

1 **Statistical-Sequential Learning in Development**

2
3
4 *Jennifer B. Misyak, Michael H. Goldstein and*
5 *Morten H. Christiansen*
6

7 8 9 **1. Introduction**

10
11 The child’s mind is embedded in a flowing, stimuli-rich world of perceived
12 regularities. Children learn to engage their surroundings skillfully, in a
13 manner reflecting knowledge of astoundingly complex structural patterns.
14 From infants to adults, studying the unsupervised, ostensibly unconscious
15 nature underlying many of these early-acquired processes has continually
16 fueled research in the fields of “implicit learning” and “statistical learning”
17 since their onset. The convention has been to trace the empirical genesis of
18 implicit learning research from the early work of Reber (1967) in the 60’s,
19 and to follow modern statistical learning developments since the seminal
20 work of Saffran and colleagues in the 90’s (Saffran, Aslin, and Newport
21 1996).

22 This chapter, though, is not about tracing the vibrant histories of the
23 implicit and statistical learning fields, but rather about interrelating the
24 accrued developmental findings from both literatures – while in the service
25 of promoting a synergistic fusion between their two approaches. For
26 despite their disparately-pursued lines of work to date, establishing this
27 common discourse may be easier to accomplish than otherwise presupposed.
28 Witness, for instance, the close overlap among operational definitions pro-
29 vided by their paradigm progenitors.¹ Implicit learning has thus been related
30 as “the process by which knowledge about complex stimulus domains is
31

32 1. While these characterizations are mostly limited to those of Reber and Saffran
33 et al., key ideas surrounding “implicit learning” and “statistical learning” may
34 also be connected to earlier theoretical contributions. The use of surface-level
35 distributional information to identify relevant structure in language, for exam-
36 ple, is a notion that has been recognized within structural linguistics and infor-
37 mation theory by Bloomfield (1933), Harris (1955), and Shannon (1948). The
38 formal metric of “transitional probability” in statistical learning segmentation
39 studies was also provided within Miller and Selfridge’s (1950) account for how
40 “the statistical dependencies between successive units form the basis for a
study of verbal context” (177). Despite methodological confounds in testing,

1 acquired largely independent of conscious awareness of either the process
 2 or the products of acquisition” (Reber and Allen 2000: 227). Reber’s
 3 initial account (1967) also included characteristics such as the “efficient
 4 responding” of the organism and the development of “a strong sensitivity
 5 to the lawfulness [...] [existing in a] stimulus array;” in other words,
 6 participants may “become sensitive to the statistical nature of their envi-
 7 ronment without using explicit or verbalizable strategies” when the stimuli
 8 they receive is “patterned” or “ordered.” He emphasized how this was
 9 “closely akin to Gibson and Gibson’s (1955) perceptual learning and is
 10 [...] a rudimentary inductive process which is intrinsic in such phenomena
 11 as language learning and pattern perception” (863).

12 Regarding statistical learning in its modern form, Saffran and collabora-
 13 tors introduced this as a “powerful mechanism for the computation of
 14 statistical properties of the language input” (Saffran, Aslin, et al. 1996:
 15 1926), emphasizing the rapidity and adeptness by which infant learners
 16 incidentally extracted relevant regularities. This was otherwise termed
 17 “the process of statistical learning, or the detection of patterns of sounds,
 18 words, and classes of words in the service of discovering underlying struc-
 19 ture” (Saffran 2002: 172) and could be relevant for learning complex (i.e.,
 20 hierarchical) linguistic forms (Saffran 2001). As subsequent findings sug-
 21 gested that such learning was widely applicable to a variety of nonlinguistic
 22 domains, this empirical definition became more general, e.g., “the ability to
 23 discover units via their statistical coherence” (Saffran 2003: 111), and con-
 24 strued more broadly by researchers, e.g., as pertaining to “the discovery of
 25 statistical structure in the environment” (Gómez 2006: 90).

26 Close terminological correspondences such as these invite parallels – or
 27 demand a more rigorous partitioning and defining of phenomena (see
 28 Perrig 2001 for analogous claims in conjunction with implicit memory
 29 and procedural knowledge). It is our contention, though, that the resem-
 30

31
 32 Hayes and Clark’s (1970) artificial segmentation experiment with adults was a
 33 forerunner to those of Saffran and colleagues. With regard to implicit learn-
 34 ing, other researchers prior to Reber had also devoted attention to the subject
 35 of unconscious cognitive processes. Among these, Jenkins (1933: 471) wrote
 36 about “‘incidental learning’ – that is to say, learning which occurs in the
 37 absence of a specific intent to remember”; and Thorndike and Rock (1934)
 38 wrote about “learning without awareness of what is being learnt or intent to
 39 learn it” in an article of that same name. Clark Hull (1920) also discussed
 40 implicit/incidental learning phenomena in his published dissertation.

1 blances in this case point to a common learning mechanism(s)² – albeit
 2 historically seen from different theoretical orientations and empirical
 3 traditions – and hereafter referred to as statistical-sequential learning.
 4 That is, much of the phenomena revealed through statistical learning and
 5 implicit learning approaches concerns the learning of sequential material
 6 and largely taps into the same probabilistically-sensitive, associative-based
 7 mechanism(s) recognized as belonging to “statistical learning” proper.³

8 One potential discrepancy, however, between the descriptions above
 9 **relate** to Reber’s notion of the process as proceeding via unconscious *rule*
 10 *abstraction*, in which the participant tacitly apprehends “a valid, if partial,
 11 representation of the actual underlying rules of the [finite-state] language”
 12 (Reber and Allen 1978: 191). Conversely, statistical learning is construed
 13 as a process driven by statistical properties of the input, which results in
 14 participants’ probabilistic knowledge of the constraints governing stimuli
 15 formation. However, the sensitivity of participants to the “lawfulness”
 16 per se in sequentially arrayed material of implicit learning experiments
 17 has been sharply contested and has given rise to statistically-based explana-
 18 tions (without symbolic or abstract rules) of the computational principles
 19 entailed by successful performance (e.g., Cleeremans 1993). Although still
 20 an active matter of debate, if one is convinced by converging evidence that
 21 such learning is indeed driven by sensitivity to the *statistical* properties of
 22 the stimuli, then this natural affinity between the two fields should be
 23

-
- 24 2. Statistical learning may involve multiple subsystems that are modality-specific
 25 and that operate in parallel over distinct perceptual dimensions (Conway and
 26 Christiansen 2006, 2009; for analogous theoretical views in the traditional
 27 implicit learning literature, see Goschke, Friederici, Kotz, and van Kampen
 28 2001; Keele, Ivry, Mayr, Hazeltine, and Heuer 2003). Additionally, it is not
 29 known whether statistical learning for adjacent and nonadjacent dependencies
 30 respectively – two types, or aspects, of statistical learning performance –
 31 entails shared or separate processing mechanism(s) in adult learners (a ques-
 32 tion raised by findings in Misyak and Christiansen, 2011); see also Friederici,
 33 Bahlmann, Heim, Schibotz, and Anwander (2006) and Pacton and Perruchet
 34 (2008). Hence, wherever wording to the effect of “a [statistical learning]
 35 mechanism” may be encountered in the text, this should be interpreted in a
 36 potentially distributive sense without necessarily inferring singularity.
- 37 3. By using the term “statistical-sequential learning” as denoting the particular
 38 convergence of many findings across the statistical and implicit learning fields
 39 with respect to at least one kind of common underlying mechanism (i.e., prob-
 40 abilistic, 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).

1 readily apparent. They both entail incidental learning of sequential patterns
2 (in spatiotemporal, temporal, or visually-arrayed distributions) that are
3 defined by statistical relations over units perceived by the learner and that
4 are processed in a manner respecting intrinsic regularities or probabilistic
5 constraints of the input.

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

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

35 In addressing developmental change, this chapter may be admittedly
36 considered unorthodox. Developmental invariance is one of the central
37 tenets, or corollaries, falling out of the theory on unconscious cognition
38 posited by Reber, in which implicit learning is viewed as recruiting upon
39 phylogenetically conserved and evolutionarily stable processes of high,
40 basic adaptive value since antiquity (“the primacy of the implicit”; Reber,

1 1993). By Reber's framework, implicit learning has been expected to
2 exhibit age independence, neurobiological robustness, little intraindividual
3 variation, and remarkable cross-species commonality. These assumptions
4 have in turn deterred many implicit learning researchers from directly
5 seeking developmental trends. And the assumption of age-independence
6 seems to even have seen its way borrowed into the canonical statistical
7 learning literature; see especially Saffran et al. (1997, but note the conflict-
8 ing evidence for developmental differences later found in Saffran 2001).

9 Such largely unconscious processes may indeed have basic and evolu-
10 tionarily old roots, as well as recruit upon mechanisms shared across
11 species. Fittingly, researchers from both implicit and statistical learning
12 fields have specifically invoked parallels to principles from the classical
13 Rescorla-Wagner model (1972) of animal learning along several lines,
14 e.g., regarding the detection of predictive co-variation of stimulus events
15 (Reber and Allen 2000), the similar subjection to perceptual constraints
16 (Creel, Newport, and Aslin 2004), the use of prediction-based estimation
17 from conditional statistics, or contingent probabilities (Aslin, Saffran,
18 and Newport 1998; Hunt and Aslin 2001; Swingley 2005), and in relation
19 to attention-based accounts of statistical learning (Pacton and Perruchet
20 2008). Despite some cross-species commonalities⁴, though, earlier claims
21 of implicit learning's (and by inference, statistical learning's) neurobiolog-
22 ical robustness across impaired populations, and reputed lack of substan-
23 tive differences across individuals, are being eroded by converging, recent
24 evidence suggesting that systematic variations do in fact exist. Even Reber
25 and Allen (2000) more recently have conceded the existence of some in-
26 dividual differences, referring back for support to findings from Reber,
27 Walkenfeld, and Hernstadt (1991); they conversely argue now for which
28 theoretical framework should be used to interpret these differences. Amid
29 such shifting ground, this chapter's ancillary aim is to reappraise the re-
30 maining, fundamental postulate of developmental invariance, with ensuing
31 implications for our understanding of the nature of statistical-sequential
32 learning and the factors by which it may be influenced.

33 The remainder of this chapter is organized into five main divisions. The
34 next section considers various paradigm implementations used in the devel-

35
36 4. There may, nonetheless, be important differences in both quantitative performance
37 and the nature of limitations on statistical-sequential learning abilities across
38 humans, non-human primates, and non-primates (Conway and Christiansen,
39 2001; Newport, Hauser, Spaepen and Aslin 2004; Saffran et al. 2008; Toro
40 and Trobalón 2005; see also related discussion by Weiss and Newport 2006).

