ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
21(3) 170–176

DOI: 10.1177/0963721412436806
http://cdps.sagepub.com
SAGE

Statistical Learning: From Acquiring Specific Items to Forming General Rules

Richard N. Aslin and Elissa L. Newport
University of Rochester

Abstract

Statistical learning is a rapid and robust mechanism that enables adults and infants to extract patterns embedded in both language and visual domains. Statistical learning operates implicitly, without instruction, through mere exposure to a set of input stimuli. However, much of what learners must acquire about a structured domain consists of principles or rules that can be applied to novel inputs. It has been claimed that statistical learning and rule learning are separate mechanisms; in this article, however, we review evidence and provide a unifying perspective that argues for a single statistical-learning mechanism that accounts for both the learning of input stimuli and the generalization of learned patterns to novel instances. The balance between instance-learning and generalization is based on two factors: the strength of perceptual and cognitive biases that highlight structural regularities, and the consistency of elements' contexts (unique vs. overlapping) in the input.

Keywords

statistical learning, rule learning, generalization, infants

Imagine that it's your 10th birthday and your parents have given you a new video game—but the instructions are missing. Surely you can figure it out. You flip on the power switch and a stream of sounds comes out of a loudspeaker and a cascade of pictures moves across the display screen. The flow of information is overwhelming. What should you attend to: the sounds or the pictures? Is it the quality of the sounds or their temporal order that matters? Is it the identity of the objects in the pictures or their specific shapes and colors that matter?

The conditions in the foregoing scenario are not unlike those of the world confronting a naïve learner. There is structure in the world, we presume, and some set of principles that determines that structure. We can't possibly learn the structure without gathering some input, yet we can't wait for every potential structure to be available in the input before we make inferences about the "rules of the game." However, there are an infinite number of structures that *could* be embedded in the input. In a video game, the sound that accompanies an alien's appearance on the screen could predict whether the alien will attack or flee. Similarly, in the natural environment, a child learning the names of objects must confront the ambiguity of what a word means: Does "doggie" refer to a type of animal, the color brown, a furry coat, or having four legs?

Statistical Learning in Language and Vision

The problem is that the learner must select the *correct* structure in a given set of data from an infinite number of *potential* structures, without waiting forever and without the aid of an instructor who can explain the principles underlying the data (Chomsky, 1965). Somewhat surprisingly, adults and even infants are quite good at extracting the organizational structure of a set of seemingly ambiguous data by merely observing (or listening to) the input. We demonstrated this powerful learning mechanism in an early study (Saffran, Aslin, & Newport, 1996) in which we investigated whether 8-month-old infants could discover the "words" in a stream of speech when the only available source of information was the probability that certain syllables occurred in specific temporal orders. The infants heard a continuous stream of speech sounds comprising four randomly ordered three-syllable nonsense "words," with no pauses between the words and no pitch- or duration-based cues to signal the location of word boundaries (see Fig. 1). What defined a given word, therefore, was the fact that the first syllable was always followed by a specific second syllable, and the second syllable was always followed by a specific third syllable; in contrast, the last syllable of each word was followed by a number of different syllables (i.e., the first syllables of *any* of the other words).

Corresponding Author:
Richard N. Aslin, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, University of Rochester, Rochester, NY 14627
E-mail: aslin@cvs.rochester.edu

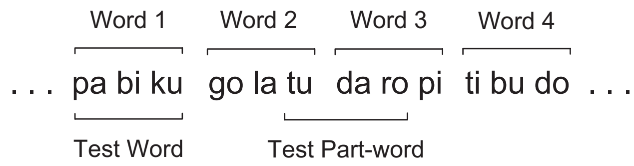

Fig. 1. Stimuli from Saffran, Aslin, and Newport (1996). In this study, infants heard a continuous stream of speech sounds comprising four randomly ordered three-syllable nonsense "words." In a post-exposure test, we investigated whether infants distinguished words from part-words (i.e., trisyllabic sequences that consisted of syllables from two different words).

