularities of the input, given as well the learner's constraints and perceptual biases. This does not deny, however, the advantages of investigating different types of statistical cues and identifying the contexts in which they may differentially aid the discovery of structure.

Accordingly, within any sequentially distributed input, there are a priori potentially many statistical cues available to the learner: (simple) frequency, co-occurrences, transitional probabilities (and "conditional probabilities," more generally, which can describe nonadjacent relationships), and higher-order conditionals (e.g., second-order, third-order, ... n^{th}-order probabilities). Researchers have also claimed psychological plausibility for other metrics, such as predictive probability (v. Rescorla 1967; "DeltaP," Shanks 1995) and normative [bi-directional] contingency (see Perruchet and Peereman 2004). Successful empirical demonstrations of statistical-sequential learning thus do not *necessarily* imply that learning must be defined or accounted for in terms of *a* particular or only a couple prominent statistical metrics already demonstrated in prior work (e.g., forward and backwards transitional probabilities; Jones and Pashler 2007; Pelucchi, Hay, and Saffran 2009a, 2009b; Perruchet and Desaulty 2008).

Despite the clear relevance of frequency information (whether simple or normalized) for statistical learning mechanisms, the research reviewed in this chapter will not focus on raw frequency as a type of statistical cue. Frequency effects are ubiquitous across many cognitive-developmental domains, and pertain to a diverse plethora of phenomena extending outside traditional implicit learning and statistical learning terrain; so studies in frequency estimation research per se (cf. Hasher and Zacks 1984) may be less likely to strongly constrain theorizing specific to the operations of statistical-sequential learning mechanisms. Instead, interested readers are directed to the helpful review compiled by Zacks and Hasher (2002) that places twenty-five years of frequency processing work within the context of a range of human behaviors—including, notably, language acquisition, statistical learning, constraint-based sentence processing, and Bayesian reasoning.

For different reasons, other interesting work in the implicit learning literature has been omitted here as well. Despite the rapprochement between various implicit *memory* and statistical learning phenomena, certain work within the former (i.e., pertaining to perceptual/conceptual priming, eye-blink conditioning, complex visual search, motor pursuit tracking) are excluded from our discussion, as they do not principally entail processing of sequentially distributed forms of information. Also not discussed here are phenomena with unclear or tenuous connections to statistical-sequential learning mechanisms, as they do not straightforwardly entail the processing of both statistical and sequential information; i.e., contextual cueing (e.g., in atypically and typically developing populations: Roodenrys and Dunn 2008; Vaidya, Huger, Howard, and Howard Jr 2007), covariation detection (in older children: Fletcher, Maybery, and Bennett 2000;

Maybery, Taylor, and O'Brien-Malone 1995), “dynamic systems control” tasks (as part of a cross-methodological investigation: Gebauer and Mackintosh, 2007), invariance learning (cf. Lewicki, Czyzewska, and Hoffman 1987), and “probabilistic classification” tasks (cf. Knowlton, Squire, and Gluck 1994).

Finally, a few caveats are in order for whether the learning of repeating sequences (in which asymmetric series alterations, as in VExP work, is a simple example) resides within the purview of statistical-sequential learning mechanisms. It may be objected, for instance, that a different kind of mechanism could potentially mediate the learning of sequentially arrayed input in fixed and repeating patterns—such as one mechanism subserving simple rote memory effects, and another mechanism ranging over more probabilistically-patterned, continuous input. Individuals certainly can and do engage in explicit memorization and recall of digits, letters, etc., with attention paid to the consecutive ordering of units—and such intentional processes are not the focus of this chapter. But the incidental, largely implicit, and (by many accounts) automatic nature of the learning processes that we review would seem to militate against this explanation, and pose difficulties for the differential application of such dual mechanisms prior to discovering the nature of the sequential input along such lines. And the parallel operation of both types of hypothetical mechanisms is less parsimonious, especially when a powerful (yet simple) mechanism can adeptly handle each. A dual-view hypothesis may be further difficult to reconcile with a larger theoretical perspective in which sequence memory is inextricably tied to sequential processing capabilities (MacDonald and Christiansen 2002), as suggested for instance by neural network models widely used in statistical and implicit learning research (e.g., Cleeremans and McClelland 1991, Dienes 1992; Keele and Jennings 1992; Mirman, Graf Estes, and Magnuson

2010; Misyak, Christiansen, and Tomblin 2010; Servan-Schreiber, Cleeremans, and McClelland 1988).

Further, the most widely cited documentation of statistical learning in the developmental literature comes from word segmentation studies (i.e., as in the original statistical learning studies described in Sec. 2), in which transitional probabilities are theorized to be computed for pairwise elements within a single sequence composed of fixed sequence-fragments. That is, the artificial trisyllabic words are fixed, contiguous orderings of certain phonemes, the latter of which are concatenated together to form a continuous input stream with probabilistic orderings among such fixed orderings. These studies naturally do not presuppose the recognition of such nonce words as units over which more probabilistic computations are performed. Locating the boundaries among such constituents is in essence the goal of the segmentation task. The continuous nature of the speech stream and the novelty of the artificial words necessitate sensitivity to differences in conditional probabilities between each of the pairwise syllables (irrespective of their absolute or graded contingency values) in order to identify the relevant word boundaries in the service of processing the speech input on-line.

It may still be, though, that some forms of fixed sequential learning, such as learning for lists of pre-individuated serial items, invoke different encoding and representation strategies—with learning for such fixed (typically "singular" or non-repeating, as well as truncated in length) sequences relying more on sensitivity to ordinal than associative information among series-internal elements. Evidence has been reviewed suggesting that humans and non-human primates may track the ordinality of relations among serial elements when presented in such a manner (see Conway and Christiansen 2001). Yet sequential statistical learning in an artificial grammar task also shows modality-specific effects paralleling known auditory-visual effects of recency and primacy in serial recall (cf. Conway and Christiansen 2009; see also discussion in Conway and Pisoni 2008). So even if learning of fixed sequences in these circumscribed contexts recruit different encoding and representation strategies from learning non-fixed sequences, they may still point to similar constraints or principles operative in both. For example, linguistic combinatorial structure is both probabilistic and deterministic. Fixed sequences composing words may be initially identified on the basis of distributional information and later comprise the units (or "chunks") for the fixed sequences constituting idioms and stock phrases, as well as the non-fixed sequences characterizing novel sentences (see McCauley and Christiansen, 2011, for a possible model). And perhaps this is so, *mutatis mutandis*, for admixtures of fixed, deterministic, and probabilistic micro- and macro-structures in other developmental domains (e.g., visual scene processing and object-parts/object-based recognition where dimensional features may be perfectly or variably correlated).

Given these arguments, this chapter does not delve into incidental learning of ordinal relations (for a recent example, see Lewkowicz 2008). As our understanding of underlying mechanisms deepens, it may be prudent to reexamine such manifestations of learning more closely with respect to the claim of shared mechanisms. However, findings from humans' learning of continuous, fixed, repeating sequences will be included in our discussions here. As a small confirmation that this may be currently the right approach, an emerging appreciation for VExP results in the statistical learning field may already be underway. For instance, Saffran and Thiessen (2007: 74) recently noted that transitional probabilities may be "only one particular example of statistical learning" if one more broadly considers evidence for the learning of regularities in one's natural environment; in this regard, they acknowledge the important findings of Canfield and Haith (1991) concerning the learning of predictive event sequences in preverbal infants.

It should also be noted that simple, repeating sequences comprise only part of the relevant literatures we review, and that sequences that are probabilistic in nature have also been studied. Thus, we turn in the next section to an overview of statistical-sequential learning in development, beginning with VExP findings and unambiguous repeating sequences, and bridging over to work on context-dependent sequences, probabilistic dependencies, and other statistical structures.

4. A sketch of the learning landscape

At their core, implicit and statistical learning literatures speak to fundamental processes underlying a diverse panoply of incidentally acquired, complex skills. Accordingly, the subsections below emphasize the broad nature of statistical-sequential learning mechanism(s) across individuals. Findings from both literatures are thereby briefly highlighted with respect to general characterizations that apply widely across human cognitive-developmental domains.

4.1. Learning fixed, continuous sequences

4.1.1. Asymmetric or simple, repeating sequences. As early as two months, infants in VExP studies show evidence for forming expectations of upcoming stimulus locations from symmetrically alternating series (i.e., in Left-Right, or Right-Left patterns) (Wentworth and Haith 1992). By three months, infants also exhibit faster and more frequent anticipatory saccades to asymmetric 2/1 repeating series (e.g., *L-L-R*) and predictive 3/1 (*L-L-L-R*, or *R-R-R-L*) patterns (Canfield and Haith 1991). Older infants (by about eight or twelve months) tested in the VExP paradigm also display some anticipatory gaze behavior for upcoming visual targets whose locations form a predictable triadic-pivot series (i.e., *ABCBABCB…* with the *"B"* location as the series' pivot point among the three locations) (Reznick, Chawarska, and Betts 2000).