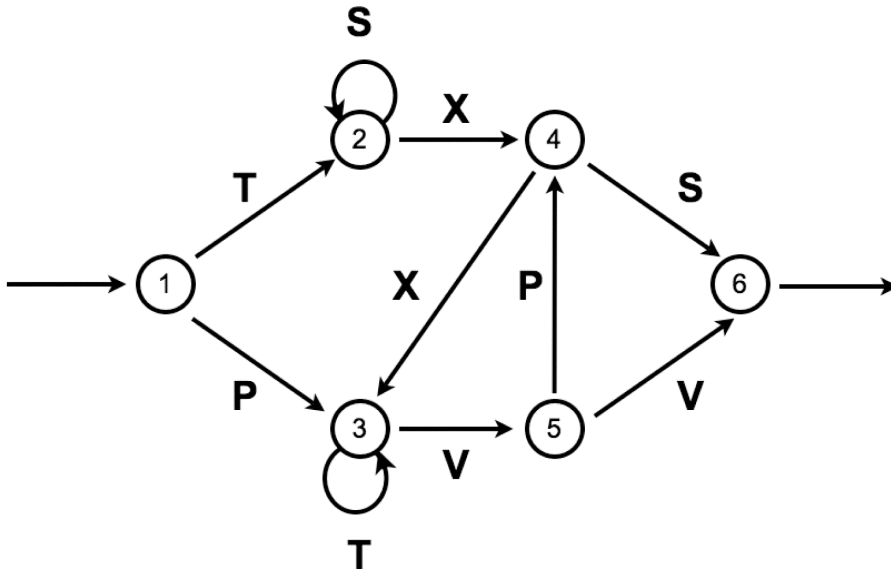
1 opmental statistical and implicit learning literatures. Subsequently in Sec. 3,
 2 attention is directed towards areas of overlap between implicit and statistical
 3 learning research, the convergence of which delimits the statistical-sequential
 4 learning phenomena discussed throughout this chapter. In Sec. 4, we high-
 5 light various aspects of infants' and children's statistical-sequential learning
 6 as they relate to the processing of sequential relations and the tracking of
 7 probabilistic dependencies. With this punctuated empirical review in place,
 8 major developmental trends are then identified and further elaborated upon
 9 with regard to potential underlying factors (Sec. 5). The conclusion then
 10 ties together prospects and future directions for one way of bridging
 11 implicit and statistical learning literatures within a developmental context.

14 **2. Common ground amid paradigmatic diversity**

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

21 In the first paradigm – artificial grammar learning (AGL; Reber 1967) –
 22 participants are typically instructed to memorize or observe exemplars
 23 presented during a training phase. Often, these exemplars are visual letter-
 24 strings (e.g., PTTVPS) generated from an artificial finite-state grammar,
 25 but they can in principle be composed of any distinctive set of stimulus
 26 tokens varying along a perceptual dimension (e.g., auditory nonwords,
 27 musical tones, shapes) that are arranged in sequence according to the
 28 grammar (see Figure 1). Importantly, participants are not apprised of the
 29 existence of any underlying regularities until the test phase, when they are
 30 informed that stimulus strings follow a set of rules specifying the particular
 31 orderings among constituents. Without being told however the precise
 32 nature of these rules, participants are then asked to classify additional
 33 strings as either “grammatical” or “ungrammatical,” relying upon intuitions
 34 or impressions of familiarity to guide their judgments. Participants typically
 35 achieve classification levels of above-chance performance on the task, even
 36 when test items comprise grammatical exemplars that were never directly
 37 encountered in training (i.e., requiring generalizations of the grammar to
 38 new strings) and despite being unable to provide verbal reports of actual
 39 patterns or rules. (Participants usually claim that they were merely “guess-
 40 ing.”) These performances have been construed as evidence for participants'

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17



18 *Figure 1.* An illustration of an artificial finite-state grammar adapted from Reber
19 and Lewis (1977). Strings are generated by starting at the leftmost node
20 and following possible paths marked by the arrows to other nodes. The
21 succession of letters associated with the arrows encountered along the
22 traced path corresponds to a grammatical string sequence. For example,
23 following the arrow from node 1 to 3, the arcing arrow back to 3, and
24 then the respective arrows to nodes 5, 4 and 6 produces the letter string
25 PTTVPS.
26

27 incidental encoding of the regularities of the grammar during training and
28 their manifestation of this knowledge through meeting the task demands
29 during test; and as alluded to earlier, are well-suited to statistical-based
30 accounts (though not always without dispute) of the computational pro-
31 cesses that mediate successful performance. Thus, participants may evince
32 knowledge for complex, statistical relationships, even in the absence of
33 reported awareness for any underlying structure and without direct inten-
34 tions to discover such regularities.

35 Although artificial grammar learning remains a fruitful paradigm within
36 both implicit learning and statistical learning work, few studies have been
37 conducted with children, especially as standard grammaticality judgments
38 require metacognitive skills not present very early in development. There
39 are a few informative exceptions, though, for work with older children
40 ages nine to eleven (Don, Schellenberg, Reber, DiGirolamo, and Wang

1 2003; Gebauer and Mackintosh 2007; van den Bos 2007), and a couple
2 studies reported in the standard statistical learning literature, with children
3 of six to seven years (Saffran 2001, 2002). Furthermore, similar method-
4 ological principles can be seen as informing the design and interpretation
5 of other additional experiments in very young children. Accordingly, classic
6 statistical learning studies with children and infants have involved familiar-
7 izing participants to carefully manipulated, frequency-balanced subsets
8 of stimuli strings, sequences, or streams from a grammatical ‘corpus’ or
9 miniature language, and then probing for sensitivity to the statistical rela-
10 tions by measuring more naturalistic behavioral responses to statistically
11 consistent and inconsistent test items (more on this below).

12 A paradigm that has been successfully extended to both adults and chil-
13 dren alike is the serial reaction-time paradigm (SRT; Nissen and Bullemer
14 1987; informative studies with children participants include Bremner,
15 Mareschal, Destrebecqz, and Cleeremans 2007; Meulemans, Van der Linden,
16 and Perruchet 1998; Thomas and Nelson 2001; and Thomas et al. 2004). In a
17 prototypical task instantiation, participants are asked to respond as quickly
18 and accurately as possible to trials of presented “targets” (e.g., illuminated
19 lights) occurring at discrete locations on a computer screen, with each
20 location mapping onto a particular response key or button. Unbeknownst
21 to participants, target appearances follow a repeating or probabilistic
22 sequence of locations. After many trials, participants become increasingly
23 adept in anticipating and responding swiftly to the targets. When there is
24 a disruption however to the predictive sequence, either through “noisy,”
25 interspersed sequence-breaks or a continuous block of trials consisting of
26 randomly-generated target locations, accuracy performance decreases and
27 response latencies increase; when target locations conform again to the
28 training sequence, participants’ reaction time (RT) performances dramati-
29 cally “recover” (e.g., Schvaneveldt and Gómez 1998; Thomas and Nelson
30 2001). Because of the indirectness of the instructions and the task demands
31 for speeded responses that discourage explicit reflection/strategizing, SRT
32 work has yielded convincing demonstrations for participants’ sensitivity
33 to violations of sequential structure and incidental learning for sequence-
34 embedded patterns.

35 For the youngest of subjects, however, other methods are necessary to
36 assess incidental sequence knowledge. In the implicit learning literature,
37 the visual expectation paradigm (VExP; Haith, Hazan, and Goodman
38 1988; Haith, Wentworth, and Canfield 1993) has been used with infants
39 as young as two months to investigate their formation of expectations for
40 upcoming visual events comprising predictable sequential patterns. A video

1 monitor displays pictures for brief durations, separated by interstimulus
2 intervals (without intervening visual input), and projected in distinct loca-
3 tions relative to the center of the infant's visual field (i.e., left versus right,
4 up-down, left-center-right or within a triadic-pivot arrangement); the location
5 and/or timing of visual events furthermore accord with either a predictive
6 or randomly ordered series. There are generally three dependent variables
7 of main interest: reaction times (RTs) for eye saccades to correctly antici-
8 pated upcoming stimulus locations in the predictive sequences (compared
9 against RTs for non-predictive sequences); the frequency of accurate antici-
10 patory (i.e., non-reactive, as opposed to elicited) saccades to target locations
11 comprising the predictive series; and any facilitation effect on RTs (i.e.,
12 shorter latency to shift fixation to a predictive location upon the onset of
13 a visual event). Importantly, infants do shift their visual fixations to the
14 location where a future picture will appear prior to the timing of that
15 picture's actual onset. Thus, it has been possible for researchers to obtain
16 a behavioral index for infants' expectations of visual event sequences
17 through measuring anticipatory RTs (assessed against an appropriate RT
18 baseline for when events unfold in a relatively unpredictable manner).
19 Such work has indicated that infants at a very early age rapidly form ex-
20 pectations based on detecting basic spatiotemporal regularities governing
21 the predictive sequences.

22 Finally, the early infant statistical learning studies (e.g., Aslin et al.
23 1998; Gómez and Gerken 1999; Saffran, Aslin et al. 1996) have employed
24 variants or adaptations of existing infant habituation-dishabituation and
25 preference methods to investigate statistical learning. They have used
26 syllables, nonwords, tones, or shapes for the elements instantiating the
27 statistical relations of their artificial training grammars or sequences. Thus,
28 for example, the word-segmentation study of Aslin et al. used a familiariza-
29 tion method to expose eight-month-old infants to a three-minute continuous
30 speech stream (e.g., *pabikugolatudaropitibudo...*) consisting of four tri-
31 syllabic nonce words concatenated together in random order without
32 immediate word repetitions and with each word occurring with a controlled
33 frequency across the stream. While there were no acoustic cues (pauses,
34 prosodic contours, etc.) marking artificial word boundaries, the edges of
35 the nonce words could be successfully segmented or "extracted out" from
36 the continuous sound sequence by utilizing statistical information govern-
37 ing sequence-element transitions: namely, that word-internal, successive
38 syllable transitions ($P = 1.00$) contain higher conditional probabilities than
39 pairwise syllables straddling word boundaries (for this study, $P = .50$). To
40 assess whether infants were in fact sensitive to such cues for segmenting

1 the stream, testing involved twelve trial presentations of two types of test
2 sequences: repetitions of either single words (e.g., *pabiku*), or repetitions
3 of part-words (e.g., *tudaro*). Infants demonstrated that they could dis-
4 criminate between the two types on the basis of the relevant distributional
5 statistics by orienting reliably longer to the direction of the loudspeaker on
6 trials in which it emanated the words of the artificial language. Capitaliz-
7 ing on the attentional and natural orienting responses of infants, differential
8 looking performance at test thus provided a measure of successful discrimi-
9 nation, and hence statistical sensitivity, in preverbal infants. This work
10 exemplifies a prototypical design structure of many infant statistical learn-
11 ing studies to date, and these in turn have contributed valuably towards
12 our understanding of statistical learning mechanisms.

15 3. Phenomenological boundaries

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

29 Accordingly, within any sequentially distributed input, there are a priori
30 potentially many statistical cues available to the learner: (simple) frequency,
31 co-occurrences, transitional probabilities (and “conditional probabilities,”
32 more generally, which can describe nonadjacent relationships), and higher-
33 order conditionals (e.g., second-order, third-order, . . . nth-order probabili-
34 ties). Researchers have also claimed psychological plausibility for other
35 metrics, such as predictive probability (v. Rescorla 1967; “DeltaP,” Shanks
36 1995) and normative [bi-directional] contingency (see Perruchet and Peere-
37 man 2004). Successful empirical demonstrations of statistical-sequential
38 learning thus do not *necessarily* imply that learning must be defined or
39 accounted for in terms of a particular or only a couple prominent statistical
40 metrics already demonstrated in prior work (e.g., forward and backwards

1 transitional probabilities; Jones and Pashler 2007; Pelucchi, Hay, and
2 Saffran 2009a, 2009b; Perruchet and Desautly 2008).