Thus, the probability that one syllable followed another within a word (the *transitional probability*) was 1.0, whereas the transitional probability of syllable pairs at word boundaries was 0.33. In a post-exposure test, the infants demonstrated their ability to learn which syllables formed words by responding differently to trisyllabic sequences that formed a word than they did to trisyllabic sequences that spanned the junctures between two words (called part-words). After listening to a continuous stream of syllables for only 2 minutes, without knowing whether or where words were present in the speech stream, the infants recognized the words—that is, they managed to discover the correct underlying structure through mere exposure.

Saffran et al. (1996) suggested the term *statistical learning* to refer to the process by which learners acquire information about distributions of elements in the input. In this experiment, the elements were the syllables and the distributions were the statistics of how likely these elements were to occur in relation to one another (see Fig. 2a). Because the frequency of syllables was equated, learners could not use this statistic to segment words. However, another statistic—one that could be used to distinguish words from other sequences of syllables in the stream of speech—was the transitional probability from one syllable to the next. If learners could keep track of this statistic for every pair of syllables in the stream, they would be able to discriminate between words and part-words. The results of the Saffran et al. study suggested that learners were indeed computing such a statistic (though without being aware of performing such a computation).

Another example of statistical learning comes from the domain of speech perception. Maye and her colleagues (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002) presented infants with syllables that came from a continuum spanning two phonetic categories (see Fig. 2b). When the frequencies of the various syllables in the exposure formed a unimodal distribution (i.e., the most frequently presented syllables were from the middle of the continuum), infants did not discriminate the difference between the categories; however, when the syllables formed a bimodal distribution (i.e., the most frequently presented syllables were from the two ends of the continuum), discrimination was reliable. Thus, as in Saffran et al. (1996), infants extracted a statistic (in this case, syllable frequencies) from a corpus of speech to make implicit decisions about a test stimulus that came from that corpus.

Subsequent experiments have shown that these remarkable statistical-learning abilities are not limited to the domain of language. Kirkham, Slemmer, and Johnson (2002) reported that after exposure to a temporal sequence of visual shapes, infants as young as 2 months of age could discriminate between familiar and novel sequences of shapes. Fiser and Aslin (2002) showed that 9-month-old infants could learn the statistical consistency with which shapes were spatially

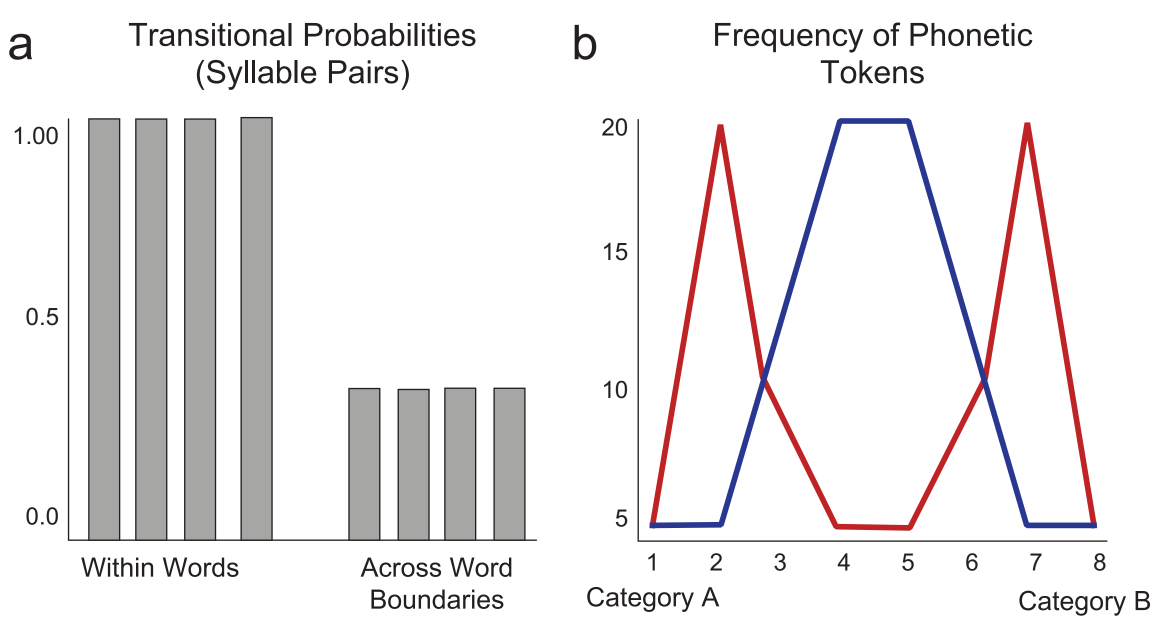

Fig. 2. Distribution of transitional probabilities of syllable pairs within and across words in stimuli from Saffran, Aslin, and Newport (1996; a) and distribution of phonetic tokens lying along a /da/–/ta/ continuum from Maye, Werker, and Gerken (2002; b). In the graph on the right, the distribution indicated by the blue line is unimodal and the distribution indicated by the red line is bimodal.