It has been noted that use of the term "expectation" need not imply explicit recognition of patterns (Reznick et al. 2000), and the VExP paradigm itself should probably be best considered along the lines of a procedural task (cf. Nelson 1995), in which skilled performance commonly reflects incidental learning and the coordination of complex sequential input with motor responses. Regarding sequence-specific knowledge, while between-subjects VExP analyses provide evidence for global probability matching (e.g., greater eye shifts back to the more frequent, or "home-side," location in 3/1 than 2/1 asymmetric conditions), within-subjects analyses also indicate sensitivity to spatiotemporal regularities inhering over and above simply the proportion of picture appearances to a given side. For example, in the 3/1 condition, infants are more likely to appropriately shift to the less frequent "target" side after the third "home-side" event appearance than after the second or first "home-side" event. Given the specific experimental design of a study, encoding may also extend beyond location to accommodate specific (visual) event content, inter-event contingencies, and temporal flow rate of the stimuli sequences (Adler, Haith, Arehart, and Lanthier 2008; Wentworth and Haith 1992; Wentworth, Haith, and Hood 2002).

VExP measures exhibit moderate internal consistency and reflect stable individual differences over the short-term in early infancy (Haith and McCarty 1990; Rose, Feldman, Jankowski, and Caro 2002). But evidence for age-related differences within the first year is partly equivocal, with both longitudinal and cross-sectional designs reporting improvements up to nine months, but no improvements (and *fewer* anticipations) between nine and twelve months (Canfield et al. 1997; Reznick et al. 2000). Rose and colleagues, using a longitudinal design, did find support for increasing anticipations from seven to twelve months with a traditional cut-off of 200 ms distinguishing anticipatory/reactive saccades, but not when employing a more conservative criterion of 150 ms. In these cases, the difficulty in establishing an appropriate cutoff amid substantial individual variability in response latencies was further compounded by increased processing speed associated with higher ages across the first year of infancy. While Reznick and colleagues have posited an underlying change in the nature of the expectations formed at twelve months, corresponding to maturation of medial temporal lobes, Canfield et al. have suggested that something as simple as motivational requirements may have been at issue—that is, what may have been an interesting visual stimuli set for younger infants, may be considerably less engaging for the oldest infants in the group and thus resulted in their underperformance. More systematic studies are needed, though, to confirm this conclusion.

In summary, *within* the two minutes or shorter period of exposure to a repeating, symmetric or asymmetric series, infants throughout the first year demonstrate remarkably rapid on-line facilitation and anticipation for basic regularities intrinsic to the independently unfolding, spatiotemporal sequences in their visual environment.

4.1.2. Context-dependent sequences. In context-dependent (i.e., nth-order) progressions, the occurrence of a sequence-element depends upon the context associated with its preceding element. For example, given a *1-2-1-3-1-2*... sequence, being able to anticipate the location after a “1” requires knowing the temporal context of whether a “2” or “3” preceded the “1.” Sequences with context-dependent transitions can be either deterministic (repeating) or probabilistic.

While four-month-olds have shown above-chance accuracy in anticipatory saccades for unambiguous, simple repeating sequences (e.g., *1-2-3*) mapping onto a triangular configuration of spatial locations, they seem unable to perform above chance for the context-dependent transitions of more ambiguous sequences (e.g., *1-2-1-3*) given roughly comparable exposure time (i.e., 27 and 32 single-location trials, respectively, which correspond to 9 and 8 sequence repetitions each) (Clohessy, Posner, and Rothbart 2001). Clohessy et al. further assessed performance for these same sequences in ten- and eighteen-month-olds. Ten-month-olds did not show anticipatory learning for the context-dependent sequence (even when exposure was doubled, i.e., expanded to 2 sessions), but eighteen-month-olds could show anticipations for both types of sequences.

Bremner et al. (2007) found successful learning for two-year olds performing on a six-element deterministic spatial sequence (e.g., *A-C-B-D-A-B*) and a subsequent generation task using an adapted *SRT* paradigm. This is quite notable because, previously, sequence learning in SRT paradigms had not been conducted with children younger than four years, as in Thomas and Nelson's (2001) study. However, the performance evidenced by the two-year olds may also be considered an important extension of the sequential learning skills evidenced at eighteen months in Clohessy et al.'s VExP study for anticipating a deterministic sequence with fewer elements.

Across two studies in children of four, six- to seven, and ten- to eleven-years, performance on a deterministic ten-element SRT task (e.g., *1-3-2-4-1-2-3-4-2-4*) reported similar learning magnitudes across age groups (Meulemans et al. 1998; Thomas and Nelson 2001). These are consistent with standard expectations in the literature that, although general processing speed improves with age, a sequence learning effect nonetheless remains comparable across age groups. However, Thomas and Nelson reported that the "number of anticipatory button presses to correct locations show[ed] evidence of developmental change" (2001: 375). Nonetheless, they refrained from forming conclusions about developmental changes in implicit learning as such, under concerns that this measure might be construed as tapping more “explicit” learning.

However, a functional magnetic resonance imaging (fMRI) study conducted later by Thomas et al. (2004), again with seven- and eleven- year-olds, reported evidence of differential neural recruitment by children and adults on a SRT task, the latter of whom performed significantly better as well on the learning index. Particularly, there were age-related differences in neural activity for premotor cortex, putamen, hippocampus, inferotemporal cortex, and parietal cortex. The sharpest age discrepancy was in greater recruitment of fronto-striatal circuitry and hippocampal activation in adults, although such activity in these regions was not significantly correlated with the magnitude of the learning effect.

4.2. Tracking probabilistic dependencies

In this subsection, we now shift attention from work on repeating sequences to studies employing non-repeating sequences or artificial grammars.

4.2.1. Adjacent dependencies. The event-related brain responses of sleeping neonates indicate that the ability to use statistical cues (such as co-occurrence frequencies) to discriminate lexical boundaries among adjacent phonemes in a continuous artificial speech stream is present as early as one-half to two days after birth (Teinonen, Fellman, Näätänen, Alku, and Huotilainen 2009). Using behavioral measures, infants by two months are also sensitive to the co-occurrence frequencies obtaining across a continuous stream of geometric shapes with reliable shape-pairings (bigrams) (Kirkham, Slemmer, and Johnson 2002).

Beyond sensitivity to co-occurrences, studies further show robust statistical segmentation processes using transitional probabilities by five-and-a-half months (E. K. Johnson and Tyler 2010). This finding suggests an earlier time date for successful learning performance analogous to that demonstrated in the premier studies of statistical learning, conducted with eight-month-olds (as elaborated in greater detail in Sec. 2.) (Aslin et al. 1998; Saffran, Aslin, et al. 1996). And at twelve months, infants appear to be able to use adjacent probabilities to concurrently track pairwise syllables and nonwords belonging to an artificial language (Saffran and Wilson 2003), and as a first step towards learning form-based categories from nonword sequences of *aX* and *bY* strings (Gómez and Lakusta 2004). Interestingly, in the latter study, infants can generalize even where there is some inconsistency in the input, e.g., an 84/16 consistent-to-inconsistent ratio (but do not show generalization in a 68/32 condition).

Skipping ahead to about six years of age, earlier word segmentation and AGL work in young child learners of six- to nine- years would suggest that statistical-sequential learning effects may be age-invariant (e.g., Don et al. 2003; Saffran et al. 1997). However, a later study by Saffran (2001) provided evidence for clear age differences. Six- to nine-year old children and adults were both trained and tested on an artificial grammar containing predictive dependencies. While all participants demonstrated significant learning, the adults consistently outperformed the children, prompting Saffran to write that the results "suggest that children may possess a limited ability to acquire syntactic knowledge via statistical information. While their performance was not as strong as the adults', the children did acquire rudimentary aspects of the phrase structure of the language" (508). Another study by van den Bos (2007) compared the learning performance of ten- and eleven-year old children with that of adults' on an artificial grammar task, varying the usefulness of the underlying structure with respect to a cover task. Although the qualitative learning effect was the same, adults in the study acquired quantitatively greater knowledge of second-order dependencies than did the children. And finally, a study (Arciuli and Simpson, in press) of visual statistical learning using a triplet segmentation task was conducted with children ranging from five to twelve years. Quantitative improvements in discrimination performance for legal triplets on a forced-choice posttest were documented with increasing age.

It is quite possible that the results from these few artificial language studies might implicate poor metacognitive judgments as the source of differences between children and adults (and between younger and older children), rather than statistical-sequential learning skill per se. On the other hand, reports of age differences in learning are consistent with other work in VExP and SRT paradigms (and canonical statistical learning paradigms in the next section). Thus it is becoming evident that an assumption of developmental invariance for statistical-sequential learning is not a foregone conclusion.

4.2.2. Nonadjacent dependencies. Studies in the implicit learning literature tend not to investigate "nonadjacent dependencies," defined as relationships where another element(s) intervenes between two dependent elements, as their primary aim. More customarily, they may investigate learning for higher-order conditionals under the assumption that the concomitant non-local dependencies embedded in the stimuli sequences are learnt through the chunking of adjacencies (though see e.g., Kuhn and Dienes 2005, Pacton and Perruchet 2008, and Remillard 2008, for findings in the implicit learning literature with adults). Thus, what is currently known about incidental learning of nonadjacencies over development is mostly limited to studies conducted within the canonical statistical learning literature.