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

17 For different reasons, other interesting work in the implicit learning
18 literature has been omitted here as well. Despite the rapprochement between
19 various implicit *memory* and statistical learning phenomena, certain work
20 within the former (i.e., pertaining to perceptual/conceptual priming, eye-
21 blink conditioning, complex visual search, motor pursuit tracking) are
22 excluded from our discussion, as they do not principally entail processing
23 of sequentially distributed forms of information. Also not discussed here are
24 phenomena with unclear or tenuous connections to statistical-sequential
25 learning mechanisms, as they do not straightforwardly entail the processing
26 of both statistical and sequential information; i.e., contextual cueing (e.g.,
27 in atypically and typically developing populations: Roodenrys and Dunn
28 2008; Vaidya, Huger, Howard, and Howard Jr 2007), covariation detec-
29 tion (in older children: Fletcher, Maybery, and Bennett 2000; Maybery,
30 Taylor, and O'Brien-Malone 1995), “dynamic systems control” tasks (as
31 part of a cross-methodological investigation: Gebauer and Mackintosh,
32 2007), invariance learning (cf. Lewicki, Czyzewska, and Hoffman 1987),
33 and “probabilistic classification” tasks (cf. Knowlton, Squire, and Gluck
34 1994).

35 Finally, a few caveats are in order for whether the learning of repeating
36 sequences (in which asymmetric series alterations, as in VExP work, is a
37 simple example) resides within the purview of statistical-sequential learn-
38 ing mechanisms. It may be objected, for instance, that a different kind of
39 mechanism could potentially mediate the learning of sequentially arrayed
40 input in fixed and repeating patterns – such as one mechanism subserving

1 simple rote memory effects, and another mechanism ranging over more
2 probabilistically-patterned, continuous input. Individuals certainly can and
3 do engage in explicit memorization and recall of digits, letters, etc., with
4 attention paid to the consecutive ordering of units – and such intentional
5 processes are not the focus of this chapter. But the incidental, largely
6 implicit, and (by many accounts) automatic nature of the learning processes
7 that we review would seem to militate against this explanation, and pose
8 difficulties for the differential application of such dual mechanisms prior to
9 discovering the nature of the sequential input along such lines. And the
10 parallel operation of both types of hypothetical mechanisms is less parsimonious,
11 especially when a powerful (yet simple) mechanism can adeptly
12 handle each. A dual-view hypothesis may be further difficult to reconcile
13 with a larger theoretical perspective in which sequence memory is inextricably
14 tied to sequential processing capabilities (MacDonald and Christiansen
15 2002), as suggested for instance by neural network models widely used in
16 statistical and implicit learning research (e.g., Cleeremans and McClelland
17 1991, Dienes 1992; Keele and Jennings 1992; Mirman, Graf Estes, and
18 Magnuson 2010; Misyak, Christiansen, and Tomblin 2010; Servan-Schreiber,
19 Cleeremans, and McClelland 1988).

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

37 It may still be, though, that some forms of fixed sequential learning,
38 such as learning for lists of pre-individuated serial items, invoke different
39 encoding and representation strategies – with learning for such fixed
40 (typically “singular” or non-repeating, as well as truncated in length)

1 sequences relying more on sensitivity to ordinal than associative informa-
 2 tion among series-internal elements. Evidence has been reviewed suggest-
 3 ing that humans and non-human primates may track the ordinality of
 4 relations among serial elements when presented in such a manner (see
 5 Conway and Christiansen 2001). Yet sequential statistical learning in an
 6 artificial grammar task also shows modality-specific effects paralleling
 7 known auditory-visual effects of recency and primacy in serial recall (cf.
 8 Conway and Christiansen 2009; see also discussion in Conway and Pisoni
 9 2008). So even if learning of fixed sequences in these circumscribed con-
 10 texts recruit different encoding and representation strategies from learning
 11 non-fixed sequences, they may still point to similar constraints or principles
 12 operative in both. For example, linguistic combinatorial structure is both
 13 probabilistic and deterministic. Fixed sequences composing words may be
 14 initially identified on the basis of distributional information and later com-
 15 prise the units (or “chunks”) for the fixed sequences constituting idioms
 16 and stock phrases, as well as the non-fixed sequences characterizing novel
 17 sentences (see McCauley and Christiansen, 2011, for a possible model). And
 18 perhaps this is so, *mutatis mutandis*, for admixtures of fixed, deterministic,
 19 and probabilistic micro- and macro-structures in other developmental
 20 domains (e.g., visual scene processing and object-parts/object-based recog-
 21 nition where dimensional features may be perfectly or variably correlated).

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

36 It should also be noted that simple, repeating sequences comprise only
 37 part of the relevant literatures we review, and that sequences that are
 38 probabilistic in nature have also been studied. Thus, we turn in the next
 39 section to an overview of statistical-sequential learning in development,
 40 beginning with VExP findings and unambiguous repeating sequences, and

1 bridging over to work on context-dependent sequences, probabilistic de-
 2 pendencies, and other statistical structures.

4 4. A sketch of the learning landscape

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

14 4.1. Learning fixed, continuous sequences

16 4.1.1. Asymmetric or simple, repeating sequences

18 As early as two months, infants in VExP studies show evidence for form-
 19 ing expectations of upcoming stimulus locations from symmetrically alter-
 20 nating series (i.e., in Left-Right, or Right-Left patterns) (Wentworth and
 21 Haith 1992). By three months, infants also exhibit faster and more
 22 frequent anticipatory saccades to asymmetric 2/1 repeating series (e.g.,
 23 *L-L-R*) and predictive 3/1 (*L-L-L-R*, or *R-R-R-L*) patterns (Canfield and
 24 Haith 1991). Older infants (by about eight or twelve months) tested in the
 25 VExP paradigm also display some anticipatory gaze behavior for upcom-
 26 ing visual targets whose locations form a predictable triadic-pivot series
 27 (i.e., *ABCBABC*... with the “*B*” location as the series’ pivot point among
 28 the three locations) (Reznick, Chawarska, and Betts 2000).

29 It has been noted that use of the term “expectation” need not imply
 30 explicit recognition of patterns (Reznick et al. 2000), and the VExP para-
 31 digm itself should probably be best considered along the lines of a proced-
 32 ural task (cf. Nelson 1995), in which skilled performance commonly reflects
 33 incidental learning and the coordination of complex sequential input with
 34 motor responses. Regarding sequence-specific knowledge, while between-
 35 subjects VExP analyses provide evidence for global probability matching
 36 (e.g., greater eye shifts back to the more frequent, or “home-side,” loca-
 37 tion in 3/1 than 2/1 asymmetric conditions), within-subjects analyses also
 38 indicate sensitivity to spatiotemporal regularities inhering over and above
 39 simply the proportion of picture appearances to a given side. For example,
 40 in the 3/1 condition, infants are more likely to appropriately shift to the

1 less frequent “target” side after the third “home-side” event appearance
 2 than after the second or first “home-side” event. Given the specific experi-
 3 mental design of a study, encoding may also extend beyond location to
 4 accommodate specific (visual) event content, inter-event contingencies,
 5 and temporal flow rate of the stimuli sequences (Adler, Haith, Arehart,
 6 and Lanthier 2008; Wentworth and Haith 1992; Wentworth, Haith, and
 7 Hood 2002).

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

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

36 4.1.2. Context-dependent sequences

38 In context-dependent (i.e., *n*th-order) progressions, the occurrence of a
 39 sequence-element depends upon the context associated with its preceding
 40 element. For example, given a 1-2-1-3-1-2... sequence, being able to antici-
 pate the location after a “1” requires knowing the temporal context of

1 whether a “2” or “3” preceded the “1.” Sequences with context-dependent
2 transitions can be either deterministic (repeating) or probabilistic.

3 While four-month-olds have shown above-chance accuracy in anticipa-
4 tory saccades for unambiguous, simple repeating sequences (e.g., *1-2-3*)
5 mapping onto a triangular configuration of spatial locations, they seem
6 unable to perform above chance for the context-dependent transitions of
7 more ambiguous sequences (e.g., *1-2-1-3*) given roughly comparable expo-
8 sure time (i.e., 27 and 32 single-location trials, respectively, which corre-
9 spond to 9 and 8 sequence repetitions each) (Clohessy, Posner, and Rothbart
10 2001). Clohessy et al. further assessed performance for these same sequences
11 in ten- and eighteen-month-olds. Ten-month-olds did not show anticipatory
12 learning for the context-dependent sequence (even when exposure was
13 doubled, i.e., expanded to 2 sessions), but eighteen-month-olds could show
14 anticipations for both types of sequences.

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

24 Across two studies in children of four, six- to seven, and ten- to eleven-
25 years, performance on a deterministic ten-element *SRT* task (e.g., *1-3-2-4-*
26 *1-2-3-4-2-4*) reported similar learning magnitudes across age groups
27 (Meulemans et al. 1998; Thomas and Nelson 2001). These are consistent
28 with standard expectations in the literature that, although general process-
29 ing speed improves with age, a sequence learning effect nonetheless remains
30 comparable across age groups. However, Thomas and Nelson reported
31 that the “number of anticipatory button presses to correct locations show[ed]
32 evidence of developmental change” (2001: 375). Nonetheless, they refrained
33 from forming conclusions about developmental changes in implicit learning
34 as such, under concerns that this measure might be construed as tapping
35 more “explicit” learning.

36 However, a functional magnetic resonance imaging (fMRI) study con-
37 ducted later by Thomas et al. (2004), again with seven- and eleven- year-
38 olds, reported evidence of differential neural recruitment by children and
39 adults on a *SRT* task, the latter of whom performed significantly better as
40 well on the learning index. Particularly, there were age-related differences in
neural activity for premotor cortex, putamen, hippocampus, inferotemporal

1 cortex, and parietal cortex. The sharpest age discrepancy was in greater
 2 recruitment of fronto-striatal circuitry and hippocampal activation in
 3 adults, although such activity in these regions was not significantly corre-
 4 lated with the magnitude of the learning effect.

5 6 4.2. Tracking probabilistic dependencies

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

9 10 4.2.1. *Adjacent dependencies*

11
12 The event-related brain responses of sleeping neonates indicate that the
 13 ability to use statistical cues (such as co-occurrence frequencies) to dis-
 14 criminate lexical boundaries among adjacent phonemes in a continuous
 15 artificial speech stream is present as early as one-half to two days after
 16 birth (Teinonen, Fellman, Näätänen, Alku, and Huotilainen 2009). Using
 17 behavioral measures, infants by two months are also sensitive to the co-
 18 occurrence frequencies obtaining across a continuous stream of geometric
 19 shapes with reliable shape-pairings (bigrams) (Kirkham, Slemmer, and
 20 Johnson 2002).

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

35 Skipping ahead to about six years of age, earlier word segmentation
 36 and AGL work in young child learners of six- to nine- years would suggest
 37 that statistical-sequential learning effects may be age-invariant (e.g., Don
 38 et al. 2003; Saffran et al. 1997). However, a later study by Saffran (2001)
 39 provided evidence for clear age differences. Six- to nine-year old children
 40 and adults were both trained and tested on an artificial grammar containing
 predictive dependencies. While all participants demonstrated significant

1 learning, the adults consistently outperformed the children, prompting
 2 Saffran to write that the results “suggest that children may possess a limited
 3 ability to acquire syntactic knowledge via statistical information. While
 4 their performance was not as strong as the adults’, the children did acquire
 5 rudimentary aspects of the phrase structure of the language” (508).
 6 Another study by van den Bos (2007) compared the learning performance
 7 of ten- and eleven-year old children with that of adults’ on an artificial
 8 grammar task, varying the usefulness of the underlying structure with
 9 respect to a cover task. Although the qualitative learning effect was the
 10 same, adults in the study acquired quantitatively greater knowledge of
 11 second-order dependencies than did the children. And finally, a study
 12 (Arciuli and Simpson, *in press*) of visual statistical learning using a triplet
 13 segmentation task was conducted with children ranging from five to twelve
 14 years. Quantitative improvements in discrimination performance for legal
 15 triplets on a forced-choice posttest were documented with increasing age.