arranged in visual scenes. In addition, recent studies have documented statistical learning of both auditory stimuli (Teinonen, Fellman, Naatanen, Alku, & Huotilainen, 2009) and visual stimuli (Bulf, Johnson, & Valenza, 2011) in newborns. Thus, statistical learning is a powerful and domain-general mechanism available early in development to infants who are naïve (i.e., uninstructed) about how to negotiate a complex learning task.

These results show that a statistical-learning mechanism enables learners to extract one or more statistics and use this information to make an implicit decision about the stimulus materials that were present in the input. This ability is important for learning which syllables form words, for estimating the number of peaks in a distribution of speech sounds, and for discovering which visual features form the parts of a scene. But this does not address the question of how learners form *rules*—abstractions about patterns that could be generalized to elements that have never been seen or heard. How do learners who are exposed to a subset of the possible patterns in their input go beyond this to infer a set of general principles or "rules of the game"?

From Statistical Learning to Rule Learning

Several studies have documented that infants can make the inductive leap from observed stimuli to novel stimuli that follow the same rules. Gomez and Gerken (1999) presented 12-month-olds with short strings of nonsense words. These words formed categories similar to nouns and verbs, and infants showed evidence of learning that a grammatical noun–verb pair that was *not* present in the exposure stimuli, but was composed of familiar words and followed the grammatical pattern, was nevertheless "familiar."

Marcus, Vijayan, Bandi Rao, and Vishton (1999) went even further. They showed that 7-month-olds who listened to three-word strings containing a repeating word in either the first two or the last two positions (i.e., AAB or ABB) were able to generalize that repetition rule to completely novel words. As in the case of statistical learning, learning of this AAB or ABB pattern of repetition is not limited to linguistic stimuli but also applies to visual stimuli and to musical sequences (Dawson & Gerken, 2009; Johnson et al., 2009; Marcus, Fernandes, & Johnson, 2007; Saffran, Pollak, Seibel, & Shkolnik, 2007).

Some researchers have claimed that statistical learning and rule learning are two separate mechanisms, because statistical learning involves learning about elements that have been presented during exposure, whereas rule learning can be applied to novel elements and novel combinations (see Endress & Bonatti, 2007; Marcus, 2000). But why do learners sometimes keep track of the specific elements in the input they are exposed to and at other times learn a rule that extends beyond the specifics of the input? An alternate hypothesis is that these two processes are in fact not distinct, but rather are different outcomes of the same learning mechanism.

For example, some stimulus dimensions are naturally more salient than others. If stimuli are encoded in terms of their salient dimensions rather than their specific details, then learners will appear to generalize a rule by applying it to all stimuli that exhibit the same pattern on these salient dimensions. Returning to the scenario of the instruction-less video game, as sounds are playing and objects are flying across the screen, it may be extremely difficult for the 10-year-old to remember the specific sounds or shapes of objects, but you know immediately that all the sounds are high pitched (or not) and that all the objects are falling (or not). These highly salient dimensions constrain the way in which the learner encodes the potential structure in the input, dramatically reducing the ambiguity about what the learner should attend to. If high-pitched sounds predict a hostile invader and falling objects provide protection, the learner can quickly induce the rules that enable longevity in the game.