In infants (as in adults), it has been demonstrated that relatively high variability in the set size from which an "intervening" middle element of a string is drawn facilitates learning of the nonadjacent relationship between the two specific, flanking elements (Gómez 2002). In other words, when exposed to artificial grammar strings of the form *aXd* and *bXe*, individuals display sensitivity to the nonadjacent dependency-pairs (i.e., the *a_d* and *b_e* relations) when the elements composing the *X* are drawn from a large set distributed across many exemplars (e.g., when /*X*/ = 18 or 24). Performance is hindered, however, when the variability of the set size for the *X* is intermediate (e.g, /*X*/ = 12) or low (e.g., /*X*/ = 2).

Gómez and collaborators have assessed young infants' learning for such nonadjacent grammars with auditory nonword stimuli (monosyllabic tokens for *a, b, d,* and *e*; bisyllabic tokens for the *X's*) instantating the string-elements. Experiments (using a familiarization method and head-turn preference procedure) involved approximately three minutes of exposure to a 2-dependency nonadjacency grammar, followed by a phase in which infants were tested on their ability to discriminate grammatical strings belonging to the familiarized grammar or a foil grammar. While twelve-month-olds were unable to successfully discriminate strings following high-variability training conditions (Gómez and Maye 2005), fifteen-, seventeen-, and eighteen-month olds were able to make the discriminations (Gómez 2002; Gómez and Maye 2005). Such performance results also obtain across a four-hour delay between familiarization and test, and across different environmental settings (i.e., when familiarized to the grammar at home and then tested in the lab; Gómez et al. 2006). Gómez and Maye also reported age-group differences in looking-time trends to grammatical test items; fifteen- and seventeen-month olds exhibited familiarity and novelty preferences, respectively, supporting the researchers' conclusion that skill in detecting nonadjacencies appears to emerge by later infancy, with more robust tracking evidenced at seventeen and eighteen months.[5]

5. Developmental changes

Age differences in the magnitude of learning effects were observed across different age groups of young children in comparison to adults when learning deterministic and probabilistic sequences in SRT tasks. Such quantitative differences between child learners and adults were also observed in AGL paradigms. In light of the assumption that children should be naturally better statistical learners than adults given sensitive period effects observed in early language acquisition (J. S. Johnson and Newport 1989; Newport 1990), these findings may be counterintuitive and surprising.

Differences between learning adjacencies and nonadjacencies within a developmental context could potentially be an extension of quantitative performance-level differences, in that, as structural complexity purportedly increases, we see later ages of proficiency documented. Complexity of artificial grammars studied in the literature thus far seem to have some correspondence with dependency length—and measures, such as topological entropy, that correlate with length—over and above simply the number of associations or predictability of the grammar (van den Bos and Poletiek 2008). However, the simplified structures employed to date do not exhaust the full range of potential structures amenable to statistical tracking (e.g., such as embedded long-distance dependencies and cross serial-dependencies evidenced in natural language). Attempts to identify facilitative contexts for acquiring remote dependencies is still in its early stages. So "dependency length" may be more of a useful starting-point, rather than a conclusive identification of the main source for statistical-sequential learning "complexity." Neural network modeling of sequential processing in natural language suggest that complexity (as corresponding to measures reflecting human processing difficulty, such as protracted on-line RTs) is not reducible to dependency length and reflects interactions between experiential variations and constraints intrinsic to the architecture of the learning mechanism (Christiansen and MacDonald 2009; MacDonald and Christiansen 2002). Nonetheless, with this initial provisional definition, the findings reviewed herein are consistent with an interpretation of improved statistical-sequential learning performance with age for incidentally detecting increasingly "complex" structures, including but not confined to explaining the behavioral emergence of nonadjacency tracking (see Figure 2).

With certain stipulations, deterministic nonadjacencies in statistical learning bear surface resemblance to the second-order context-dependent sequences in the implicit learning literature. That is to say, a context-dependent sequence with 2^{nd} order relations (e.g., *1-2-1-3-1-2-1-3*...; which is the precise structure studied by Clohessy et al. 2001, noted earlier) parallels in its embedded relations the nonadjacency grammar-strings of type *2_3* or *3_2* as investigated in infants by Gómez and colleagues (see Sec. 4.2.2.) (n.b., set-size $\overline{X} = 1$, which is a "zero variability," potentially learning-conducive condition; Onnis, Christiansen, Chater, and Gómez 2003; Onnis, Monaghan, Christiansen, and Chater 2004). Interestingly, robust learning for these kinds of sequential regularities appears around 18 months in *both literatures*. Indeed, as mentioned earlier, much implicit learning work generally does not discriminate whether second-order context-dependent sequence relations are represented in the same manner as nonadjacencies of the kind studied in the statistical learning literature (although theoretically both forms of learned representations would be consistent with statistical-sequential learning mechanisms).

5.1. Transient cognitive constraints

While various cognitive constraints might be operative in accounting for patterns of developmental performance differences, here we concentrate on one promising factor that has already received some attention in the literature and that has particular utility for explaining the emergence of nonadjacency skills. The provisional hypothesis is that infants begin with limitations in the information that they can process in time, and are thus restricted in the amount of elements that can be effectively related with one another within the distance of this processing window. Following Santelmann and Jusczyk (1998) and others, the number of intervening elements (or syllable/word constituents) between dependencies comprised of similar units is used as a working definition of "distance" (precluding for now the issue of precise temporal duration). As the processing space expands during later development, this allows the infant to efficiently exploit and integrate more of an element's preceding context to discover appropriate nonadjacent relations that would otherwise be obscured.

Indeed, a narrow processing window may act as an initial "filter" to constrain the problem space of potential mappings (Newport 1988, 1990), thus focusing the infant's attention on more basic, local dependencies that can be later applied over longer distances, when the temporal window grows. Another related possibility is that a narrow window may act as an initial "amplifier" for detecting the covariation of input elements, because smaller sampling of a distribution increases the likelihood of observing correlations that are more extreme in magnitude than the true associations (Kareev 1995; Kareev, Lieberman, and Lev 1997). Given the structure of language, such memory-based constraints (in contrast to those of adults) might paradoxically contribute toward superior performance in language acquisition for child learners (Newport's "less is more" hypothesis). Computationally, "less is more" has a parallel in one of the two methods used in Elman's (1993) "starting small" simulations, in which Elman manipulated the resetting of a simple recurrent network's context units in order to simulate the child's initially reduced and then growing window for relating dependencies. Such a procedure enabled the network, thus "handicapped," to learn a complex language corpus that it had previously failed to master without recourse to such developmental limitations (or without having received incrementally staged input that externally mimicked such constraints). Conway, Ellefson, and Christiansen (2003) further investigated the effects of “starting small” in artificial grammar learning experiments with adults. In support of a starting small hypothesis, they documented a learning advantage for participants trained with incrementally staged input on complex visual grammars.

The notion that initial developmental constraints (or staged starting-small input) might scaffold the acquisition of more complex dependency-forms or tracking-skills than otherwise possible (or in a relatively quicker or more robust manner) is not unique to “purely” cognitive or linguistic phenomena, and may cut across perceptual development, too. For instance, it has been similarly postulated (with some support by neural-connectionist simulations) that the early limitations in human newborns’ visual acuities may actually *promote* the subsequent development of binocular disparity sensitivities emerging around four months (Dominguez and Jacobs 2003). More generally, such a hypothesis is consistent with categorization schemas of asynchronously developing experience-expectant brain systems in mammals (Greenough, Black, and Wallace 1987).

Empirically, the developmental timing of nonadjacency-related skills at eighteen months in statistical learning paradigms parallels emerging sensitivity by this same age in natural language learning for sensitivity to morphosyntactic relationships obtaining over one to three intervening syllables (Santelmann and Jusczyk 1998). This is consistent as well with Gómez and Maye’s (2005) suggested interpretation of their developmental results (described in Sec 4.2.2.). At the neural level, synaptic pruning might also be a mechanism for such gains in memory performance (and can also be explored computationaly in neural networks via “selective pruning” of nodes and connection weights) (cf. Quartz and Sejnowski 1997). Thus, the hypothesis of a limited temporal processing window has current empirical and theoretical support (see also Goldstein et al. 2010), and can be investigated in greater detail from various aspects within a computational perspective.

On the latter note, however, Rohde and Plaut's (2003) neural network simulations ("less is less") failed to replicate Elman's findings, thus questioning computational support for the hypothesis. However, our explanation here dodges these particular concerns. That is to say, although intriguing and unresolved, the issue of whether early limitations in processing space would be *beneficial per se* for the infant (under certain circumstances, perhaps) is not decisive to the validity of whether performance differences in statistical-sequential learning, especially for the emergence of nonadjacency tracking, can be traced to such a window. What is important to our exposition, then, is the idea that such transient cognitive limitations do appear to exist and that the notion of a temporal processing window, as expounded above, may offer a powerful framework for organizing the existing developmental data to date.