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

25 26 4.2.2. *Nonadjacent dependencies*

27
28 Studies in the implicit learning literature tend not to investigate “non-
 29 adjacent dependencies,” defined as relationships where another element(s)
 30 intervenes between two dependent elements, as their primary aim. More
 31 customarily, they may investigate learning for higher-order conditionals
 32 under the assumption that the concomitant non-local dependencies em-
 33 bedded in the stimuli sequences are learnt through the chunking of adja-
 34 cencies (though see e.g., Kuhn and Dienes 2005, Pacton and Perruchet
 35 2008, and Remillard 2008, for findings in the implicit learning literature
 36 with adults). Thus, what is currently known about incidental learning of
 37 nonadjacencies over development is mostly limited to studies conducted
 38 within the canonical statistical learning literature.

39 In infants (as in adults), it has been demonstrated that relatively high
 40 variability in the set size from which an “intervening” middle element of

1 a string is drawn facilitates learning of the nonadjacent relationship
 2 between the two specific, flanking elements (Gómez 2002). In other words,
 3 when exposed to artificial grammar strings of the form aXd and bXe , indi-
 4 viduals display sensitivity to the nonadjacent dependency-pairs (i.e., the
 5 a_d and b_e relations) when the elements composing the X are drawn
 6 from a large set distributed across many exemplars (e.g., when $|X| = 18$
 7 or 24). Performance is hindered, however, when the variability of the set
 8 size for the X is intermediate (e.g., $|X| = 12$) or low (e.g., $|X| = 2$).

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

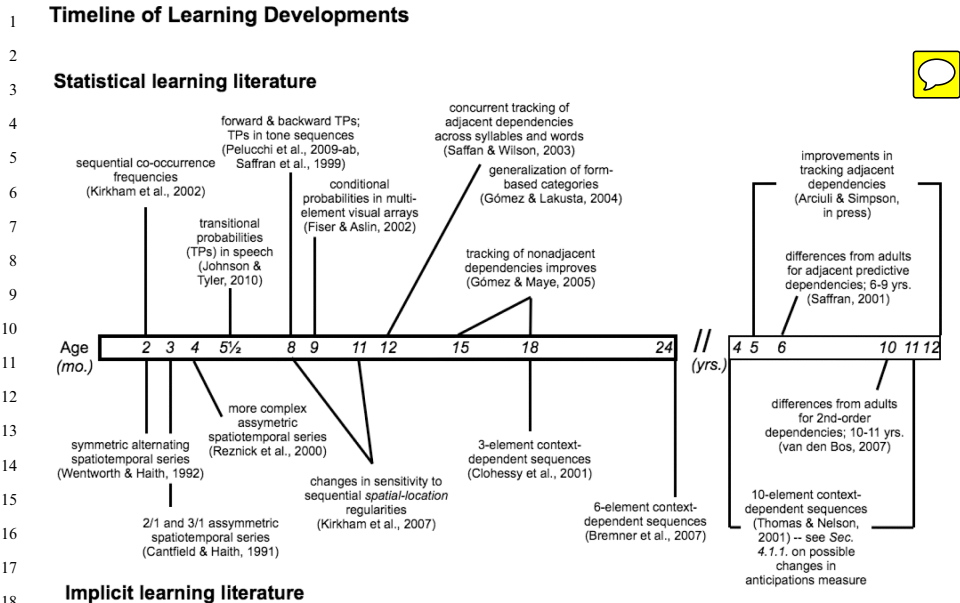
31
 32 5. This interpretation is informed by the influential Hunter-Ames model (Hunter
 33 and Ames 1988) in which greater stimulus complexity or partial encoding by
 34 infants is predicted to elicit familiarity preferences (longer looking/listening
 35 times for test stimuli that are consistent with the training exemplars) rather
 36 than the opposite pattern (i.e., a preference for attending longer to the novel,
 37 or inconsistent, test items). While the generality for interpreting preference
 38 patterns has not been definitively established (and is thus open to dispute),
 39 the observation that twelve-month olds were unable to demonstrate learning
 40 for the nonadjacent grammar (under the same experimental conditions) as
 the older infants further underscores Gómez and Maye's conclusion.

1 5. Developmental changes

2
3 Age differences in the magnitude of learning effects were observed across
4 different age groups of young children in comparison to adults when learn-
5 ing deterministic and probabilistic sequences in SRT tasks. Such quantita-
6 tive differences between child learners and adults were also observed in
7 AGL paradigms. In light of the assumption that children should be natu-
8 rally better statistical learners than adults given sensitive period effects
9 observed in early language acquisition (J. S. Johnson and Newport 1989;
10 Newport 1990), these findings may be counterintuitive and surprising.

11 Differences between learning adjacencies and nonadjacencies within a
12 developmental context could potentially be an extension of quantitative
13 performance-level differences, in that, as structural complexity purportedly
14 increases, we see later ages of proficiency documented. Complexity of arti-
15 ficial grammars studied in the literature thus far seem to have some corre-
16 spondence with dependency length – and measures, such as topological
17 entropy, that correlate with length – over and above simply the number
18 of associations or predictability of the grammar (van den Bos and Poletiek
19 2008). However, the simplified structures employed to date do not exhaust
20 the full range of potential structures amenable to statistical tracking (e.g.,
21 such as embedded long-distance dependencies and cross serial-dependencies
22 evidenced in natural language). Attempts to identify facilitative contexts
23 for acquiring remote dependencies is still in its early stages. So “depend-
24 ency length” may be more of a useful starting-point, rather than a
25 conclusive identification of the main source for statistical-sequential learn-
26 ing “complexity.” Neural network modeling of sequential processing in
27 natural language suggest that complexity (as corresponding to measures
28 reflecting human processing difficulty, such as protracted on-line RTs)
29 is not reducible to dependency length and reflects interactions between
30 experiential variations and constraints intrinsic to the architecture of the
31 learning mechanism (Christiansen and MacDonald 2009; MacDonald
32 and Christiansen 2002). Nonetheless, with this initial provisional defini-
33 tion, the findings reviewed herein are consistent with an interpretation of
34 improved statistical-sequential learning performance with age for incident-
35 ally detecting increasingly “complex” structures, including but not confined
36 to explaining the behavioral emergence of nonadjacency tracking (see
37 Figure 2).

38 With certain stipulations, deterministic nonadjacencies in statistical
39 learning bear surface resemblance to the second-order context-dependent
40 sequences in the implicit learning literature. That is to say, a context-



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

33 dependent sequence with 2nd order relations (e.g., 1-2-1-3-1-2-13... ;
 34 which is the precise structure studied by Clohessy et al. 2001, noted earlier)
 35 parallels in its embedded relations the nonadjacency grammar-strings of
 36 type 2_3 or 3_2 as investigated in infants by Gómez and colleagues (see
 37 Sec. 4.2.2.) (n.b., set-size $\bar{X} = 1$, which is a “zero variability,” potentially
 38 learning-conducive condition; Onnis, Christiansen, Chater, and Gómez
 39 2003; Onnis, Monaghan, Christiansen, and Chater 2004). Interestingly,
 40 robust learning for these kinds of sequential regularities appears around

1 18 months in *both literatures*. Indeed, as mentioned earlier, much implicit
 2 learning work generally does not discriminate whether second-order context-
 3 dependent sequence relations are represented in the same manner as non-
 4 adjacencies of the kind studied in the statistical learning literature (although
 5 theoretically both forms of learned representations would be consistent
 6 with statistical-sequential learning mechanisms).

8 5.1. Transient cognitive constraints

9
 10 While various cognitive constraints might be operative in accounting for
 11 patterns of developmental performance differences, here we concentrate
 12 on one promising factor that has already received some attention in the
 13 literature and that has particular utility for explaining the emergence of
 14 nonadjacency skills. The provisional hypothesis is that infants begin with
 15 limitations in the information that they can process in time, and are thus
 16 restricted in the amount of elements that can be effectively related with
 17 one another within the distance of this processing window. Following
 18 Santelmann and Jusczyk (1998) and others, the number of intervening
 19 elements (or syllable/word constituents) between dependencies comprised
 20 of similar units is used as a working definition of “distance” (precluding
 21 for now the issue of precise temporal duration). As the processing space
 22 expands during later development, this allows the infant to efficiently
 23 exploit and integrate more of an element’s preceding context to discover
 24 appropriate nonadjacent relations that would otherwise be obscured.

25 Indeed, a narrow processing window may act as an initial “filter” to
 26 constrain the problem space of potential mappings (Newport 1988, 1990),
 27 thus focusing the infant’s attention on more basic, local dependencies
 28 that can be later applied over longer distances, when the temporal window
 29 grows. Another related possibility is that a narrow window may act as an
 30 initial “amplifier” for detecting the covariation of input elements, because
 31 smaller sampling of a distribution increases the likelihood of observing
 32 correlations that are more extreme in magnitude than the true associations
 33 (Kareev 1995; Kareev, Lieberman, and Lev 1997). Given the structure of
 34 language, such memory-based constraints (in contrast to those of adults)
 35 might paradoxically contribute toward superior performance in language
 36 acquisition for child learners (Newport’s “less is more” hypothesis). Com-
 37 putationally, “less is more” has a parallel in one of the two methods used
 38 in Elman’s (1993) “starting small” simulations, in which Elman mani-
 39 pulated the resetting of a simple recurrent network’s context units in order
 40 to simulate the child’s initially reduced and then growing window for

1 relating dependencies. Such a procedure enabled the network, thus “handi-
 2 capped,” to learn a complex language corpus that it had previously failed
 3 to master without recourse to such developmental limitations (or without
 4 having received incrementally staged input that externally mimicked such
 5 constraints). Conway, Ellefson, and Christiansen (2003) further investigated
 6 the effects of “starting small” in artificial grammar learning experiments
 7 with adults. In support of a starting small hypothesis, they documented a
 8 learning advantage for participants trained with incrementally staged input
 9 on complex visual grammars.

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

22 Empirically, the developmental timing of nonadjacency-related skills
 23 at eighteen months in statistical learning paradigms parallels emerging
 24 sensitivity by this same age in natural language learning for sensitivity to
 25 morphosyntactic relationships obtaining over one to three intervening
 26 syllables (Santelmann and Jusczyk 1998). This is consistent as well with
 27 Gómez and Maye’s (2005) suggested interpretation of their developmental
 28 results (described in Sec 4.2.2.). At the neural level, synaptic pruning
 29 might also be a mechanism for such gains in memory performance (and
 30 can also be explored computationally in neural networks via “selective
 31 pruning” of nodes and connection weights) (cf. Quartz and Sejnowski
 32 1997). Thus, the hypothesis of a limited temporal processing window has
 33 current empirical and theoretical support (see also Goldstein et al. 2010),
 34 and can be investigated in greater detail from various aspects within a
 35 computational perspective.