Salient perceptual dimensions can also constrain the statistical patterns that learners most readily acquire. Temporal proximity is a powerful constraint: Learners rapidly acquire the statistical patterns among elements that immediately follow each other. Moreover, infants are particularly attentive to the immediate repetition of a stimulus (Marcus et al., 1999), even infants as young as 1–2 days of age (Gervain, Macagno, Cogoi, Pena, & Mehler, 2008). However, temporal proximity does not always dominate learning. Adults automatically attend to the musical octave of a sequence of tones, and this (more than temporal proximity) can constrain how they learn the statistical relationships between tones. Learners acquire the melodic patterns among tones in the same octave even if they do not immediately follow one another, whereas melodic patterns among interleaved and adjacent tones in different octaves are not acquired (Creel, Newport, & Aslin, 2004; Dawson & Gerken, 2009). More generally, gestalt principles of perceptual grouping (e.g., temporal proximity and perceptual similarity) serve as important constraints on the element groupings whose statistical regularities are most readily learned (Creel et al., 2004; Endress, Nespor, & Mehler, 2009). These constraints influence whether adults learn statistical regularities among elements that are temporally adjacent or that span an intervening element (Gebhart, Newport, & Aslin, 2009; Newport & Aslin, 2004).

Rule Learning Without Perceptual Cues

Although perceptual cues can serve as powerful constraints on statistical learning, perceptual salience is not how most rules are defined in the natural environment. For example, all chairs have some perceptual similarity, but it is the *function* of a chair, not its form, that defines it. Similarly, in language, verbs do not sound alike, and they do not consistently sound different from nouns. What allows a naïve learner to induce a general rule that applies to a set of elements rather than just one instance but has no perceptual basis? One possibility is that learners are sensitive to contexts that signal this important distinction: They acquire rules when patterns in the input indicate that several elements occur interchangeably in the same

	Syllable B			
Syllable A	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwi li	wiwiwe
ji	ji ji di	ji ji je	ji ji li	ji ji we
de	dededi	dedeje	dedeli	dedewe

Fig. 3. The design of Marcus, Vijayan, Bandi Rao, and Vishton (1999). The two sets of four words used by Gerken (2006) are highlighted in red and blue.

contexts, but acquire specific instances when the patterns apply only to the individual elements. For example, Xu and Tenenbaum (2007) have shown that if children hear the word "glim" applied to three different dogs, they will infer that "glim" means *dog*. In contrast, if "glim" is used three times to refer to the same dog, children interpret it as the dog's name. The same contrast between learning items and learning rules can occur for syllable and word sequences.

Gerken (2006) has made this argument by reconsidering and modifying the design of the Marcus et al. (1999) rule-learning experiment (see Fig. 3). Marcus et al. presented 16 different AAB strings in the learning phase of their experiment. Notice in Figure 3 that four strings ended in *di,* four ended in *je,* four ended in *li,* and four ended in *we.* Thus, infants could have learned the general AAB rule, or they could have learned a more specific pattern: that every string ended in *di, je, li,* or *we*. The more consistent or reliable cue was the repetition of the first two syllables—the AAB rule—because it applied to every string, whereas the "ends in *di* (or *je*, or *li*, or *we*)" rule applied to only one-fourth of the strings.

Gerken (2006) asked whether infants presented with a subset of the 16 strings from the Marcus et al. (1999) study would favor the "repetition of the first two syllables" rule or the "ends in *di, je, li,* or *we*" rule. Infants who heard only four AAB strings that ended in the same syllable (e.g., *di* in the leftmost column of Fig. 3) were tested on two equally plausible rules: (1) all strings involve an AAB repetition, and (2) all strings end in *di*. These infants failed to generalize the first rule to a novel string that retained the AAB pattern but did not end in *di*. In contrast, infants who heard only four AAB strings lying along the diagonal in Figure 3 replicated the Marcus et al. result. Because these strings shared an AAB pattern but ended in four *different* syllables, only the AAB rule was reliable.

In recent work, we (Reeder, Newport, & Aslin, 2009, 2010) demonstrated a similar phenomenon—and described some of the principles for its operation—in the learning of an artificial-language grammar. In our experiments, adult learners were presented with sentences made up of nonsense words that came from three different grammatical categories (A, X, and B), much like subjects, verbs, and direct objects in sentences such as "Bill ate lunch." Depending on the experiment, the input included sentences in which all of the words within a particular category occurred in the same contexts (e.g., words X_1, X_2, and X_3 all occurred after any of the A words and before any of the B words), or the input included only sentences in which the X words occurred in a limited number of overlapping A-word or B-word contexts.