5.2. Changes in underlying neural structures

Work reviewed for the statistical-sequential learning of context-dependent sequences suggested differential recruitment of cortical and subcortical structures between children and adults, with the latter showing greater hippocampal activation (although this was not associated with the magnitude of the learning effect per se). Statistical-sequential learning, as a form of procedural learning, is likely to involve the participation of the basal ganglia through cortico-striatal circuits (based on supportive molecular and neuropsychological evidence, as well as theoretical views; Ackermann 2008; Christiansen, Kelly, Shillcock, and Greenfield 2010; Conway and Pisoni 2008; Lieberman 2002); and continually receives much attention in implicit learning accounts, especially in relation to putative performance "dissociations" among impaired populations. It has also been suggested that the basal ganglia may play a role in speech and language development via feedback-driven vocal learning, such as in the socially guided statistical learning of phonological patterns through contingent interactions between caregivers and prelinguistic, vocalizing infants (Goldstein and Schwade 2008).

However, the age-related differences in activation patterns with respect to hippocampal involvement are less clear. Minimally, they are interesting in that they dovetail with recent neuroimaging work implicating a potential role for both striatum (right caudate activation) and hippocampus in the on-line statistical learning performance of adults, using a visual triplet-segmentation task with glyphs from Sabean and Ndjuka syllabaries (Turk-Browne, Scholl, Chun, and Johnson 2009). The researchers have postulated potentially different roles/pathways for each, with the hippocampus putatively mediating more abstract learning, and the striatum involved in more specific associative encoding.

Of further interest, the age-related differences in hippocampal activation patterns might play a role in explaining conflicting findings of the effect of sleep on children's statistical-sequential learning. Within the implicit learning literature, Fischer, Wilhelm and Born (2007) had reported a sleep-dependent deterioration effect on children's learning of second-order contingencies from an SRT task. They had assessed the level of learning retention by seven- to eleven-year-olds when retested on the SRT sequence after either a period of sleep or an equal interval of wakefulness. The difference in average reaction times to grammatical versus ungrammatical SRT trials in the test blocks was the study's dependent measure of learning. Twenty- to thirty-year olds adults exhibited an improvement in this RT learning measure when retested following sleep, but a decrement in learning when retested in the wake condition. Compared to adults, however, the magnitude of learning decreased in the group of children following sleep, while remaining unchanged at retest in the wake condition.

This sleep-dependent deterioration in learning for the children contrasts with results from Gómez, Bootzin, and Nadel's (2006) statistical learning study, in which naps promoted better statistical learning generalization of nonadjacencies for infants. That is, fifteen-month-olds who napped within a 4-hour interval between a familiarization period and testing were able to discriminate between nonadjacency pairings that were either consistent or inconsistent with the artificial grammar generating the string they were presented with on the first test trial, but did not exhibit veridical recall for the specific pairings they had been acquainted with during familiarization. In contrast, fifteen-month olds without an intervening nap displayed veridical discrimination for prior dependency-strings, but not generalization of the nonadjacent structure to novel pairings. These results may nonetheless fit with those in Fischer et al.'s (2007) study, because the performance decrement in Fischer et al. was observed for veridical dependencies, rather than abstracted relations to new forms based on prior grammar probabilities.

Furthermore, in speculating on the differential effects of sleep in adults and children, Fischer et al. state:

> Moreover, for different learning tasks, a competitive interference between striatal and hippocampal systems has been shown (Packard and Knowlton 2002; Schroeder, Wingard, and Packard 2002; Poldrack et al. 2001). In this framework, opposite effects of sleep on implicit learning in children and adults might reflect that sleep in children leads to a preferential strengthening of hippocampal aspects of the memory representation, whereas sleep in adults strengthens caudate involvement. (224)

In support of this hypothesis, Fischer et al. note that whereas the children and adults in their study did not differ in amounts of rapid-eye-movement (REM) sleep, children's sleep was characterized by a greater amount of slow-wave sleep (SWS). In turn, SWS is associated with hippocampus-dependent memory consolidation (see the review by Marshall and Born 2007). Greater caudate involvement may lead to enhanced procedural learning, thus explaining adults' gains in SRT learning performances. However, strengthening hippocampal aspects of memories may not impact the implicit/indirect indices of learning tapped by the SRT, thus explaining the lack of gains in children after the sleep interval. (Recall again that in the Thomas et al. (2004) study (reported in Sec. 4.1.2.), adults' learning showed evidence of greater reliance on hippocampal recruitment than children's during on-line task performance, but this difference was not significantly linked to the SRT learning measure).

What other learning outcome then might be affected by strengthening hippocampal-dependent memories? No other measures of learning beyond procedural ones reflected by SRT performance were reported in Fischer et al. (2007). However, if Turk-Browne et al. (2009) are correct in connecting the hippocampus to the formation of more abstract representations in their statistical learning study, and if infants' napping comprises strengthening of hippocampal-dependent associations, this might account for the enhancement specific to *generalization* of the nonadjacencies to the novel pairings in Gómez et al. (2006). It would also be consistent with hypotheses put forward by Gómez et al. at a more cognitive-level of description. Namely, in speculating as to why sleep enhances generalization after learning in infants, Gómez et al. proposed three possibilities: 1) preferential weighting of abstract vs. specific information changes after sleep, 2) infants forget details of the items implementing the specific nonadjacent-pairings after sleep, and 3) sleep prolongs the learning-dependent processing necessary for abstraction to later occur. There is not conclusive evidence here for putting the matter to rest, but for now it provides further room for much speculation (and thoughts to sleep on).

5.3. Early perceptual development

Because statistical-sequential learning is closely mediated by perceptual features of the input and by modality constraints (Conway and Christiansen 2005, 2006, 2009; Conway et al. 2007), part of a thorough picture may likely include then the manner in which perceptual systems develop in response to early environmental input and are recruited through experience. Early perceptual learning, at least for certain gross-level acoustic or prosodic patterns, begins *in utero* during the last trimester (e.g., DeCasper and Spence 1986; see review of Gómez and Gerken 2000, Box 1, as well as computational simulations of scaffolding from prenatal "filtered" stimuli in Christiansen, Dale, and Reali 2010), suggesting a developmental trajectory that may have a substantial prenatal initial progression or foundation.

Postnatally, the evidence reviewed herein indicates that visual and spatiotemporal sequential learning for co-occurrence frequencies are present at two months (e.g., as in VExP findings and the statistical learning study by Kirkham et al. 2002) and older (e.g., by nine months; Emberson, Misyak, Schwade, Christiansen, and Goldstein 2008), with further statistical learning of regularities within visual arrays documented at nine months (Fiser and Aslin, 2002). In the auditory domain, natural language acquisition skills (some of which likely recruit upon statistical-sequential learning mechanisms) have also been investigated in very early infancy. Rudimentary, auditory sequential learning abilities appear present as early as one-half to two days after birth (Teinonen et al. 2009), and evidence of learning transitional probabilities embedded within linguistic stimuli are seen by five-and-a-half months (E. K. Johnson and Tyler 2010). However, auditory statistical learning for non-speech stimuli such as tones has seemingly not been documented in infants younger than eight months (Saffran et al. 1999). Systematic comparative work (using comparable procedures and stimuli) does not currently exist for deriving firm conclusions about potential modality-driven performance patterns and differences in infancy (Misyak, Emberson, Schwade, Christiansen, and Goldstein 2009); although one study in children indicates somewhat better learning of predictive dependencies in the auditory versus the visual modality (Saffran 2002). Thus, the timeline gaps in these cases probably reflect the fact that studies have typically not been motivated by comparative developmental trends—at least not with regard to modality or performance-level differences—rather than reflecting absences of ability per se.

It is further unknown how learning in infants and children may differ among specific dimension-modality pairings (i.e., comparisons for learning between visual and auditory stimuli, occurring in arrayed or sequential format, when regularities are also encoded along the dimension of location or variable timings). Kirkham, Slemmer, Richardson and Johnson (2007), however, have some work suggesting that sequential spatial-location regularities (of more complex form than those studied in the VExP task) may be more difficult for infants than other forms of sequence learning tasks and that proficiency may manifest later in perceptual development. We are also far from systematic investigations in infancy/childhood of feature dimension pairings. For instance, color and shape were always perfectly correlated in Fiser and Aslin's (2002) and Kirkham et al.'s (2002) visual statistical/sequential learning studies, thus preempting developmental considerations for the kind of phenomena investigated by Turk-Browne and colleagues (Turk-Browne et al. 2008; Turk-Browne and Scholl 2009) in adults regarding "bound object representations" and spatiotemporal generalization abilities.