36 On the latter note, however, Rohde and Plaut’s (2003) neural network
 37 simulations (“less is less”) failed to replicate Elman’s findings, thus question-
 38 ing computational support for the hypothesis. However, our explanation
 39 here dodges these particular concerns. That is to say, although intriguing
 40 and unresolved, the issue of whether early limitations in processing space

1 would be *beneficial per se* for the infant (under certain circumstances,
2 perhaps) is not decisive to the validity of whether performance differences
3 in statistical-sequential learning, especially for the emergence of non-
4 adjacency tracking, can be traced to such a window. What is important
5 to our exposition, then, is the idea that such transient cognitive limitations
6 do appear to exist and that the notion of a temporal processing window,
7 as expounded above, may offer a powerful framework for organizing the
8 existing developmental data to date.

9 10 5.2. Changes in underlying neural structures

11
12 Work reviewed for the statistical-sequential learning of context-dependent
13 sequences suggested differential recruitment of cortical and subcortical struc-
14 tures between children and adults, with the latter showing greater hippo-
15 campal activation (although this was not associated with the magnitude
16 of the learning effect *per se*). Statistical-sequential learning, as a form of
17 procedural learning, is likely to involve the participation of the basal
18 ganglia through cortico-striatal circuits (based on supportive molecular
19 and neuropsychological evidence, as well as theoretical views; Ackermann
20 2008; Christiansen, Kelly, Shillcock, and Greenfield 2010; Conway and
21 Pisoni 2008; Lieberman 2002); and continually receives much attention in
22 implicit learning accounts, especially in relation to putative performance
23 “dissociations” among impaired populations. It has also been suggested
24 that the basal ganglia may play a role in speech and language develop-
25 ment via feedback-driven vocal learning, such as in the socially guided
26 statistical learning of phonological patterns through contingent interac-
27 tions between caregivers and prelinguistic, vocalizing infants (Goldstein
28 and Schwade 2008).

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

39 Of further interest, the age-related differences in hippocampal activation
40 patterns might play a role in explaining conflicting findings of the effect

1 of sleep on children's statistical-sequential learning. Within the implicit
2 learning literature, Fischer, Wilhelm and Born (2007) had reported a
3 sleep-dependent deterioration effect on children's learning of second-order
4 contingencies from an SRT task. They had assessed the level of learning
5 retention by seven- to eleven-year-olds when retested on the SRT sequence
6 after either a period of sleep or an equal interval of wakefulness. The
7 difference in average reaction times to grammatical versus ungrammatical
8 SRT trials in the test blocks was the study's dependent measure of learn-
9 ing. Twenty- to thirty-year olds adults exhibited an improvement in this
10 RT learning measure when retested following sleep, but a decrement in
11 learning when retested in the wake condition. Compared to adults, how-
12 ever, the magnitude of learning decreased in the group of children follow-
13 ing sleep, while remaining unchanged at retest in the wake condition.

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

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

31 Moreover, for different learning tasks, a competitive interference between
32 striatal and hippocampal systems has been shown (Packard and Knowlton
33 2002; Schroeder, Wingard, and Packard 2002; Poldrack et al. 2001). In this
34 framework, opposite effects of sleep on implicit learning in children and
35 adults might reflect that sleep in children leads to a preferential strengthen-
36 ing of hippocampal aspects of the memory representation, whereas sleep in
37 adults strengthens caudate involvement. (224)

38 In support of this hypothesis, Fischer et al. note that whereas the children
39 and adults in their study did not differ in amounts of rapid-eye-movement
40 (REM) sleep, children's sleep was characterized by a greater amount of

1 slow-wave sleep (SWS). In turn, SWS is associated with hippocampus-
 2 dependent memory consolidation (see the review by Marshall and Born
 3 2007). Greater caudate involvement may lead to enhanced procedural learn-
 4 ing, thus explaining adults' gains in SRT learning performances. However,
 5 strengthening hippocampal aspects of memories may not impact the
 6 implicit/indirect indices of learning tapped by the SRT, thus explaining
 7 the lack of gains in children after the sleep interval. (Recall again that in
 8 the Thomas et al. (2004) study (reported in Sec. 4.1.2.), adults' learning
 9 showed evidence of greater reliance on hippocampal recruitment than
 10 children's during on-line task performance, but this difference was not
 11 significantly linked to the SRT learning measure).

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

30

31 5.3. Early perceptual development

32

33 Because statistical-sequential learning is closely mediated by perceptual
 34 features of the input and by modality constraints (Conway and Christiansen
 35 2005, 2006, 2009; Conway et al. 2007), part of a thorough picture may
 36 likely include then the manner in which perceptual systems develop in
 37 response to early environmental input and are recruited through experience.
 38 Early perceptual learning, at least for certain gross-level acoustic or pro-
 39 sodic patterns, begins *in utero* during the last trimester (e.g., DeCasper
 40 and Spence 1986; see review of Gómez and Gerken 2000, Box 1, as

1 well as computational simulations of scaffolding from prenatal “filtered”
2 stimuli in Christiansen, Dale, and Reali 2010), suggesting a developmental
3 trajectory that may have a substantial prenatal initial progression or
4 foundation.

5 Postnatally, the evidence reviewed herein indicates that visual and spatio-
6 temporal sequential learning for co-occurrence frequencies are present at
7 two months (e.g., as in VExP findings and the statistical learning study by
8 Kirkham et al. 2002) and older (e.g., by nine months; Emberson, Misyak,
9 Schwade, Christiansen, and Goldstein 2008), with further statistical learning
10 of regularities within visual arrays documented at nine months (Fiser and
11 Aslin, 2002). In the auditory domain, natural language acquisition skills
12 (some of which likely recruit upon statistical-sequential learning mecha-
13 nisms) have also been investigated in very early infancy. Rudimentary,
14 auditory sequential learning abilities appear present as early as one-half
15 to two days after birth (Teinonen et al. 2009), and evidence of learning
16 transitional probabilities embedded within linguistic stimuli are seen by
17 five-and-a-half months (E. K. Johnson and Tyler 2010). However, audi-
18 tory statistical learning for non-speech stimuli such as tones has seemingly
19 not been documented in infants younger than eight months (Saffran et al.
20 1999). Systematic comparative work (using comparable procedures and
21 stimuli) does not currently exist for deriving firm conclusions about potential
22 modality-driven performance patterns and differences in infancy (Misyak,
23 Emberson, Schwade, Christiansen, and Goldstein 2009); although one study
24 in children indicates somewhat better learning of predictive dependencies
25 in the auditory versus the visual modality (Saffran 2002). Thus, the time-
26 line gaps in these cases probably reflect the fact that studies have typically
27 not been motivated by comparative developmental trends – at least not
28 with regard to modality or performance-level differences – rather than
29 reflecting absences of ability per se.

30 It is further unknown how learning in infants and children may differ
31 among specific dimension-modality pairings (i.e., comparisons for learning
32 between visual and auditory stimuli, occurring in arrayed or sequential
33 format, when regularities are also encoded along the dimension of location
34 or variable timings). Kirkham, Slemmer, Richardson and Johnson (2007),
35 however, have some work suggesting that sequential spatial-location regu-
36 larities (of more complex form than those studied in the VExP task) may
37 be more difficult for infants than other forms of sequence learning tasks
38 and that proficiency may manifest later in perceptual development. We
39 are also far from systematic investigations in infancy/childhood of feature
40 dimension pairings. For instance, color and shape were always perfectly

1 correlated in Fiser and Aslin's (2002) and Kirkham et al.'s (2002) visual
2 statistical/sequential learning studies, thus preempting developmental con-
3 siderations for the kind of phenomena investigated by Turk-Browne and
4 colleagues (Turk-Browne et al. 2008; Turk-Browne and Scholl 2009) in
5 adults regarding "bound object representations" and spatiotemporal gener-
6 alization abilities.

7 In contrast to fairly-developed auditory abilities at birth (Lasky and
8 Williams 2005; Saffran, Werker, and Werner 2006), the visual system under-
9 goes more dramatic changes during the first year. As newborns, preferential
10 looking visual acuity estimates are at approximately 1 cycles per degree
11 (cpd; equivalent to 20/600 Snellen), developing to 3 cpd (20/200 Snellen)
12 at 3 months, and reaching about 12 cpd (20/50 Snellen) by the end of the
13 first year (Birch, Gwiazda, Bauer Jr, Naegele, and Held 1983; Courage
14 and Adams 1990; Dobson and Teller 1978). Such early, transient limita-
15 tions in the detail of infants' visual fields may thus narrow their perceptual
16 focus to close-range visual stimuli and in turn support a more sequentially
17 constrained format upon visual images they perceive – which could in
18 turn necessarily favor visual statistical learning of temporally-distributed
19 sequences over that of spatially-arrayed sequences in early processing.
20 There are differential effects of temporal and spatial formats for auditory
21 and visual statistical learning in adults, with visual-temporal conditions
22 eliciting poorest performances (Conway and Christiansen 2009); but it has
23 not been established whether and how such performance patterns might be
24 shaped by, and/or possibly depart from, early perceptual experiences/biases.
25 One intriguing possibility, therefore, is that the kind of early processing
26 constraint alluded above, in tandem with an abundance of such early
27 visual experiences during the first months, may temporarily place young
28 infants' visual statistical learning of sequences above or on par with
29 analogous learning for auditory sequences. Another complimentary idea
30 is that the structure of prelinguistic social interaction with caregivers
31 shapes infant attention in ways that facilitate specific forms of statistical
32 learning (Goldstein et al. 2010). Such comparative learning among modal-
33 ities is an ongoing matter of investigation, and if the former hypotheses
34 bear out, it would form a surprising counterpoint to auditory superiority
35 performance patterns observed in adults (Conway and Christiansen 2005,
36 2009; Saffran 2002).

37 Beyond the development of sequential learning through particular sensory
38 experiences (and by implication, for distributed modality-constrained
39 subsystems of statistical learning; see *Note 2*), there may further be a

40

1 possible role for general principles in early neural development that facili-
 2 tate and maintain so-called “entrenched” perceptual “discrimination”
 3 abilities (cf. Scott, Pascalis, and Nelson 2007). That is, it has been pro-
 4 posed that such mechanisms may be broader (more domain-general) than
 5 traditionally supposed. For instance, the language-specific “narrowing”
 6 characteristic of infants’ later babbling as a product of increasing exposure
 7 to the ambient language, might also be driven or facilitated by experiences
 8 that are supralinguistic in some form; and more specifically, with respect
 9 to incorporated phonological patterns and articulatory/acoustic features,
 10 can be shaped by contingent parental feedback (Goldstein, King, and
 11 West 2003; Goldstein and West 1999) and “socially guided statistical learn-
 12 ing” (Goldstein and Schwade 2008). Given, though, the evidence for
 13 gradiency effects in many “discrete” categorization performance tasks (e.g.,
 14 Dale, Kehoe, and Spivey 2007; McMurray, Tanenhaus, Aslin, and Spivey
 15 2003; see also Spivey 2007), an account encompassing these principles
 16 may also be heavily context-sensitive and ultimately entail a probabilistic
 17 cue-weighting explanation/mechanism that need not be *perceptually specific*
 18 in its *explanatory range* or extension. Assuming the requisite experiences
 19 for shaping such behavioral response patterns, these may in turn provide
 20 a “representational platform” or weighted biases over which related input
 21 features or cues for statistical-sequential learning mechanisms may be
 22 integrated. Furthermore, regarding so-called perceptual “enhancement”
 23 processes, it is likely that *the statistical distribution of the featural input*
 24 itself may play a large role in such phenomena (cf. Maye, Weiss, and Aslin
 25 2008; Maye, Werker, and Gerken 2002).