Adult learners are surprisingly sensitive to these differences. Our results showed that participants' tendency to generalize depended on the precise degree of overlap among word contexts that they heard in the input, and also on the consistency with which a particular A or B word was missing from possible X-word contexts. Adults generalize rules when the shared contexts are largely the same, with only an occasional absence of overlap (i.e., a "gap"). However, when the gaps are persistent, adults judge them to be legitimate exceptions to the rule and no longer generalize to these contexts. Thus, similar to the results of Gerken (2006), our findings showed that it was the consistency of context cues that led learners to generalize rules to novel strings, and it was the inconsistency of context cues that kept learners from generalizing and led them to treat some strings as exceptions.

The key point here is that in terms of the reliability of context cues, statistical learning and rule learning are not different mechanisms (see Orban, Fiser, Aslin, & Lengyel, 2008). When there are strong perceptual cues, such as the repetition of elements in an AAB sequence, a statistical-learning mechanism can compute the regularities of the repetitions (i.e., they are either present or absent) or of the elements themselves (e.g., the particular syllables). And, as hypothesized by Gerken (2006) and Reeder et al. (2009, 2010), even when there are no perceptual cues, the consistency of how the context cues are distributed across strings of input determines whether a rule is formed—enabling generalization to novel strings—or whether specific instances are learned. According to this hypothesis, statistical learning is a *single* mechanism whose outcome applies either to elements that have been experienced or to generalization beyond experienced elements, depending on the manner and consistency with which elements are patterned in the learner's input. Importantly, this balance of learning is accomplished without instruction, through mere exposure to structured input.

Language Universals and Statistical Learning

Perceptual salience and the patterning of context cues are not the only factors that can influence what learners acquire via a statistical-learning mechanism. An extensive literature in linguistics has argued that languages of the world display a small number of universal patterns—or a few highly common patterns, out of many that are possible—and has suggested that language learners will fail to acquire languages that do not exhibit these regularities (Chomsky, 1965, 1995; Croft, in press; Greenberg, 1963). Recently, a number of studies using artificial grammars have indeed shown that both children and adults will more readily acquire languages that observe the universal or more typologically common patterns found in natural languages.

For example, Hudson Kam and Newport (2005, 2009) and Austin and Newport (2011) presented adults and children with miniature languages containing inconsistent, probabilistically occurring forms (e.g., nouns were followed by the nonsense word *ka* 67% of the time and by the nonsense word *po* the remaining 33% of the time). This type of probabilistic variation is not characteristic of natural languages, but it does occur in the speech of nonnative speakers who make grammatical errors. Adult learners in these experiments matched the probabilistic variation they had heard in their input when they produced sentences using the miniature language, but young children formed a regular rule, producing *ka* virtually all of the time, thereby restoring to the language the type of regularity that is more characteristic of natural languages.

Other artificial-language studies (Culbertson & Legendre, 2010; Culbertson, Smolensky, & Legendre, 2011; Fedzechkina, Jaeger, & Newport, 2011; Finley & Badecker, 2009; Tily, Frank, & Jaeger, 2011) have shown that even adult learners preferentially learn languages that follow universal linguistic patterns and often alter the languages to be more in line with these universals. In adult learners, these alterations are very small, but such changes can accumulate over generations of learners, shifting languages gradually through time (Tily et al., 2011).

It is not always clear *why* learners acquire certain types of patterns more easily than others (and why languages therefore more commonly exhibit these patterns). Some word orders place prominent words in more consistent positions across different types of phrases; other patterns are more internally regular or conform better to the left-to-right biases of auditory processing. A full understanding of the principles underlying these learning outcomes awaits further research. What is clear, however, is that statistical learning is not simply a veridical reproduction of the stimulus input. Learning is shaped by a number of constraints on perception and memory, at least some of which may apply not only to languages but also to nonlinguistic patterns.