In contrast to fairly-developed auditory abilities at birth (Lasky and Williams 2005; Saffran, Werker, and Werner 2006), the visual system undergoes more dramatic changes during the first year. As newborns, preferential looking visual acuity estimates are at approximately 1 cycles per degree (cpd; equivalent to 20/600 Snellen), developing to 3 cpd (20/200 Snellen) at 3 months, and reaching about 12 cpd (20/50 Snellen) by the end of the first year (Birch, Gwiazda, Bauer Jr, Naegele, and Held 1983; Courage and Adams 1990; Dobson and Teller 1978). Such early, transient limitations in the detail of infants' visual fields may thus narrow their perceptual focus to close-range visual stimuli and in turn support a more sequentially constrained format upon visual images they perceive—which could in turn necessarily favor visual statistical learning of temporally-distributed sequences over that of spatially-arrayed sequences in early processing. There are differential effects of temporal and spatial formats for auditory and visual statistical learning in adults, with visual-temporal conditions eliciting poorest performances (Conway and Christiansen 2009); but it has not been established whether and how such performance patterns might be shaped by, and/or possibly depart from, early perceptual experiences/biases. One intriguing possibility, therefore, is that the kind of early processing constraint alluded above, in tandem with an abundance of such early visual experiences during the first months, may temporarily place young infants' visual statistical learning of sequences above or on par with analogous learning for auditory sequences. Another complimentary idea is that the structure of prelinguistic social interaction with caregivers shapes infant attention in ways that facilitate specific forms of statistical learning (Goldstein et al. 2010). Such comparative learning among modalities is an ongoing matter of investigation, and if the former hypotheses bear out, it would form a surprising counterpoint to auditory superiority performance patterns observed in adults (Conway and Christiansen 2005, 2009; Saffran 2002).

Beyond the development of sequential learning through particular sensory experiences (and by implication, for distributed modality-constrained subsystems of statistical learning; see *Note 2*), there may further be a possible role for general principles in early neural development that facilitate and maintain so-called "entrenched" perceptual "discrimination" abilities (cf. Scott, Pascalis, and Nelson 2007). That is, it has been proposed that such mechanisms may be broader (more domain-general) than traditionally supposed. For instance, the language-specific "narrowing" characteristic of infants' later babbling as a product of increasing exposure to the ambient language, might also be driven or facilitated by experiences that are supralinguistic in some form; and more specifically, with respect to incorporated phonological patterns and articulatory/acoustic features, can be shaped by contingent parental feedback (Goldstein, King, and West 2003; Goldstein and West 1999) and "socially guided statistical learning" (Goldstein and Schwade 2008). Given, though, the evidence for gradiency effects in many "discrete" categorization performance tasks (e.g., Dale, Kehoe, and Spivey 2007; McMurray, Tanenhaus, Aslin, and Spivey 2003; see also Spivey 2007), an account encompassing these principles may also be heavily context-sensitive and ultimately entail a probabilistic cue-weighting explanation/mechanism that need not be *perceptually specific* in its *explanatory range* or extension. Assuming the requisite experiences for shaping such behavioral response patterns, these may in turn provide a "representational platform" or weighted biases over which related input features or cues for statistical-sequential learning mechanisms may be integrated. Furthermore, regarding so-called perceptual "enhancement" processes, it is likely that *the statistical distribution of the featural input* itself may play a large role in such phenomena (cf. Maye, Weiss, and Aslin 2008; Maye, Werker, and Gerken 2002).

In sum, empirical investigation of many of the newly hypothesized links, as described throughout this section, to statistical-sequential learning across development must be awaited. As of yet, despite promising potential, there are no systematic cross-sectional or longitudinal data for informing our understanding of patterns and trajectories in learning performance across modalities and/or with respect to related perceptual phenomena.

6. *Future developmental strides for the merger of statistical and implicit learning work*

A synthesis of findings from across the implicit and statistical learning literatures suggests that these two fields may have much to synergistically offer one another. It also indicates that their convergence may be especially fruitful for exploring issues related to infant/childhood cognitive constraints, underlying neural mechanisms, and early perceptual development. Furthermore, despite the orthodox assumption of age invariance, it appears that the possibility of developmental changes across studies should merit much stronger consideration than they presently have to date. Abandoning the *presumption* of developmental invariance might also provide the impetus for much-needed longitudinal and cross-sectional designs. There are conspicuous age gaps that reflect the nature of existing work in the area. Much work concentrates on early infancy, but does not clearly connect learning continuously across childhood. By going beyond documenting the age of successful learning for different statistical-sequential skills towards providing more detailed developmental trajectories, the projected merger of research findings across implicit and statistical learning paradigms will not only become truly developmental, but may perhaps flourish even more prominently past its early formal youth.

Notes

[1] While these characterizations are mostly limited to those of Reber and Saffran et al., key ideas surrounding "implicit learning" and "statistical learning" may also be connected to earlier theoretical contributions. The use of surface-level distributional information to identify relevant structure in language, for example, is a notion that has been recognized within structural linguistics and information theory by Bloomfield (1933), Harris (1955), and Shannon (1948). The formal metric of "transitional probability" in statistical learning segmentation studies was also provided within Miller and Selfridge's (1950) account for how "the statistical dependencies between successive units form the basis for a study of verbal context" (177). Despite methodological confounds in testing, Hayes and Clark's (1970) artificial segmentation experiment with adults was a forerunner to those of Saffran and colleagues. With regard to implicit learning, other researchers prior to Reber had also devoted attention to the subject of unconscious cognitive processes. Among these, Jenkins (1933: 471) wrote about "'incidental learning'—that is to say, learning which occurs in the absence of a specific intent to remember"; and Thorndike and Rock (1934) wrote about "learning without awareness of what is being learnt or intent to learn it" in an article of that same name. Clark Hull (1920) also discussed implicit/incidental learning phenomena in his published dissertation.

[2] Statistical learning may involve multiple subsystems that are modality-specific and that operate in parallel over distinct perceptual dimensions (Conway and Christiansen 2006, 2009; for analogous theoretical views in the traditional implicit learning literature, see Goschke, Friederici, Kotz, and van Kampen 2001; Keele, Ivry, Mayr, Hazeltine, and Heuer 2003). Additionally, it is not known whether statistical learning for adjacent and nonadjacent dependencies respectively—two types, or aspects, of statistical learning performance—entails shared or separate processing mechanism(s) in adult learners (a question raised by findings in Misyak and Christiansen, 2011); see also Friederici, Bahlmann, Heim, Schibotz, and Anwander (2006) and Pacton and Perruchet (2008). Hence, wherever wording to the effect of "a [statistical learning] mechanism" may be encountered in the text, this should be interpreted in a potentially distributive sense without necessarily inferring singularity.

[3] By using the term "statistical-sequential learning" as denoting the particular convergence of many findings across the statistical and implicit learning fields with respect to at least one kind of common underlying mechanism (i.e., probabilistic, associative-based and sequential), we are not suggesting that the merger of findings from the two fields cannot be construed as forming other meaningful overlaps (e.g., with respect to more "implicit" learning processes).

[4] There may, nonetheless, be important differences in both quantitative performance and the nature of limitations on statistical-sequential learning abilities across humans, non-human primates, and non-primates (Conway and Christiansen, 2001; Newport, Hauser, Spaepen and Aslin 2004; Saffran et al. 2008; Toro and Trobalón 2005; see also related discussion by Weiss and Newport 2006).

[5] This interpretation is informed by the influential Hunter-Ames model (Hunter and Ames 1988) in which greater stimulus complexity or partial encoding by infants is predicted to elicit familiarity preferences (longer looking/listening times for test stimuli that are consistent with the training exemplars) rather than the opposite pattern (i.e., a preference for attending longer to the novel, or inconsistent, test items). While the generality for interpreting preference patterns has not been definitively established (and is thus open to dispute), the observation that twelve-month olds were unable to demonstrate learning for the nonadjacent grammar (under the same experimental conditions) as the older infants further underscores Gómez and Maye's conclusion.

References

Ackermann, Hermann 2008 Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences, 31*, 265-272.

Adler, Scott A., Marshall M. Haith, Denise M. Arehart, and Elizabeth C. Lanthier 2008 Infants' visual expectations and the processing of time. *Journal of Cognition and Development, 9*, 1-25.

Arciuli, Joanne, and Ian C. Simpson in press Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*. doi: 10.1111/j.1467-7687.2009.00937.x

Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport 1998 Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

Birch, E. E., J. Gwiazda, J. A. Bauer Jr, J. Naegele, and R. Held 1983 Visual acuity and its meridional variations in children aged 7-60 months. *Vision Research, 23*, 1019-1024.

Bloomfield, Leonard 1933 *Language*. Chicago: University of Chicago Press.

Bremner, Andrew J., Denis Mareschal, Arnaud Destrebecqz, and Axel Cleeremans 2007 Cognitive control of sequential knowledge in 2-year-olds: Evidence from an incidental sequence-learning and -generation task. *Psychological Science, 18*, 261-266.

Canfield, Richard L., and Marshall M. Haith 1991 Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology, 27*, 198-208.

Canfield, Richard L., Elliott G. Smith, Michael P. Brezsnyak, Kyle L. Snow, Richard N. Aslin, Marshall M. Haith, Tara S. Wass, and Scott A. Adler 1997 Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development, 62*, i-160.

Christiansen, Morten H., Rick Dale, and Florencia Reali 2010 Connectionist explorations of multiple-cue integration in syntax acquisition. In S. P. Johnson (Ed.), *Neoconstructivism: The new science of cognitive development* (pp. 87-108). New York: Oxford University Press.

Christiansen, Morten H., M. Louise Kelly, Richard C. Shillcock, and Katie Greenfield 2010 Impaired artificial grammar learning in agrammatism. *Cognition, 116,* 382-393.