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

33 34 35 **6. Future developmental strides for the merger of statistical and** 36 **implicit learning work**

37
38 A synthesis of findings from across the implicit and statistical learning
 39 literatures suggests that these two fields may have much to synergistically
 40 offer one another. It also indicates that their convergence may be especially

1 fruitful for exploring issues related to infant/childhood cognitive con-
 2 straints, underlying neural mechanisms, and early perceptual development.
 3 Furthermore, despite the orthodox assumption of age invariance, it appears
 4 that the possibility of developmental changes across studies should merit
 5 much stronger consideration than they presently have to date. Abandoning
 6 the *presumption* of developmental invariance might also provide the impetus
 7 for much-needed longitudinal and cross-sectional designs. There are con-
 8 spicuous age gaps that reflect the nature of existing work in the area.
 9 Much work concentrates on early infancy, but does not clearly connect
 10 learning continuously across childhood. By going beyond documenting
 11 the age of successful learning for different statistical-sequential skills
 12 towards providing more detailed developmental trajectories, the projected
 13 merger of research findings across implicit and statistical learning para-
 14 digms will not only become truly developmental, but may perhaps flourish
 15 even more prominently past its early formal youth.

17 References

- 18
 19
 20 Ackermann, Hermann
 21 2008 Cerebellar contributions to speech production and speech per-
 22 ception: Psycholinguistic and neurobiological perspectives. *Trends*
 23 *in Neurosciences*, 31, 265–272.
- 24 Adler, Scott A., Marshall M. Haith, Denise M. Arehart, and Elizabeth C. Lanthier
 25 2008 Infants' visual expectations and the processing of time. *Journal*
 26 *of Cognition and Development*, 9, 1–25.
- 27 Arciuli, Joanne, and Ian C. Simpson
 28 **in press** Statistical learning in typically developing children: The role of
 29 age and speed of stimulus presentation. *Developmental Science*.
 30 doi: 10.1111/j.1467-7687.2009.00937.x
- 31 Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport
 32 1998 Computation of conditional probability statistics by 8-month-old
 33 infants. *Psychological Science*, 9, 321–324.
- 34 Birch, E. E., J. Gwiazda, J. A. Bauer Jr, J. Naegele, and R. Held
 35 1983 Visual acuity and its meridional variations in children aged 7–60
 36 months. *Vision Research*, 23, 1019–1024.
- 37 Bloomfield, Leonard
 38 1933 *Language*. Chicago: University of Chicago Press.
- 39 Bremner, Andrew J., Denis Mareschal, Arnaud Destrebecqz, and Axel Cleeremans
 40 2007 Cognitive control of sequential knowledge in 2-year-olds: Evi-
 41 dence from an incidental sequence-learning and -generation
 42 task. *Psychological Science*, 18, 261–266.

- 1 Canfield, Richard L., and Marshall M. Haith
 2 1991 Young infants' visual expectations for symmetric and asymmetric
 3 stimulus sequences. *Developmental Psychology*, 27, 198–208.
- 4 Canfield, Richard L., Elliott G. Smith, Michael P. Brezsnayak, Kyle L. Snow,
 5 Richard N. Aslin, Marshall M. Haith, Tara S. Wass, and Scott
 6 A. Adler
 7 1997 Information processing through the first year of life: A longitudi-
 8 nal study using the visual expectation paradigm. *Monographs of
 9 the Society for Research in Child Development*, 62, i–160.
- 10 Christiansen, Morten H., Rick Dale, and Florencia Reali
 11 2010 Connectionist explorations of multiple-cue integration in syntax
 12 acquisition. In S. P. Johnson (Ed.), *Neoconstructivism: The new
 13 science of cognitive development* (pp. 87–108). New York: Oxford
 14 University Press.
- 15 Christiansen, Morten H., M. Louise Kelly, Richard C. Shillcock, and Katie
 16 Greenfield
 17 2010 Impaired artificial grammar learning in agrammatism. *Cognition*,
 18 116, 382–393.
- 19 Christiansen, Morten H. and Maryellen C. MacDonald
 20 2009 A usage-based approach to recursion in sentence processing.
 21 *Language Learning*, 59 (Suppl. 1), 126–161.
- 22 Cleeremans, Axel
 23 1993 *Mechanisms of implicit learning: A connectionist model of sequence
 24 processing*: MIT Press.
- 25 Cleeremans, Axel, and James L. McClelland
 26 1991 Learning the structure of event sequences. *Journal of Experimental
 27 Psychology: General*, 120, 235–253.
- 28 Clohessy, Anne B., Michael I. Posner, and Mary K. Rothbart
 29 2001 Development of the functional visual field. *Acta Psychologica*,
 30 106, 51–68.
- 31 Conway, Christopher M., and Morten H. Christiansen
 32 2001 Sequential learning in non-human primates. *Trends in Cognitive
 33 Sciences*, 5, 539–546.
- 34 Conway, Christopher M., and Morten H. Christiansen
 35 2005 Modality-constrained statistical learning of tactile, visual, and
 36 auditory sequences. *Journal of Experimental Psychology: Learning,
 37 Memory, and Cognition*, 31, 24–39.
- 38 Conway, Christopher M., and Morten H. Christiansen
 39 2006 Statistical learning within and between modalities: Pitting abstract
 40 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.

- 1 Conway, Christopher M., Michelle R. Ellefson, and Morten H. Christiansen
 2 2003 When less is less and when less is more: Starting small with
 3 staged input. In Rick Alterman and David Kirsh (Eds.), *Proceed-*
 4 *ings of the 25th Annual Cognitive Science Society* (pp. 270–275).
 Boston, MA: Cognitive Science Society.
- 5 Conway, Christopher M., Robert L. Goldstone, and Morten H. Christiansen
 6 2007 Spatial constraints on visual statistical learning of multi-element
 7 scenes. In D. S. McNamara and J. G. Trafton (Eds.), *Proceed-*
 8 *ings of the 29th Annual Cognitive Science Society* (pp. 185–190).
 9 Austin, TX: Cognitive Science Society.
- 10 Conway, Christopher M., and Pisoni, D. B.
 11 2008 Neurocognitive basis of implicit learning of sequential structure
 12 and its relation to language processing. In Guinevere F. Eden
 13 and D. Lynn Flowers (Eds.), *Annals of the New York Academy*
of Sciences, 1145, 113–131.
- 14 Courage, Mary L., and Russell J. Adams
 15 1990 Visual acuity assessment from birth to three years using the
 16 acuity card procedure: Cross-sectional and longitudinal samples.
 17 *Optometry and Vision Science*, 67, 713–718.
- 18 Creel, Sarah C., Elissa L. Newport, and Richard N. Aslin
 19 2004 Distant melodies: Statistical learning of nonadjacent depen-
 20 dencies in tone sequences. *Journal of Experimental Psychology:*
Learning, Memory, and Cognition, 30, 1119–1130.
- 21 Curtin, Suzanne, Toben H. Mintz, and Morten H. Christiansen
 22 2005 Stress changes the representational landscape: Evidence from
 23 word segmentation. *Cognition*, 96, 233–262.
- 24 Dale, Rick, Caitlin Kehoe, and Michael J. Spivey
 25 2007 Graded motor responses in the time course of categorizing
 26 atypical exemplars. *Memory and Cognition*, 35, 15–28.
- 27 DeCasper, A., and Spence, M.
 28 1986 Prenatal maternal speech influences on newborns' perception of
 29 speech sounds. *Infant Behavior and Development*, 9, 133–150.
- 30 Dienes, Zoltán
 31 1992 Connectionist and memory-array models of artificial grammar
 32 learning. *Cognitive Science*, 16, 41–79.
- 33 Dobson, Velma, and Davida Y. Teller
 34 1978 Visual acuity in human infants: A review and comparison of
 35 behavioral and electrophysiological studies. *Vision Research*, 18,
 1469–1483.
- 36 Dominguez, Melissa, and Robert A. Jacobs
 37 2003 Does visual development aid visual learning? In P. T. Quinlan
 38 (Ed.), *Connectionist models of development: Developmental pro-*
 39 *cesses in real and artificial neural networks* (pp. 257–278). NY:
 Psychology Press.
- 40

- 1 Don, Audrey J., E. Glenn Schellenberg, Arthur S. Reber, Kristen M. DiGirolamo,
 2 and Paul P. Wang
 3 2003 Implicit learning in children and adults with Williams syndrome.
 4 *Developmental Neuropsychology*, 23, 201–225.
- 5 Ellis, Nick C. (Ed.)
 6 1994 *Implicit and explicit learning of languages*. London: Academic
 7 Press.
- 8 Elman, Jeffrey L.
 9 1993 Learning and development in neural networks: The importance
 10 of starting small. *Cognition*, 48, 71–99.
- 11 Emberson, Lauren L., Jennifer B. Misyak, Jennifer A. Schwade, Morten H.
 12 Christiansen, and Michael H. Goldstein
 13 2008 Face-to-face: Visual statistical learning with complex natural
 14 stimuli. *XVIIth International Conference on Infant Studies* (ab-
 15 stracts ed., pp. 259). Vancouver, Canada.
- 16 Fischer, Stefan, Ines Wilhelm, and Jan Born
 17 2007 Developmental differences in sleep's role for implicit off-line
 18 learning: Comparing children with adults. *Journal of Cognitive
 19 Neuroscience*, 19, 214–227.
- 20 Fiser, József, and Richard N. Aslin
 21 2002 Statistical learning of new visual feature combinations by infants.
 22 *Proceedings of the National Academy of Sciences*, 99, 15822–
 23 15826.
- 24 Fletcher, Janet, Murray T. Maybery, and Sarah Bennett
 25 2000 Implicit learning differences: A question of developmental level?
 26 *Journal of Experimental Psychology: Learning, Memory, and
 27 Cognition*, 26, 246–252.
- 28 Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schibotz, and
 29 Alfred Anwander
 30 2006 The brain differentiates human and non-human grammars: Func-
 31 tional localization and structural connectivity. *Proceedings of the
 32 National Academy of Sciences*, 103, 2458–2463.
- 33 Gebauer, Guido F., and Nicholas J. Mackintosh
 34 2007 Psychometric intelligence dissociates implicit and explicit learn-
 35 ing. *Journal of Experimental Psychology: Learning, Memory, and
 36 Cognition*, 33, 34–54.
- 37 Goldstein, Michael H., Andrew P. King, and Meredith J. West
 38 2003 Social interaction shapes babbling: Testing parallels between
 39 birdsong and speech. *Proceedings of the National Academy of
 40 Sciences*, 100, 8030–8035.
- Goldstein, Michael H., and Jennifer A. Schwade
 2008 Social feedback to infants' babbling facilitates rapid phonologi-
 cal learning. *Psychological Science*, 19, 515–523.