Summary and Future Directions

Studies of statistical learning have revealed a remarkably robust mechanism that extracts distributional information in different domains and across development. There remain two fundamental challenges for the future: (1) to provide a comprehensive theory of the statistical computations that suffice to explain such learning, and (2) to understand the neural mechanisms that support statistical learning and determine whether and how these mechanisms change over development (see Abla, Katahira, & Okanoya, 2008; Abla & Okanoya, 2008; Friederici, Bahlmann, Helm, Schubotz, & Anwander, 2006; Gervain et al., 2008; Karuza et al., 2011; McNealy, Mazziotta, & Dapretto, 2006, 2010; Teinonen et al., 2009; Turk-Browne, Scholl, Chun, & Johnson, 2009).

Recommended Reading

Aslin, R. N., & Newport, E. L. (2008). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions* (pp. 15–29). Mahwah, NJ: Erlbaum. An opinion piece about what the original research on statistical learning allows us to conclude about language learning.

Baker, M. C. (2002). *The atoms of language: The mind's hidden rules of grammar*. New York, NY: Basic Books. A summary of the basic principles of linguistics and how they apply across different languages.

Gerken, L. A. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition, 115*, 362–366. A recent study of infant language learning that illustrates how different sources of structural information are weighted.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*, 191–218. An overview of the early phases of language development in human infants.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*, 93–106. An illustration of how statistical learning operates on natural-language materials.

Johnson, S. P. (2010). How infants learn about the visual world. *Cognitive Science, 34*, 1158–1184. A review of studies of visual statistical learning in human infants.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The preparation of this article was supported in part by National Institutes of Health Grants HD-37082 to both authors and DC-00167 to E. L. N.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*, 952–964.

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia, 46*, 2787–2795.

Austin, A., & Newport, E. L. (2011). *When children learn more than they are taught: Regularization in child and adult learners*. Unpublished manuscript, University of Rochester, Rochester, NY.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121*, 127–132.

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

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1119–1130.

Croft, W. (in press). Typology and universals. In M. Aronoff & J. Rees-Miller (Eds.), *The Blackwell handbook of linguistics* (2nd ed.). Oxford, England: Basil Blackwell.

Culbertson, J., & Legendre, G. (2010, November). *Investigating the evolution of agreement systems using an artificial language learning paradigm*. Paper presented at the Western Conference on Linguistics, Fresno, CA.

Culbertson, J., Smolensky, P., & Legendre, G. (2011). *Learning biases predict a word order universal*. Unpublished manuscript, Johns Hopkins University, Baltimore, MD.

Dawson, C., & Gerken, L. A. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition, 111*, 378–382.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*, 247–299.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13*, 348–353.

Fedzechkina, M., Jaeger, T., & Newport, E. (2011). Functional biases in language learning: Evidence from word order and case-marking interaction. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 318–323). Austin, TX: Cognitive Science Society.

Finley, S., & Badecker, W. (2009). Vowel harmony and feature-based representations. *Journal of Memory and Language, 61*, 423–437.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA, 99*, 15822–15826.

Friederici, A. D., Bahlmann, J., Helm, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, USA, 103*, 2458–2463.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review, 16*, 486–490.

Gerken, L. A. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74.

Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, M. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, USA, 105*, 14222–14227.

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N. Z., Marcus, G. F., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy, 14*, 2–18.

Karuza, E., Newport, E. L., Aslin, R. N., Davis, S., Tivarus, M., & Bavelier, D. (2011, October). *Neural correlates of statistical learning in a word segmentation task: An fMRI study*. Paper presented at Neurobiology of Language Conference, Annapolis, MD.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.

Marcus, G. F. (2000). Pabiku and Ga Ti Ga: Two mechanisms infants use to learn about the world. *Current Directions in Psychological Science, 9*, 145–147.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*, 387–391.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning in seven-month-old infants. *Science, 283*, 77–80.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience, 26*, 7629–7639.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2010). The neural basis of speech parsing in children and adults. *Developmental Science, 13*, 385–406.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance 1: Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

Orban, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, USA, 105*, 2745–2750.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2009). The role of distributional information in linguistic category formation. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2564–2569). Austin, TX: Cognitive Science Society.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2010). Novel words in novel contexts: The role of distributional information in form-class category learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2063–2068). Austin, TX: Cognitive Science Society.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*, 669–680.

Teinonen, T., Fellman, V., Naatanen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*, 21.

Tily, H., Frank, M. C., & Jaeger, T. F. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1364–1369). Austin, TX: Cognitive Science Society.

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

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.