Christiansen, Morten H. and Maryellen C. MacDonald 2009 A usage-based approach to recursion in sentence processing. *Language Learning, 59 (Suppl. 1)*, 126-161.

Cleeremans, Axel 1993 *Mechanisms of implicit learning: A connectionist model of sequence processing*: MIT Press.

Cleeremans, Axel, and James L. McClelland 1991 Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235-253.

Clohessy, Anne B., Michael I. Posner, and Mary K. Rothbart 2001 Development of the functional visual field. *Acta Psychologica, 106*, 51-68.

Conway, Christopher M., and Morten H. Christiansen 2001 Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*, 539-546.

Conway, Christopher M., and Morten H. Christiansen 2005 Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 24-39.

Conway, Christopher M., and Morten H. Christiansen 2006 Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905-912.

Conway, Christopher M., and Morten H. Christiansen 2009 Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*, 561-580.

Conway, Christopher M., Michelle R. Ellefson, and Morten H. Christiansen 2003 When less is less and when less is more: Starting small with staged input. In Rick Alterman and David Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society* (pp. 270-275). Boston, MA: Cognitive Science Society.

Conway, Christopher M., Robert L. Goldstone, and Morten H. Christiansen 2007 Spatial constraints on visual statistical learning of multi-element scenes. In D. S. McNamara and J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 185-190). Austin, TX: Cognitive Science Society.

Conway, Christopher M., and Pisoni, D. B. 2008 Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. In Guinevere F. Eden and D. Lynn Flowers (Eds.), *Annals of the New York Academy of Sciences, 1145,* 113-131.

Courage, Mary L., and Russell J. Adams 1990 Visual acuity assessment from birth to three years using the acuity card procedure: Cross-sectional and longitudinal samples. *Optometry and Vision Science, 67*, 713-718.

Creel, Sarah C., Elissa L. Newport, and Richard N. Aslin 2004 Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1119-1130.

Curtin, Suzanne, Toben H. Mintz, and Morten H. Christiansen 2005 Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96*, 233-262.

Dale, Rick, Caitlin Kehoe, and Michael J. Spivey 2007 Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition, 35*, 15-28.

DeCasper, A., and Spence, M. 1986 Prenatal maternal speech influences on newborns' perception of speech sounds. *Infant Behavior and Development, 9*, 133-150.

Dienes, Zoltán 1992 Connectionist and memory-array models of artificial grammar learning. *Cognitive Science, 16*, 41-79.

Dobson, Velma, and Davida Y. Teller 1978 Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research, 18*, 1469-1483.

Dominguez, Melissa, and Robert A. Jacobs 2003 Does visual development aid visual learning? In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 257-278). NY: Psychology Press.

Don, Audrey J., E. Glenn Schellenberg, Arthur S. Reber, Kristen M. DiGirolamo, and Paul P. Wang 2003 Implicit learning in children and adults with Williams syndrome. *Developmental Neuropsychology, 23*, 201-225.

Ellis, Nick C. (Ed.) 1994 *Implicit and explicit learning of languages*. London: Academic Press.

Elman, Jeffrey L. 1993 Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99.

Emberson, Lauren L., Jennifer B. Misyak, Jennifer A. Schwade, Morten H. Christiansen, and Michael H. Goldstein 2008 Face-to-face: Visual statistical learning with complex natural stimuli. *XVIth International Conference on Infant Studies* (abstracts ed., pp. 259). Vancouver, Canada.

Fischer, Stefan, Ines Wilhelm, and Jan Born 2007 Developmental differences in sleep's role for implicit off-line learning: Comparing children with adults. *Journal of Cognitive Neuroscience, 19*, 214-227.

Fiser, József, and Richard N. Aslin 2002 Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*, 15822-15826.

Fletcher, Janet, Murray T. Maybery, and Sarah Bennett 2000 Implicit learning differences: A question of developmental level? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 246-252.

Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schibotz, and Alfred Anwander 2006 The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458-2463.

Gebauer, Guido F., and Nicholas J. Mackintosh 2007 Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 34-54.

Goldstein, Michael H., Andrew P. King, and Meredith J. West 2003 Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100*, 8030-8035.

Goldstein, Michael H., and Jennifer A. Schwade 2008 Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*, 515-523.

Goldstein, Michael H., Heidi R. Waterfall, Arnon Lotem, Joseph Y. Halpern, Jennifer A. Schwade, Luca Onnis, and Shimon Edelman 2010 General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences, 14*, 249-258.

Goldstein, Michael H., and Meredith J. West 1999 Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology, 113*, 52-58.

Gómez, Rebecca L. 2002 Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gómez, Rebecca L. 2006 Dynamically guided learning. In Y. Munakata and M. H. Johnson (Eds.), *Attention and performance XXI: Processes of change in brain and cognitive development* (pp. 87-110). Oxford: Oxford University Press.

Gómez, Rebecca L., Richard R. Bootzin, and Lynn Nadel 2006 Naps promote abstraction in language-learning infants. *Psychological Science, 17*, 670-674.

Gómez, Rebecca L., and LouAnn Gerken 1999 Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gómez, Rebecca L., and LouAnn Gerken 2000 Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178-186.

Gómez, Rebecca L., and Laura Lakusta 2004 A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*, 567-580.

Gómez, Rebecca L., and Jessica Maye 2005 The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*, 183-206.

Goschke, Thomas, Angela D. Friederici, Sonja A. Kotz, and Anja van Kampen 2001 Procedural learning in Broca's aphasia: Dissociation between the implicit acquisition of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience, 13*, 370-388.

Greenough, William T., James E. Black, and Christopher S. Wallace 1987 Experience and brain development. *Child Development, 58*, 539-559.

Haith, Marshall M., Cindy Hazan, and Gail S. Goodman 1988 Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development, 59*, 467-479.

Haith, Marshall M., and Michael E. McCarty 1990 Stability of visual expectations at 3.0 months of age. *Developmental Psychology, 26*, 68-74.

Haith, Marshall M., Naomi Wentworth, and Richard L. Canfield 1993 The formation of expectations in early infancy. In C. Rovee-Collier and L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8). Norwood, NJ: Ablex.

Harris, Zellig S. (1955). From phoneme to morpheme. *Language, 31*, 190-222.

Hasher, Lynn, and Rose T. Zacks 1984 Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372-1388.

Hayes, J. R., and H. H. Clark 1970 Experiments in the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 221-234). New York: Wiley.

Hull, Clark Leonard 1920 Quantitative aspects of the evolution of concepts. An experimental study. *Psychological Monographs, 28*.

Hunt, Ruskin H., and Richard N. Aslin 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.

Hunter, Michael A., and Elinor W. Ames 1988 A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5*, 69-95.

Jenkins, John G. 1933 Instruction as a factor in 'incidental' learning. *The American Journal of Psychology, 45*, 471-477.

Johnson, Elizabeth K., and Peter W. Jusczyk 2001 Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548-567.

Johnson, Elizabeth K., and Michael D. Tyler 2010 Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339-345.

Johnson, Jacqueline S., and Elissa L. Newport 1989 Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Jones, James, and Harold Pashler 2007 Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review, 14*, 295-300.

Kareev, Yaakov 1995 Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263-269.

Kareev, Yaakov, Iris Lieberman, and Miri Lev 1997 Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*, 278-287.

Keele, Steven W., Richard Ivry, Ulrich Mayr, Eliot Hazeltine, and Herbert Heuer 2003 The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316-339.

Keele, Steven W., and Peggy J. Jennings 1992 Attention in the representation of sequence: Experiment and theory. *Human Movement Science, 11*, 125-138.

Kirkham, Natasha Z., Jonathan A. Slemmer, and Scott P. Johnson 2002 Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35-B42.

Kirkham, Natasha Z., Jonathan A. Slemmer, Daniel C. Richardson, and Scott P. Johnson 2007 Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78*, 1559-1571.

Knowlton, Barbara J., Squire, Larry R., and Gluck, Mark A. 1994 Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106-120.

Kuhn, Gustav, and Zoltán Dienes 2005 Implicit learning of nonlocal musical rules: Implicitly learning more rhan chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1417-1432.

Lasky, Robert E., and Amber L. Williams 2005 The development of the auditory system from conception to term. *Neoreviews, 6*, 141-152.

Lewicki, Pawel, Maria Czyzewska, and Hunter Hoffman 1987 Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 523-530.

Lewkowicz, David J. 2008 Perception of dynamic and static audiovisual sequences in 3- and 4-month-old infants. *Child Development, 79*, 1538-1554.

Lieberman, Philip 2002 On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology, 35*, 36-62.

MacDonald, Maryellen C., and Morten H. Christiansen 2002 Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*, 35-54.

Marshall, Lisa, and Jan Born 2007 The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*, 442-450.

Maybery, Murray T., Margaret Taylor, and Angela O'Brien-Malone 1995 Implicit learning: Sensitive to age but not IQ. *Australian Journal of Psychology, 47*, 8-17.

Maye, Jessica, Daniel J. Weiss, and Richard N. Aslin 2008 Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, Jessica, Janet F. Werker, and LouAnn Gerken 2002 Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

McCauley, Stewart M., and Morten H. Christiansen 2011 *Learning simple statistics for language comprehension and production: The CAPPUCCINO model*. Submitted manuscript.