- 1 Goldstein, Michael H., Heidi R. Waterfall, Arnon Lotem, Joseph Y. Halpern,
 2 Jennifer A. Schwade, Luca Onnis, and Shimon Edelman
 3 2010 General cognitive principles for learning structure in time and
 space. *Trends in Cognitive Sciences*, 14, 249–258.
- 4 Goldstein, Michael H., and Meredith J. West
 5 1999 Consistent responses of human mothers to prelinguistic infants:
 6 The effect of prelinguistic repertoire size. *Journal of Comparative*
 7 *Psychology*, 113, 52–58.
- 8 Gómez, Rebecca L.
 9 2002 Variability and detection of invariant structure. *Psychological*
 10 *Science*, 13, 431–436.
- 11 Gómez, Rebecca L.
 12 2006 Dynamically guided learning. In Y. Munakata and M. H. Johnson
 13 (Eds.), *Attention and performance XXI: Processes of change in*
 14 *brain and cognitive development* (pp. 87–110). Oxford: Oxford
 University Press.
- 15 Gómez, Rebecca L., Richard R. Bootzin, and Lynn Nadel
 16 2006 Naps promote abstraction in language-learning infants. *Psycho-*
 17 *logical Science*, 17, 670–674.
- 18 Gómez, Rebecca L., and LouAnn Gerken
 19 1999 Artificial grammar learning by 1-year-olds leads to specific and
 abstract knowledge. *Cognition*, 70, 109–135.
- 20 Gómez, Rebecca L., and LouAnn Gerken
 21 2000 Infant artificial language learning and language acquisition.
 22 *Trends in Cognitive Sciences*, 4, 178–186.
- 23 Gómez, Rebecca L., and Laura Lakusta
 24 2004 A first step in form-based category abstraction by 12-month-old
 25 infants. *Developmental Science*, 7, 567–580.
- 26 Gómez, Rebecca L., and Jessica Maye
 27 2005 The developmental trajectory of nonadjacent dependency learn-
 28 ing. *Infancy*, 7, 183–206.
- 29 Goschke, Thomas, Angela D. Friederici, Sonja A. Kotz, and Anja van Kampen
 30 2001 Procedural learning in Broca’s aphasia: Dissociation between the
 31 implicit acquisition of spatio-motor and phoneme sequences.
Journal of Cognitive Neuroscience, 13, 370–388.
- 32 Greenough, William T., James E. Black, and Christopher S. Wallace
 33 1987 Experience and brain development. *Child Development*, 58, 539–
 559.
- 34 Haith, Marshall M., Cindy Hazan, and Gail S. Goodman
 35 1988 Expectation and anticipation of dynamic visual events by 3.5-
 36 month-old babies. *Child Development*, 59, 467–479.
- 37 Haith, Marshall M., and Michael E. McCarty
 38 1990 Stability of visual expectations at 3.0 months of age. *Develop-*
 39 *mental Psychology*, 26, 68–74.
- 40

- 1 Haith, Marshall M., Naomi Wentworth, and Richard L. Canfield
 2 1993 The formation of expectations in early infancy. In C. Rovee-
 3 Collier and L. P. Lipsitt (Eds.), *Advances in infancy research*
 4 (Vol. 8). Norwood, NJ: Ablex.
- 5 Harris, Zellig S.
 6 1955 From phoneme to morpheme. *Language*, 31, 190–222.
- 7 Hasher, Lynn, and Rose T. Zacks
 8 1984 Automatic processing of fundamental information: The case of
 9 frequency of occurrence. *American Psychologist*, 39, 1372–1388.
- 10 Hayes, J. R., and H. H. Clark
 11 1970 Experiments in the segmentation of an artificial speech analogue.
 12 In J. R. Hayes (Ed.), *Cognition and the development of language*
 13 (pp. 221–234). New York: Wiley.
- 14 Hull, Clark Leonard
 15 1920 Quantitative aspects of the evolution of concepts. An experimen-
 16 tal study. *Psychological Monographs*, 28.
- 17 Hunt, Ruskin H., and Richard N. Aslin
 18 2001 Statistical learning in a serial reaction time task: Access to sepa-
 19 rable statistical cues by individual learners. *Journal of Experimen-*
 20 *tal Psychology: General*, 130, 658–680.
- 21 Hunter, Michael A., and Elinor W. Ames
 22 1988 A multifactor model of infant preferences for novel and familiar
 23 stimuli. *Advances in Infancy Research*, 5, 69–95.
- 24 Jenkins, John G.
 25 1933 Instruction as a factor in ‘incidental’ learning. *The American*
 26 *Journal of Psychology*, 45, 471–477.
- 27 Johnson, Elizabeth K., and Peter W. Jusczyk
 28 2001 Word segmentation by 8-month-olds: When speech cues count
 29 more than statistics. *Journal of Memory and Language*, 44, 548–
 30 567.
- 31 Johnson, Elizabeth K., and Michael D. Tyler
 32 2010 Testing the limits of statistical learning for word segmentation.
 33 *Developmental Science*, 13, 339–345.
- 34 Johnson, Jacqueline S., and Elissa L. Newport
 35 1989 Critical period effects in second language learning: The influence
 36 of maturational state on the acquisition of English as a second
 37 language. *Cognitive Psychology*, 21, 60–99.
- 38 Jones, James, and Harold Pashler
 39 2007 Is the mind inherently forward looking? Comparing prediction
 40 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.

- 1 Kareev, Yaakov, Iris Lieberman, and Miri Lev
 2 1997 Through a narrow window: Sample size and the perception of
 3 correlation. *Journal of Experimental Psychology: General*, 126,
 4 278–287.
- 5 Keele, Steven W., Richard Ivry, Ulrich Mayr, Eliot Hazeltine, and Herbert Heuer
 6 2003 The cognitive and neural architecture of sequence representa-
 7 tion. *Psychological Review*, 110, 316–339.
- 8 Keele, Steven W., and Peggy J. Jennings
 9 1992 Attention in the representation of sequence: Experiment and
 10 theory. *Human Movement Science*, 11, 125–138.
- 11 Kirkham, Natasha Z., Jonathan A. Slemmer, and Scott P. Johnson
 12 2002 Visual statistical learning in infancy: Evidence for a domain
 13 general learning mechanism. *Cognition*, 83, B35–B42.
- 14 Kirkham, Natasha Z., Jonathan A. Slemmer, Daniel C. Richardson, and Scott P.
 15 Johnson
 16 2007 Location, location, location: Development of spatiotemporal
 17 sequence learning in infancy. *Child Development*, 78, 1559–1571.
- 18 Knowlton, Barbara J., Squire, Larry R., and Gluck, Mark A.
 19 1994 Probabilistic classification learning in amnesia. *Learning and
 20 Memory*, 1, 106–120.
- 21 Kuhn, Gustav, and Zoltán Dienes
 22 2005 Implicit learning of nonlocal musical rules: Implicitly learning
 23 more than chunks. *Journal of Experimental Psychology: Learn-
 24 ing, Memory, and Cognition*, 31, 1417–1432.
- 25 Lasky, Robert E., and Amber L. Williams
 26 2005 The development of the auditory system from conception to
 27 term. *Neoreviews*, 6, 141–152.
- 28 Lewicki, Pawel, Maria Czyzewska, and Hunter Hoffman
 29 1987 Unconscious acquisition of complex procedural knowledge. *Jour-
 30 nal of Experimental Psychology: Learning, Memory, and Cognition*,
 31 13, 523–530.
- 32 Lewkowicz, David J.
 33 2008 Perception of dynamic and static audiovisual sequences in 3- and
 34 4-month-old infants. *Child Development*, 79, 1538–1554.
- 35 Lieberman, Philip
 36 2002 On the nature and evolution of the neural bases of human lan-
 37 guage. *American Journal of Physical Anthropology*, 35, 36–62.
- 38 MacDonald, Maryellen C., and Morten H. Christiansen
 39 2002 Reassessing working memory: A comment on Just & Carpenter
 40 (1992) and Waters & Caplan (1996). *Psychological Review*, 109,
 35–54.
- 41 Marshall, Lisa, and Jan Born
 42 2007 The contribution of sleep to hippocampus-dependent memory
 43 consolidation. *Trends in Cognitive Sciences*, 11, 442–450.

- 1 Maybery, Murray T., Margaret Taylor, and Angela O'Brien-Malone
 2 1995 Implicit learning: Sensitive to age but not IQ. *Australian Journal*
 3 *of Psychology*, 47, 8–17.
- 4 Maye, Jessica, Daniel J. Weiss, and Richard N. Aslin
 5 2008 Statistical phonetic learning in infants: Facilitation and feature
 6 generalization. *Developmental Science*, 11, 122–134.
- 7 Maye, Jessica, Janet F. Werker, and LouAnn Gerken
 8 2002 Infant sensitivity to distributional information can affect phonetic
 9 discrimination. *Cognition*, 82, B101–B111.
- 10 McCauley, Stewart M., and Morten H. Christiansen
 11 2011 *Learning simple statistics for language comprehension and produc-*
 12 *tion: The CAPPUCCINO model*. Submitted manuscript.
- 13 McMurray, Bob, Michael K. Tanenhaus, Richard N. Aslin, and Michael J. Spivey
 14 2003 Probabilistic constraint satisfaction at the lexical/phonetic inter-
 15 face: Evidence for gradient effects of within-category VOT on
 16 lexical access. *Journal of Psycholinguistic Research*, 32, 77–97.
- 17 Meulemans, Thierry, Martial Van der Linden, and Pierre Perruchet
 18 1998 Implicit sequence learning in children. *Journal of Experimental*
 19 *Child Psychology*, 69, 199–221.
- 20 Miller, George A., and Jennifer A. Selfridge
 21 1950 Verbal context and the recall of meaningful material. *The Amer-*
 22 *ican Journal of Psychology*, 63, 176–185.
- 23 Mirman, Daniel, Katharine Graf Estes, and James S. Magnuson
 24 2010 Computational modeling of statistical learning: Effects of transi-
 25 tional probability versus frequency and links to word learning.
 26 *Infancy*, 15, 471–486.
- 27 Misyak, Jennifer B., and Morten H. Christiansen
 28 2011 Statistical learning and language: An individual differences study.
 29 *Language Learning*. doi: 10.1111/j.1467-9922.2010.00626.x
- 30 Misyak, Jennifer B., Morten H. Christiansen, and J. Bruce Tomblin
 31 2010 Sequential expectations: The role of prediction-based learning in
 32 language. *Topics in Cognitive Science*, 2, 138–153.
- 33 Misyak, Jennifer B., Lauren L. Emberson, Jennifer A. Schwade, Morten H.
 34 Christiansen, and Michael H. Goldstein
 35 2009 Face-to-face, word-for-word: Comparing infants' statistical learn-
 36 ing of visual and auditory sequences. Poster session presented
 37 at the *Biennial Meeting of the Society for Research in Child*
 38 *Development*. Denver, CO.
- 39 Nelson, Charles A.
 40 1995 The ontogeny of human memory: A cognitive neuroscience per-
 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.