McMurray, Bob, Michael K. Tanenhaus, Richard N. Aslin, and Michael J. Spivey 2003 Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research, 32*, 77-97.

Meulemans, Thierry, Martial Van der Linden, and Pierre Perruchet 1998 Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*, 199-221.

Miller, George A., and Jennifer A. Selfridge 1950 Verbal context and the recall of meaningful material. *The American Journal of Psychology, 63*, 176-185.

Mirman, Daniel, Katharine Graf Estes, and James S. Magnuson 2010 Computational modeling of statistical learning: Effects of transitional probability versus frequency and links to word learning. *Infancy, 15,* 471-486.

Misyak, Jennifer B., and Morten H. Christiansen February 7, 2011 Statistical learning and language: An individual differences study. *Language Learning*. doi: 10.1111/j.1467-9922.2010.00626.x

Misyak, Jennifer B., Morten H. Christiansen, and J. Bruce Tomblin 2010 Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2*, 138-153.

Misyak, Jennifer B., Lauren L. Emberson, Jennifer A. Schwade, Morten H. Christiansen, and Michael H. Goldstein 2009 Face-to-face, word-for-word: Comparing infants' statistical learning of visual and auditory sequences. Poster session presented at the *Biennial Meeting of the Society for Research in Child Development*. Denver, CO.

Nelson, Charles A. 1995 The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology, 31*, 723-738.

Newport, Elissa L. 1988 Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*, 147-172.

Newport, Elissa L. 1990 Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen, and Richard N. Aslin 2004 Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*, 85-117.

Nissen, Mary Jo, and Peter Bullemer 1987 Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

Onnis, Luca, Morten H. Christiansen, Nick Chater, and Rebecca L. Gómez 2003 Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In R. Alterman and D. Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society* (pp. 886-891). Boston, MA: Cognitive Science Society.

Onnis, Luca, Padraic Monaghan, Morten H. Christiansen, and Nick Chater 2004 Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In K. Forbus, D. Gentner and T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1047-1052). Chicago, Illinois: Cognitive Science Society.

Pacton, Sébastien, and Pierre Perruchet 2008 An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 80-96.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran 2009a Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*, 244-247.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran 2009b Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*, 674-685.

Perrig, W. J. 2001 Implicit memory, cognitive psychology of. In: N. J. Smelser and P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences,* Vol. 11, Elsevier, Amsterdam (2001), pp. 7241-7245.

Perruchet, Pierre, and Stéphane Desaulty 2008 A role for backward transitional probabilities in word segmentation? *Memory & Cognition, 36*, 1299-1305.

Perruchet, Pierre, and Sébastien Pacton 2006 Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Perruchet, Pierre, and Ronald Peereman 2004 The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17*, 97-119.

Quartz, Steven R., and Terrence J. Sejnowski 1997 The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences, 20*, 537-596.

Reber, Arthur S. 1967 Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Reber, Arthur S. 1993 *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reber, Arthur S., and Rhianon Allen 1978 Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition, 6*, 189-221.

Reber, Arthur S., and Rhianon Allen 2000 Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf and B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227-247). Amsterdam: John Benjamins Publishing Company.

Reber, Arthur S., and Selma Lewis 1977 Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition, 5*, 333-361.

Reber, Arthur S., Faye F. Walkenfeld, and Ruth Hernstadt 1991 Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888-896.

Remillard, Gilbert 2008 Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology, 61*, 400-424.

Rescorla, Robert A. 1967 Pavlovian conditioning and its proper control procedures. *Psychological Review, 74*, 71-80.

Rescorla, Robert A., and Allan R. Wagner 1972 A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Reznick, J. Steven, Katarzyna Chawarska, and Stephanie Betts 2000 The development of visual expectations in the first year. *Child Development, 71*, 1191-1204.

Rohde, Douglas L. T., and David C. Plaut 2003 Less is less in language acquisition. In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 189-232). NY: Psychology Press.

Roodenrys, Steven, and Naomi Dunn 2008 Unimpaired implicit learning in children with developmental dyslexia. *Dyslexia, 14*, 1-15.

Rose, Susan A., Judith F. Feldman, Jeffery J. Jankowski, and Donna M. Caro 2002 A longitudinal study of visual expectation and reaction time in the first year of life. *Child Development, 73*, 47-61.

Saffran, Jenny R. 2001 The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.

Saffran, Jenny R. 2002 Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Saffran, Jenny R. 2003 Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport 1996 Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, Jenny R., Marc Hauser, Rebecca Seibel, Joshua Kapfhamer, Fritz Tsao, and Fiery Cushman 2008 Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition, 107*, 479-500.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport 1999 Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Saffran, Jenny R., Elissa L. Newport, and Richard N. Aslin 1996 Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, Jenny R., Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and Sandra Barrueco 1997 Incidental language learning: Listening (and learning) out of the comer of your ear. *Psychological Science, 8*, 101-105.

Saffran, Jenny R., and Erik D. Thiessen 2003 Pattern induction by infant language learners. *Developmental Psychology, 39*, 484-494.

Saffran, Jenny R., and Erik D. Thiessen 2007 Domain-general learning capacities. In E. Hoff. and M. Shatz (Eds.), *Handbook of Language Development* (pp. 68-86). Cambridge: Blackwell.

Saffran, Jenny R., Janet F. Werker, and Lynne A. Werner 2006 The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler and D. Kuhn (Eds.), *Handbook of Child Development* (pp. 58-108). New York: Wiley.

Saffran, Jenny R., and Diana P. Wilson 2003 From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273-284.

Santelmann, Lynn M., and Peter W. Jusczyk 1998 Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*, 105-134.

Scott, Lisa S., Olivier Pascalis, and Charles A. Nelson 2007 A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science, 16*, 197-201.

Schvaneveldt, Roger W., and Rebecca L. Gómez 1998 Attention and probabilistic sequence learning. *Psychological Research, 61,* 175-190.

Servan-Schreiber, David, Axel Cleeremans, and James L. McClelland 1988 Encoding sequential structure in simple recurrent networks, *CMU Technical Report CMU-CS-335-87*: Computer Science Department, Carnegie-Mellon University.

Shanks, David R. 1995 *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shannon, C. E. 1948 A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379-423, 623-656.

Spivey, Michael J. 2007 *The continuity of mind*. New York: Oxford University Press.

Swingley, Daniel 2005 Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86-132.

Teinonen, Tuomas, Vineta Fellman, Risto Näätänen, Paavo Alku, and Minna Huotilainen 2009 Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*, 21.

Thomas, Kathleen M., Ruskin H. Hunt, Nathalie Vizueta, Tobias Sommer, Sarah Durston, Yihong Yang, and Michael S. Worden 2004 Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience, 16*, 1339-1351.

Thomas, Kathleen M., and Charles A. Nelson 2001 Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology, 79*, 364-387.

Thorndike, E. L., and Rock, R. T. 1934 Learning without awareness of what is being learned or intent to learn it. *Journal of Experimental Psychology, 17*, 1-19.

Toro, Juan M., and Josep B. Trobalón 2005 Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*, 867-875.

Turk-Browne, Nicholas B., Philip J. Isola, Brian J. Scholl, and Teresa A. Treat 2008 Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 399-407.

Turk-Browne, Nicholas B., and Brian J. Scholl 2009 Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance, 35, 195-202*.

Turk-Browne, Nicholas B., Brian J. Scholl, Marvin M. Chun, and Marcia K. Johnson 2009 Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934-1945.

Vaidya, Chandan J., Marianne Huger, Darlene V. Howard, and James H. Howard Jr 2007 Developmental differences in implicit learning of spatial context. *Neuropsychology, 21*, 497-506.

van den Bos, Esther 2007 *Implicit artificial grammar learning: Effects of complexity and usefulness of the structure*. Unpublished PhD dissertation, Leiden University, Netherlands.

van den Bos, Esther, and Fenna H. Poletiek 2008 Effects of grammar complexity on artificial grammar learning. *Memory & Cognition, 36*, 1122-1131.

Weiss, Daniel J., and Elissa L. Newport 2006 Mechanisms underlying language acquisition: Benefits from a comparative approach. *Infancy, 9*, 241-257.

Wentworth, Naomi, and Marshall M. Haith 1992 Event-specific expectations of 2- and 3-month-old infants. *Developmental Psychology, 28*, 842-850.

Wentworth, Naomi, Marshall M. Haith, and Roberta Hood 2002 Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy, 3*, 303-322.

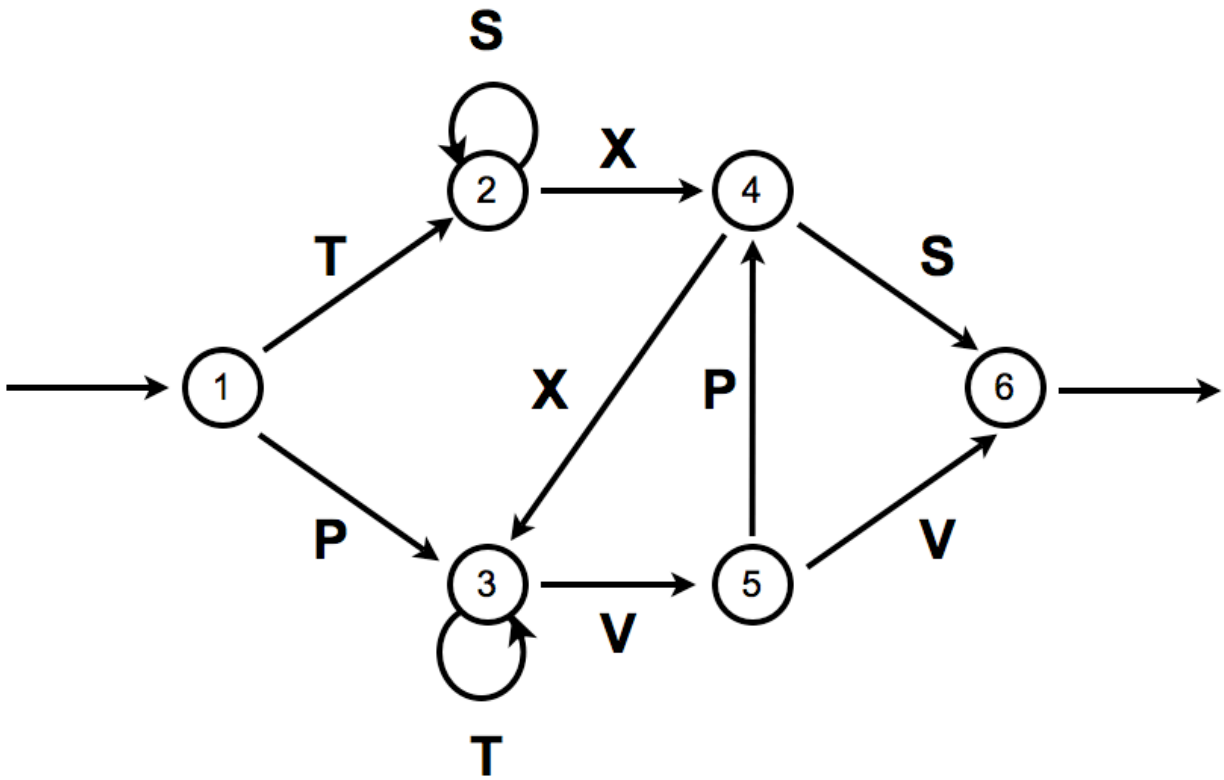
Zacks, Rose T., and Lynn Hasher 2002 Frequency processing: A twenty-five year perspective. In P. Sedlmeier and T. Betsch (Eds.), *Frequency processing and cognition* (pp. 21-36). New York: Oxford University Press

Figure Caption

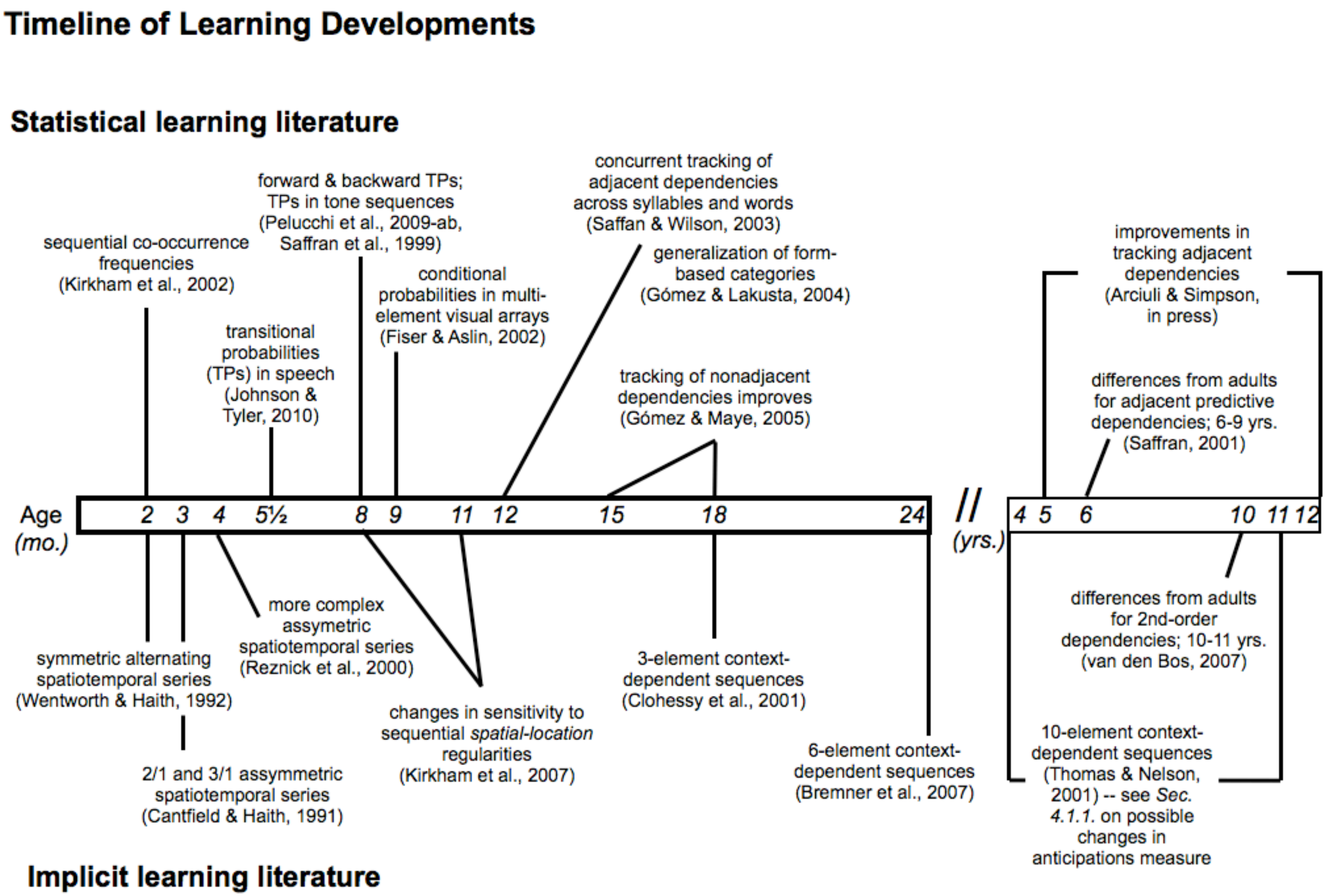
Figure 1. An illustration of an artificial finite-state grammar adapted from Reber and Lewis (1977). Strings are generated by starting at the leftmost node and following possible paths marked by the arrows to other nodes. The succession of letters associated with the arrows encountered along the traced path corresponds to a grammatical string sequence. For example, following the arrow from node 1 to 3, the arcing arrow back to 3, and then the respective arrows to nodes 5, 4 and 6 produces the letter string PTTVPS.

Figure 2. A provisional timeline of learning that indicates behavioral developments documented within the statistical learning (top) and implicit learning (bottom) literatures. As the timeline is intended to illuminate potential developmental changes, findings of learning sensitivity that are established exclusively from neurophysiological measures are omitted (e.g., ERP measures in Teinonen et al. 2009). Neural sensitivity can be evidenced even without behavioral discrimination on standard performance measures (Turk-Browne et al. 2009), thus making neurophysiological comparisons to quantitative behavioral assessments (especially null findings) less straightforward. Nonetheless, developmental trends suggested by neuroscience data, as discussed in Sec. 6.2., may be particularly fruitful for understanding neural mechanisms involved in statistical-sequential learning.

1



2



Timeline of Learning Developments
Statistical learning literature
sequential co-occurrence frequencies (Kirkham et al., 2002)
transitional probabilities (TPs) in speech (Johnson & Tyler, 2010)
forward & backward TPs; TPs in tone sequences (Pelucchi et al., 2009-ab, Saffran et al., 1999)
conditional probabilities in multi-element visual arrays (Fiser & Aslin, 2002)
concurrent tracking of adjacent dependencies across syllables and words (Saffan & Wilson, 2003)
generalization of form-based categories (Gómez & Lakusta, 2004)
tracking of nonadjacent dependencies improves (Gómez & Maye, 2005)
improvements in tracking adjacent dependencies (Arciuli & Simpson, in press)
differences from adults for adjacent predictive dependencies; 6-9 yrs. (Saffran, 2001)
Age (mo.)
2 3 4 5½ 8 9 11 12 15 18 24
//
(yrs.)
4 5 6 10 11 12
symmetric alternating spatiotemporal series (Wentworth & Haith, 1992)
2/1 and 3/1 assymmetric spatiotemporal series (Cantfield & Haith, 1991)
more complex assymetric spatiotemporal series (Reznick et al., 2000)
changes in sensitivity to sequential *spatial-location* regularities (Kirkham et al., 2007)
3-element context-dependent sequences (Clohessy et al., 2001)
6-element context-dependent sequences (Bremner et al., 2007)
differences from adults for 2nd-order dependencies; 10-11 yrs. (van den Bos, 2007)
10-element context-dependent sequences (Thomas & Nelson, 2001) -- see *Sec. 4.1.1.* on possible changes in anticipations measure
Implicit learning literature