- 1 Newport, Elissa L.
 2 1990 Maturational constraints on language learning. *Cognitive Science*,
 3 14, 11–28.
- 4 Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen, and Richard N. Aslin
 5 2004 Learning at a distance II. Statistical learning of non-adjacent
 6 dependencies in a non-human primate. *Cognitive Psychology*, 49,
 7 85–117.
- 8 Nissen, Mary Jo, and Peter Bullemer
 9 1987 Attentional requirements of learning: Evidence from performance
 10 measures. *Cognitive Psychology*, 19, 1–32.
- 11 Onnis, Luca, Morten H. Christiansen, Nick Chater, and Rebecca L. Gómez
 12 2003 Reduction of uncertainty in human sequential learning: Evidence
 13 from artificial language learning. In R. Alterman and D. Kirsh
 14 (Eds.), *Proceedings of the 25th Annual Cognitive Science Society*
 15 (pp. 886–891). Boston, MA: Cognitive Science Society.
- 16 Onnis, Luca, Padraic Monaghan, Morten H. Christiansen, and Nick Chater
 17 2004 Variability is the spice of learning, and a crucial ingredient for
 18 detecting and generalizing in nonadjacent dependencies. In K.
 19 Forbus, D. Gentner and T. Regier (Eds.), *Proceedings of the*
 20 *26th Annual Cognitive Science Society* (pp. 1047–1052). Chicago,
 21 Illinois: Cognitive Science Society.
- 22 Pacton, Sébastien, and Pierre Perruchet
 23 2008 An attention-based associative account of adjacent and nonadja-
 24 cent dependency learning. *Journal of Experimental Psychology:*
 25 *Learning, Memory, and Cognition*, 34, 80–96.
- 26 Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran
 27 2009a Learning in reverse: Eight-month-old infants track backward
 28 transitional probabilities. *Cognition*, 113, 244–247.
- 29 Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran
 30 2009b Statistical learning in a natural language by 8-month-old infants.
 31 *Child Development*, 80, 674–685.
- 32 Perrig, W. J.
 33 2001 Implicit memory, cognitive psychology of. In: N. J. Smelser and
 34 P. B. Baltes (Eds.), *International Encyclopedia of the Social and*
 35 *Behavioral Sciences*, Vol. 11, Elsevier, Amsterdam (2001), pp.
 36 7241–7245.
- 37 Perruchet, Pierre, and Stéphane Desaulty
 38 2008 A role for backward transitional probabilities in word segmenta-
 39 tion? *Memory & Cognition*, 36, 1299–1305.
- 40 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 process-
 ing. *Journal of Neurolinguistics*, 17, 97–119.

- 1 Quartz, Steven R., and Terrence J. Sejnowski
 2 1997 The neural basis of cognitive development: A constructivist
 3 manifesto. *Behavioral and Brain Sciences*, 20, 537–596.
- 4 Reber, Arthur S.
 5 1967 Implicit learning of artificial grammars. *Journal of Verbal Learn-*
 6 *ing and Verbal Behavior*, 6, 855–863.
- 7 Reber, Arthur S.
 8 1993 *Implicit learning and tacit knowledge: An essay on the cognitive*
 9 *unconscious*. New York: Oxford University Press.
- 10 Reber, Arthur S., and Rhianon Allen
 11 1978 Analogic and abstraction strategies in synthetic grammar learn-
 12 ing: A functionalist interpretation. *Cognition*, 6, 189–221.
- 13 Reber, Arthur S., and Rhianon Allen
 14 2000 Individual differences in implicit learning: Implications for the
 15 evolution of consciousness. In R. G. Kunzendorf and B. Wallace
 16 (Eds.), *Individual differences in conscious experience* (pp. 227–
 17 247). Amsterdam: John Benjamins Publishing Company.
- 18 Reber, Arthur S., and Selma Lewis
 19 1977 Implicit learning: An analysis of the form and structure of a
 20 body of tacit knowledge. *Cognition*, 5, 333–361.
- 21 Reber, Arthur S., Faye F. Walkenfeld, and Ruth Hernstadt
 22 1991 Implicit and explicit learning: Individual differences and IQ.
 23 *Journal of Experimental Psychology: Learning, Memory, and*
 24 *Cognition*, 17, 888–896.
- 25 Remillard, Gilbert
 26 2008 Implicit learning of second-, third-, and fourth-order adjacent
 27 and nonadjacent sequential dependencies. *The Quarterly Journal*
 28 *of Experimental Psychology*, 61, 400–424.
- 29 Rescorla, Robert A.
 30 1967 Pavlovian conditioning and its proper control procedures. *Psy-*
 31 *chological Review*, 74, 71–80.
- 32 Rescorla, Robert A., and Allan R. Wagner
 33 1972 A theory of Pavlovian conditioning: Variations in the effective-
 34 ness of reinforcement and nonreinforcement. In A. H. Black
 35 and W. F. Prokasy (Eds.), *Classical conditioning II: Current*
 36 *research and theory* (pp. 64–99). New York: Appleton-Century-
 37 Crofts.
- 38 Reznick, J. Steven, Katarzyna Chawarska, and Stephanie Betts
 39 2000 The development of visual expectations in the first year. *Child*
 40 *Development*, 71, 1191–1204.
- 41 Rohde, Douglas L. T., and David C. Plaut
 42 2003 Less is less in language acquisition. In P. T. Quinlan (Ed.), *Con-*
 43 *nectionist models of development: Developmental processes in real*
 44 *and artificial neural networks* (pp. 189–232). NY: Psychology
 45 Press.

- 1 Roodenrys, Steven, and Naomi Dunn
 2 2008 Unimpaired implicit learning in children with developmental
 3 dyslexia. *Dyslexia*, 14, 1–15.
- 4 Rose, Susan A., Judith F. Feldman, Jeffery J. Jankowski, and Donna M. Caro
 5 2002 A longitudinal study of visual expectation and reaction time in
 6 the first year of life. *Child Development*, 73, 47–61.
- 7 Saffran, Jenny R.
 8 2001 The use of predictive dependencies in language learning. *Journal*
 9 *of Memory and Language*, 44, 493–515.
- 10 Saffran, Jenny R.
 11 2002 Constraints on statistical language learning. *Journal of Memory*
 12 *and Language*, 47, 172–196.
- 13 Saffran, Jenny R.
 14 2003 Statistical language learning: Mechanisms and constraints. *Current*
 15 *Directions in Psychological Science*, 12, 110–114.
- 16 Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport
 17 1996 Statistical learning by 8-month-old infants. *Science*, 274, 1926–
 18 1928.
- 19 Saffran, Jenny R., Marc Hauser, Rebecca Seibel, Joshua Kapfhamer, Fritz Tsao,
 20 and Fieri Cushman
 21 2008 Grammatical pattern learning by human infants and cotton-top
 22 tamarin monkeys. *Cognition*, 107, 479–500.
- 23 Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport
 24 1999 Statistical learning of tone sequences by human infants and
 25 adults. *Cognition*, 70, 27–52.
- 26 Saffran, Jenny R., Elissa L. Newport, and Richard N. Aslin
 27 1996 Word segmentation: The role of distributional cues. *Journal of*
 28 *Memory and Language*, 35, 606–621.
- 29 Saffran, Jenny R., Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and
 30 Sandra Barrueco
 31 1997 Incidental language learning: Listening (and learning) out of the
 32 comer of your ear. *Psychological Science*, 8, 101–105.
- 33 Saffran, Jenny R., and Erik D. Thiessen
 34 2003 Pattern induction by infant language learners. *Developmental*
 35 *Psychology*, 39, 484–494.
- 36 Saffran, Jenny R., and Erik D. Thiessen
 37 2007 Domain-general learning capacities. In E. Hoff. and M. Shatz
 38 (Eds.), *Handbook of Language Development* (pp. 68–86). Cam-
 39 bridge: Blackwell.
- 40 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.

- 1 Santelmann, Lynn M., and Peter W. Jusczyk
 2 1998 Sensitivity to discontinuous dependencies in language learners:
 3 Evidence for limitations in processing space. *Cognition*, 69, 105–
 4 134.
- 5 Scott, Lisa S., Olivier Pascalis, and Charles A. Nelson
 6 2007 A domain-general theory of the development of perceptual
 7 discrimination. *Current Directions in Psychological Science*, 16,
 8 197–201.
- 9 Schvaneveldt, Roger W., and Rebecca L. Gómez
 10 1998 Attention and probabilistic sequence learning. *Psychological
 11 Research*, 61, 175–190.
- 12 Servan-Schreiber, David, Axel Cleeremans, and James L. McClelland
 13 1988 Encoding sequential structure in simple recurrent networks,
 14 *CMU Technical Report CMU-CS-335-87*: Computer Science
 15 Department, Carnegie-Mellon University.
- 16 Shanks, David R.
 17 1995 *The psychology of associative learning*. Cambridge: Cambridge
 18 University Press.
- 19 Shannon, C. E.
 20 1948 A mathematical theory of communication. *Bell Systems Technical
 21 Journal*, 27, 379–423, 623–656.
- 22 Spivey, Michael J.
 23 2007 *The continuity of mind*. New York: Oxford University Press.
- 24 Swingley, Daniel
 25 2005 Statistical clustering and the contents of the infant vocabulary.
 26 *Cognitive Psychology*, 50, 86–132.
- 27 Teinonen, Tuomas, Vineta Fellman, Risto Näätänen, Paavo Alku, and Minna
 28 Huotilainen
 29 2009 Statistical language learning in neonates revealed by event-
 30 related brain potentials. *BMC Neuroscience*, 10, 21.
- 31 Thomas, Kathleen M., Ruskin H. Hunt, Nathalie Vizueta, Tobias Sommer, Sarah
 32 Durston, Yihong Yang, and Michael S. Worden
 33 2004 Evidence of developmental differences in implicit sequence learn-
 34 ing: An fMRI study of children and adults. *Journal of Cognitive
 35 Neuroscience*, 16, 1339–1351.
- 36 Thomas, Kathleen M., and Charles A. Nelson
 37 2001 Serial reaction time learning in preschool- and school-age children.
 38 *Journal of Experimental Child Psychology*, 79, 364–387.
- 39 Thorndike, E. L., and Rock, R. T.
 40 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. *Per-
 ception & Psychophysics*, 67, 867–875.

- 1 Turk-Browne, Nicholas B., Philip J. Isola, Brian J. Scholl, and Teresa A. Treat
2 2008 Multidimensional visual statistical learning. *Journal of Experimental*
3 *Psychology: Learning, Memory, and Cognition*, 34, 399–407.
- 4 Turk-Browne, Nicholas B., and Brian J. Scholl
5 2009 Flexible visual statistical learning: Transfer across space and
6 time. *Journal of Experimental Psychology: Human Perception*
7 *and Performance*, 35, 195–202.
- 8 Turk-Browne, Nicholas B., Brian J. Scholl, Marvin M. Chun, and Marcia K.
9 Johnson
10 2009 Neural evidence of statistical learning: Efficient detection of
11 visual regularities without awareness. *Journal of Cognitive Neuro-*
12 *science*, 21, 1934–1945.
- 13 Vaidya, Chandan J., Marianne Huger, Darlene V. Howard, and James H.
14 Howard Jr
15 2007 Developmental differences in implicit learning of spatial context.
16 *Neuropsychology*, 21, 497–506.
- 17 van den Bos, Esther
18 2007 *Implicit artificial grammar learning: Effects of complexity and*
19 *usefulness of the structure*. Unpublished PhD dissertation, Leiden
20 University, Netherlands.
- 21 van den Bos, Esther, and Fenna H. Poletiek
22 2008 Effects of grammar complexity on artificial grammar learning.
23 *Memory & Cognition*, 36, 1122–1131.
- 24 Weiss, Daniel J., and Elissa L. Newport
25 2006 Mechanisms underlying language acquisition: Benefits from a
26 comparative approach. *Infancy*, 9, 241–257.
- 27 Wentworth, Naomi, and Marshall M. Haith
28 1992 Event-specific expectations of 2- and 3-month-old infants. *Devel-*
29 *opmental Psychology*, 28, 842–850.
- 30 Wentworth, Naomi, Marshall M. Haith, and Roberta Hood
31 2002 Spatiotemporal regularity and interevent contingencies as infor-
32 mation for infants' visual expectations. *Infancy*, 3, 303–322.
- 33 Zacks, Rose T., and Lynn Hasher
34 2002 Frequency processing: A twenty-five year perspective. In P.
35 Sedlmeier and T. Betsch (Eds.), *Frequency processing and cogni-*
36 *tion* (pp. 21–36). New York: Oxford University Press.
- 37
38
39
40